BMC Microbiology

Research article

Avian papillomaviruses: the parrot *Psittacus erithacus* papillomavirus (PePV) genome has a unique organization of the early protein region and is phylogenetically related to the chaffinch papillomavirus

Ruth Tachezy[1], Annabel Rector[2], Marta Havelkova[1], Elke Wollants[2], Pierre Fiten[2], Ghislain Opdenakker[2], A Bennett Jenson[3], John P Sundberg[4] and Marc Van Ranst*[2]

Address: [1]Department of Experimental Virology, Institute of Hematology and Blood Transfusion, U Nemocnice 1, 128 22 Prague, Czech Republic, [2]Laboratory of Clinical and Epidemiological Virology, and Laboratory of Molecular Immunology, Department of Microbiology and Immunology, Rega Institute for Medical Research, University of Leuven, Belgium, [3]Cervical Cancer Research Institute, The Western Pennsylvania Hospital Foundation, 4818 Liberty Avenue, Pittsburgh, PA 15224, USA and [4]The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609, USA

E-mail: Ruth Tachezy - rutach@ukht.cz; Annabel Rector - annabel.rector@uz.kuleuven.ac.be; Marta Havelkova - martahavelkova@hotmail.com; Elke Wollants - elke.wollants@uz.kuleuven.ac.be; Pierre Fiten - pierre.fiten@rega.kuleuven.ac.be; Ghislain Opdenakker - ghislain.opdenakker@uz.kuleuven.ac.be; A Bennett Jenson - abjenson@yahoo.com; John P Sundberg - jps@jax.org; Marc Van Ranst* - marc.vanranst@uz.kuleuven.ac.be

*Corresponding author

Published: 10 July 2002

BMC Microbiology 2002, **2**:19

Received: 21 April 2002
Accepted: 10 July 2002

This article is available from: http://www.biomedcentral.com/1471-2180/2/19

Abstract

Background: An avian papillomavirus genome has been cloned from a cutaneous exophytic papilloma from an African grey parrot (*Psittacus erithacus*). The nucleotide sequence, genome organization, and phylogenetic position of the *Psittacus erithacus* papillomavirus (PePV) were determined. This PePV sequence represents the first complete avian papillomavirus genome defined.

Results: The PePV genome (7304 basepairs) differs from other papillomaviruses, in that it has a unique organization of the early protein region lacking classical E6 and E7 open reading frames. Phylogenetic comparison of the PePV sequence with partial E1 and L1 sequences of the chaffinch (*Fringilla coelebs*) papillomavirus (FPV) reveals that these two avian papillomaviruses form a monophyletic cluster with a common branch that originates near the unresolved center of the papillomavirus evolutionary tree.

Conclusions: The PePV genome has a unique layout of the early protein region which represents a novel prototypic genomic organization for avian papillomaviruses. The close relationship between PePV and FPV, and between their *Psittaciformes* and *Passeriformes* hosts, supports the hypothesis that papillomaviruses have co-evolved and speciated together with their host species throughout evolution.

Background

Papillomaviruses are a large group of pathogens that cause epithelial proliferations in a wide spectrum of vertebrate species. More than 100 different human papillomaviruses (HPV) have been isolated [1,2]. Such an extensive genotype variety has not yet been detected in nonhuman species, although papillomavirus genomes have been isolated from many species where careful investigational efforts were made [3,4]. Most papillomaviruses appear to be species-specific or at least restricted to infection of closely related animals within the same genus. Papillomavirus genomes have been cloned from 20 mammalian host species. Thus far, only two avian papillomavirus genomes were cloned, one from a chaffinch and the second from a parrot papilloma. To date no complete avian papillomavirus genome had been sequenced.

In a large survey of 25,000 captured chaffinches (*Fringilla coelebs*) in the Netherlands, papillomas were found on the foot or tarsometatarsus (the bare part of the leg) of 1.3% of the birds [5]. The DNA of a *Fringilla* PV (FPV) was isolated from such skin papillomas, and two partial sequences of FPV, totaling about 900 basepairs, were determined [6,7]. Papillomavirus particles have also been observed using electron microscopy in greenfinches (*Carduelis chloris*) [8], and in canaries (*Serinus canarius*) [9].

Cutaneous papillomas were observed on the palpebrae, commisure of the beak, and the head of a captive African grey parrot (*Psittacus erithacus timneh*) [10]. The *Psittacus erithacus* papillomavirus genome was cloned [11], and abbreviated PePV in accordance with the nomenclature guidelines for nonhuman papillomaviruses [3]. PePV DNA was also detected in one other oral papilloma of an African grey parrot, and was not present in an oral papilloma of an Amazon parrot (*Amazona ochrocephala*), nor in 24 cloacal papillomas of Amazon parrots, macaws (*Ara sp.*), conures (*Aratinga sp.*) and cockatoos (*Cacatua moluccensis*) [12,13].

This report describes the first complete nucleotide sequence of an avian papillomavirus from a cutaneous lesion of an African grey parrot: PePV.

Results and discussion

Complete sequence of the PePV genome

The complete nucleotide sequence of PePV contains 7304 basepairs (bp) (Fig. 1), and has a GC-content of 49.3%. The size of the PePV genome is the second-smallest of the animal papillomaviruses, after bovine papillomavirus type 4 (BPV-4) (7265 bp) [14]. The part of the PePV E1 ORF containing the *Sal*I cloning site is homologous and colinear with the corresponding region in other PVs. This indicates that no sequences have been lost during the establishment of the PePV clone. The position of the first nucleotide of the PePV genome corresponds to the start codon of the first major open reading frame in the early protein region.

Open reading frame organization

All papillomaviruses have their open reading frames (ORFs) on the same coding (sense) strand of their circular double-stranded DNA genome. Usually, a papillomavirus genome contains seven major ORFs coding for five early (E) proteins E6, E7, E1, E2 and E4, followed by two late (L) capsid proteins L2 and L1, and a non-coding upstream regulatory region (URR). The layout of the PePV genome is different from the organization of other characterized PV genomes (Fig. 2). PePV does not contain classic E6 or E7 ORFs. Instead, it contains an E8 ORF in front of the E1 ORF, followed by an E9 ORF which overlaps with the aminoterminal part of the E1. The E8 ORF has the capacity to code for a 177 amino acid (aa) protein with a predicted molecular weight of 19.6 kilodaltons (kDa), and the E9 ORF encodes a 195 aa protein of 22.7 kDa. Both E8 and E9 do not show recognizable homology with any other known papillomavirus or non-papillomavirus proteins in GenBank. Manual alignment of the PePV E8 with a series of E7 proteins revealed two Cys-X-X-Cys motifs separated by 23 amino acids instead of the usual 28 or 29 residues. In the aminoterminal part of E8, a stretch of amino acids (DNLLCHESSMDD) is similar to the putative cellular division motif involved in retinoblastoma tumor suppressor protein (pRb) binding. In PePV, this pRb-binding motif is more related to the pRb-binding domain of the large T (LT) antigen of polyomaviruses than that of the E7 of other papillomaviruses. However, alignment of PePV E8 with SV40 LT did not reveal further similarities. The E8 is maybe part of a remnant of a very ancient common evolutionary origin of the early region proteins of avian papillomaviruses and the LT proteins of polyomaviruses. Computational motif searches (using SMART, InterPro and Prosite algorithms) failed to detect biologically significant sites, patterns or conserved protein motifs in PePV E8 and E9. Using the PSORT algorithm for the prediction of the subcellular location of proteins [15,16], E8 was predicted to be cytoplasmic, and E9 to be either nuclear or cytoplasmic. The cellular existence and function of E8 and E9 remains unknown.

Unusual ORFs have also been described in the genomes of the Europen elk papillomavirus, deer papillomavirus, and reindeer papillomavirus, which contain a transforming gene (E9) between the E2 and L2 ORFs [17]. The subgroup B bovine papillomaviruses (BPV-3, -4, and -6) are another group of papillomaviruses that lack an E6 (but do have an E7) [14,18]. Instead of an E6, there is a BPV-4 E8 ORF that encodes a 42-residue polypeptide that can transform NIH3T3 cells [19]. No discernable homology between BPV-4 E8 and PePV E8 sequences was detected. It

```
   1  ATGCGAACAA TGCGCTACCA CTACCCCACG GAGACGGACG ACAGCAGTCC TGATGAAGGT GAGACTGGCA ACCATGCTGT AACCCACTTG CTAATGCAGC
 101  TGCAGGAACA ACTGCATGCA CTTAACTATC CCACAGATGA TAGCGACGAC TCCACAGACG GTGAGGAATT AGTATTCCGA TTAACCGTGG AGGAAAGCGC
 201  AAGCGAGGAC GACAATGACA GCCGTCAAAG CCTGTCCGTA GACGATGTGG GGGCAGAGGT AGATGTAGAG GTGGAAGGGG CATACGGTGG TGTTGCATCG
 301  GATAATTTAC TATGTCATGA AAGTATGGAT GATCCCGAGT ACAGCGGTGC CAGCGTCGGG AGTCGCCCTG ACGGGTATGA TGAAAGGGCA CCCTGGAAAT
 401  GCACAATATG TGGGAGACCT GTCACTCCAC AAGAGCTAGC TACCTTCGGG GTAGTCAATC CCTGGAATAA GCAGGGAGTG TGCACCGTGT GCTTTCACGG
 501  GCAACAAGAG AGGTTCAACA GCATCTGGGG TTAATGGACT TCATTACCAT TGAAGCTGAA GATACCAATA GTTCAGACAG CAGCCAATGT GAGGAAGAGG
 601  AAGATAAGGA CGATTCTGCA TTAGGGACAT TTATAGACGA TAGTAATGGA AACGACGAAC CCGACAATAT CAGCCATGTG CAATTGTTGC ACTCCCAGCG
 701  ATCACAAGAC TTTGAGATTA TACCTAAGCG AGTGGCAGGA CATTATACCC GTAAGCGACG AAGAACACGG TCCCCCCACG GTGAGAGCGA TATCATTCGA
 801  TGCCAGCACG GCGGACATGG AGACCCAAAG AAAACTGTGC AAGTATTGTC TCATAGTCCT CCTAAAACGA TTAGGCAGTA TTTTACGTCT GGGGATACTA
 901  ATAGACCCCA TAAAGATACT TCGGAGGGAT GCACAGGGTC AGGGGCCACG CACACCACAC GGGTTATTAC CGCGCAGGTG CACGTTCCTG CTGAACCTAC
1001  GTATACGGCT GGAGGAAATT CTCCTGCCAG AACTCCCCTT AGACAGATTG ATCCTAATGG TCGACCTCCA GGACGTGTCA CACCTGTATC GCCTCCTGTC
1101  CCCCACCGTA GACGAAACGT ATCTGCAGGT GGGACACATT ATGGAGAGAA TGGAGCACGC CTGCCGCCGC CGTATGAAAC GGCATCAGAA ACGAGGCAGA
1201  CTGAAGGACA GAAACTACTT CAACGCTGCC TAGTTTCAAA AAATAAGACA TTAACTGCGT TAGCGGTGTT TAAAGAACTA TATACGGCTA GCTTTACGGA
1301  GGTTACAAGG ACATTTAAAA GCGATAAAAC GCAAAGTTAT GAATGGGTGT TCATGGTGTT GGGGTGTTCA CATATTGCAT TAGAAGCTGT AAAAGATGTA
1401  TTAATACATA ATACGGAACA TGTAATCCTT GATATAGATC CATATAAACA TCTGGGAGTA TATTACGTAG GATTTACAGT TAGCAAAAGC AGGGAGGGGC
1501  TTTTGCGGTT TCTAAAACAA CATAATATAT TTACCGAAAA TGTAGTGCTA TCTAATCCGC CAAATAAACG CTCCGTGCTA TCCGCGCTAT TCTTTGATAA
1601  GTTAGTACAG GTAAGCGGGG ACAAACCACA GTGGATGATA GATATAATAA CCTCTGGGGA CAAGGGTGGT GAAGGCTTTG AATTAAGTAA AATGATACAA
1701  TGGGCTCTGG ACAACAATAT GTACGATGAA GGGGCAATAG CATATAACTA TGCGTTATTA GCTGATACGG ATCTGAATGC ACAATTATGG CTTAAGCATA
1801  ATTCGCAGGC TAAATATGTA CGGGACGCTG CTACAATGTG TAGACACTAT AGAAGGGGGC AAATGCAAGC GATAGGGGTA ATGGAACATT TGGCTACACG
1901  TATGCGGGAA TATGCAGATA GCGATATAGA AGAGGGCTGG AAACGTATAA TTGTGTTCTT ACGATATCAA CATGTAGACC ATCATACATT TATAAATGAT
2001  TTAAAATACT GGATTGTAAA TAGACCAAAA CGTAGTACAA TCGCTATAGT GGGAATACCG GACAGTGGGA AAAGCATGTT CGGGATGTCA CTCATACAAT
2101  TTCTAGATGG ACGGGTGTTA AGCTTCTCTA ATCATAAATC GCATTTCTGG TTACAACCAT TGTCAGAAAC ACGTTATGCA TTAGTGGACG ATGTAACATG
2201  GCCTGCATGG GACTATATGG ATGTGTATAT GCGAAATGCA TTAGACGGTA ACCCTATATG TATCGATTGT AAACATAGAG CACCAATACA AACAAAGTGT
2301  CCGCCGTTGT TATTAACGAG CAATTACGAC CCTAGGGAGC GTGGGACAGG GGCGGAAAAC AGCTACCGAT ACCTACTAAG TAGAATTACC TTTATGTCAT
2401  TTAATAGGAG TATCCCATGT ATTGGGGGAC AGCCGAGATT CCTAATTTCA CCAGCGGACT GGAGATCATT CATGCTTAAG TTCAGGAAAG AACTGGACAT
2501  CAACCTAACA GACCTTGATT ATGGAGGGGC TACGGGAGAG CCTGGAGAGA CTGCAGAGGA GAGAGGCTGA GATACTAGAG CAGGACCCAA CTGATCTGCA
2601  GACAATAACT GAATATTGGG AAAACGTTAA GAAACAACAC CTGCTGCTAT ATGCAGCTGG ACAAAAAGGT TATAAACAAC TAGGCCTGCA ACGCGTACCA
2701  CCACTACACG TAAGCGAACA GGAGGCAAGG GACGGTATTC TGATGGTCGT CTTATTGAGG TCTCTCCTAG GTACACCACA TGCCCTACGC ACGTGGAGTC
2801  TAGGGGAGTG GGGGCCCAGA CTGTTCCGGA CCCCACCGGA CGGTCTAAAG TTCGGGCCGC ATACTGTACG GGTATTTTAT TGTAATGATC CAAGTACGGA
2901  AACCGAATAC CCATACTGGG ATAGTTATCT ATTTTATGAT CCCACTACAG GGGAGTGGAC AGAGGGTATA GGGGGTTATG ATAACGTAGG TATATGGCAT
3001  GAAACCATTA ATGGGCGAGG GTATCATATG ATATGGAGGG ATGAAGCGCG TAGAGTGTGT GGGGGTAATC AAGTAACATG GGAACTTTCC ACTAGCGACC
3101  GTGACTCGCT AGACGGTATA CCGCCACCGC TACTGGAATC AACGCGTGTG GAGTCACCCG AACCAGCCAC AGAGCCGGAG ACACCATCTC CACCACAACT
3201  GTCAAACTAC GCCTTAGGCG ACCCGCCAAA CATACGAGGC GGTTCAAATT CGGCGCGCAC CCGATCTCGA CGAATCCGTA CAAGAGGTAC GGGTACTGAC
3301  ACACCAACAG GCATCCGTCC AGAGGACGTG GGGACGGCGA GAACAACGGT CCGGGGTGGT GGGACGCGAC TAGACCGCCT CATCGCGGAG GCTAAGGACC
3401  CACCTGGGCT GTGCTTTGTC GGCCGGACAG GCCAGCTCAA GACCATACGG TACCGGGTGC AGACCGGACC GTATAACGTA ACACGCATAA GCACTACCTG
3501  GCACTGGATA GGGGACGGGG AACATTTGTC ACGTATGATT ATTCTGTTCA ATGATTCGCA TCAGCGGGAA GTATTTGCAA GAGCGTTTCG GGTCGTATCG
3601  GGTGTACGGG TCTATAAGGT CTCATTGTCC GGAATATAAG GATGGTTGCG TACAGACGGT TCCTTCTCCA GCTGCCTGTT CCTCCTACCT CATTTCCGTA
3701  TTATTACGGG CCGCGTTTTT TATTACATTC TGCTGCTACC TCGTCCTCTT TGTCCTCTAC TGTGTCCCCC TCCTCCGCTC GTACTATATC GAGGCGGCGG
3801  AGAGCCGCAG CGGACGACCT ATGGCGTAAA TGCCAATATG GGGATTGTCC AGACGATGTC CGTCAACGTT ATACTCAGAC TACTATAGCT GATAAAATAC
3901  TCCAATGGGG AAGTGCCCTA GCATATTTAG GTGGCCTTGC GGTCGGGACA GGCCGTGGCG GTGGTGGTGT TCGGTTGGGT AGTGGGGCGT CTGCAGTGCC
4001  CCGACCCACG GCCCCTGATA CTACAATTCC GTTTACCCGT CCTGTTGTAC CCGATAGTGC GGCCGCCGTT CCTGAGGGTG GCATAGGGAC TGTTCCTGCT
4101  GACAGACCAT TCCAGGTGCC CACTGCCAAC GTTCCTGTGG ATGTGGTACC TGTGTCCGGG ACCGAGCATG CTAATACATT ACCTCCTAAT ACGTTTGTGA
4201  ATCCTGCCTT TGAGGGGGAT TTAGATTCCT CATCCTCTAT CGATAGTGTA GTTATAGGCG ATTTGGTTCG TAGTGAGGAT GCCCCACGGA CACGGGGTGA
4301  CACCTTTATT ATGGAACATG AATTTGTCGG CCCTATTAGT GAGGAGCGTA GGCCGTACTT CCCTACTCTT AATAGTAATC CTACACAACC TTTCGAGGAA
4401  ATCGAGATGG TTACTTTCGG TAGTACTGCT GATGAGGATT CTGCAATAGT CGGTCGGGAG CCTAGCACAA GCACTCCTGT CACTTTTGGG ACACGTGGCA
4501  CTGCAGGACG GCCTCGTGGA TACACTCTTG ATGTAACCAT CAGTAATCCC GTATATGATG AGGCCCTAGA TATTGACAGG TTATTTCAAG AAGGTTTGCA
4601  GGAGTGGGTG GATAGTGATA TTATCTCTCC CGATGCCCCC GGCATTCCTC TTGGGGATCC CTCATATGCA ACAGCGTCCT TCGGAACACG GTTGCAGGTA
4701  TCTAGACCGG GACAGCTCCC TGGTATCCGC CTCCGCAGTG GTAGACGGTT GCAGATACCG GTATTGTTTA CAGGGGACTT GTCGTCCATA GCTCCCGATC
4801  TGGAATTACA GCCCTTGCAG CCTGTTGGTG CAACCGGCAC TGTCGTTTCT AACAGTGGTG CTGCAGAAAC AGTGTTCTCT GCTGCAGACA CAATAGGCTG
4901  GGACGGTCAG AATATTTCGG CCTCTACTGG TATTATAGAC ATTGGTCCCG TGTCTGGGGA CGACCTACCG TTTTTTGAAA TCCCCCTTGA TGATCCCTTT
5001  CCTGAAATGG AGATTGTGGA GGAGAGTGAG ACTAATACTA CTCCTTATAC TTTATCTGAT ATCTCGGTTG TGGATACTAC TTATCCGTTC CCTGCTATCA
5101  CATCTGCGGC GTACCCGTCT GTGAGCATTC AGGAGGATGG GGGGATTGTC GTCTACCCGA CTCCTGCCGC ACCGGTGTCG TATGGTGGGC TCGTCAGTAT
5201  GGACCCTAAC TCGTTGTTCT GGTTTCTGCT CCGCCGCCGT CGGCGCCGCC GTCGTACCAC TAAACGTATC CTTCTCAACA GATGAGTGCT GCTGGGCCTG
5301  CTCCTGCGTT ACCATCGGCA TTGTATATTC CTAATGCTGC GCCTCTACAA CCACCCCTAT TTACTACGGA CGACTTTGTT TCCCCTACGG ACTATGTGTA
5401  TCACGTAAAT ACGGGACGTC TTTTGATGGT CGGTAACCCA TACTTTTCTG TCCCTGATGC TGATAAGGAC CGTGCAGCGG TTCCTAAGGT GTCTGGTAAT
5501  CAATATAGGG TGTTCAGATT GAAGCTTCCT GATCCTAATG ATCAGTTTGA CCTCCCGGAC GGTCTGTTTG ACCCGGAGAA GTTTCGATAT GTATGGCAAC
5601  TTGTAGGCCT TGAGGTTTGC CGTGGTCAGC CTCTGGGTGT GGGCATTTCC GCGGCCCCGG CCTTTAATAA GGGTCGTGAT GTTGAAAGCC CTGCACGTTT
5701  AGTTGCGGAC GATGCCACGC GAGAGGATGA CAATCGCGTG AGTGTCGGTC TTGACCCCAA ACAGAACCAG ATGTTAATTG TCGGTTGTGC CCCAGCATAT
5801  GGTCAGCACT GGGGCAAGGC AACTCCGTGT CCGGATGACA CATTGGATAC CCAGTGTCCA CCTATAGAAC TGATTAGTAG CACATTGCAG GATGGTGACA
5901  TGTGCGATAT TGGCTTCGGG TGCATGGACT TTGCAGCCTT GGCCGCCAAT ACGTCCGATA TACCCTTGGA ACTCATTAAC ACTGTTAGTA AGTACCCAGA
6001  CTGGATCCGG ATGCATAATG ATCCTAAGGG CGATTGCTGT TTCTTTCTAA TGCGTAGAGA ACAGTTGTAT GCACGACACA TGTGGCAACA TTCTGGTGGC
6101  ATCGGTGAGG CCATACCGAG TGTTTATCTT AATACCTCGT TTACGAGTAC TAATAACTGT GCTTACATGT GTGTTCCTTC CGGGTCTGTA TACACCTCTG
6201  ATACCCAGTT GTTTAATCGG CCGTACTGGC TGTCCAAGGC GCAAGGTCCT AACAACGGCG TTTGTTGGGG TGATGATCTG TTCATTACTG TGTTGGACAA
6301  TACGCGGGGT GGGGTCATGA ACATTTCTAC GAAACCTACG GATAGTGGGG ATGTGTATAA ACCTTCGGAC TTCCGTGAAT ATGTCCGACA TGTAGAGGAA
6401  TACGAATTAT CCTGTGTGTT ACGGCTATGT AAAGTGCCCC TCTCCCCAGA TGTTCTTGCC TCTCTCTACC GTGCTGTCCC CCATGTGCTG GGCCGTTGGG
6501  GTATTTCCGA GTACCCACAG GCCGATACTA CCCCGGAGGA TAAATATCGG TATATCAGTT CACAGGCAAC CCGATGTCCA CTACCTGCTG CAGATACGCC
6601  TACGCCGGTG CAGGATCCGT GGGCGGATAT GACTTTCTGG ACTGTCGATT GCACGTCCCG CATTTCCCCC GAATTACCGC GTTTTCCCCT CGGTCGCAAG
6701  TTCCTAGCTT TGCCCGGACC CCGTCCGGCA ACCCCATTAT ATGGAAAACG TTCTGCTACT GCCGCAGCCC TCACGGGTGC AGCTGGTGTG CGATCTGCTG
6801  GCGTCCGCTC AGGCGTGCGC ACTGCGAAGC GTAGGCGGAG GTAATGGCGT CGCTTATATG ATATCCAATG TTTTTAACGT TACATAAATA TATGCATATC
6901  TTGTTACATG CCTGTTCCTG TATTCATTGT ACCCTCGTCA ACCCCTTAGT TGACAGCCCC TTATACGCCC ACATTTCGAT GTGTTGAGAC ATTTCCTGGG
7001  GCTTATGTTT GCGCGTTCAA CCGTTCGGTT GACATTCGTT GACAGCTGCA ATTGCTCGAA TGAATAGAAG GCGCGGCGGC AGCAACCGTC TGGCACCGTT
7101  CCGACTTGAA AAAGGTACAC CCTTATATAA TATATATATT ATATCATCAG ATATAAAGTT TTTACCACTC TCAACCCTAT GGTTGACATA ACGTATATTG
7201  GATAGGAACC GTAACCAATA GCGGTTCGTA CAGGGGGGTC CTTCGCTTCC GGATCCCATA TACGGGAGAT ATCATAAGAA TAAATAGCTG AGGGGAATAA
7301  CAGT
```

Figure 1
Nucleotide sequence of the *Psittacus erithacus* papillomavirus type 1 (PePV), GenBank Accession number AF502599.

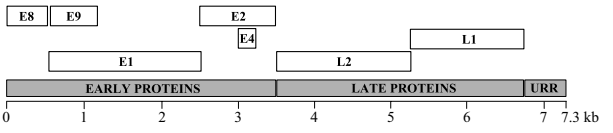

Figure 2
Linear representation of the ORFs of the *Psittacus erithacus* papillomavirus genome. NCR; non-coding region.

seems that E6 functions are either not required by some papillomaviruses, or that they are performed by another viral (or host) protein. It remains to be established if this unique organization of the early region is typical for PePV or common to other/all avian papillomaviruses.

Upstream regulatory region (URR)

In PePV, the noncoding region or upstream regulatory region (URR) between the stop codon of L1 and the first ATG of E8 is only 460 bp long (nucleotides (nt) 6845–7304). Only one typical palindromic E2-binding site (E2BS) with the consensus sequence ACC-N_6-GGT is found at nt 7214–7225. Two additional atypical putative E2BS (ACC-N_4-GGT) are found at nt 7020–7029 and 7174–7183. The URR also contains a polyadenylation site (AATAAA; nt 7279–7284) located 16 nucleotides 5' of a CA dinucleotide, necessary for the processing of the L1 and L2 capsid mRNA transcripts.

PePV sequence similarity to other papillomaviruses

The sequence similarity between PePV-1 and HPV-1 (a benign cutaneous PV), HPV-5 (an epidermodysplasia verruciformis-associated PV), HPV-16 (a prototypic mucosal high-risk PV), and bovine papillomavirus type 1 (BPV-1, a fibropapillomavirus) was investigated by pairwise alignments of the corresponding ORFs and their proteins (Table 2). PePV showed only low similarity to other papillomaviruses. Maizel-Lenk (dot) matrix plots illustrate that similarity can only be observed in the conserved parts of E1 and L1 (Fig. 3).

Phylogenetic analysis

In order to compare the PePV sequence with that of the chaffinch (FPV), we retrieved two partially overlapping partial sequences in the FPV L1 ORF (GenBank accession numbers K02020 and U29669) and one piece of FPV E1 sequence (K02019) from GenBank [7,20]. A total of 312 amino acids (132 E1 and 180 L1 residues) could be compared for both viruses. The similarity at the amino acid level between PePV and FPV was 68% in the E1 region, and 47% in the L1 region.

To define the relationships of PePV with FPV and with other papillomaviruses, we constructed a phylogenetic tree based on a compound E1/L1 312 amino acid sequence alignment of 50 human and animal papillomaviruses. The resulting neighbour-joining phylogenetic tree (fig. 4) clusters different papillomavirus groups, largely according to their tissue tropism and oncogenic potential [1,21,22]. The two avian papillomaviruses form a monophyletic cluster with a common branch that originates close to the unresolved center of the unrooted papillomavirus evolutionary tree, near to the origin of the branch that groups the cutaneous papillomaviruses associated with epidermodysplasia verruciformis (EV). The avian papillomaviruses occupy a unique position among the other known papillomaviruses, with whom they are only distanly related.

Divergence between PePV and FPV coincides with divergence of their Psittaciformes and Passeriformes host species

Mammals and birds diverged around 310 million years (Myr) ago during the late Paleizoic Era [23]. The start of the first avian differentiation into *Paleognathae* (larger, flightless birds such as ostrich, rheas, cassowary, emu, and kiwi), *Galliformes* (turkey, chicken), *Anseriformes* (duck, goose) and other *Neognathae* (all other extant modern birds) is currently thought to have coincided with the Mesozoic breakup of the world-continent Pangaea at the Jurassic-Cretaceous boundary 146 Myr ago [24]. Most of the extant avian orders were establishing themselves at the Cretaceous-Tertiary boundary 65 Myr ago [25]. The *Psittacus erithacus* belongs to the family of the *Psittacidae* in the order of the *Psittaciformes* (parrots), whereas the chaffinch (*Fringilla coelebs*) belongs to the family of the *Fringillidae* in the order of the *Passeriformes* (perching birds).

The rates of nucleotide substitutions between two homologous sequences can be used as a measure for the time elapsed since the two sequences diverged (i.e. the molecular clock concept). When we know the papillomavirus mutation rate, we can approximate the divergence time between PePV and FPV. We earlier calculated a papillomavirus mutation rate based on the divergence of the *Felis domesticus* PV (FdPV-1) and the canine oral papillomavirus (COPV) since the divergence of their *Felidae* and the *Canidae* host species 38–50 Myr ago, to be 0.73 to 0.96×10^{-8} nucleotide substitutions per site per year [26]. We constructed a 921 bp pairwise alignment between PePV and FPV for 441 bp in E1 and 480 bp in L1 where the nucleotide sequence of FPV was available in GenBank (corresponding to nt 1908–2348 and 6107–6586 in PePV). In this 921 bp alignment, 385 differences between PePV and FPV were observed. Using the estimations of the mutation rates derived from the feline/canine papillomavirus divergence, this corresponds to PePV and FPV diverging from

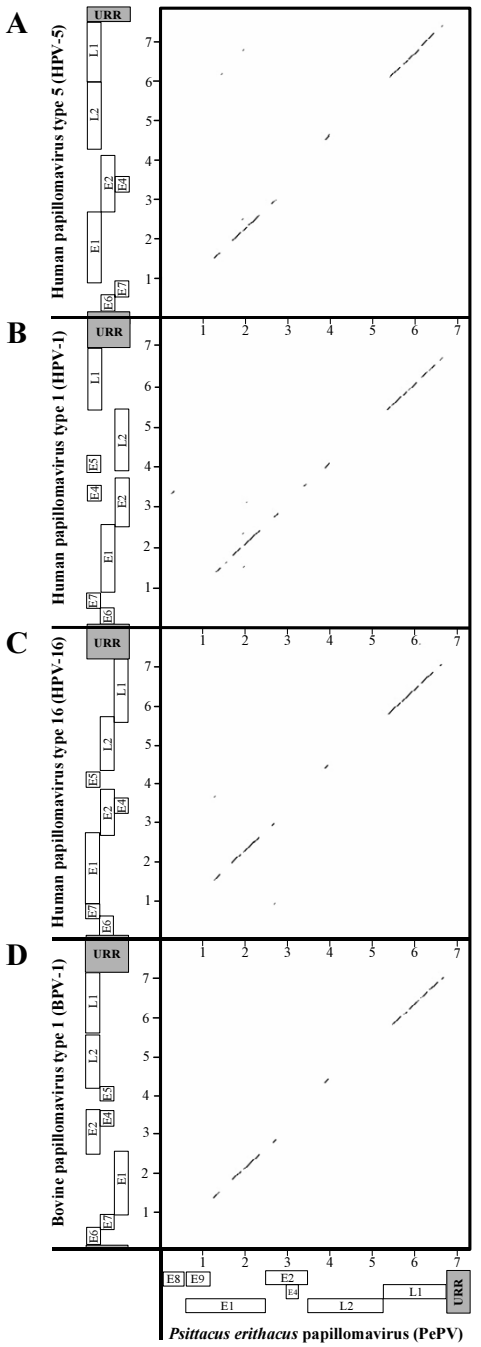

Figure 3
Dot plot matrix (Maizel-Lenk plot) aligning PePV with HPV-5 (A), HPV-1 (B), HPV-16 (C), and BPV-1 (D),

each other 44 to 57 Myr ago. Since these calculations were based on alignments of the most conserved regions in the papillomaviral genome, this calculation is likely an underestimation of the true divergence time. This would place the PePV/FPV divergence at about the same time that their *Psittaciformes* and *Passeriformes* host species were diverging from each in the Late Cretaceous or Early Tertiary period. The high level of congruence between divergences in the papillomavirus phylogenetic tree and the divergence of their host species lineages supports co-speciation. Co-speciation was also hypothesized to be a prominent feature in mammalian and avian herpesvirus evolution [27].

Papillomaviruses in inbred species: emerging infectious pathogens?

Although African grey parrots are not yet officially listed as an endangered species, their survival is threatened by the same factors that most other exotic parrot species face, and they are monitored by the World Wildlife Fund for conservation concerns. In nature, the range of the African Grey parrot extends from Guinea-Bissau and Sierra Leone to southern Cameroon, Congo, Uganda, nortwestern Tanzania, and southwestern Kenya. Human-caused destruction and fragmentation of their habitats in the tropical rain forests cause an increase in inbreeding and reduced heterozygosity. The illegal parrot smuggling trade also causes the natural population size to decrease, and the ensuing import restrictions lead animal breeders to a higher degree of inbreeding in the captive population. Reduced diversity in the major histocompatibility complex (MHC) genes due to inbreeding and genetic bottlenecks may contribute to an increased sensitivity to emerging infectious pathogens, as has been observed in exotic felids [28]. Whenever a population goes through a demographic and genetic reduction, papillomaviruses seem to become more prevalent. This has been documented in endangered exotic felid species, such as the snow leopard, where papillomaviruses are causing an increasing number of cutaneous squamous cell carcinomas [29]. Also the Florida manatee, one of the most endangered marine mammals in American coastal waters, currently suffers from an epidemic of viral papillomatosis [30]. We have previously described this phenomenon in pygmy chimpanzees (*Pan paniscus*) and in Greenlandic Inuits and Navajo Indians, where species-specific papillomaviruses (PpPV and HPV-13, respectively) cause oral focal epithelial hyperplasia, a disease rarely encountered in non-inbred populations [31,32].

Papillomaviruses are ancient viruses that infect amniotes

The amniotes (*Amniota*) are a clade that includes the dinosaurs and most of the extant land-dwelling vertebrates, namely mammals, birds and reptiles. They evolved 360 to 286 Myr ago during the Carboniferous Period in the late

Table 1: Position of the open reading frames of PePV.

	Nucleotide position					
ORF	**Start ORF**	**First ATG**	**Stop codon**	**Bases[1]**	**Amino acids[2]**	**MW(kDa)[3]**
E8	7288	1	532	549	177	19.6
E9	637	646	1231	594	195	22.7
E1	471	534	2568	2097	678	76.8
E2	2509	2521	3637	1128	372	42.1
E4	3218	No ATG	3371	153	51	5.9
L2	3318	3642	5283	1665	547	58.8
L1	5264	5282	6842	1578	520	57.3

[1] Number of bases from the first base of the ORF until the last base of the last codon before the stop codon. [2] Number of amino acids which would produce the predicted proteins if translation starts at the first ATG of the ORF, except for E4 which does not have a start codon and therefore hypothetically begins at the first in-frame amino acid. [3] Predicted molecular weight (in kilodaltons) of the putative protein.

Table 2: Percentage nucleotide (amino acid) similarity of the different PePV ORFs with corresponding ORFs of HPV-1, HPV-5, HPV-16, and BPV-1.

PePV ORF	BPV-1	HPV-1	HPV-5	HPV-16
E1	38 (32)	38 (32)	39 (31)	41 (33)
E2	28 (26)	39 (34)	31 (25)	34 (27)
E4	NA[1]	NA[1]	NA[1]	NA[1]
L2	28 (27)	30 (30)	29 (28)	30 (27)
L1	44 (39)	46 (42)	44 (45)	47 (41)

[1] NA: Not alignable because of insufficient similarity

Paleozoic Era. Papillomaviruses have currently been characterized in more than 20 mammalian species, and in 2 avian species. Papillomavirus particles have also been described in a a Bolivian side-neck turtle reptile species [33]. Since papillomaviruses have been described in mammals, birds and reptiles, and were never found in amphibians or fish, it is tempting to speculate that the host-specificity of papillomaviruses would encompass the amniotes. This means that species-specific papillomaviruses could potentially infect more than 20,000 living species, living in virtually every habitat of the planet. We know that papillomaviruses have been detected throughout the world, even in non-gregarious hosts. This wide geographic distribution cannot be attributed to transmission as an airborne infection (with the possible exception of pulmonary fibromatosis in European elks) [34], since transmission of papillomaviruses requires close direct cutaneous or mucosal contact. Together with the viral species-specificity and the genomic stability of their double-stranded DNA, this requirement for close physical contact makes it unlikely that interspecies transmission in recent history can account for the global presence of a spectrum of papillomaviruses in many amniotes. Assuming that, like humans, all of the more than 20,000 species in the amniotes clade have their own set of species-specific genotypes (humans have more than 100 HPV genotypes), papillomaviruses could be the oldest, largest, and most diverse viral family.

Materials and methods

DNA sequencing

The PePV genome was cloned in the *Sal*I restriction enzyme site of pBR322 [11]. Subclones were prepared by partial digestion of the PePV-insert with *Sau*3AI. The *Sau*3AI restriction fragments were ligated with dephosphorylated *Bam*HI-cut pUC19. After transformation of MAX Efficiency DH5α *E. coli* (Life Technologies/Invitrogen, Carlsbad, CA), the bacteria were incubated for blue-

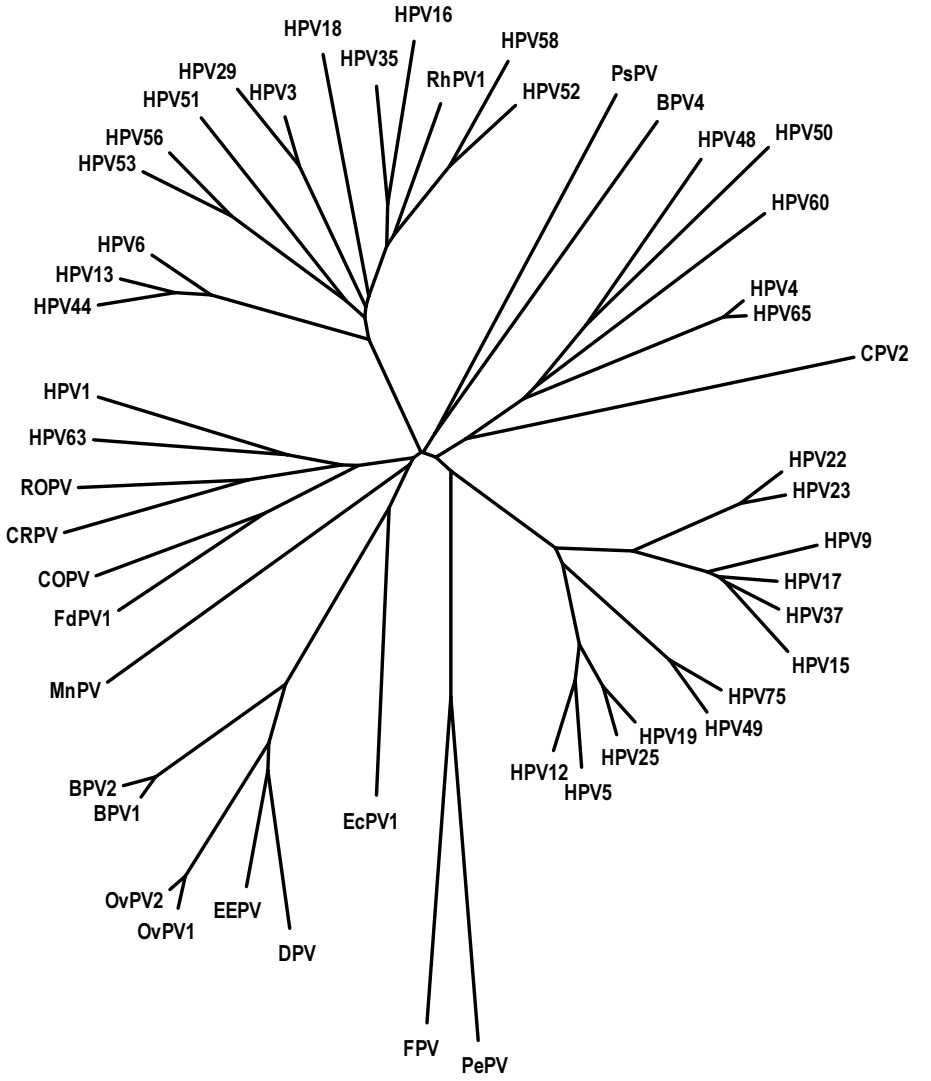

Figure 4
Phylogenetic analysis of a 312 amino acid alignment (132 E1 and 180 L1 residues, corresponding to nt. 2015–2410 in HPV-1 E1, and nt. 6292–6831 in HPV-1 L1) of 50 human and animal papillomaviruses. Papillomaviruses included (with their GenBank accession numbers) were FPV (K02019, K02020 and U29669), bovine BPV 1 (NC_001522), BPV 2 (NC_001521), BPV4 (X05817), canine oral COPV (NC_001619), cottontail rabbit CRPV (NC_001541), deer DPV (NC_001523), *Equus caballus* EcPV (AF498323), European elk EEPV (NC_001524), *Felis domesticus* FdPV 1 (AF480454), HPV 1 (NC_001356), HPV 3 (NC_001588), HPV 4 (NC_001457), HPV 5 (NC_001531), HPV 6 (NC_000904), HPV 9 (NC_001596), HPV 12 (NC_001577), HPV 13 (NC_001349), HPV 15 (NC_001579), HPV 16 (NC_001526), HPV 17(NC_001580), HPV 18 (NC_001357), HPV 19 (NC_001581), HPV 22 (NC_001681), HPV 23 (NC_001682), HPV 25 (NC_001582), HPV 29 (NC_001685), HPV 35 (X74477), HPV 37 (NC_001687), HPV 44 (NC_001689), HPV 48 (NC_001690), HPV 49 (NC_001591), HPV 50 (NC_001691), HPV 51 (NC_001533), HPV 52 (NC_001592), HPV 53 (NC_001593), HPV 56 (NC_001594), HPV 58 (NC_001443), HPV 60 (NC_001693), HPV 63 (NC_001458), HPV 65 (NC_001459), HPV 68 (X67161), HPV 75 (Y15173), *Mastomys natalensis* MnPV (NC_001605), Ovine OvPV 1 (NC_001789), OvPV 2 (NC_001790), *Psittacus erithacus* PePV (AF502599), Canine papillomavirus type 2 CPV2 (unpublished), *Phocoena spinipinnis* PsPV 1 (NC_003348), Rhesus RhPV 1 (NC_001678), and rabbit oral ROPV (NC_002232).

white colony screening on agar plates containing X-gal and IPTG. Ten *Sau*3AI-subclones with a PePV-insert ranging in size from 250 to 1900 basepairs were investigated. Plasmid DNA was extracted using the QIAGEN Midi Plasmid Purification Kit (QIAGEN, Hilden, Germany). Nucleotide sequencing was started using pBR322-specific primers or the universal primers in the multiple cloning site of pUC19. Primer walking sequencing was performed, using 59 sequencing primers to cover the complete genome on both strands. Sequencing was performed on an ABI Prism 310 Genetic Analyzer (Perkin-Elmer Applied Biosystems, Foster City, CA, USA) at the Leuven and Prague core DNA sequencing facilities. Chromatrogram sequencing files were inspected with Chromas 2.2 (Technelysium, Helensvale, Australia), and contigs were prepared using SeqMan II (DNASTAR, Madison, WI).

DNA sequence submission

The nucleotide sequence data reported in this paper were deposited in GenBank using the National Center for Biotechnology Information (NCBI, Bethesda, MD) BankIt v3.0 submission tool [http://www3.ncbi.nlm.nih.gov/BankIt/] under accession number AF502599.

DNA and protein sequence analysis

DNA and protein similarity searches were performed using the NCBI WWW-BLAST (Basic local alignment search tool) server on GenBank DNA database release 118.0 [35]. Molecular weight of the putative proteins was calculated using the Molecular Biology Shortcuts (MBS) ProtCALC program [http://www.justbio.com/protcalc/]. The subcellular location of proteins was predicted using the PSORT II server at the National Institute for Basic Biology in Okazaki, Japan [http://psort.nibb.ac.jp][15,16]. Protein motif searches were performed using SMART (Simple Modular Architecture Research Tool) at the European Molecular Biology Laboratory (EMBL) in Heidelberg [http://smart.embl-heidelberg.de][36], InterPro (release 4.0) at the European Bioinformatics Institute (EBI) [http://www.ebi.ac.uk/interpro][37], and Prosite (release 17.7) at the proteomics server of ExPASy (Expert Protein Analysis System) of the Swiss Institute of Bioinformatics (SIB) [http://www.expasy.ch/prosite][38]. Pairwise sequence alignments were calculated using the GAP-program on the Sequence Analysis Server at Michigan Technological University [http://genome.cs.mtu.edu/align/align.html]. Maizel-Lenk dot matrix plots were calculated via the DotMatrix module on the 'Molecular Toolkit' server of Colorado State University [http://arbl.cvmbs.colostate.edu/molkit/dnadot/index.html] using a (for PV full genome alignment optimal) window size of 115 nucleotides and a mismatch allowance of 60/115. Multiple sequence alignments were prepared using CLUSTALW [39], and corrected in the GENEDOC alignment editor [40]. Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 2.1 [41].

Acknowledgments

This work was supported by a grant of the Flemish Fund for Scientific Research (FWO-Vlaanderen), Brussels, Belgium, and by grant N° 6077-3 of the Granting Agency of the Ministry of Health of the Czech Republic. Annabel Rector received a fellowship of the University of Leuven.

References

1. Van Ranst M, Tachezy R, Delius H, Burk RD: **Taxonomy of the human papillomaviruses.** *Papillomavirus Report* 1993, **4:**61-65
2. Van Ranst M, Tachezy R, Burk RD: **Human papillomavirus nucleotide sequences: What's in stock ?** *Papillomavirus Report* 1994, **6:**65-75
3. Sundberg JP, Van Ranst M, Burk RD, Jenson AB: **Host range, epitope conservation, and molecular diversity.** *In Human Papillomavirus Infections in Dermatovenereology (Edited by: Von Krogh G, Gross G) Florida, CRC Press* 1996, 47-68
4. Sundberg JP, Van Ranst M, Montali R, Homer BL, Miller WH, Rowland PH, Scott DW, England JJ, Dunstan RW, Mikaelian I, Jenson AB: **Feline papillomas and papillomaviruses.** *Vet Pathol* 2000, **37:**1-10
5. Lina PHC, van Noord MJ, de Groot FG: **Detection of virus in squamous papillomas of the wild bird species Fringilla coelebs.** *J Natl Cancer Inst* 1973, **50:**567-571
6. Osterhaus AD, Ellens DJ, Horzinek MC: **Identification and characterization of a papillomavirus from birds (Fringillidae).** *Intervirology* 1977, **8:**351-359
7. Moreno-Lopez J, Ahola H, Stenlund A, Osterhaus A, Pettersson U: **Genome of an avian papillomavirus.** *J Virol* 1984, **51:**872-875
8. Sironi G, Gallazzi D: **Papillomavirus infection in greenfinches (Carduelis chloris).** *Zentralbl Veterinarmed* 1992, **39:**454-458
9. Dom P, Ducatelle R, Charlier G: **Papillomavirus-like infections in canaries (Serinus canarius).** *Avian Pathol* 1993, **22:**797-803
10. Jacobson ER, Mladinich CR, Clubb S, Sundberg JP, Lancaster WD: **Papilloma-like virus infection in an African gray parrot.** *J Am Vet Med Assoc* 1983, **183:**1307-1308
11. O'Banion MK, Jacobson ER, Sundberg JP: **Molecular cloning and partial characterization of a parrot papillomavirus.** *Intervirology* 1992, **33:**91-96
12. Sundberg JP, Junge RE, O'Banion MK, Basgall EJ, Harrison G, Herron AJ, Shivaprasad HL: **Cloacal papillomas in psittacines.** *Am J Vet Res* 1986, **47:**928-932
13. Latimer KS, Niagro FD, Rakich PM, Campagnoli RP, Ritchie BW, McGee ED: **Investigation of parrot papillomavirus in cloacal and oral papillomas of psittacine birds.** *Vet Clin Path* 1997, **26:**158-163
14. Patel KR, Smith KT, Campo MS: **The nucleotide sequence and genome organization of bovine papillomavirus type 4.** *J Gen Virol* 1987, **68:**2117-2128
15. Reinhardt A, Hubbard T: **Using neural networks for prediction of the subcellular location of proteins.** *Nucleic Acids Res* 1998, **26:**2230-2236
16. Nakai K: **Protein sorting signals and prediction of subcellular localization.** *Adv Protein Chem* 2000, **54:**277-344
17. Eriksson A, Stewart AC, Moreno-Lopez J, Pettersson U: **The genomes of the animal papillomaviruses European elk papillomavirus, deer papillomavirus, and reindeer papillomavirus contain a novel transforming gene (E9) near the early polyadenylation site.** *J Virol* 1994, **68:**8365-8373
18. Jackson ME, Pennie WD, McCaffery RE, Smith KT, Grindlay GJ, Campo MS: **The B subgroup bovine papillomaviruses lack an identifiable E6 open reading frame.** *Mol Carcinog* 1991, **4:**382-387
19. O'Brien V, Campo MS: **BPV-4 E8 transforms NIH3T3 cells, upregulates cyclin A and cyclin A-associated kinase activity and de-regulates expression of the cdk inhibitor p27Kip1.** *Oncogene* 1998, **17:**293-301
20. Chan SY, Delius H, Halpern AL, Bernard HU: **Analysis of genomic sequences of 95 papillomavirus types: uniting typing, phylogeny, and taxonomy.** *J Virol* 1995, **69:**3074-3083
21. Van Ranst M, Kaplan JB, Burk RD: **Phylogenetic classification of human papillomaviruses: correlation with clinical manifestations.** *J Gen Virol* 1992, **73:**2653-2660

22. Van Ranst M, Kaplan JB, Sundberg JP, Burk RD: **Molecular evolution of the human papillomaviruses.** *In Molecular Basis of Virus Evolution (Edited by: Gibbs A, Calisher CH and Garcia-Arenal) Cambridge University Press* 1995
23. Kumar S, Hedges SB: **A molecular timescale for vertebrate evolution.** *Nature* 1998, **392:**917-920
24. Hedges SB, parker PH, Sibley CG, Kumar S: **Continental breakup and the original diversification of birds and mammals.** *Nature* 1996, **381:**226-229
25. Cooper A, Penny D: **Mass survival of birds across the Cretaceous-Tertiary boundary: molecular evidence.** *Nature* 1997, **275:**1109-1113
26. Tachezy R, Duson G, Rector A, Jenson AB, Sundberg JP, Van Ranst M: **Cloning and genomic characterization of Felis domesticus papillomavirus type 1 (FdPV-1).** 2002
27. McGeoch DJ, Dolan A, Ralph AC: **Toward a comprehensive phylogeny for mammalian and avian herpesviruses.** *J Virol* 2000, **74:**10401-10406
28. O'Brien SJ, Yuhki N: **Comparative genome organization of the major histocompatibility complex: lessons from the Felidae.** *Immunol Rev* 1999, **167:**133-144
29. Joslin JO, Garner M, Collins D, Kamaka E, Sinabaldi K, Meleo K, Montali R, Sundberg JP, Jenson AB, Ghim SJ, Davidow B, Hargis AM, West K, Clark T, Haines D: **Viral papilloma and squamous cell carcinomas in snow leopards (Uncia uncia).** *Proc AAZV/IAAAM Conf* 2000, 155-158
30. Bossart GD, Ewing RY, Lowe M, Sweat M, Decker SJ, Walsh CJ, Ghim SJ, Jenson AB: **Viral papillomatosis in Florida manatees (Trichechus manatus latirostris).** *Exp Mol Pathol* 2002, **72:**37-48
31. Van Ranst M, Fuse A, Sobis H, De Meurichy W, Syrjanen SM, Billiau A, Opdenakker G: **A papillomavirus related to HPV type 13 in oral focal epithelial hyperplasia in the pygmy chimpanzee.** *J Oral Pathol Med* 1991, **20:**325-331
32. Van Ranst M, Fuse A, Fiten P, Beuken E, Pfister H, Burk RD, Opdenakker G: **Human papillomavirus type 13 and pygmy chimpanzee papillomavirus type 1: comparison of the genome organizations.** *Virology* 1992, **190:**587-596
33. Jacobson ER, Gaskin JM, Clubb S, Calderwood MB: **Papilloma-like virus infection in Bolivian side-neck turtles.** *J Am Vet Med Assoc* 1982, **181:**1325-1328
34. Moreno-Lopez J, Morner T, Petterson U: **Papillomavirus DNA associated with pulmonary fibromatosis in European elks.** *J Virol* 1986, **57:**1173-1176
35. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403-410
36. Letunic I, Goodstadt L, Dickens NJ, Doerks T, Schultz J, Mott R, Ciccarelli F, Copley RR, Ponting CP, Bork P: **Recent improvements to the SMART domain-based sequence annotation resource.** *Nucleic Acids Res* 2002, **30:**242-244
37. Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M, Bucher P, Cerutti L, Corpet F, Croning MD, Durbin R, Falquet L, Fleischmann W, Gouzy J, Hermjakob H, Hulo N, Jonassen I, Kahn D, Kanapin A, Y Karavidopoulou, Lopez R, Marx B, Mulder NJ, Oinn TM, Pagni M, Servant F, CJ Sigrist, Zdobnov EM: **InterPro: an integrated documentation resource for protein families, domains and functional sites.** *Bioinformatics* 2000, **16:**1145-1150
38. Falquet L, Pagni M, Bucher P, Hulo N, Sigrist CJ, Hofmann K, Bairoch A: **The PROSITE database, its status in 2002.** *Nucleic Acids Res* 2002, **30:**235-238
39. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22:**4673-4680
40. Nicholas KB, Nicholas HB, Deerfield DW: **GeneDoc: Analysis and Visualization of Genetic Variation.** *Embnet.news* 1997, **4:**14
41. Kumar S, Tamura K, Jakobsen IB, Nei M: **MEGA2: molecular evolutionary genetics analysis software.** *Bioinformatics* 2001, **17:**1244-1245