VIRUS GENES 1:2, 123-133, 1988

Genome of the European Elk Papillomavirus (EEPV)

ANNIKA ERIKSSON,[1] HARRI AHOLA,[1,2] ULF PETTERSSON,[2] AND JORGE MORENO-LOPEZ[1]

[1]*Department of Veterinary Microbiology (Virology), Swedish University of Agriculture Sciences, Biomedical Center, Box 585, S-751 23 Uppsala, Sweden*

[2]*Department of Medical Genetics, Uppsala University, Biomedical Center, Box 589, S-751 23 Uppsala, Sweden*

Key words: papilomavirus, European elk, EEPV genome organization, Sequences, L3 ORF

Abstract

The genome of the European elk papillomavirus (EEPV) was found to be 8,095 base pairs (bp) long and its genetic organization was similar to that of other papillomaviruses. Ten open reading frames (ORFs), designated E1-E7 and L1-L3, were identified in the genome, all located on one strand. The presence of the L3 ORF is rare among the papillomaviruses and to date has only been identified in the genomes of EEPV, the deer papillomavirus (DPV) and the Cottontail papillomavirus (CRPV). The ORF is well conserved between DPV and EEPV with regard to both length and sequence. Potential promoter regions were identified at the 5′-end of the E6 ORF, at the 3′-end of the E1 ORF and downstream of the L1 ORF. Furthermore, two potential polyadenylation signals were found, one located in the long control region (LCR), downstream of the L1 ORF, and another preceding the L2 ORF. The EEVP genome is closely related to the genome of the DPV, the most highly conserved regions being ORFs E1 (70%), E5 (69%), and L1 (74%).

Introduction

Papillomaviruses belong to the Papovaviridae family and consist of a protein capsid surrounding a circular double-stranded DNA of approximately 8,000 base pairs (bp). They have been isolated and characterized from many different animal species (for references, see 1). The papillomaviruses induce proliferation of either the epithelial layer (papillomas), or the fibroblastic layer (fibromas), or both these layers (fibropapillomas). The majority of the papillomaviruses cause papillomas, but a few also cause fibromas or fibropapillomas.

We have focused our attention on viruses from the two latter groups. Warts induced by the European elk papillomavirus (EEPV) and the deer papillomavirus (DPV) are mainly fibromas, while the reindeer papillomavirus (RPV) and the bovine papillomavirus types 1 and 2 (BPV-1 and 2) give rise to fibropapillomas (2-5). EEPV also causes benign lung fibromas in elks (9). All of these viruses transform cultured murine fibroblast cells and induce sacromas in hamsters (2-4,6-8). Moreover, their genomes are more closely related to one another than to other papillomaviruses. Under stringent hybridization conditions EEPV DNA hybridizes to the genomes of DPV and RPV, but not with the BPV-1 genome that is more distantly related to EEPV (4).

The biological properties of EEPV have been described previously (2) and recently the nucleotide sequence of the early (transforming) region of EEPV has been published (10). We report here the complete nucleotide sequence of the EEPV genome. A molecular analysis of the late region made possible a comparison of regions which are equivalent to previously characterized regions in BPV-1, BPV-2, DPV and human papillomavirus type 6 and 16 (HPV-6 and 16) (11-16).

EEPV was isolated from the epithelial layer of a cutaneous fibropapilloma. The viral DNA was cloned into the plasmid vector pBR322 and subclones were constructed to facilitate sequence analysis as described before (2,10). The nucleotide sequence, obtained by the technique of Maxam and Gilbert (17), was analyzed using the University of Wisconsin Genetic Computer Group (UWGCG) software (18). The strategy for sequencing the late region is shown in Fig. 1 and the complete 8,095 bp nucleotide sequence of EEPV starting with the first nucleotide of the *Hpa*1 recognition site is presented in Fig. 2.

The sequence of EEPV DNA revealed that all open reading frames (ORFs) longer than 400 bp are located on one strand (Fig. 3). Ten ORFs were identified, seven in the early region (E1-E7) and three in the late region (L1-L3) (Table 1 and Fig. 3). Three additional potential ORFs were identified, one 309 bp long (ORFX) located between the E5 and L2 ORFs, and two located in the long control region

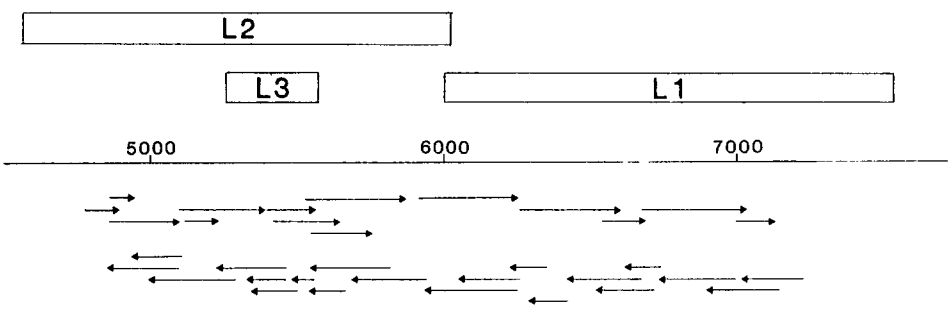

Fig. 1. The strategy for sequencing the late region of EEPV. The arrows show the direction in which the sequence was determined.

```
   1  GTTAACAATC ACCAGATCTT GCCCGTTTTT GTGAGCGGGA AAGCTGGTTA CAGGTTTATA   60
  61  TAAAAAGGCC CACCGCACAA GTTTTCACAG ACGGTTCAGG ATACTTCTAA TACATGCATG  120
 121  TGTGGCGAAT GCTATGCATA CCTCACCTGC ATCTGGTGCA AGAAGGGCTT AGATAAGGTA  180
 181  GATGCAAAGC GATGCCATGA AAAAAAAATA AGAATAGCGT GCAGGAACGG AAAACATTGT  240
 241  GCTGTCTGTA CATCTTGCCT GGAAAATGGG CTGTACCTTG AAAGGTCCCT TTTTCCTGGG  300
 301  CGACCCATCT ACCCTGGAGA CCTGTATGAG CCCGATCCAT GGGTCATGTT CAACGACATT  360
 361  AGATGCATGT ATTGTGGTGG ATGCCTAACC CGCGACGAAA AAGAGAGACA CAGACTGTTT  420
 421  TGTGAAGACT TCTGGATATT CAGGCATCAG GTGCGGGGAC GTTGCTATCT CTGCACCAGG  480
 481  CATGGTTCAC GGCCCCCGTA CAAAGAAACA CCTGCCGCCG TATGAATCAC CTCCCCTCAC  540
 541  ACTGCTCCTA GAGCCAGTTG CTCCGGTGCA ACAGACAGGC ATTCAGGCAC CGCAGAGGAA  600
 601  GCCACCTTCC CAGAAAGGAC ACAAAAAAGG ACACAAGAAA GTTTATTCTG TGACTGTGCC  660
 661  TTGCAATGGA TGTGACAAAA ACCTGGAATT TTGTGCAAGA ACTTCCAGCG CCACCATCTT  720
 721  AACGCTGCAA AACCTCCTGC TGAAAGACCT AGACTTCCTG TGCTCTACCT GCGAGACCAA  780
 781  CCATGGCTGA AACTGCAGGT AGCTCGGGGC AGGGGGGGGG AGCTTATATC TGCTTTGAAG  840
 841  CCGACTGTAG CGACTCTGAT ACAGAGGTTG ATTCACCTGT ACAATGCTCT GATTCAAGTG  900
 901  ATGAGGATCT AGTAGATAAT GCCAATATCG TTCCGGGAAA CCACCTGGAG TTGTTCCAAA  960
 961  CGCAGGAAAA AGAGGCGGGA GAAAGACAGA TTTCGCTTTT GAAAAGAAAA TTCTGTTTGA 1020
1021  GCCCGGGAAC CTCAGAGGTC GAGGAGCTTA GTCCTGGGCT TGCCGGAATC AGAATCTCTC 1080
1081  CGCCAAAGCG AAATCCGGTG GTTAGGAGAA GGCTTTTTGA CGCAGGTGGG AGAGACGCCG 1140
1141  TGCGAACACC GCGTGATCAT GAAGTTAATA GTTCTCCTGA ACCCAGGAGT CAGGTACAGT 1200
1201  CGGGAAGTAG CAGTAGGTCT TGGGAGGGAC ATCTGGAATC CATTAACGAG CCTGCTAGTG 1260
1261  ACGGCAACAT GGCCGCCGTG ATGCACAAGT TGTTCAAGAC TTTGTACATC GCGGGTTTTG 1320
1321  GGGAGATAAC ACGCGTCTTT CAAAGTGATA AAACTAACAA TAATCAGTGG GTGATAGCAG 1380
1381  CCCATGGCGC ATCAGAGGTG CTTTATGCCG CAAGCTTTGA AATACTGAGC AAACACTGCA 1440
1441  GCTACCTGCA GGCGTCTAGG AAGGTGCATG AGACAGGAAG CATGTCTTTG TTCTTAGCTG 1500
1501  TCTTCAATGT TGGGAAGAGT AGGGAGACTG TCAGAAAACT AATTTCAGGT GTCTTAAACA 1560
1561  CCCCGTGTAG CCGCCTACTA TTGCAACCGC CGAAAATTCG TGGACTATGT CCTGCTTTAT 1620
1621  TTTGGTTTAA GTTGGGGCTC TCCCCAGCAA CACAGACGCA CGGTACGACT CCGGACTGGA 1680
1681  TTAAGCAGCA GACCAATGTG GCCTATAATA CTGGGGAGGC CTCTAAATTT GATTTTGGCA 1740
1741  CAATGGTACA GTGGGCATAT GACCACCGGC TAACAGAGGA GTGCAAAATT GCATATCAAT 1800
1801  ATGCAAAATG TGCAGGTACA GACCTAAATG CGAAAGCATT TCTTGCAAGT ACCAATCAGG 1860
1861  CACGGCTGGT CAAGGACTGC TGTACTATGG TGAAACATTA CCTGAGAGCT GAAGAGCAGT 1920
1921  CATTAACCAT TTCTGCTTTT ATTAAAAGGA GATGCGATAA TGCAACTGGA AAAGGCAGTT 1980
1981  GGTTGAGCAT TATGAATCTG TTAAAGTTTC AAGGCATCGA GCCCATTAAC TTTGTAAATG 2040
2041  CCTTGAAACC ATGGCTGAAA GGCACCCCAA AACATAATTG CATAGCAATT GTAGGACCCC 2100
2101  CAAATAGTGG GAAGTCTCTT CTGTGCAATA CCCTCATGTC GTTTCTGGGA GGAAAGGTAC 2160
2161  TGACGTTTGC CAACCACTCC AGCCACTTCT GGTTAGCGCC CCTTACCGAC TGTAGGGTCG 2220
2221  CCTTGATAGA TGATGCCACG CATGCGTGCT GGAGATACTT TGACACATAT CTCAGAAATG 2280
2281  TACTTGACGG TTATCCAGTT TGTATTGACA GAAAGCACAA ATCCGCTGTG CAGCTCAAAG 2340
2341  CCCCTCCCCT TTTGCTAACC AGTAATATTG ATGTGCATGC AGATGAAAAG TATTTCTATC 2400
2401  TGCAAAGTAG AGTCAAAACC TTCTATTTCA AGGAGCCGTG CCCTGCGTCT GATACTGGTG 2460
2461  AGCCCCTTTT CTTTATTACT GATGCTGACT GGAAAAATTT TTTTGAAAGG CTATGGGAGC 2520
2521  GATTAGATCT CAGCGACCAA GAGGACGAGG TTGATGAAGA TGAGTGCAGC CAGCGATCAT 2580
2581  TTACTTGCAG CGCAAGAAAC ACAGATGCAA TGCATTGAGA AAGATAGTCG CCTGTTACAG 2640
2641  GATCATGCAT GCTATTGGGG GGCAGTAAGA AGGGAAAAAC TGTTATTATA TGCAGCGAGA 2700
2701  ACAAAGGGGT TAAAAACAAT TGGGTGTGTG CCTGTGCCTC CTTGTTCTGT TACTGCAGAG 2760
2761  CAAGCGAAGC AAGCAATATG CATGCAATTG ATTGTGGAGG AATTACTGCA CAGTCCATGG 2820
2821  GCCAAAGAAC CATGGTCCCT TACAGACCTA AGCTGGGAGA GATATCAGGC TGCCCCAAAA 2880
2881  GGGTGTTTGA AAAAAGGCGC CAGAGTGGTG GAAGTGGAGT ATGATGGGAA CTCTTCTAAT 2940
2941  AAGACTTGGT ATACAGCTTG GAGTACAGTG TACGTGCGCG GAACGGAAGA GGAGGGCTGG 3000
3001  GAGACTGCTG TCTGTGCTGC AGACGGACAG GGCATTTATT ATTGCGCCGG GATGAGCAGT 3060
3061  AAGGTGTACT TTGAAACCTT TGAAACTGAT GCCCGCAGAT GGAGCAGGAC GGGGCACTGG 3120
3121  ACTGTGAGGG ATAACGATGT GATATATCAT TCAACCTTTG GTGCACCCCC TCACTCTAGA 3180
3181  AACGACAGAG ACTGCATCGA AGGATTCTGG AGCGACGCCG GGGAGCGTAG AGGCTCGAGA 3240
3241  GGGTCCGACA CAACCGACAG AGCCCTGCCT TACCCTGCTG CTCGACAATC CCCCATTTGT 3300
3301  CGCCCCGTCA GAACTGGCGA AAACCGGAGT CGGGCCGTTC ACCGCCAGGC TCCCTACAGC 3360
3361  GCACCATCAT CCCCGGGGAG TTCCGTGGGC CCCGATTCCC CCTCCGAGAG CTCGCGCCAG 3420
3421  GTACCGCTGG TTTTGCTACC AGGACCATCA GATCCAGCGC CGCCGTCGCC GGACTCTACA 3480
3481  GACGTAATCG CAGAGGGTGA CAAGGAACCT GAGCGGTTCA GCATTCTCTC AAAACCAGGT 3540
3541  GGGCAGCCAT GTCTGATACT TAGTGGAAAC GGAAACCAAG CTAAGTGCTA TCGTTTCCGC 3600
3601  TGCAAGAGAT ATTTCAGAGA ACACTATCAG CACATAACGA CCACCTGGTG GACTGTAGGA 3660
3661  GAGCGAGGAT CTGAAAGGCA CGGAGATGCC TGTGTGCTGG TGACATTCAA AGACAGTTCC 3720
3721  CAGAGAGGGG TGTTTTTGAA GCGAGTGCCT TTGCCACCTG GAATGCGCGC GCAGGCACTT 3780
3781  ACAATGATTG CGGACTTTTG AGAATGTGAC TCTGGTAGCC GCAATACTGT GCAGTTTGCT 3840
3841  CACCTTCATG TTTTACCTGT GTAGCACAGA CTGCTGCATG CTGTGATGAC ATACGGTTTG 3900
```

Fig. 2. The complete 8095 nucleotide sequence of EEPV, starting with the first nucleotide in the *Hpa*1 recognition sequence. The sequence data for the early (transforming) region is from Ahola et al. (10).

```
3901  CTTTTGTTTC TGGGGCTCAC ATTTGGACTA CAGCTGATGC TACTTGTCTT TCTGCTGTTT  3960
3961  TTCTTTCTCG TATGGTGGGA CCAGTTTGGC TGCCGTTGTG AAAACATGCA GTTGTAAATA  4020
4021  GTGTATATTG AGGTGTAGAT ATTCATTTGA TCCTGTACAT ACATTTTCTC TATTTTTTTA  4080
4081  AAAAATGCTG GTTGATAAAC ATACATAGGT CACAAACAGG TCATTCCATA ACGTACAACA  4140
4141  TACTTTTTAT AGGCGTGGAG TCTCTTATCT CTGTCCCTTT GTTTCTCTCT TTTGAACTGT  4200
4201  TATCCGAGTC AGCAAGTGCC ATTTTTGTTG CCATTTTCTT CAAGTGCCAA TCTCCTAACT  4260
4261  GCATCCCCAA GAGCTGGTGT AACGTCAGAA GAACAGTCGA GAATCTAGCC TTTGAAAGAC  4320
4321  TTTCAACCCT CAAACAGACC ATGTGCATAC CTCCTGTAAC AGCAAAGCTG CCACGACATT  4380
4381  GAAGCCTGCA TGAAATATTG TTTACTATTG TTTTTGCTGC TATTGCTGGG GCAGTGGAAT  4440
4441  CCAATGTGGG TATTGTTACT AATTGTCTGG TTGTCCATTG TACTTTTATT TGAGTTGCAT  4500
4501  TTTGTGGAAC ACTTCACATG ACAGGCTGCA CTGCAGCGCC TTCATCTCAT CCCTAAATTT  4560
4561  AATAAACCTT CCCCTATTTA ACCCCTACCA TGGCGCCTCG GCGAGTAAAG CGTGCAAATG  4620
4621  TCTATGACCT GTATCGCACA TGCAAGCAGG CAGGCACCTG TCCTCCGGAT GTGATACCTA  4680
4681  AGGTGGAAGG GAAGACGATA GCAGACAAGA TATTGCAGTA TGGAAGCATG GGCGTTTATT  4740
4741  TAGGCGGCCT AGGCATTGGA ACAGGTTCTG GAAAGCCAGG AACAGGAGGC TATATTCCAC  4800
4801  TCAGAGGTGG GGGCTCTACC ACTTCACTAT CAAGCAAACC TTTTGCTGGG GGGATACCCT  4860
4861  TAGAAACCTT AGAAGGGATA GGGGCATTCC GGCCTGGCAT AGTGGAAGAT GCAGGGCCTG  4920
4921  CTTTAGAAGG CATTCTTCCT GACGCACCAG CAGTTGTCAC TCCTGAGGCA GTGCCAGTGG  4980
4981  ATGAGGGGTT AAGTGGGCTA GATATTTCCA GGGAATTAAG CCAGGAACAA ATTCTCAGCT  5040
5041  TTCTCCACCC TGAGGGTCCG GATGATATTG CAGTACTTGA GGTAAGGCCA ACAGAACATG  5100
5101  ATCAGGCACA TTTGCTGTCT ACAAGCACAC ACCCAAATCC ACTGTTTCAG GGTCCTGTAC  5160
5161  AGCAGGCACG AATTATTGCA GAAACATCTG GTGCAGAAAA CGTTTTTGTG GGTGGAAGTG  5220
5221  GCATTGGAAG CAATGCAGGA GAGGACATTG AACTGACACT GTTTGCTGAA CCAAGGACAA  5280
5281  GTACACCTGA GGTGCCTATT AAACGTTCTC GGGGCATTTT CAATTGGTTT AGCAGGCGCT  5340
5341  ACTATACACA GGTACCTGTC GAAGACCCAG ACGAGATTGC TGCTGCAGGC TCGTATGTCT  5400
5401  TTGAGAATCC TGTATACGAT TCAAAGGCGT TCAAACCTGC GCAGCAGCCG GACATTACTC  5460
5461  TACAGGATGA AGCTTCTGTC ACTGGGCGGG ACGCTGCAAG ATTGCTGGCG GGACCCTCGG  5520
5521  GCAGGATTGG GTGGAGTCGT ATCACACGAC CCACTAGTCT TGGAACACGT AGTGGCGTGA  5580
5581  GGGTAGGCCC TCTTTATCAT TTACGATCCT CTTTCAGCAC TATCCATAGT CCTGAGACAA  5640
5641  TAGAGCTAAT ACCCACAGTA CTTGAGGATG ATACTGAGGT GCTTACAGGT GTTCCTGAGA  5700
5701  GAGACACTGG TTTTGATGAT GTGGATTTGG ACAGTATAGC AAGTGACAGT CCATTACTAC  5760
5761  CTGAGCGGCA TCACCTTGCT TTTGGAGCAA GGCGGTCTCA CATTCCAATT GTGGCACGAC  5820
5821  CAGGTGTTCA AACTGGTACA GTGATTGATA CACGTCAGAT GGCTGAAAAC TCTGTTTACG  5880
5881  TGTCGGACAA TGGAGGACAG GAGTCACAGC AGACGCCCAC TGTGGTAATC AATGGCAACA  5940
5941  TTAATGTGTC CATGGAATAT TTTAGGCATT ACTATTTGCA CCCTAGCCTT CTAGGTCGCA  6000
6001  AACGAAAACG TCTATTCGGT TAATGTTTTA CAGATGGCGT TCTGGCAGCC TAGTCAAAGG  6060
6061  CTATACCTGC CTCCCACACC TGTGACAAAG GTGCTGTGCT CTGAGCAATA TATTAGGCGT  6120
6121  AAGGACGTAT TTTATCACGG GGAGACGGAG CGCATGCTCA CTGTAGGGCA TCCATATTAT  6180
6181  GAAATTAAAC AATCAGGGTC TGGGAAAACC ATTCCAAAGG TTTCACCTAA TCAATATCGT  6240
6241  GTTTTTCGGA TCTTACTGCC GGATCCCAAC CAGTTTGCTC TTCCAGATAA AGCCATGTAT  6300
6301  GACCCAAGTA AGGAAAGGCT AGTCTGGGCT GTTGTGGGGG TACAGGTGTC TAGAGGACAA  6360
6361  CCTTTAGGTG GCTCTGTTTC AGGACATTCG TATCAGAACA CTCTGATTGA TGCGGAGAAT  6420
6421  GTTAGTAAAA AGGTAAATGC ACAGGGCACA GATGACAGGA AGCAGGGAGG CATGGACGTC  6480
6481  AAGCAACAGC AAATTCTACT GCTAGGATGC ACCCCAGCTA TTGGTGAGTA TTGGACAACT  6540
6541  GCTAGGCCCT GCGTTACAGA TAGGCCAGAG ACTGGCTCCT GCCCCCCTAT AGAACTAAAA  6600
6601  AACAAACCTA TAGAAGATGG TGATATGATG GATATTGGCT TTGGTGCAGC TAATTTCAAA  6660
6661  GAGTTAAATG CCACAAAGTC AGATCTCCCT TTAGATATTG CAAAAGATAT TTGTTTGTAT  6720
6721  CCTGATTATT TAAAGATGAC TGAAGAAGCG GCTGGCAACA GTATGTTTTT TTTTGCTCGG  6780
6781  AAAGAACAAG TTTATGTTCG CCACATCTGG TCGCGTGGGG GTACCGACAA AGAAATGCCT  6840
6841  CCAGAGGCAT ACTTTCTGAA GCCAAAGGGT GGGGACCAAA CACAGAAAAT GCCTAGTATT  6900
6901  CTTTTTGGAG TGCCAAGTGG CAGTTTAGTT TCTACAGATG GACAATTGTT TAATAGACCT  6960
6961  TACTGGCTGT TTCGTGCACA GGGCATGAAT AATGGCATAT GCTGGCTTAA TCAACTGTTT  7020
7021  GTTACTGTTG GTGACAATAC AAGAGGAACC ACATTAACCA TTACGGTGCC TACATCCGGG  7080
7081  TCCCCACTCA CTGAATATGA CACGAGCAAA TTTAATGTTT TTCAAAGGCA TGTTGAAGAA  7140
7141  TATAAGCTTG CCTTTGTATT TCAGCTTTGC TCTGTCACTC TAAGTCCAGA AACCGTCTCA  7200
7201  CATCTCCAGG GGTTAATGCC TTCGATCCTG GAACACTGGG ATATTAACAT GCAGCCTCCT  7260
7261  ACGTCCTCGA TTCTTGAGGA TACTTACAGA TATCTTGAAT CACCTGCTAC TAAATGTGCA  7320
7321  GATAATGTAA CCCCTATGGG ACCTGAAGAT CCCTATGCTG GTTTAAAGTT TTGGGAGGTG  7380
7381  AATCTAAAAG AAAGGTTGTC TCTTGATCTT GATCAATTTC CTCTGGGACG GCGTTTTCTT  7440
7441  GCGCAGCAAG GATTAGGGTG CAGTACTAGA AAGAGGGTTG CACCTGTCCC TAAGGTCACC  7500
7501  GAAAAAAGGA TTGTTAGGAA AAGAAGAAAG GGGAATTAAG GGCATGAAAT CTTAAAAACT  7560
7561  GCTGTGTTTG CTAAATAAAT GCAATTTTTC TTATGTGTCA AGAGTTTATG TGTCATGTCC  7620
7621  TGCTGTTCAG TCCAACTTGC ACCACACCCG GTGCTGGCAT CTGATTAGAC GCAGTGTCAG  7680
7681  CAGCTTTATG AAAAGCAGAC ACTTGGCTAG ACACACAGGC GCCTGGCGCC CTCATCGAAT  7740
7741  TGGCGCACCG CTGGCGTTCG GGATCAAATT TCCTCTACCG CTGCCGGTTG TTAAAGCGCC  7800
7801  CTTCCTGTAC CGTTCCCGGT AGCGCCTCTT CTCTCCCTTC AGCGCTACCG CTCCCGGTGT  7860
```

Fig. 2. (continued)

```
7861  GCATGGTAAG TAGGCGGTCA TTGTCGAAGA GAACTGGTAA GCAAGTCCGA ACAAGAAAAA   7920
7921  TGCTTGGCGC AACGCTGACG GTAGTCGCTA CCGTCCGCGG TGCTCGCTTT TCTAAGAAAT   7980
7981  GCTCAAACGG TCTCTTGCTA GCTCTGCTCC TATTGGCTGT GCTGAAATTA CTCACGCCGC   8040
8041  TTTGCCTGTA CCGTGAACGG TTTTGAATCC TACTTTTTCT CAGGGAATGA TTGTT  8095
```

Fig. 2. (continued)

(LCR), 263 and 371 bp long respectively (LCR1 and LCR2). The L1 and L2 ORFs in EEPV have their counterparts in all papillomaviruses studied so far, while the L3 ORF has only been identified in the genomes of DPV and the Cottontail rabbit papillomavirus (CRPV) (14,19). The L3 ORF is 321 bp long in both EEPV and DPV, and parts of these two ORFs show a high degree of homology (Fig. 4), whereas no significant homology was found when the L3 ORF of EEPV was compared to that of CRPV. The L1 and L2 ORFs have been shown to code for viral structural proteins (20), but no functional activity has yet been assigned to the L3 ORF.

When compared with other papillomavirus genomes, EEPV shows a higher degree of homology to DPV than to any other papillomaviruses. The ORFs L1 (74%), E1 (70%), and E5 (69%) are best conserved (Table 2a). The extent of homology between the L1 ORF of EEPV and either DPV, BPV-1, BPV-2, HPV-6, or HPV-16 was determined and is shown in Table 2b. The homologies of other ORFs between EEPV, DPV, and BPV-1 are listed in Table 2a. A noteworthy feature of the L1 ORFs in the papillomaviruses sequenced to date is the presence of a conserved postulated splice acceptor site preceding the putative initiator ATG in this ORF

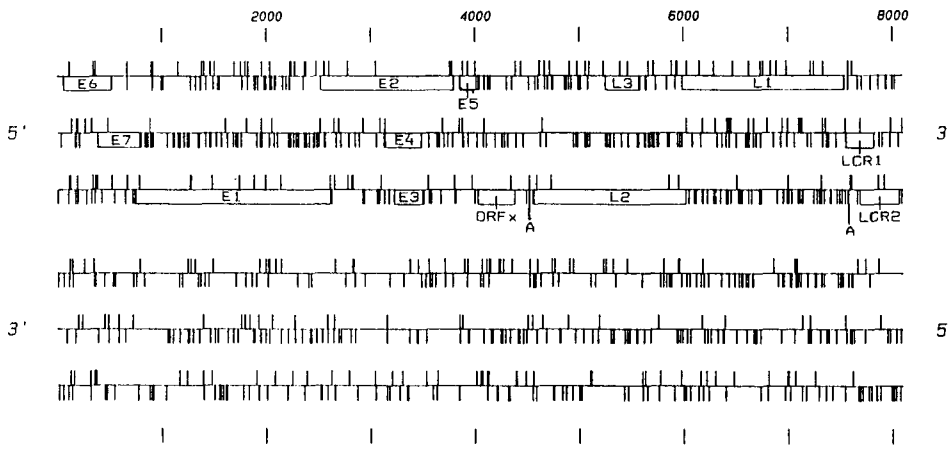

Fig. 3. Genomic organization of EEPV, including all ATG initiation codons (vertical bars above the lines) and termination codons (vertical bars beneath the lines). The ORFs in the genome are boxed. Potential polyadenylatons signals are indicated with A. The sequence data for the early (transforming) region is from Ahola et al. (10).

Table 1. Location and coding capacity of ORFs in the genome of EEPV

ORF	Positions	First ATG	AA* From First ATG	AA*
E1	753–2615	783	611	621
E2	2527–3798	2554	415	424
E3	3231–3497	—	—	89
E4	3143–3484	—	—	114
E5	3865–4014	3886	43	50
E6	64– 522	118	135	153
E7	389– 787	482	102	133
ORFX	4032–4379	4341	13	116
L1	5995–7536	6034	501	514
L2	4566–6020	4590	477	485
L3	5257–5577	5395	61	107
LCR1	7556–7819	7688	88	44
LCR2	7692–8063	7863	124	67

*AA = amino acids

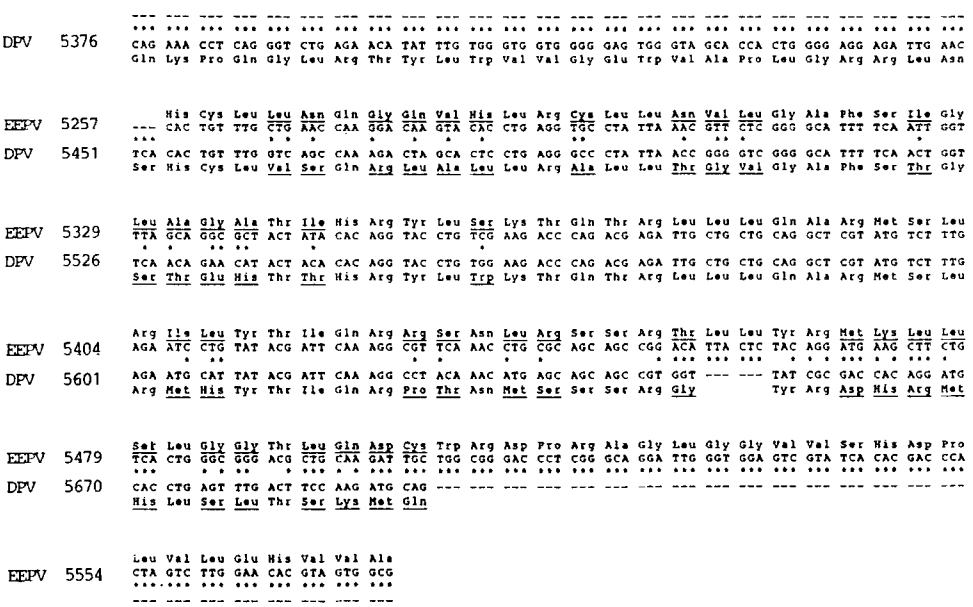

Fig. 4. A comparison between the L3 ORF in DPV and EEPV. Differences at nucleotide sequence level are indicated by asterisks, nonhomologous amino acids are underlined.

Table 2a. Degree of homology at the nucleotide (NT) and amino acid (AA) sequence levels between different ORFs in EEPV, DPV and BPV-1, measured from the first ATG where applicable.

	DPV		BPV-1	
ORF	NT (%)	AA (%)	NT (%)	AA (%)
E1	70	71	57	54
E2	65	56	55	44
E3	—	—	53	28
E4	67	58	52	37
E5	69	72	60	53
E6	58	54	46	43
E7	54	46	46	40
L1	74	73	66	72
L2	60	64	43	37
L3	38	28	—	—

The sequence data was obtained from references 10–12, 14.

Table 2b. The degree of homology between L1 ORFs in EEPV, DPV, BPV-1, BPV-2, HPV-6 and HPV-16, measured from the first ATG in each ORF.

L1	NT (%)	AA (%)
DPV	74	73
BPV-1	66	72
BPV-2	66	71
HPV-6	52	46
HPV-16	56	46

The sequence data was obtained from reference 11–16.

(Table 3). This splice acceptor site has indeed been shown to be used in the case of BPV-1 by S1-analysis (22).

The G + C content of the EEPV genome is 47.7%. However, the bases are not evenly distributed in the genome. For example, the E3 and E4 ORFs (bp 3143–3497) have a G + C content of 62%. In the middle of the genome, including the E5 ORF, there is a T-rich area with a T-content of 35% (bp 3865–4566). A similar T-rich area has been observed in DPV DNA (14).

In the genome of BPV-1, promoter regions have been identified near positions 89 (P_{89}), 2443 (P_{2443}), and 3080 (P_{3080}) (for references see 21). Recently, three additional promoter regions were identified in BPV-1 at positions 7185 (P_{7185}), 7940 (P_{7949}), and within the region 7214–7256, designated the major late promoter (P_L) (21,22). Potential promoter regions were identified in the EEPV genome at positions 87, corresponding to P_{89} in BPV-1 (Fig. 5) and at 2392 corresponding to P_{2443} in BPV-1. A ten base pair repeated sequence (bp 7591–7600 and bp 7606–7615) was identified in EEPV. This has 8 of 10 nucleotides (nt) in common with the

Table 3. Comparison between the nucleotide sequences at the beginning of L1 ORFs in different papillomaviruses (11–16, 19, 27–32).

EEPV	6016	T C G G T T A A T G T T T T A C A G	A T G*
DPV	6237	T T A T C T G A T G T T T T G C A G	A T G
BPV-1	5591	G C C T A A T T T T T T T T G C A G	A T G
BPV-2	1399**	G C C T A A T T T T T T T T G C A G	A T G
CRPV	5810	C T A T C T T T T T A C T T G C A G	A T G
HPV-1a	5413	A T G T A T A A T G T T T T T C A G	A T G
HPV-5	5899	T T G T G A T T T G C A T T C G A G	A T G
HPV-6b	5771	T T C C C T T A T T T T T T T C A G	A T G
HPV-8	5833	T G T G A T T T T G C A T T A C A G	A T G
HPV-11	5753	T T C C C T T A T T T T T T A C A G	A T G
HPV-16	5619	T A C C A T A T T T T T T T T C A G	A T G
HPV-18	5412	C T T T A A C C T C C T C T T G G G	A T G
HPV-33	5576	T T C C A T A T T T T T T T A C A G	A T G

The vertical line preceding the first ATG in the ORFs, indicates potential splice junctions.
*This is the second ATG in the L1 ORF of the HPV-1a, HPV-8 and HPV-16 genomes.
**Base position 1 is the A of the first ATG in the L2 ORF of BPV-2.

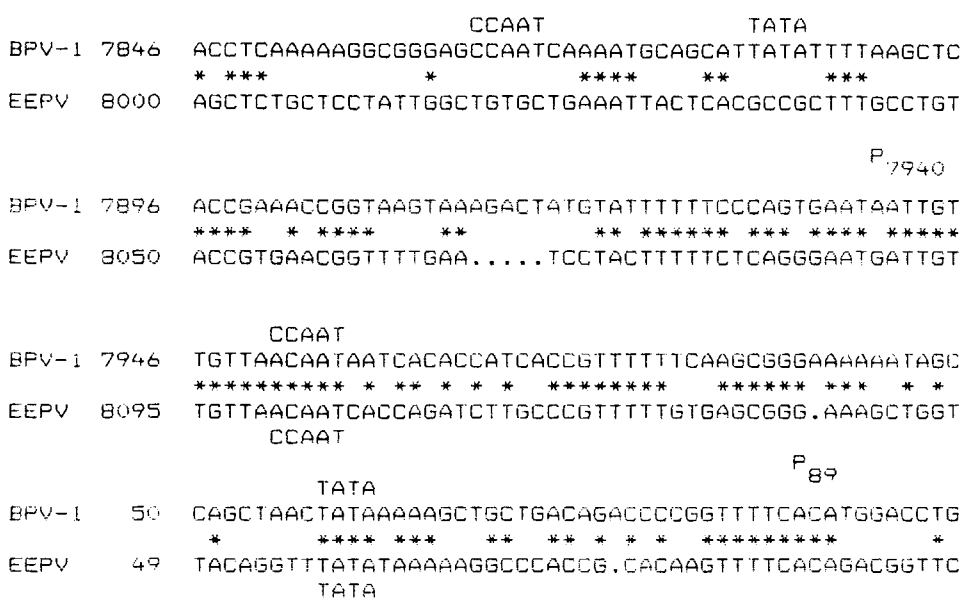

Fig. 5. A sequence comparison between the promoter regions P_{7940} and P_{89} in BPV-1 and the corresponding regions of EEPV.

flanking area of P_{7185} (bp 7180–7189) in the BPV-1 genome. Moreover, a similar repeated sequence is present in DPV (bp 7811–7826 and 7836–7851). Upstream of the P_L region in BPV-1 there is a sequence homologous to the simian virus 40 (SV40) late promoter element GGTACCTAACC that has been shown to be important for efficient utilization of the SV40 major late transcriptional start site (21, 23, 24). It is interesting that this sequence homology is present in EEPV at approximately the corresponding location as in BPV-1, whereas no such sequence is found in this part of the genome in DPV. No area similar to the BPV-1 promoter region P_{3080} has been identified in the genome of EEPV.

Two potential polyadenylation (pA) signals (AATAAA) are present in the EEPV genome (Fig. 3). One is located immediately upstream of the L2 ORF (bp 4561) (putative early pA signal) and the other is located downstream of the L1 ORF (bp 7575) (putative late pA signal). Potential pA signals have been found at the corresponding locations in the genomes of DPV and BPV-1 (11, 14).

The repeated motif $ACCGN_4CGGT$ that has recently been shown to be part of two E2-binding regions in BPV-1 (25, 26), is present in five copies in the LCR of EEPV (Table 4).

Taxonomically, elk and deer belong to the Cervidae family, and they show a similar host response to papillomavirus infections, i.e., mainly fibromas develop. Besides the biological similarities between EEPV and DPV, it is interesting that the molecular analysis indicates a higher degree of evolutionary relationship between them than with the bovine fibropapillomaviruses BPV-1 and BPV-2. On the basis of the sequence homolgy and the characteristic host response to infection, EEPV and DPV could be classed as a subgroup of fibromaviruses to distinguish them from the fibropapillomaviruses.

Acknowledgments

We are grateful to Dr. S. Burnett for valuable comments during the preparation of the manuscript. This study was supported by grants from the Swedish Council for Forestry and Agricultural Research, the Swedish Cancer Society and the Swedish National Board for Technical Development.

Table 4. Putative E2 binding sites in EEPV.

Position	ACCGNNNNCGGT
7777	ACCGCTGCCGGT
7809	ACCGTTCCCGGT
7847	ACCGCTCCCGGT
7950	ACCGTCCGCGGT
8050	ACCGTGAACGGT

References

1. Pfister H. Biology and biochemistry of papillomaviruses, Rev Physiol Biochem Pharmacol, *99*:111-181, 1984.
2. Stenlund A., Moreno-Lopez J., Ahola H., & Pettersson U. European elk papillomavirus; characterization of the genome, induction of tumors in animals, and transformation in vitro, J Virol *48*:370-376, 1983.
3. Groff D.E., Sundberg J.P., & Lancaster W.D. Extrachromosomal deer fibromavirus DNA in deer fibromas and virus-transformed mouse cells, Virology *131*:546-550, 1983.
4. Moreno-Lopez J., Ahola H., Eriksson A., Bergman P., & Pettersson U. Reindeer papillomavirus transforming properties correlate with a highly conserved E5 region, J Virol *61*:3394-3400, 1987.
5. Olson C., Gordon D.E., Robl M.G., & Lee K.P. Oncogenicity of bovine papillomavirus, Arch Environ Health *19*:827-837, 1969.
6. Koller L.D., & Olson C. Attempted transmission of warts from man, cattle, and horses and of deer fibroma, to selected hosts. J Invest Dermatol *58*:366-368, 1972.
7. Dvoretzky I., Shober R., Chattopadhyay S.K., & Lowy D.R. A quantitative in vitro focus assay for bovine papilloma virus. Virology *103*:369-375, 1980.
8. Friedman J.C., Levy J.P., Lasneret J., Thomas M., Boiron M., & Bernard J. Induction de fibromes sous-cutanes chez le hamster dore par inoculation dextraits acellularies de papillomes bovins. C R Acad Sci, Paris, *257*:2328-2331, 1963.
9. Moreno-Lopez J., Mörner T., Pettersson U. Papillomavirus DNA associated with pulmonary fibromatosis in European elk, J Virol *57*:1173-1176, 1986.
10. Ahola H., Bergman P., Ström A.C., Moreno-Lopez J., & Pettersson U. Organization and expression of the transforming region from the European elk papillomavirus (EEPV), Gene *50*:195-205, 1986.
11. Chen E.Y., Howley P.M., Levinson A.D. & Seeburg P.H. The primary structure and genetic organization of the bovine papillomavirus type 1 genome, Nature *299*:529-534, 1982.
12. Ahola H., Stenlund A., Moreno-Lopez J., & Pettersson U. Sequences of bovine papillomavirus type 1 DNA—functional and evolutionary implications, Nucl Acid Res *11*:2639-2650, 1983.
13. Potter H.L.Jr., & Meinke W.J. Nucleotide sequence of bovine papillomavirus type 2 late region, J Gen Virol *66*:187-193, 1985.
14. Groff D.E., & Lancaster W.D. Molecular cloning and nucleotide sequence of deer papillomavirus, J Virol *56*:85-91, 1985.
15. Schwartz A., Durst M., Demankowski C., Lattermann O., Zech R., Wolfsperger E., Suhai S., & zur Hausen H. DNA sequence and genome organization of genital human papillomavirus type 6b, EMBO J *2*:2341-2348, 1983.
16. Seedorf K., Krämmer G., Durst M., Suhai S., & Röwekamp W.G. Human papillomavirus type 16 DNA sequence, Virology *145*:181-185, 1985.
17. Maxam A., & Gilbert W. Sequencing end-labeled DNA with base-specific chemical cleavage, Methods enzymol *65*:499-560, 1980.
18. Devereux J., Haeberli P., & Smithies O. A comprehensive set of sequence analysis programs for the VAX, Nucl Acids Res *12*:387-395, 1984.
19. Giri I., Danos O., Yaniv M. Genomic structure of the cottontail rabbit (Shope) papillomavirus. Proc Natl. Acad Sci USA *82*:1580-1584, 1985.
20. Pilacinski W.P., Glassman D.L., Krzyzek A., Sadowski P.L., & Robbins A.K. Cloning and expression in Escherichia Coli of the bovine papillomavirus L1 and L2 open reading frames. Biotechnology *1*:356-360, 1984.
21. Baker C.C., & Howley P.M. Differential promoter utilization by the bovine papillomavirus in transformed cells and productively infected wart tissues, EMBO J *6*:1027-1035, 1987.
22. Stenlund A., Bream G.L., & Botchan M.R. A promoter with an internal regulatory domain is part of the origin of replication in BPV-1. Science *236*:1666-1671, 1987.

23. Brady J., Radonovich M., Vodkin M., Natarajan V., Thoren M., Das G., Janik J., & Salzman P. Site-specific base substitution and deletion mutations that enhance or suppress transcription of the SV40 major late RNA. Cell *31*:625-633, 1982.
24. Brady J., Radonovich M., Thoren M., Das G., & Salzman N.P. Simian virus 40 major late promoter: an upstream DNA sequence required for efficient in vitro transcription, Mol Cell Biol *4*:133-141, 1984.
25. Spalholz B., Lambert P.F., Yee C.L., & Howley P.M. Bovine papillomavirus transcriptional regulation: localization of the E2-responsive elements of the long control region, J Virol *61*:2128-2137, 1987.
26. Moskaluk C., & Bastia D. The E2 "gene" of bovine papillomavirus encodes an enhancer-binding protein, Proc Natl Acad Sci USA *84*:1215-1218, 1987.
27. Danos O., Katinka M., & Yaniv M. Human papillomavirus 1a complete DNA sequence: a novel type of genome organization among papovaviridae, EMBO J *1*:231-236, 1982.
28. Zachow K.R., Sherrow R.S., & Faras A.J. Nucleotide sequence and genome organization of human papillomavirus type 5, Virology *158*:251-254, 1987.
29. Fuchs P.G., Iftner T., Weninger J., & Pfister H. Epidermodysplasia verruciformis-associated human papillomavirus 8: genomic sequence and comparative analysis, J Virol *58*:626-634, 1986.
30. Dartmann K., Schwartz E., Gissman L., & zur Hausen H. The nucleotide sequence and genome organization of human papillomavirus type 11, Virology *151*:124-130, 1986.
31. Cole, S.T. & Danos O. Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome, J Mol Biol *193*:599-608, 1986.
32. Cole S.T., & Streeck R.E. Genome organization and nucleotide sequnce of human papillomavirus type 33, which is associated with cervical cancer, J Virol *58*:991-995, 1986.