Group I Sequences

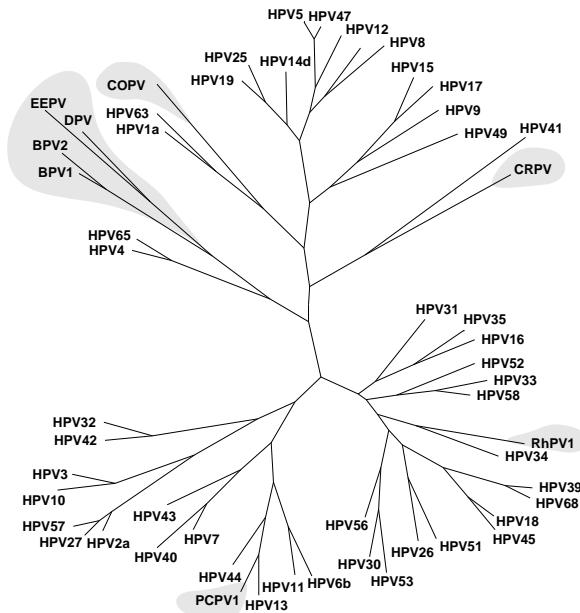|	
COPV	CRPV
EEPV	DPV
BPV1	BPV2
BPV4	PCPV1
MnPV	RhPV

INTRODUCTION

Papillomaviruses infect a wide range of hosts including ungulates, birds, rodents, dogs, marsupials and primates. However, only a subset of these animal papillomavirus types have been cloned and characterized, while an even smaller number have been sequenced over the entire genome. Group I of this compendium currently includes those animal types whose complete genomes have been sequenced: COPV, CRPV, EEPV, DPV, BPV-1, BPV-2, BPV-4, PCPV-1, MnPV, and RhPV. Distinctive clinical pathologies have been associated with the various categories of papillomaviruses. Researchers have linked the ungulate papillomaviruses, EEPV, DPV, BPV-1, BPV-2, and BPV-4, with fibropapillomas, fibrosarcomas, and epithelial neoplasms; either alone or in conjunction with the other lesions. Conversely, the rodent papillomaviruses, CRPV and MnPV, strictly infect cutaneous tissue. Canine oral papillomavirus (COPV) infection occurs mainly in mucosal tissue, although it has been observed in cutaneous lesions in close proximity to mucosal structures. Members of the non-human primate group, RhPV and PCPV-1, cause mucosal infections. Other members of this group cause cutaneous infection; both types of infection parallel human pathology. In addition to these shared pathological characteristics, cases of sequence similarity and homology also exist within the four groups.

Primates

To date, the complete genomes of two primate papillomaviruses have been sequenced: PCPV-1 and RhPV-1. RhPV-1 was isolated and cloned from a metastatic tumor of a rhesus monkey. Ostrow et al. reported a 71% prevalence of RhPV1 DNA within a mating cohort of monkeys including a male with penile carcinoma and subsequent lymph node metastasis [1]. These data indicate that the most likely mode of RhPV transmission is sexual activity. Notable sequence similarity to the sexually transmitted Group A human anogenital papillomaviruses (alignments of RhPV DNA are found with Group A viruses) further substantiates this conclusion. RhPV's oncogenic potential is indicated by its ability to cooperatively transform primary epithelial cells with activated *Ha-ras*, independent of dexamethasone treatment [1]. The physical state of RhPV DNA is integrated with high copy number (roughly 100 copies per cell) [1]. Regulation of the viral genome is likely to be dependent on the glucocorticoid hormones, as the glucocorticoid response element (GRE) sequences of HPV-6, HPV-16, HPV-11 and RhPV differ by only a single nucleotide [1].

PCPV-1, the other primate type which has been completely sequenced, was recently isolated from a pygmy chimp (*Pan paniscus*) afflicted with a high incidence of focal epithelial hyperplasia (FEH) [2]. PCPV-1 hybridized to HPV-13 under stringent conditions; subsequent analysis revealed a similarity of 85% (alignments of PCPV-1 are found with Group B viruses) [2]. The fraction of nucleotide substitutions that represent nonsynonymous changes (amino-acid replacing substitutions)

versus synonymous changes ("silent" substitutions) is remarkably low, suggesting that cross- species transmission may best account for the close similarity to HPV-13 (Part III). Compared to the human genital papillomaviruses, the PCPV-1 genome does not seem to contain a homologous GRE or the homologous splice/acceptor pair potentially used to generate a truncated E6 protein product [2]. Several other primate papillomaviruses have recently been cloned and partially sequenced: a venereal papillomavirus in an Abyssinian colobus monkey, a cutaneous papillomavirus on the feet of a black and white colobus monkey and an Abyssinian colobus monkey, and an oral papillomavirus in *Pan troglodyte*, a monkey related to the pygmy chimp [2].

Rodents

CRPV was the first papillomavirus to be studied in depth. In 1933, Shope et al. isolated CRPV DNA from large horny warts of cottontail rabbits, thus establishing the link between papillomavirus infection and cutaneous papillomas in this animal [3]. CRPV infects epithelial tissue exclusively in both wild and domestic rabbits [3]. The virus has been shown to induce cutaneous papillomas in domestic rabbits under experimental conditions [3]. Malignant progression occurs in up to 25% of infected cottontail rabbits and up to 75% of infected domestic rabbits [3]. Because of its oncogenic potential, CRPV is a potential model for viral-induced multistage transformation, a progression mediated by genetic susceptibility of the host and environmental factors. The most distinctive characteristic of the CRPV genome is the length of the E6 coding region. This coding region is roughly twice as long as any of the E6 proteins sequenced thus far [3].

Only two mouse papillomaviruses have been cloned and sequenced: MnPV and MmPV. In 1978, Muller et al. isolated and cloned MnPV. It was derived from *Mastomys natalensis*, the South African mouse species known to have a very high incidence of both keratoacanthomas and squamous carcinomas [4]. Subsequently, Tan et al. sequenced and published the complete genome in 1994. Unique characteristics of this genome include the unusually large hinge region of the E2 transactivator relative to other papillomavirus genomes and the absence of a homologous E5 coding region [4]. Following the identification of MnPV, the European harvest mouse (*Mastomys minutus*) papillomavirus, MmPV, was isolated and cloned. Viral DNA was suspected to be present in mice diagnosed with a diverse spectrum of skin tumors including papillomas, trichoepitheliomas, and sebaceous carcinomas [5]. Subsequent detection of MmPV DNA in a supercoiled unintegrated form led to the isolation and partial sequencing of this rodent papillomavirus [5]. This partial genome has not been included in this compendium.

Birds

Two avian papillomaviruses have been cloned and characterized: FPV derived from epithelial warts from two chaffinch species (*Fringilla coelebs*) found in different geographical locations [6], one from Sweden and the other from the Netherlands, and PePV derived from a cutaneous lesion on the head of an African grey parrot (*Psittacus erithacus timneh*) [7]. PePV DNA hybridized with European chaffinch papillomavirus DNA at low stringency, indicating that it is a unique avian type [7]. Short portions of the both the PePV and FPV genome have been sequenced [6,7].

Dogs

Canine oral papillomavirus mainly infects the oral cavity, although it has been observed in lesions of conjunctival epithelium, eyelid, and skin around the nose and mouth [8]. In addition, canine papillomavirus has been detected in cutaneous papillomas [8].

Ungulates

Substantial sequence data is available for the ungulate papillomavirus group. Full genomic sequences exist for BPV-1, BPV-2 and BPV-4, EEPV, DPV, and partial genomes for BPV-3, BPV-6 and RPV. The bovine papillomaviruses can be classified into two groups, subgroup A (BPV-1,

BPV-2 and BPV-5) and subgroup B (BPV-3, BPV-4 and BPV-6), based on the tissues they infect [9]. Using this criteria, both DPV and EEPV can be grouped with the Subgroup A viruses. In addition to these characterized types, papillomavirus DNA has been detected in cutaneous fibropapillomas on sheep [10], in papillomas on the udders of goats [10], in cutaneous lesions on pigs [10] and in equine oral, genital, ocular, and cutaneous lesions [11].

The subgroup A viruses infect both dermal fibroblasts and squamous epithelial cells [12]. BPV-1, isolated from a Swedish cow, has been linked specifically to frond-like fibropapillomas which occur on the teats, penis, and nose and equine sarcoida, a benign, naturally occurring fibroblastic tumor in horses [12, 13]. BPV-2 is associated with fibropapillomas of the head, neck, and alimentary canal [12]. "Rice grain" lesions of the teat and the udder are characteristic of BPV-5 [12]. RPV was cloned from a cutanteous fibropapilloma on a Swedish reindeer (*Rangifer tarandus*) [14]. EEPV, isolated from a Swedish wild elk, (*Alces a. alces*) causes fibromas and fibropapillomas [15]. DPV infection is unique compared to the clinical profiles of the other Subgroup A viruses; infection results in fibrobproliferation without epithelial proliferation [16]. Previously called deer fibromavirus, DPV was first isolated from the American white tailed deer (*Odocoileus virginianus*) [16].

The subgroup B viruses are commonly associated with true epithelial papillomas [12]. BPV-3 was isolated from cutaneous epithelial papillomas [9]. Teat-frond epithelial papillomas are characteristic of BPV-6 [9]. BPV-4, cloned from alimentary epithelial papillomas, can progress to malignancy when infected cattle feed on bracken [9]. Because of its oncogenic potential, more research has focused on BPV-4 than any of the other subgroup B viruses.

In addition to differences in host tissue restriction, several other characteristics distinguish the subgroups of the bovine papillomaviruses. First, Subgroup B viruses have smaller genomes (7.2 kB) than subgroup A viruses (7.9 kB). Second, the analogous position of the Subgroup A E6 ORF is occupied by the Subgroup B E8 ORF [9]. This coding region encodes a protein which strongly resembles the E5 transforming protein of the Subgroup A viruses [9].

Many clinical and sequence similarites are prevalent in the ungulate group. EEPV, the bovine subgroup A viruses, RPV and DPV all induce tumors in young hamsters [14,15,16]. These same viruses transform the mouse cell line NIH 3T3 in vitro, whereas all but DPV transform C127 [16]. The E5 regions of BPV-1, DPV, RPV and EEPV are very hydrophobic and both the BPV-1 and EEPV genome exists in an episomal form in the transformed cell in high copy number [15].

BPV-1 Molecular Biology

BPV-1 plasmid replication is dependent upon the expression of 6 or 8 of the early orfs (only E3 and E4 do not appear to play a role) [17]. Molecular regulation is highly complex: thus far, seven promoters, several complicated splice patterns, and eighteen distinct mRNA species have been identified. Six of the seven promoters, P_{89}, P_{890}, P_{2443}, P_{3080}, P_{7185}, and P_{794}, are active in transformed cells [17]. Conversely, the major late promoter, P_{7250}, is active only in differentiating keratinocytes of a fibropapilloma or papilloma [17]. Transcription of the structural proteins origininates at P7250. Multiple interacting elements encoded in the E2 region act to regulate transcription. (The E2 model of transcription regulation is a characteristic feature of papillomaviruses.) Three E2 regulatory proteins have been identified: two transcription repressors, E2-TR and the E2∧E8 fusion product, and the full length E2 transactivator. These proteins bind the motif, $ACCN_6GGT$, which exist in many copies in the genome, particularly in the LCR. E2 responsive elements 1 (E2RE1) and 2 (E2RE2), expression enhancers, are activated by the E2 transactivator [17]. Lambert et al. suggest that the relative abundance of the positive- and negative-acting E2 proteins determines the level of viral gene expression [17].

The capacity of BPV-1 to transform rodent cells in culture has primarily been attributed to the proteins encoded by the E5 and E6 orfs. The putative E5 transformation pathway involves binding of E5 to a 16-kDa cellular protein and the subsequent loss of cell-cycle control [18]. It is possible that the E6 mechanism of transformation may be linked to the alteration of gene expression through nucleic acid-binding [18].

[1] Ostrow, R.S., LaBresh,K.V., and Faras, A.J. Characterization of the complete RhPV 1 genomic sequence and an integration locus from a metastatic tumor. *Virology* **181**: 424–429 (1991)

[2] Van Ranst, M., Fuse, A., Fiten, P., Beuken, E., Pfister, H., Burk, R.D., and Opdenakker, G. Human papillomavirus type 13 and pygmy chimpanzee papillomavirus type 1: comparison of the genome organizations. *Virology* **190**: 587–596 (1992)

[3] Giri, I., Danos, O., and Yaniv, M. Genomic structure of the cottontail rabbit (Shope) papillomavirus. *P.N.A.S.* **82**: 1580–1584 (1985)

[4] Tan, C., Tachezy, R., Van Ranst, M., Chan, S., Bernard, H., and Burk, R. The *Mastomys natalensis* papillomavirus: nucleotide sequence, genome organization, and phylogenetic relationship of a rodent papillomavirus involved in tumorigenesis of cutaneous epithelia. *Virology* **198**: 534–541 (1994)

[5] O'Banion, M.K., Reichmann, M.E., and Sundberg, J.P. Cloning and characterization of a papillomavirus associated with papillomas and carcinomas in the european harvest mouse. *J Virol* **62**: 226–233 (1988)

[6] Moreno-Lopez, J., Ahola, H., Stenlund, A., Osterhaus, A., and Pettersson, U. Genome of an avian papillomavirus. *J Virol* **51**: 872–875 (1984)

[7] O'Banion, M.K., Jacobson, E.R., and Sundberg, J. Molecular cloning and partial characterization of a parrot papillomavirus. *Intervirology* **33**: 91–96 (1992)

[8] Pfister, H., and Meszaros,J. Partial characterization of a canine oral papillomavirus. *Virology* **104**: 243–246 (1980)

[9] Jackson, M.E., Pennie, W.D., McCaffery, R.E., Smith, K.T., Grindlay, G.J., and Campo, M.S. The B subgroup bovine papillomaviruses lack an identifiable E6 open reading frame. *Molecular Carcinogenesis* **4**: 382–387 (1991)

[10] Lancaster, W.D., and Olson, C. Animal papillomaviruses. *Microbiological Reviews* **46**: 191–207 (1982)

[11] O'Banion, M.K., Reichmann, M.E., and Sundberg, J.P. Cloning and characterization of an equine cutaneous papillomavirus. *Virology* **152**: 100–109 (1986)

[12] Coggins, L.W., Ma, J.Q., Slater, A.A., and Campo, M.S. Sequence homologies between bovine papillomavirus genomes mapped by a novel low-stringency heteroduplex method. *Virology* **143**: 603–611 (1985)

[13] Amtmann, E., Muller, H., and Sauer, G. Equine connective tissue tumors contain unintegrated bovine papilloma virus DNA. *J Virol* **35**: 962–964 (1980)

[14] Moreno-Lopez, J., Ahola, H., Eriksson, A., Bergman, P., and Pettersson, U. Reindeer papillomavirus transforming properties correlate with a highly conserved E5 region. *J Virol* **61**: 3394–3400 (1987)

[15] Ahola, H., Bergman, P., Strom, A.C., Moreno-Lopez, J., and Pettersson, U. Organization and expression of the transforming region from the European elk papillomavirus (EEPV). *Gene* **50**:195–205 (1986)

[16] Groff, D.E., and Lancasterk, W.D. Molecular cloning and nucleotide sequence of deer papillomavirus. *J Virol* **56**: 85–91 (1985)

[17] Lambert, P.F., Baker, C.C., and Howley, P.M. The genetics of bovine papillomavirus type 1. *Annu. Rev. Genet.* **22**: 235–258 (1988)

[18] Campo, S. Cell transformation by animal papillomaviruses. *J. Gen. Virol.* **73**: 217-222 (1992)

```
LOCUS       COPV       8607 bp ds-DNA           VRL       01-AUG-1993
DEFINITION  Canine oral papillomavirus (COPV), complete genome.
ACCESSION   L22695
KEYWORDS    complete genome.
SOURCE      Canine oral papillomavirus DNA.
REFERENCE   1  (bases 1 to 8607)
  AUTHORS   Delius,H., Van Ranst,M.A., Jenson,B.A., Zur Hausen,H. and
            Sundberg,J.P.
    TITLE   Canine oral papillomavirus genomic sequence: A unique noncoding
            region between e2 and l2
  JOURNAL   Unpublished (1993)
COMMENT     Canine oral papillomavirus (COPV) infection occurs mainly in
            mucosal tissue, although it has been observed in cutaneous lesions
            in close proximity to mucosal structures.  COPV has also been
            isolated from canine cutaneous papillomas (Pfister and Meszaros
            Vir. 104: 243-246).  Temporary benign lesion arise four to eight
            weeks following infection (Pfister and Meszaros Vir. 104: 243-
            246).  Pfister and Meszaros (Vir. 104: 243-246) suggest that COPV
            may be oncogenic "under conditions of impaired host reactivity."
            The canine oral papillomavirus genome seems to contain an
            insertion relative to other papillomaviruses approximately 1 kbp
            long between the E2 and L2 reading frames.  Under nonstringent
            condition, Pfister and Meszaros (Vir. 104: 243-246) identified
            similar regions between HPV-1 and COPV.  The authors further
            suggest COPV may be similar to BPV-2 as the pattern of "homology"
            between COPV and HPV-1 is analogous to that observed between
            HPV-1 and BPV-2.
BASE COUNT    2630 a   1674 c   1905 g   2398 t
ORIGIN
        1 aaaaggtgtg ttctcttatt gtagctaaca acaatcttac ttacagtaaa attccaagac
 E6 start ->
       61 cgatttcggt cctggcaact gtttcggtcg gtatatatag cATGttttgg ggggcactgt
                                                   E6 cds ->
      121 tatcaatgga gcgcccgacg tcggtgagag atctttgcat gtctctaaag ctctctttgc
      181 ttgatctgtc gcttgcttgc aaattttgtg gcaataatat aacaaatata gaaaagctgc
      241 tttttgataa agctggtttt cagttaatct ggcgagaaa caacgcattt ggatgctgtc
      301 agtactgtgc aagagtctgc agcgttgtgg agcaatgttt tggaagccac agacacttga
      361 cttctgagga gcttgtcaac gtaacaaaga ccttgcagca gcttagtctt agatgtttag
      421 gatgcctcag tattctgagt gaggcggaca aagaactatg tgctgaattg aacgattttt
      481 ctgtggtcag ggggaaaacT AGgggcttgt gttcgctgtg ccgattacca ccATGAttgg
               E7 orf start ->                        E7 cds ->
                                                                  <- E6 end
      541 gcaatgcgca acccttttgg atattgtgct gacagagcag ccggagccga tagacttgca
      601 atgctatgaa caattaccat cgtctgacga ggaggaggag gaggaggagc caactgaaaa
      661 aaatgtttac agaatagagg ctgcctgtgg attttgtggg aaaggggtga ggttttttttg
      721 tctgtctcaa aaagaggatc tgcgtgtgct gcaggtcact ttgcTGAgcc tcagcctggt
                                                  E1 orf start ->
      781 gtgcaccacc tgtgtgcaga ccgccaagct tgaccATGgc ggcTAGaaaa ggtactgact
                                          E1 cds ->           <- E7 end
      841 ctgagactga ggatggtggt tgggtactaa tagaggcaga ttgtagtgag gtagactctg
      901 cagatgaaac cagtgaaaat gcaagtaatg tctctgatct tgtagacaat gcgagcattg
      961 cagaaacaca gggactttcc ctgcaactgt tcaacagca agagctgact gaatgtgaag
     1021 agcagttgca gcagctaaaa cgaaagtttg tacaaagccc gcaatctcgg gatttgtgta
     1081 gccttagtcc gcaattggca agcattagct taacgccacg gacgtctaaa aaggttaaaa
     1141 agcagctgtt tgcaactgat agtggcattc agtccagcaa tgaagctgat gattctcttg
     1201 aggggcaaag acaggtagaa ccgttgccgg gtcgggaaga aaatggcgcc gatgcattgt
     1261 ttaaagtgag ggataagcgc gcctttttgt attcaaaatt taaatcttcg tttggaataa
     1321 gctttacaga tttaactaga gttataata gtgataaaac ttgcagctcg gattgggtag
     1381 tatgtcttta tcatgtatct gatgatagaa gagaggcagg aaaaacatta ttgcaggatc
     1441 attgtgaata ttttttttttg cattcaatgg ggttttgtac tttgttatta ttatgtttgt
     1501 ttgtgcctaa gtgtagaaat actttgttta aattatgtag aagtttattt catataagta
     1561 atgtacagat gttggctgat cctcctaaaa ctagaagtgcc tgcagttgca ttatattggt
     1621 ataaaaaagg gtttgcatca ggtacattta cacacggaga gttgccaagt tggatagctc
     1681 agcagacact aataacacat catttagctg cagagaaaac ctttgatttg agtgagatgg
```

```
1741 ttcagtgggc ttatgataat gatctgaaag acgagtctga aattgcatac aaatatgcag
1801 cattagcaga aacagatgaa aatgctttag cttttttaaa gtctaataac cagcctaaac
1861 atgtaaaaga ctgtgcaaca atgtgcagat attataaaaa agctgaaatg aaaagattaa
1921 gtatgtctca gtggatagac gaaagatgca aggctactga tgatggtcca ggtgattgga
1981 aggaagttgt gaaattttta agacatcaag ggatagaatt tattttgttt ttggcagact
2041 ttaaaagatt tttgagaggt aggcctaaaa aaaattgcct tgtattctgg ggtcctccaa
2101 atacaggcaa gtctatgttt tgcatgagcc tgcttagttt tttgcacgga gtagttattt
2161 catatgtcaa tagcaaaagt cattttttggc tgcaacctct tacagagggg aaaatgggtc
2221 tgttagatga tgccactagg ccttgctggc tctatataga cacttatttg agaaatgctc
2281 tagatggcaa tacatttagt gttgattgca agcacaaagc gcctttgcaa ctaaaatgcc
2341 cgcctctgct gattactact aatgtcaatg tttgtggaga tgaaaaattt aaatatcttc
2401 gcagcagatg ctctttcttt cattttccac aagaatttcc tttggatgac aatggaaatc
2461 ctggctttca gttaaatgac caaagctggg cttctttttt taaaaggttc tggaaacatt
2521 tagattTAAg tgaccctgaa gacgggaag ATGgagaaac tcagcgaggc cttagactta
     E2 orf start ->                      E2 cds ->
2581 ctgcaagagg aactactgag tctgtaTGAg cagaatagcc aaagtcttgc agaccaatca
                                    <- E1 end
2641 aggcactggt cattgctcag aaaagagcaa gtcctacttt attatgccag aggcaagggc
2701 ataatgagga taggcatgca gcctgtgcct ccacagtctg tgtctcaagc caaagctaag
2761 caggccatag agcagtcact ttacatagac agcttgttac actcaaagta tgcaaatgaa
2821 ccgtggacac tatgcgatac aagcagggag aggttggttg cagaacctgc atacaccttc
2881 aaaaaaggtg aaagcagat tgatgtcaga tatggtgaca gtgaggaaaa cattgtcaga
2941 tatgtattgt ggctggatat ctattaccag gatgagtttg cacctgggaa aaagcacat
3001 ggcaagctag atcacaaagg actctcatac atgcatggga ctcagcaggt gtattatgtg
3061 gactttgaag aggaggccaa caaatatagc gagactggga aatatgagat tcTAAaccaa
                                                       E4 orf start ->
3121 cccactacta ttcccaccac cagtgccgcc ggaacctccg gaccggaact ccccggtcac
3181 tccgcctcgg ggtccggtgc ctgttccctt accccaggaa aagggccgtc acggcggcct
3241 ggacggaggt cgtcgcggtt ccccagaagg tcaggaggac gaggaagact cggacgagga
3301 ggaagcggag aattaccccc ccagccgcag ccgtcctcgt cgtggtcgcc gccgtctcca
3361 caacaagtgg gatcaaaaca tcaactacga accaccagca gcgccggagg acgactggga
3421 agacttctgc aagaagctta cgatcccca gttcttgttt TAGctgggga tcctaatagt
                                                       <- E4 end
3481 ttaaaatgca taagatatag attaagtcat aagcataggg ggttatattt gggggccagc
3541 acgacgtgga aatggacatc aggcggggat ggagcatcta agcatgaccg gggcagtgcg
3601 cggatgctgt tagcatttt aagtgatcaa caacgggagg actttatgga cagagtgact
3661 tttcctaagt ctgtgcgagt atttcgggga gggttagatg agttaTAAgg gagggagggg
                a   1-kbp-long insertion relative to      <- E2 end
          other papillomaviruses begins approximately here /\
3721 gggatggtgt ggtgaagggg accaaaaaaa aaaggttaac aggtgtagta aggtgttggg
3781 acatataact gaaagggtgg ttcataagaa tttcagccag agctttaata tttacagagt
3841 caacaaacgt tagtctctta gcctaaaaat ttaagcagtt taaagcataa gcaaacaatc
3901 tttcattatt tttcttttttt tttttttttt tttagtgtgt agacttcaca gaagagggca
3961 aaacaaaaac attgtgtgta gtgttagata acagtcatat tttagctacc atagtcttat
4021 tttttttttct ctgttttgtt ttaagtcacc atttattcac tgcgaaaata ccatcatata
4081 ttcttagcct agctatattc atcagtttta gaccattatc cattagttgc taagttagtg
4141 ttaggtaaca acgcttaggt aaacaacaga aaagatcata gtgtttacaa caacaacaag
4201 atagttaaag tgaacaacga caagtccgac agtccaaccc tactttgtgt ttatttgtct
4261 ctcgtatagt ataaggcata tagtatcagt taagttagaa aatgtgtaaa taagccagat
4321 atatcgatat ccaattaggc aagtgaaact tcaatatgtt tattgcagtc tccaaattta
4381 gtttagtttg atatgttgtc tgattagatt atattgcagt tgggagaata tcaagaacac
4441 accatccaca gacaaatgca ttgacatcgg aacaaaaaga aagaaaagaa gtagaaaatg
4501 ggaatgaaag tttgaatgaa tgctgttaat tttgttccat ttgtacctgt tatagctatc
4561 tctttttttgg tcgatccttt gtgatcctct gtactatact gtgaaaacca gtggaattca
4621 gaaactctac tacttataag ctgtccccc cacctgctct ttcatcatac tgtaaaactg
4681 tttactgttc aactacaata ctgttatact gttttctggg caccttctct ctcactagca
4741 aaacaatggt tcactgaaaa atccagtata ttgtaaatat tgtatataca aagttttgaa
4801 attgttcaca actgtataac ctgtagaaca tctctgcatt tcttcacatg tcttacatag
4861 gataacacac atttgcaaat tgttctgagt tgttgcatac tccataagtt ccttttgaag
4921 ttcctctttt ttttttttgt agaaatcctt aagttttttt atattgtacc ggcagtcaca
4981 gtaggtgttt gaatctagaa taactacatc atacatacca tcaaatcaaa ccatcagtga
5041 aacatcatca gctttaatgt caaaagcaat catagcatta gcactgttgc ataagtccaa
5101 aaactactcg accccccctac ccgcttatct gaccaagagt ctgtgtgcgg acatctgagc
5161 acattgtaaa tacatttaac cgctaaaacc ccgtgcagca aaggaaagac aaaaaaaagt
```

```
5221 ccccgctgtc acaaccgtga cagcaaagaa agacaaaaaa aaaTAAaaaa catccctcgc
                                                    L2 orf start ->
5281 ttacgcaATG gcattgatca ggaaaagacg cgcagcccct caagatatat accctgcttg
     L2 cds ->
5341 taaagtgtcc aacacttgcc ccgctgatat tttgaataaa atggagcaaa atacgcttgc
5401 agataaaatc ctcaaatatg gtagtgctgg tgtttttttg gggggtctag gaatatcaac
5461 aggcaaaggg gtaggggggc gcacaggtta cattcctttg ggaggaacag aaagtggagt
5521 gggtgtaggc acaaggggtca caacaataag acctactgtc ccaataagca gtgtgggggtc
5581 tcctgacttt attcctgtag atgcagtaga ccctctaggg cctgcagtca tacccccaga
5641 aagatttcct atagcagtag aggatccttt tactctacct ccaccacgtt tcccaactgc
5701 agtagaagaa gatgtaattg agctgcagcc tattccaggc ccctcatctg aaatcccact
5761 cgctggccct aagattacca ctgatgctca gcctgcaata ttggaagtca taccagagac
5821 taggcctcct aaagtaatca ctaggcatca gtacagtaac cctgcatttg aggtgtccat
5881 tacttctaac tctggtgcag gggagtcttc tgcttctgac catgtgctgg tggaaggttt
5941 ttctggtggc cactcaattg gtgaacacat tccattgcaa gaccttgcac ccagcaggcc
6001 ttcattctct gagaccatag aagatgaaac tgctttttagc agcagtactc ctaaacaagg
6061 ctctagatct gaaaggccta aaagttacta taataggcga agatatcagc aagtacaagt
6121 tactgaccct gtgtttattt caagaccacg gtcacttgtc acgtttgata acccagcctt
6181 tgatgaatct gttgacctga tatttgaaag agatgttgca gaaataactg cagcacctca
6241 tgcagacttt acagatatca caaagctcac aaagcctgca tatcacagag gcccatctgg
6301 ccatgtccgt gtcagtaggc ttggacatag agctaatata aaaactagaa gtggtcttac
6361 aatagggcca caaagccatt tttactacga tgtcagcagc attgaccctg cagaatcttt
6421 tgagctgcag gcacttggca atgtatccag tgctgaacaa acaggagaag cagtaatctc
6481 ctctggcaca ggagactttg aaattataag ccttgaagac agtattttgg aatcctacaa
6541 tgatgaggat ttaatagacg tgtttgagga tgtagctaga gatttgcatt tattagttgg
6601 agaaagaagg cagcaaccga tccaagttca acgttacata aagcctttt cttttgttaa
6661 tgagggagta cacataattc acccaggatc tgagtcagat tttttggctgc ctcctgtaac
6721 gcctgacagc acacctgcaa tagtgattga cattttggac tcctctgcag attactatct
6781 gcatccaagt ttaaTAAaaa aacgcaaacg caaacatttt ttttttTAAt ttgcagATGg
                L1 orf start ->                                  L1 cds ->
                                                            <- L2 end
6841 cggtttggct tcctgcacag aataaatttt accttccacc acagcccagc accaaggtct
6901 taagcacgga tgaatatgtc tccagaacaa atattttta tcatgctagc agtgaacgtc
6961 ttcttactgt ggggcacccl ttttatgaaa tttataaaga agaacgttct gaagaggtta
7021 tagttcctaa agtatctcct aatcagtacc gggtattccg cttgctactt ccagaccctal
7081 acaattttgc atttggagat aagtcattat ttgatcctga aaaagaaaga cttgtttggg
7141 gcttaagagg attagaaata ggtaggggggc aaccattagg tataagtgtt acgggtcatc
7201 caacatttga cagatacaat gatgtagaaa acccaaacaa aaatcttgct ggacatggag
7261 gtggaacaga cagcagggtt aacatgggtt tagaccctaa acaaactcag atgtttatga
7321 tagggtgcaa accagcttta ggtgaacact ggtctttaac tagatggtgc acaggacagg
7381 tacacactgc aggacaatgt ccaccaatag aactgagaaa cacaacaata gaagatggag
7441 atatggtaga tatagggttt ggtgcaatgg attttaaggc tttgcagcat tataagtcag
7501 gagttccaat tgacatagta aattctgcat gcaaatatcc agactacctc aaaatggcaa
7561 atgagcctta tggagataga tgtttttttt ttgtaagaag agagcaactg tatgccagac
7621 atattatgtc cagatctggc acacaaggtt tagaaccagt ccccaaagat acctatgcaa
7681 caagagaaga caataacata ggaacaacta attacttctc cacacctagt ggctctctgg
7741 tttctagtga gggacaactg tttaacaggc cttactgaat ccagcgctcg cagggcaaga
7801 ataatgggat tgcatgggc aatcagctgt ttttaacagt agtggacaac acacgaggaa
7861 ctcccttaac tataaacata gggcaacaag acaagccaga agaaggaaat tatgttcctt
7921 catcatacag aacctacctc agacatgttg aagaatatga agtaagcata attgtgcagc
7981 tgtgcaaagt taagctgtcc cctgaaaatc tagcaataat tcatactatg gatcctaata
8041 ttattgagga ttggcaccta aatgtcactc ctccatctgg tactttagat gacacatata
8101 ggtacataaa ctctcttgct actaagtgcc ctactaatat aacctccaaaa actaacgttg
8161 atcctttgc agactttaaa ttttgggaag tagatcttaa agataaaatg actgaacagt
8221 tagaccaaac tccactgggt cgcaaatttt tattccagac aaatgtgtta cgtcctagat
8281 ctgtaaaagt acgttctacc tcgcacgttt ctgtcaaacg aaaagctgtg aaacgcaaac
8341 gcaaaTAAtg tgtcattgat tacttgtgaa taaacagata attatttatg tccagttgtt
             <- L1 end
8401 gtggtcattg tttactgact gaccggcacc gcaccctgca catattgcac acagcaccag
8461 caaaggcagg ctaactcaga caagccggca cctgaattaa gcttttaatc tttttaatct
8521 taaaaatccc tttaatcttt tggagcgACC GTTATTGGTt tggagtgacg cccggacatt
                                       -> E2 bind
8581 cctgacaaga ccggattcgt tcgaccg
```

```
LOCUS       CRPV         7868 bp ds-DNA  Circular  VRL      31-AUG-1987
DEFINITION  Cottontail rabbit (Shope) papillomavirus (CRPV), complete genome.
ACCESSION   K02708
KEYWORDS    complete genome.
SOURCE      Cottontail Rabbit Papillomavirus DNA recovered from a
            papilloma of wild cottontail rabbit (provided by G.Orth),
            clone pIG30.
REFERENCE   1  (bases 1 to 7868)
  AUTHORS   Yaniv,M., Danos,O. and Giri,I.
  TITLE     Genomic structure of the cottontail rabbit (Shope) papillomavirus
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 82, 1580-1584 (1985)
COMMENT     CRPV was the first papillomavirus to be studied in depth.  In
            1933, Shope et al. isolated CRPV DNA from large horny warts of
            cottontail rabbits, thus establishing the link between
            papillomavirus infection and cutaneous papillomas in cottontail
            rabbits.  CRPV infects epithelial tissue exclusively in both wild
            and domestic rabbits.  The virus has been shown to induce
            cutaneous papillomas in domestic rabbits under experimental
            conditions.  Malignant progression occurs in up to 25% of infected
            cottontail rabbits and up to 75% of infected domestic rabbits
            (Giri et al. P.N.A.S. 82: 1580-1584).  Because of its oncogenic
            potential, CRPV is a potential model for viral-induced multistage
            transformation, a progression mediated by genetic susceptibility
            of the host and environmental factors.  The most distinctive
            characteristic of the CRPV genome is the length of the E6 coding
            region.  This coding region is roughly twice as long as any of the
            E6 proteins sequenced thus far.
                Compared to BPV-1, HPV1a and HPV-6b, CRPV showed the highest
            sequence similarity to HPV-1a.  This may reflect either a more
            recent evolutionary split between the HPV-1a and the rodent virus
            or a convergence resulting from shared target tissue (Giri et al.
            P.N.A.S. 82: 1580-1584).  Giri et al. (P.N.A.S. 82: 1580-1584)
            suggest that the low level of conservation in the E1 NH$_2$-terminal
            domain of the four PVs relative to the remainder of the E1 region
            may indicate either less evolutionary pressure or divergent
            function.  Additional notable sequence similarities include the E1
            region and the carboxyl domain of the large T antigen of SV40 and
            polyoma, an 11 residue region in E7 and the large T antigen of
            SV40, E6 and the beta chain of the ATP synthase, the carboxyl
            terminal of E2 and the human c-mos proto-oncogene.
BASE COUNT   2252 a   1725 c   1929 g   1962 t
ORIGIN      1063 bp upstream of unique EcoRI site (on sense strand).
        1 gctaacaata attaagaaac atgtaatggc cagaaACCGA TATCGGTTGC TGGCACTGTA
                                                     -> 32 bp direct repeat
                                          -> E2 bind <-      signal ->
       61 TATCTGAGAT CGCATCGGTT GCTGGCACTG TGTATCTGAG ATCGCaacgc attgccagga
                   <-->     32 bp direct repeat       ->
          <-                              E6 ORF start ->
      121 atttctgcaT ATAAgacaag aaacttagag cagATGgaga actgcctgcc acgctcgcta
          TATA box ->  <-               E6 cds ->
      181 gagaagctgc agcaaatatt acaaatatca ttggaggact tgccgtttgg ttgtatattt
      241 tgcgggaaat tgcttggggc tgcagaaaaa caattgttca aatgcacggg gctatgcatt
      301 gtatggcata aagggtggcc gtatgggacc tgcagagact gcactgtatt gtcttgtgct
      361 ttggatcttt attgtcacct tgctcttact gctcctgctt tggaggctga agcgctggtt
      421 ggtcaggaaa tatctagctg gttcatgcgt tgtacagttt gcggaagaag attaactatt
      481 ccagaaaaga ttgaattaag agctagaaat tgcacgcttt gttgtattga taaaggtcaa
      541 tatttccagt ggagggggtca ttgcagttct tgcaaactgt cagaccaagt tgatttgggg
      601 ggctatcccc cgagtcccgg cagtcgctgc gggggaatgtg acgagtgttg cgtcccggac
      661 ctgacacatc taactccggt ggatctggag gaacttggat tatatccagg ccccgaagga
      721 acctatccgg atttagttga cctagggcca ggcgtttttg gggaagaaga cgaggagggg
      781 ggtgggctgt ttgacagctt cgaggaggag gatcctggac ccaaccagtg tgggtgtttt
      841 ttttgcacca gctatccgtc cggaacaggt gatacagata taaatcaggg accggcagga
      901 gctgcaggga ttgcactgca gtcagatcca gtctgtttct gtgagaattg tattaacttc
      961 acagaattta gaTGATAGta tttctgctat cctgtgcgca gggctgcttc tttatctttt
```

```
                E7 ORF start ->
                      <- E6 end
1021 ttctttatat actactgttt ttccttctgt actggctttta tcgaattctg caagATGata
                                                        E7 cds ->
1081 ggcagaactc ctaagcttag tgagctggtt ttaggtgaaa ctgctgaagc gcttagtctg
1141 cattgcgacg aagcattaga gaatttaagt gatgatgatg aggaggatca tcaagataga
1201 caggtgttca tagaaaggcc ctatgcagtg tccgtgccat gtaagcgctg taggcaaact
1261 atcagcttcg tctgcgtctg tgctccagaa gccataagaa ccttgaatcg actgctatcc
1321 gcatcgcttt ccctggtgtg cccggagtgt tgTAAcTGAa aATGgctgaa ggtacagacc
                      E1 ORF start ->       -> E1 cds
                                    <- E7 end
1381 ctttagatga ctgtgggggg ttcttagaca cggaagcgga ctgtttagac tgtgacaacc
1441 ttgaggagga cctgacagag ctgtttgatg ctgacactgt aagcagttta ctagatgata
1501 cagatcaggt gcaggaaaat tccctggaac ctttttcagca tcatgaggcg actgagacct
1561 tgaaaagcat agagcatctc aagagaaagt atgtcgatag tcctgataag agcctgggta
1621 tcgacaactc cgtcaatgcc ttgagtccaa gattacaagc tttctcactg tcaggacaaa
1681 aaaaggctgt taaaaagaga cttttcggta ctgacggaga tgaagctgct tctggtgctg
1741 agtcgttaca ggtagaatcg ggatttgggt ctcaacaaag cgtatcagat acacctgtga
1801 ctgacatttt aaatgcaaat acagcaagag tcaaacattt gttgttattt aggcaagctc
1861 acagtgttag cttttcggag ctcaccagaa catttcaaag tgacaagact atgagttggg
1921 attgggtagg tgggctggcg gacattcatg taagcgtgtt ggagagcttg cagacatctc
1981 tgagaagtca ttgcgtatat gttcagtatg atctcaattt tgcagagaca aatgcttcat
2041 ctctgctgct gctcctgaga tttaaagcac aaaaatgtag ggacggggtt aaagcgctgc
2101 tatcccaatt gttgggagtt caagatctaa aagtttttatt agaacctcca aaaacaagga
2161 gtgtcgctgt tgcattgttc tggtacaaaa gggcgatggt ttctgggtt tttagctacg
2221 gtccaatgcc tgaatggata acgcagcaga caaatgttaa ccatcaaatg ttgcaggaaa
2281 agccgtttca gttgtctgtc atggtccagt gggcatatga taaccacctt caggatgaaa
2341 gtagtattgc atacaagtat gcaatgctcg ctgaaactga tgagaatgca agagcgtttc
2401 tagcttctaa ttctcaggcg aagtatgtta gggactgttg caacatggtc agactctatt
2461 taagagcaga aatgagacag atgaccatgt ctgcatggat aaactacaga ttggatggga
2521 tgaacgatga tggggattgg aaggtggtcg tgcattttct gcggcaccaa cgagtggagt
2581 tcatacccttt catggtgaag ctgaaggcct tcctaagagg aacaccaaaa aaaaattgca
2641 tggtgtttta tgggccacca aatagtggga agtcatattt ttgcatgagc ctcataagat
2701 tacttgcagg acgggtcttg tcgtttgcaa acagcagaag ccatttttgg ctgcaaccat
2761 tagcagacgc caagctagcg ctcgtggatg atgctacatc cgcgtgctgg gatttcattg
2821 atacatacct cagaaatgcc cttgatggca atccccatcc ggtggacctg aagcacaagg
2881 caccaataga gattaagtgc cctcccctcc tgataaccac aaatgtggac gtcaaatcag
2941 atgatagatg gagatactta tttagtagaa tttgtgtgtt taactttttg caagaattgc
3001 ccattagaaa tgggacacct gtgtatgaat aaatgatgc aaactggaaa tctttttta
3061 aaaggttctg gtccacctta gaacTAAgcg acccggaaga cgaggggtgac gATGgaggct
                      E2 ORF start ->                E2 cds ->
3121 ctcagccagc gcttagactc catacaggag gaacttctca gtctctaTGA gaaggagagc
                                                        <- E1 end
3181 acgagtttgg agtcccagct acagcactgg aacttactaa gaaaagaaca ggtccttttta
3241 catttctgta aaaaacacgg gatcaggcaa ctgggctaca cgcctgtccc gtctcttctt
3301 acctcacagg aatgtgcaaa gcaagccaTA Gaaatggtgc tgtacattga aagcctactc
                      E4 orf ->
3361 aggtccccgt attcagATGa gccatggaca ttgcaggata ccagtagaga aaggttcgaa
            E4 cds ->
3421 agccctccgc aaaagacatt caaaaagaac ccagctattg ttgaggttta ctatgatggt
3481 gacagaggga acaacaatga atacacactg tggggtatat ttattattgg gaacgctgat
3541 ggggagtggg ttaagactga aagtggagtg gactatagag ggatttatta tgtggactct
3601 gaaggaaact atgtgtatta tgtggacttc tcaaccgacg cgggacgttt tgctgctaat
3661 ggacactatg acgtggtgtt tcaaaacatg cgcctctctt cttctgtcac cagctcccc
3721 cagccgctgg tcagtgcccc tgaagacacc gtccccgaag aggcccccga cagtgcagtg
3781 cccgccgctc aaaagaaaac agggcccaaa accacgcgta cactgggcag acgaaggtca
3841 aggtcaccag gggtgcaacg aaggccggca aagcaacgaa aacaggccgc cccggacgaa
3901 gcggattctg ctgccgggga catcagaccg cctgctccag aggacgttgg acgaagaact
3961 acgacggttg gaagaacacc tcccgggcgg aatagacggc ttcgcgagct taTAAcagaa
                                                        <- E4 end
4021 gctagcgatc cgcccgtgat ttgcttgaaa gggggcaca accagcttaa gtgcttaagg
4081 tatcgcctta aaagcaagca ctcctcacta ttcgactgca taagcactac ttggagctgg
4141 gttgacacaa cgagcacatg caggctaggt agcgggcgca tgcttataaa gtttgcggac
4201 tctgagcagc gcgataagtt tcttagcagg gtcccactcc catcaacaac gcaggtgttt
```

```
4261 tTAGggaatt tttATGggct tTAGtgacgt gtatgcatgT AAcccattcc catcagcagc
     E5 orf -> E5 cds ->          <- E2 end     L2 ORF ->
4321 ttttgtaacg caacgttttt ttgtaccaat aaatcttgca catacgcaaa aggtgtcATG
                                                                L2 cds ->
4381 gttgcacggt cacgaaaacg cagggctgca ccacaagaca tttatccaac atgcaaaatt
4441 gctggcaatt gcccagctga catacagaat aaatttgaaa acaaaacaat tgcagataaa
4501 attctgcagt atggcagcct tggtgttttc tttggaggac ttggaatcag tagtgccgga
4561 ggttctgggg gtcgacTAGg gtacacccca ttatctggag ggggggacg tgtcatagca
                        <- E5 end
4621 gcagccccag taagacctcc cataacaaca gaatctgtag gccctctaga tatagtgcct
4681 gaggtagctg atcctggggg tcctactcta gtgtcactac atgaactgcc tgcagaaaca
4741 ccatatgtat caagcacaaa tgttacaggg gatggcgcag cagagcccct tccagctggt
4801 catgggggaa gccagatttc agacgtcaca tctggtacat ccggcacagt gtccagaaca
4861 cacattaata accctgtatt cgaggctcca atgaccggtg atcaggatgt ctcggatgtg
4921 catgtatttg ctcactctga aagtagtata actatcaacc aaacagaaaa cacgggcgga
4981 gagctaatag agatggtccc cctcagacac cccccctcgca gtgagggaga tttcagggaa
5041 acatccttca gcacaagcac accaatccct gatagatctg cgttgcgatc tataaacgta
5101 gctagcagaa gatatcagca ggtacaagta gaaaaccctg ctttcctgaa caggcccagg
5161 gaactggtgc aattcgaaaa cacatttgac aatcctgctt ttgtggatga tgagcaacta
5221 agcctccttt ttgaacagga tctagacacc gtagttgcaa ctccagatcc tgcgttccag
5281 gatgttgtgc gtttaagtag gcctagtttc actcaatcca gagctggtag ggttcgcgtc
5341 agccgcctgg ggaggacgct aacaatgcaa acacgcagtg gtaaggcctt tgggcctgcc
5401 aaaacacttct actatgagct ttcaagcata gcagaaggtc ccgagccaga catcctcatc
5461 cccgaatcag aacaggaaac atcattcaca gatgccacat ctaaagacac acaacaggaa
5521 gcagaagtgt atgcagatgg ttcaaccctg gaaactgaca catcggcaga tgaaaacttg
5581 acacttgtct tttcagacag aggcaggggt cagggttcac atgtacctat tccaggcaag
5641 tccacaattg ggggtcctgt aaatattggg gacagcaaat actatactct gaaccctgga
5701 gaaactacaa gctttgaagc agatgtaatt tcacctgttt tcatatttga gggtaacgca
5761 gatggcactt actatcTAGa ggaacctcta cgaaagaaaa gacgcaaatc tatctttttta
             L1 ORF start ->
5821 cttgcagATG gcagtgtggc tgtctacgca gaaTAAgttt tacctgccgc ctcagcctgt
     L1 cds ->                           <- L2 end
5881 cacaaagata cctagcacgg atgaatacgt tactcggaca aacgtctttt attatgcatc
5941 cagtgaccgg ctactcacag tgggacatcc atactatgaa atacgtgata aaggcaccat
6001 gcttgttcca aaggtttctc caaaccaata cagagtattt agaatcaaac tccctgaccc
6061 taacaagttt gcatttggtg acaagcaact atatgatcca gagaaggaac ggcttgtgtg
6121 gtgccttaga ggtattgagg tcaatcgtgagg ccagcctcta ggagtcagtg tcacaggaaa
6181 ccctatcttt aataaaatttg atgatgtcga gaatcccaca aagtattaca ataaccatgc
6241 agaccagcaa gactacagaa aaagcatggc gttcgacccc aagcaagtgc agctgttaat
6301 gcttggatgc gtccctgcca caggagaaca ctgggctcag gcaaagcagt gtgcagagga
6361 tccaccacaa cagaccgact gtcccccccat tgaactagtg aacactgtta tagaagatgg
6421 ggacatgtgt gaaataggct ttggggcaat ggaccataaa acattgcagg ccagtttatc
6481 agaggttccc cttgagttag cacagtcaat cagcaagtat ccagactatc taaaaatgca
6541 aaaagatcag tttggggatt ctatgttctt ttatgctaga agagagcaga tgtatgctag
6601 acatttcttc agcagggcag gaggggacaa ggaaaatgtg aagagcaggg cctacataaa
6661 acgcacacag atgcaggag aggcaaatgc caacattgca actgacaatt actgcatcac
6721 acctagtgga tctttggtct ctagcgattc acaggttttt aaccgtgcat attggctcca
6781 gaaagctcaa ggcatgaaca tggagtttg ctgggacaat caaatttttg tgactgtggt
6841 agataacacc aggggtacaa tattaagtct tgtcacaaaa tccaaggagc aaatcaagaa
6901 gacccatgga aaaacagtac attttttcttc ctatctaagg catgtggagg agtatgaact
6961 gcaatttgtg ctccagctat gtaaggtcaa gttaacaccc gaaaacctat catacctaca
7021 tagcatgcac ccaacaatca tagataattg gcaattgtca gtgtcagctc agcccagtgg
7081 aacgctagaa gaccagtaca gatacctgca gtccattgca accaaatgtc cacccccaga
7141 acctcccaaa gaaaacactg acccatataa aaactataag ttttgggaag tagatttgtc
7201 tgagaagcta tctgatcagc tagatcagta tccacttggc agaaagttcc taaatcaaag
7261 tggcctgcaa agaattggta caaaaagacc tgcacctgca cctgttagta ttgtgaaatc
7321 atctaaacgc aagagacgta ctTAAttgta tattattgga tgtgtattta atacctcata
                         <- L1 end
7381 acttataaat atactaAATA AAgtttaaaa ccgatgcata ctagtattac ttctttttttg
                 signal ->
7441 tgaagctctt tgcgACCGCA CCCGGTggca tgtgatgaca catcacacag taagccttga
                 -> E2 bind
7501 gagaacgcca gaggttgtaa acatcctgtt ccggctggca ctgagcctga tcgcaggcag
7561 tgatcttaac cgtaagcgtt ctcattaatt ctgggtgtatt tccaagaaaa acgttatcgc
```

```
7621 ttttggccac cgtactcggt gtcagcaaat accaagccaa gatctttttt ggtgccaaaa
7681 atctaagcaA CCGCTCCCGG TcccggcaAC CGCTCCCGGT CTTGCAACCG CCCACGTTCG
              -> E2 bind              ->       32 bp direct repeat      <-
                                      -> E2 bind <-
7741 ACCGCTCCCG GTCTTGCAAC CGCCCACGTT CGACCGCTCC CGGTcttgca accgccaggt
      ->          32 bp direct repeat       <-
      -> E2 bind <-                      -> E2 bind <-
7801 gtgcatgact gtaagtattg actcacACCG CATGCGGTgt tatgcacttt gatttaatga
                                   -> E2 bind
7861 tggttgtt
```

//

```
LOCUS       EEPV         8095 bp ds-DNA   Circular   VRL       15-JUN-1988
DEFINITION  European Elk papillomavirus (EEPV), complete genome.
ACCESSION   M15953
KEYWORDS    circular; complete genome.
SOURCE      European Elk papillomavirus DNA isolated from a wild elk
            (Alces a. alces) in Sweden.
REFERENCE   1  (bases 1 to 4780; 7144 to 8095)
  AUTHORS   Ahola,H., Bergman,P., Stroem,A.C., Moreno-Lopez,J. and Petterson,U.
  TITLE     Organization and expression of the transforming region from the
            European elk papillomavirus
  JOURNAL   Gene 50, 195-205 (1986)
REFERENCE   2  (bases 1 to 8095)
  AUTHORS   Eriksson,A.
  TITLE     The genome of the Elk papilloma virus
  JOURNAL   Unpublished (1987) Swedish Univ. of Agricultural Science, Uppsala
COMMENT     Substantial sequence data is available for the ungulate
            papillomavirus group.  Full genomic sequences exist for BPV-1,
            BPV-2 and BPV-4, EEPV, DPV, and partial genomes for BPV-3, BPV-6
            and RPV.  The Bovine papillomaviruses can be classified into two
            groups, subgroup A (BPV-1, BPV-2 and BPV-5) and subgroup B (BPV-3,
            BPV-4 and BPV-6), based on the tissues they infect (Jackson et al.
            Mol. Carc. 4: 382-387).  Using this criteria, both DPV and EEPV
            can be grouped with the subgroup A viruses.  EEPV is associated
            with the category of papillomaviruses causing fibromas and
            fibropapillomas.  Extensive study has focused on this category,
            including BPV-1, DPV (deer papillomavirus), and RPV (reindeer
            papillomavirus), because they are able to transform NIH 3T3 mouse
            cells.

            Sequence analysis of DPV, EEPV, RPV, and BPV-1 reveal a
            significant degree of similarity between the four viruses,
            although the closest relationships seems to exist between DPV,
            EEPV and RPV.  Indeed, Ahola et al. (Gene 50: 195-205) speculate
            that DPV, EEPV and RPV consistute a distinct subgroup amoung the
            animal papilomaviruses.  EEPV, the bovine subgroup A viruses, RPV
            and DPV all induce tumors in young hamsters (Moreno-Lopez et al. J.
            Virol. 61: 3394- 400; Ahola et al. Gene 50: 195-205; Groff et al.
            J. Virol. 56: 85- 91).  These same viruses transform the mouse cell
            line NIH 3T3 in vitro, whereas all but DPV transform C127 (Groff
            et al. J Virol 56: 85-91).  The E5 regions of BPV-1, DPV, RPV and
            EEPV are very hydrophobic and both the BPV-1 and EEPV genome exists
            in an episomal form in the transformed cell in high copy number
            (Ahola et al. Gene 50: 195-205). The highest similarity exists
            between the E1 and E5 ORFs of DPV and EEPV, 70%.  The similarity
            between these regions for EEPV and BPV-1 was 10% lower,
            approximately 60% (Ahola et al. Gene 50: 195-205).  (Ahola et al.
            (Gene 50: 195-205) note that the E5 ORF is a common feature of
            papillomaviruses capable of transformation and pathological
            symptoms including fibromas and fibropapillomas.)  An additional
            feature shared between these viruses as well as HPV-1a and an avian
            papillomavirus is an Hpa1 cleavage site in the noncoding region.
            Ahola et al. (Gene 50: 195-205) interpret the conservation of this
            site as suggesting its location in a regulatory element.

        Ahola et al. (Gene 50: 195-205) identified several common
            transcription and processing features in EEPV and BPV-1.  One, the
            presence of analogous TATA motifs within the promoter 56 basepairs
            downstream from the HpaI site.  Two, an additional putative
            promoter region at nucleotide 2414 (this motif has also been mapped
            in DPV). And three, similar putative splice sites.

            Sequence in computer-readable form for [1],[2] kindly provided by
            U.Petterson, 25-Aug-1987.
BASE COUNT     2159 a   1814 c   2047 g   2075 t
ORIGIN      3 bp upstream of HpaI site.
        1 GTTAACAATC Accagatctt gcccgttttt gtgagcggga aagctggtta caggtttaTA
```

```
direct repeat<-                                                    signal ->
inverted rpt ->      <-
       61 TAAAAaggcc caccgcacaa gttttcacag acggttcagg atacttctaa tacatgcATG
E6 orf start ->                                                   E6 cds ->
      121 tgtggcgaat gctatgcata cctcacctgc atctggtgca agaagggctt agataaggta
      181 gatgcaaagc gatgccatga aaaaaaaata agaatagcgt gcaggaacgg aaaacattgt
      241 gctgtctgta catcttgcct ggaaaatggg ctgtaccttg aaagGTccct ttttcctggg
                                                          5' sj /\
      301 cgacccatct accctggaga cctgtatgag cccgatccat ggGTcatgtt caacgacatt
                                                          5' sj /\
      361 agatgcatgt attgtggtgg atgccTAAcc cgcgacgaaa aagagagaca cagactgttt
                       E7 orf start  ->
      421 tgtgaagact ctggatatT caggcatcAG gtgcggggac gttgctatct ctgcaccAGg
                                      /\ 3' sj                           /\ 3' sj
      481 cATGgttcac ggcccccgta caaagaaaca cctgccgccg taTGAatcac ctcccctcac
     E7 cds ->                                               <- E6 end
      541 actgctcctA Gagccagttg ctccggtgca acagacaggc attcaggcac cgcagaggaa
                       /\ 3' sj
      601 gccaccttcc cagaaaggac acaaaaaagg acacaagaaa gtttattctg tgactgtgcc
      661 ttgcaatgga tgtgacaaaa acctggaatt ttgtgcaaga acttccagcg ccaccatctt
      721 aacgctgcaa aacctcctgc tgaaagaccT AGacttcctg tgctctacct gcgagaccaa
                             E1 orf start ->
      781 ccATGGcTGA aactgcagGT agctcggggc aggggggggg agcttatatc tgctttgaag
       E1 cds ->         5' sj /\
                 <- E7 end
      841 ccgactGTag cgactctgat acagagGTtg attcacctgt acaatgctct gattcaagtg
          5' sj /\                5' sj /\
      901 atgaggatct agtagataat gccaatatcg ttccgggaaa ccacctggag ttgttccaaa
      961 cgcaggaaaa agaggcggga gaaagacaga tttcgctttt gaaaagaaaa ttctgtttga
     1021 gcccgggaac ctcagaggtc gaggagctta gtcctgggct tgccggaatc agaatctctc
     1081 cgccaaagcg aaatccggtg gttaggagaa ggcttttttga cgcaggtggg agagacgccg
     1141 tgcgaacacc gcgtgatcat gaagttaata gttctcctga acccaggagt caggtacagt
     1201 cgggaagtag cagtagGTct tgggagggac atctggaatc cattaacgag cctgctagtg
                       5' sj /\
     1261 acggcaacat ggccgccgtg atgcacaagt tgttcaagac tttgtacatc gcgggttttg
     1321 gggagataac acgcgtcttt caaagtgata aaactaacaa taatcagtgg gtgatagcag
     1381 cccatggcgc atcagaggtg ctttatgccg caagctttga aatactgaca aaacactgca
     1441 gctacctgca ggcgtctagg aaggtgcatg agacaggaag catgtctttg ttcttagctg
     1501 tcttcaatgt tgggaagagt agggagactg tcagaaaact aatttccaggt gtcttaaaca
     1561 ccccgtgtag ccgcctacta ttgcaaccgc cgaaaattcg tggactatgt cctgctttat
     1621 tttggtttaa gttggggctc tccccagcaa cacagacgca cggtacgact ccggactgga
     1681 ttaagcagca gaccaatgtg gcctataata ctggggaggc ctctaaattt gattttggca
     1741 caatggtaca gtgggcatat gaccaccggc taacagagga gtgcaaaatt gcatatcaat
     1801 atgcaaaatg tgcaggtaca gacctaaatg cgaaagcatt tcttgcaagt accaatcagg
     1861 cacggctggt caaggactgc tgtactatgg tgaaacatta cctgagagct gaagagcagt
     1921 cattaaccat ttctgctttt attaaaagga gatgcgataa tgcaactgga aaaggcagtt
     1981 ggttgagcat tatgaatctg ttaaagtttc aaggcatcga gcccattaac tttgtaaatg
     2041 ccttgaaacc atggctgaaa ggcaccccaa aacataattg catagcaatt gtaggacccc
     2101 caaatagtgg gaagtctctt ctgtgcaata ccctcatgtc gtttctggga ggaaaggtac
     2161 tgacgtttgc caaccactcc agccacttct ggttagcgcc ccttaccgac tgtagggtcg
     2221 ccttgataga tgatgccacg catgcgtgct ggagatactt tgacacatat ctcagaaatg
     2281 tacttgacgg ttatccagtt tgtattgaca gaaagcacaa atccgctgtg cagctcaaag
     2341 cccctcccct tttgctaacc agTAATATTg atgtgcatgc agatgaaaag tatttctatc
                          signal ->
     2401 tgcaaagtag agtcaaaacc ttctatttca aggagccgtg ccctgcgtct gatactggtg
     2461 agccccttttt ctttattact gatgctgact ggaaaaattt ttttgaaagg ctatgggagc
     2521 gatTAGatct cagcgaccaa gaggacgagG TtgATGaaga tgagtgcagc cagcgatcat
     E2 orf start ->              5' sj /\    -> E2 cds
     2581 ttacttgcag cgcaagaaac acagatgcaa tgcatTGAga aagatagtcg cctgttacag
                                                <- E1 end
     2641 gatcatgcat gctattgggg ggcagtaaga agggaaaaac tgttattata tgcagcgaga
     2701 acaaaggggt taaaaacaat tgggtgtgtg cctgtgcctc cttgttctgt tactgcagag
     2761 caagcgaagc aagcaatatg catgcaattg attgtggagg aattactgca cagtccatgg
     2821 gccaaagaac catggtccct tacagaccta agctgggaga gatatcaggc tgccccaaaa
```

```
2881 gggtgtttga aaaaaggcgc cagagtggtg gaagtggagt atgatgggaa ctccttctaat
2941 aagacttggt atacagcttg gagtacagtg tacgtgcgcg gaacggaaga ggagggctgg
3001 gagactgctg tctgtgctgc agacggacag ggcatttatt attgcgccgg gatgagcagt
3061 aaggtgtact ttgaaacctt tgaaactgat gcccgcagat ggagcaggac ggggcactgg
3121 actgtgaggg ataacgatgT GAtatatcat tcaacctttg gtgcacccc tcactctAGa
                   E4 orf start  ->                                /\ 3' sj
                   NH2 terminus unknown
3181 aacgacagag actgcatcga aggattctgg agcgacgccg gggagcgTAG aggctcgaga
                                        E3 orf start ->
3241 gggtccgaca caaccgacag agccctgcct taccctgctg ctcgacaatc ccccatttgt
3301 cgccccgtcA Gaactggcga aaaccggagt cgggccgttc accgccagGC tccctacAGc
            /\ 3' sj                                              /\ 3' sj
3361 GCAccatcat cccccgggag ttccgtgggc cccgattccc cctccgagag ctcgcgccag
3421 gtaccgctgg ttttgctacc aggaccatca gatccagcgc cgccgtcgcc ggactctaca
3481 gacgTAAtcg cagagggTGA caaggaacct gagcggttca gcattctctc aaaaccaggt
         <- E4 end      <- E3 end
3541 gggcagccat gtctgatact tagtggaaac ggaaaccaag ctaagtgcta tcgtttccgc
3601 tgcaagagat atttcagaga acactatcag cacataacga ccacctggtg gactgtagga
3661 gagcgaggat ctgaaaggca cggagatgcc tgtgtgctgg tgacattcaa agacagttcc
3721 cagagagggg tgtttttgaa gcgagtgcct ttgccacctg gaatgcgcgc gcaggcactt
3781 acaatgattg cggactttTG Agaatgtgac tctggtagcc gcaatactgt gcagtttgct
                         <- E2 end
3841 caccttcatg ttttacctgt gTAGcacaga ctgctgcatg ctgtgATGac atacggtttg
                  E5 orf start ->                   E5 cds ->
3901 ctttttgtttc tggggctcac atttggacta cagctgatgc tacttgtctt tctgctgttt
3961 ttctttctcg tatggtggga ccagtttggc tgccgttgtg aaaacatgca gttgTAAata
                                                                  <- E5 end
4021 gtgtatattg aggtgtagat attcatttga tcctgtacat acattttctc tattttttta
4081 aaaaatgctg gttgataaac atacataggt cacaaacagg tcattccata acgtacaaca
4141 tactttttat aggcgtggag tctcttatct ctgtcccttt gtttctctct tttgaactgt
4201 tatccgagtc agcaagtgcc atttttgttg ccattttctt caagtgccaa tctcctaact
4261 gcatccccaa gagctggtgt aacgtcagaa gaacagtcga gaatctagcc tttgaaagac
4321 tttcaaccct caaacagacc atgtgcatac ctcctgtaac agcaaagctg ccacgacatt
4381 gaagcctgca tgaaatattg tttactattg tttttgctgc tattgctggg gcagtggaat
4441 ccaatgtggg tattgttact aattgtctgg ttgtccattg tacttttatt tgagttgcat
4501 tttgtgtggaac acttcacatg acaggctgca ctgcagcgcc ttcatctcat ccctaaattt
4561 AATAAActt cccctattta acccctaccA TGgcgcctcg gcgagtaaag cgtgcaaatg
L2 orf start ->                     L2 cds ->
 signal ->
4621 tctatgacct gtatcgcaca tgcaagcagg caggcacctg tcctccggat gtgataccta
4681 aggtggaagg gaagacgata gcagacaaga tattgcagta tggaagcatg ggcgtttatt
4741 taggcggcct aggcattgga acaggttctg gaaagccagg aacaggaggc tatattccac
4801 tcagaggtgg gggctctacc acttcactat caagcaaacc ttttgctggg gggatacccct
4861 tagaaacctt agaagggata ggggcattcc ggcctggcat agtggaagat gcagggcctg
4921 ctttagaagg cattcttcct gacgcaccag cagttgtcac tcctgaggca gtgccagtgg
4981 atgaggggtt aagtgggcta gatatttcca gggaattaag ccaggaacaa attctcagct
5041 ttctccaccc tgagggtccg gatgatattg cagtacttga ggtaaggcca acagaacatg
5101 atcaggcaca tttgctgtct acaagcacac acccaaatcc actgtttcag ggtcctgtac
5161 agcaggcacg aattattgca gaaacatctg gtgcagaaaa cgttttttgtg ggtggaagtg
5221 gcattggaag caatgcagga gaggacattg aactgacact gtttgctgaa ccaaggacaa
5281 gtacacctga ggtgcctatt aaacgttctc ggggcatttt caattggttt agcaggcgct
5341 actatacaca ggtacctgtc gaagacccag acgagattgc tgctgcaggc tcgtatgtct
5401 ttgagaatcc tgtatacgat tcaaaggcgt tcaaacctgc gcagcagccg gacattactc
5461 tacaggatga agcttctgtc actgggcggg acgctgcaag attgctggcg ggaccctcgg
5521 gcaggattgg gtggagtcgt atcacacgac ccactagtct tggaacacgt agtggcgtga
5581 gggtaggccc tctttatcat ttacgatcct cttttcagcac tatccatagt cctgagacaa
5641 tagagctaat acccacagta cttgaggatg atactgaggt gcttacaggt gttcctgaga
5701 gagacactgg ttttgatgat gtggatttgg acagtatagc aagtgacagt ccattactac
5761 ctgagcggca tcaccttgct tttggagcaa ggcggtctca cattccaatt gtggcacgac
5821 caggtgttca aactggtaca gtgattgata cacgtcagat ggctgaaaac tctgtttacg
5881 tgtcggacaa tggaggacag gagtcacagc agacgcccac tgtggtaatc aatggcaaca
5941 ttaatgtgtc catggaatat tttaggcatt actatttgca ccctagcctt cTAGgtcgca
                                                          L1 orf start ->
6001 aacgaaaacg tctattcggt TAAtgttttta cagATGgcgt tctggcagcc tagtcaaagg
```

```
                              <- L2 end
                                 L1 cds ->
6061 ctatacctgc ctcccacacc tgtgacaaag gtgctgtgct ctgagcaata tattaggcgt
6121 aaggacgtat tttatcacgg ggagacggag cgcatgctca ctgtagggca tccatattat
6181 gaaattaaac aatcagggtc tgggaaaacc attccaaagg tttcacctaa tcaatatcgt
6241 gtttttcgga tcttactgcc ggatcccaac cagtttgctc ttccagataa agccatgtat
6301 gacccaagta aggaaaggct agtctgggct gttgtggggg tacaggtgtc tagaggacaa
6361 cctttaggtg gctctgtttc aggacattcg tatcagaaca ctctgattga tgcggagaat
6421 gttagtaaaa aggtaaatgc acagggcaca gatgacagga agcagggagg catggacgtc
6481 aagcaacagc aaattctact gctaggatgc accccagcta ttggtgagta ttggacaact
6541 gctaggccct gcgttacaga taggccagag actggctcct gcccccctat agaactaaaa
6601 aacaaaccta tagaagatgg tgatatgatg gatattggct ttggtgcagc taatttcaaa
6661 gagttaaatg ccacaaagtc agatctccct ttagatattg caaaagatat ttgtttgtat
6721 cctgattatt taaagatgac tgaagaagcg gctggcaaca gtatgttttt ttttgctcgg
6781 aaagaacaag tttatgttcg ccacatctgg tcgcgtgggg gtaccgacaa agaaatgcct
6841 ccagaggcat actttctgaa gccaaagggt ggggaccaaa cacagaaaat gcctagtatt
6901 ctttttggag tgccaagtgg cagtttagtt tctacagatg gacaattgtt aatagacct
6961 tactggctgt ttcgtgcaca gggcatgaat aatggcatat gctggcttaa tcaactgttt
7021 gttactgttg gtgacaatac aagaggaacc acattaacca ttacggtgcc tacatccggg
7081 tccccactca ctgaatatga cacgagcaaa tttaatgttt ttcaaaggca tgttgaagaa
7141 tataagcttg cctttgtatt tcagctttgc tctgtcactc taagtccaga aaccgtctca
7201 catctccagg ggttaatgcc ttcgatcctg gaacactggg atattaacat gcagcctcct
7261 acgtcctcga ttcttgagga tacttacaga tatcttgaat cacctgctac taaatgtgca
7321 gataatgtaa ccccctatggg acctgaagat ccctatgctg gtttaaagtt ttgggaggtg
7381 aatctaaaag aaaggttgtc tcttgatctt gatcaatttc ctctgggacg gcgtttttctt
7441 gcgcagcaag gattagggtg cagtactaga aagagggttg cacctgtccc taaggtcacc
7501 gaaaaaagga ttgttaggaa aagaagaaag gggaatTAAg ggcatgaaat cttaaaaact
                                                    <- L1 end
7561 gctgtgtttg ctaAATAAAt gcaattttttc TTATGTGTCA agagtTTATG TGTCAtgtcc
             signal ->                 ->direct<-         ->direct <-
                                          repeat            repeat
7621 tgctgttcag tccaacttgc ACCACACCCG GTgctggcat ctgattagac gcagtgtcag
                          -> E2 bind
7681 cagctttatG AAAAGCagac acttggctag acacacaggc gcctggcgcc ctcatcgaaT
inverted repeat ->     <-                                             ->
7741 TGGCGCACCG CTGGCGTTCG ggatcaaatT TCCTCTACCG CTGCCGGTTG TTAAagcgcc
      direct repeat      <-           ->  direct repeat    <-
                                           -> E2 bind <-
                                                          ->      <-direct repeat
7801 cTTCCTGTAC CGTTCCCGGT AGcgcctctt ctctcccttc agCGCTACCG CTCCCGGTGT
      -> direct repeat    <-                     -> direct repeat
            -> E2 bind <-                            -> E2 bind <-
7861 GCatggtaag taggcggtca ttgtcgaaga gaactggtaa gcaagtccga acaagaaaaa
      <-
7921 tgcTTGGCGC AACGCTGACG GTAGtCGCTA CCGTCCGCGG TGCTCGCTTT TCtaagaaat
         -> direct repeat    <- -> direct repeat   <-
       inverted repeat ->     <-    -> E2 bind <-
                   inverted rpt ->    <-          ->     <-inverted rpt
7981 gctcaaacgg tctcttgcta gctctgctcc tattggctgt gctgaaatta ctcacgccgc
8041 tttgcctgtA CCGTGAACGG Ttttgaatcc tactttttct cagggaaTGA TTGTT
            -> E2 bind  <-                                 -> direct repeat
                                      inverted repeat ->      <-
```

```
LOCUS       DPV          8374 bp ds-DNA   Circular  VRL       10-NOV-1986
DEFINITION  Deer papillomavirus (DPV), complete genome.
ACCESSION   M11910
KEYWORDS    complete genome.
SOURCE      Deer papillomavirus DNA from an American white-tailed deer fibroma,
            clone pDEG98.
REFERENCE   1  (bases 1 to 8374)
  AUTHORS   Groff,D.E. and Lancaster,W.D.
    TITLE   Molecular cloning and nucleotide sequence of deer papillomavirus
  JOURNAL   J. Virol. 56, 85-91 (1985)
COMMENT     Previously called deer fibromavirus, DPV was first isolated from
            the American white tailed deer (\it Odocoileus virginianus)
            (Groff et al. J Virol 56: 85-91). Substantial sequence data is
            available for the ungulate papillomavirus group.  Full genomic
            sequences exist for BPV-1, BPV-2 and BPV-4, EEPV, DPV, and partial
            genomes for BPV-3, BPV-6 and RPV.  The Bovine papillomaviruses can
            be classified into two groups, subgroup A (BPV-1, BPV-2 and BPV-5)
            and subgroup B (BPV-3, BPV-4 and BPV-6), based on the tissues they
            infect (Jackson et al. Mol. Carc. 4: 382-387). Using this criteria,
            both DPV and EEPV can be grouped with the subgroup A viruses.  Many
            clinical and sequence similarites are prevalent in the ungulate
            group.  EEPV, the bovine subgroup A viruses, RPV and DPV all induce
            tumors in young hamsters (Moreno-Lopez et al.  J Virol 61: 3394-
            400; Ahola et al. Gene 50: 195-205; Groff et al. J Virol 56: 85-
            91).  These same viruses transform the mouse cell line NIH 3T3 in
            vitro, whereas all but DPV transform C127 (Groff et al. J Virol 56:
            85-91).  And, the E5 regions of BPV-1, DPV, RPV and EEPV are very
            hydrophobic (Ahola et al. Gene 50: 195-205).  DPV infection,
            however, is unique compared to the clinical profiles of the other
            subgroup A viruses; infection results in fibroblastic proliferation
            without epithelial proliferation (Groff et al. J Virol 56: 85-91).

            The bovine papillomavirus, BPV-1, also causes fibroblastic
            proliferation and shares a high degree of sequence similarity with
            DPV.  In addition, both viruses are oncogenic.  Sequence analysis
            revealed that DPV and BPV-1 have the highest degree of similarity
            in the E1 and L1 ORF's.  Groff et al. calculated a homology of over
            90% between the 3' half of the E1 ORF.  The authors propose that
            the trans-acting DNA replication function encoded in this region is
            analagous for the two papillomaviruses because both BPV-1 DNA and
            DPV are maintained in episomal form in transformed cells.  (A high
            degree of similarity was also found to exist between the BPV-1
            and DPV E1 ORFs and the nucleotide-binding domains of the simian
            virus  40 large T antigen suggesting potential homology between the
            two regions.)  The L1 ORF's of DPV and BPV-1 indicate further
            homology between DPV and BPV-1 in protein-protein interactions in
            the capsid.  Groff et al. (J. Virol. 56: 85-91) suggest that the
            similarity in these regions may be indicators of host viability.

            Groff et al. (J. Virol. 56: 85-91) identified four areas of
            significant difference between the DPV and BPV-1 genomes: one, E5
            and E8 ORFs are absent in DPV, two, the presence of a unique L3 ORF
            a in the deer papillomavirus, three, the intergenic NCR-1 region in
            BPV-1 is substantially larger than that of DPV while the NCR-2 in
            DPV lacks an ORF and is 253 basepairs longer than that of BPV-1,
            and four, transformation mechanisms.

            Draft entry and sequence in computer readable form for [1] kindly
            provided by D.E.Groff, 24-JUN-1986.
            The sense strand is shown below.  The open reading frames were
            located by computer analysis and comparison with BPV-1.
BASE COUNT    2189 a   1874 c   2105 g   2206 t
ORIGIN      3 bp upstream of HpaI site; 0.65 mu.
        1 gttaacaata accagacctt agccgttttt gggtgagcgg gaaagatggg ttacaggttc
       61 TATAAAgca ccacaccgca caaggttgct atcacttgtc actctgctca gacccttcct
  E6 orf start ->
```

```
signal ->
  121 tctgcATGtc tgctgattac tatgaacatc tatactgtgt attttgctac tgtgttcttg
    E6 cds ->
  181 gaaaggtgga agctcgccga tgctatgaca aaaaaattag aacagtggta agaggagggc
  241 tcagatgtgc agtttgcact gcatgcttgg aaaaagggct ctatctggaa agagtgctga
  301 atgcgcctca acctgtatat cagGGgtcca ttgaagagcc tgatcctttc attcaaaaag
                         5' sj /\
  361 cctgcataag atgcatgtac tgtgggggaa tactgacccg tgacgaaaag gacaggcaca
  421 gatattttga agagctttac gTGAtattca ggaatcaggt tcttggcaga tgctacacct
                    E7 orf start ->
  481 gtactaggcA TGgcatgtgc tcggcccctt accgggcgaa cgctaccggc TGAtgaatca
        E7 cds ->                                          <- E6 end
  541 ccttgcttaa cattgatttt ggAGccagtc tcgggagaag cagccaagaa cagtacacca
                         /\ 3' sj
  601 gtcgttgtgg ataagcctgg aaaaccgccc cctaaacgcc accgaagaca gtataatgtg
  661 actgtttcct gcaacgactg tgacaagcgt ctgaacttct ctgtcaaaac tacctgcagc
  721 acaatactca ccctgcagca actccTGAca gaggacctgg atttcctgtg ttctttctgt
                    E1 orf start ->
  781 gaggccaaga ATGgaTAAag aaaatgcagG Tagctctggg gttggggggg attcttttat
            E1 cds ->       <- E7 end
                             5' sj /\
  841 cctctttgag gcagaatgct cagatacaga ttctgaatca cctgcccaag GTgaatctac
                                                           5' sj /\
  901 tgatgaggat ttactagata atgccactgc cgttccggga aaccacctgg agctcttcca
  961 aactcaggaa aaagaggcgg gagaaagaca gatttcaatt ttgaaaagaa aactgtgttt
 1021 aagcccttgc tctgctgact ctgaggtgga gcagctcaag tcctgggctt gctgtcataa
 1081 gtatcacacc tcggaagcga atcccgttgt tagacgcagg ctttttcgaaa gaggtgatcc
 1141 aggcggtgct aacacacctg tgaaccatga agctgacaat ttttctccgt caggactgca
 1201 ggtacagtct ggggaaaata ggtggagcca ggaaaaggga aaaggggggag tttcgcccgt
 1261 gcctagctca gctgagccaa atatgccgc ctgcatacag aaattgttca agactctcta
 1321 catcgcctcc catggggaga tcactcgtgt attccaaagt aataagactg ttaaccatca
 1381 gtgggtgatt ctggcatatg gagttagtga ggtgttgtat tctgctagct ttgatctctt
 1441 tggtaaacag tgtaactgcc tgcaaacgtc cagaaaggtt catgaaaaag ggagtatttc
 1501 tgtttaccgt tgtatgttca atgttgccaa agtagagat acagtgcaga aattaatgac
 1561 cacaattctg aatgttaccg cgggcaacct cctcctacag cccctaaaa tcagaggtct
 1621 cgggcctgcc ctattctggt ttaagctcac actgtcacct gctaccttaa cccatggtac
 1681 cacaccggaa tggatacagc aggcaactaa tgttgccagc aatactggag aggcggctaa
 1741 atttgatttta ggaactatgg tgcagtgggc ttatgaccat ggtttcacgg aggagtcgaa
 1801 aattgcttat gaatatgctc tgtgtgctgg gagcgactgc aatgccaaag cattttttggc
 1861 aagcactagc caggcccgtt tggtcaaaga ctgctgcacc atggtgagac attacctgcg
 1921 tgctgaggta caggccctga caatgtcagg ttaTATAAAA aggcggtgtg atcaaactgc
                         signal ->
 1981 aggaagtggc agctggctct ctatcatgaa tttgcttaaa tatcatggga tagaacatat
 2041 acagtttgtg aatgcattaa agccttggtt aaaaggcatt cccaaatata actgcattac
 2101 aattgttggc ccgcctaaca gtgggaagtc actcctttgc aactctttga tagcatttct
 2161 tggcggcaag gtgcttacat ttgcaaacca ccacagccac ttttggctcg cacccttagc
 2221 ggactgccga gttgctttaa ttgatgatgc taccacagct tgctggaggt actttgacac
 2281 acacctcaga aatgtgttgg atggctaccc attcggtatt gatagaaaac acaataccgc
 2341 tgttcaaatg aaagcccctc ccctcttagt aaccagTAAT ATTGatgtgc atgcagagga
                         signal ->
 2401 aaagtatttc tattcgcaca gcagagttaa gccgttttac ttcaaggagc cgtgccctgc
 2461 ttcagacaat ggtgagccta tgttttctat aactgatgct gattggaaac atttctttga
 2521 aaggctatgg ggacgtttag accTGAgcga tcaagaggac gaggttgacg acgATGagtg
                    E2 orf start ->                       E2 cds ->
 2581 cagccaaaga acagttactt gcagcgcaag aaacgcaaat gacattaatT GAaaaagata
                                                           <- E1 end
 2641 gcacagattt gaaagatcac atagactctt ggggtccggt caggagagag catggtttgc
 2701 tttatgctgc cagacacaaa ggcttgattt ggctcggttt gaaccctgtg ccaccatgct
 2761 ctgtgaagtg cttagaagct cggcaagcaa ttgagatgca gcttctgggg aacagcttaa
 2821 aggagagccc atggtgcaat gagccatggt cactgtgtga cttaagctgg ggacgctatc
 2881 aagcgcctcc agcagaaact ttgaaaaaag gcgccagact ggttgaagtg agtatgatg
 2941 ggagctccac tAATAAAact tggtataccg cttggaattc attgtacttg cgcaaaccgg
            signal ->
 3001 atgaggaggg ctgggagacg gcgactggtg gtgcagacgc agacggtctc ttctatacta
```

```
3061 ccatgtccgg tacacgggtt tattatgagc tctttgaaag agatgcagcc agatacagca
3121 ctacagggac ttggactgTG Agggataacg atcgtactta tcactcacat tctgcgccct
                   E4 orf start ->
                   NH2 terminus unknown
3181 cccactctAG agagaccatc gaaggactgt ggaactccgg gggccgtgaa agaggcagac
                   /\ 3' sj
3241 ccaccaactc gcccgaccgc gccgtgcttc acactcctcc tggaggcaac accgttcacg
3301 gtcccgtcag agcttgcgaa aaccggggtc ggtccattaa ccgcccgact ccctacagca
3361 caccacagtc cccgaggagt ggcgtggggcc ccgataccac ctccccgctg ccgagcccgg
3421 taccgcagaa cccccggtgc gtatctctac ccgacggttt tggacgaggg gaggaggata
3481 acccgccgtc gccagatcaa cacgacgTAA tccccaaccc ccagccgaaa gaaccgcggt
                                          <- E4 end
3541 ttagcctatt tggctcttca ggtgggctgc cctgtcttct aattagtgga actgggaacc
3601 aagtcaagtg ctattccttc cgcgtgaaaa gatggcatag ggacaagtat caccactgca
3661 cgaccacctg gtgggcagtc ggggagcagg gatctgaaag accaggcgat gccacagtga
3721 tcgtcacctt caaagaccaa agtcagagat caatgtttct gcagcaggtg cccttaccgc
3781 ctggtatgtc agcacatgga gtgactatga ctgttgactt tTGAgtctgt tgtgctactg
                                               <- E2 end
3841 gctgctatct gctgcagcat cgcgagtctg atactttacc tgtggaaaat ggaaacgctg
3901 catgctgtaa agcatgaacc atcctgggct gtttctgttt ctgggactgg tctttggagt
3961 gcagttgcta ctgttagtat ttattttgtt tttcttcttt gtatggtggg atcagtttgg
4021 gtgtaagtgt gaaaacttcc acatgtaaat ccctttcctg catggaggga tcaggtatgg
4081 tgtctctgtg aacacttta gatgtaaata atcatttgat ccaaagcttt gtaaatagca
4141 ataggtgcaa tttaaggagc attatttact ggcgcatcgg gcagatttct gtgtcatccc
4201 aaggagcctt ctctctctct ctagttctct cacgtggtac agcataatcc tccttccttt
4261 atttattttt gatattattc atattattca TATATAgagc tattattatc caaggtcagg
                         signal ->
4321 tactgttatc caaggtcagc aagatcagtt cccaagagtt ttcttctttc gaccctcttc
4381 gagttggact gatacggact atggactatg gactaaaact TAAgtgcgtg tgcgacgatc
                                            L2 orf start ->
4441 accagggggc cattcttgtt ttgtttttttt tctttttgttg ttttgttctc gagtaacctc
4501 aagcctttgt cagacagaac cattagtaca gcggtgccat atctttgaac tgccaaaaaa
4561 aagacacata agcctgcatg aaactttggt tgttattgtt gctgctatta ttattaggtc
4621 attggacacc agtttgggtt atattttgtt tgctgtggtc gagcttccta tttttgttac
4681 attctgtttt cataacttttt tgacaggggct gtttcagctt cccaatcaaa gtttaagtgt
4741 ctgtgtttac AATAAAtgca ccATGccacc attaaagcgt gtgaagcgtg caaatccata
                         L2 cds ->
             signal ->
4801 tgatctctac aggacgtgca agaggtggaa gtgcccactg atgtcattcc tgaaggtgga
4861 gggaaaaact gttgctgata aatactgcag tatgggcagt atgggtgtat actggcggct
4921 ggcattgcat gggagtggcc gcccaaccca aggtgggtat gttcctttga ggggaggtgg
4981 gtcctctaca tctctttcta gcaggggaag tgggtcctct acatctattt ctaggccctt
5041 tgctggaggg attcctttgg aaacgctgga aacagtgggt gcttttcgcc ctggcatcat
5101 agaggaggtt gctccgacac tggaaggtgt cctacctgat gctcctgcag tggtgactcc
5161 agaggctgta cctgtggatc agggcctgag tggcttggat gtggcaaggg aagtcacaca
5221 agaaagcctc atcacttttt tgcaaccaga agggccagat gatatagcag tgctggaact
5281 taggcccaca gagcatgatc aaaacacacct gatttccacc tcaacgcacc caaacccttt
5341 gtttcatgca cctattcagc agagcagcat taTAGcagaa acctcagggt ctgagaacat
                         L3 orf start ->
5401 atttgtgggt ggtgggggag tgggtagcac cactggggag gagattgaac tcacactgtt
5461 tggtcagcca aagactagca ctcctgaggg ccctattaac cggggtcggg gcattttcaa
5521 ctggttcaac agaacatact acacacaggt acctgtggaa gacccagacg agattgctgc
5581 tgcaggctcg tATGtctttg agaatgcatt atacgattca aaggcctaca acatgagca
             L3 cds ->
5641 gcagccgtgg ttatcgcgac cacaggatgc acctgagttt gacttccaag atgcagTGAg
                                                              <- L3 end
5701 gctacttcaa ggaccctcgg gtcgtgtggg gtggagcaga attattaggc ctacctcaat
5761 aggcacacgt tcaggggtac gggtggggcc tttgtatcac ctgcgtcagt cgttcagcac
5821 tattgatgag cctgaaacaa tagaattgat tcctagcact gttgatgaag gaggaggtttt
5881 gactggagta cctgagtcgg ctgaaggacc tgatgcagaa tattcagata ttgatcttca
5941 aagtatagga agcgatgagc cccttttagg acaggcatc atctatcctt tggtgggcgg
6001 aggacaaata ttcctgtgca tgcacagggc cccagtgggt tggtcctcag ggacatacat
6061 taatcatgaa ggacaaagta gggatgatgg tgagtacgtg atagacaatg gaggacaatc
6121 gaacatcacg cctactgtag ttattgatgg atcaattgct ttgtctttgg aatattttag
```

```
6181 gcattactac ctgcacccaa gtctgcTAAg acgcaagcgc aaacgcaacc caatatttat
                         L1 orf start ->
6241 cTGAtgtttt gcagATGgcg ttctggcagc ctggtcaagc gctatacctg cctccaacac
          <- L2 end
              L1 cds ->
6301 ctgtgacaaa ggttctttgc tcagagcagt acattaacgt acgggatata ttttatcatg
6361 gggaaacaga gcgcatgctc accagtgggt ccatcctttc tcttgaggtg acacagaagc
6421 acacgactgt ccctaaagtg tctccaaatc aatacaggt gtttagagtc gcattacctg
6481 atcctaatca gtttgcttta cctgataagg cccttcataa ccctagtaaa gaaaggttag
6541 tgtgggcagt tgtgggtgtc caggtgtcaa gagggcagcc cctgggtggg gaggtaagag
6601 ggcattccta cttcaatact ttcttggatg cagagaatgt cagtaaaaaa gtaactgctc
6661 agggcaccga tgaccgtaag caagcaggta tggatacaaa acagcaacag gtgctgatgc
6721 ttggatgcac tcccgctata ggtgagtact ggacaaaggc tcgtccatgt gtaacggaca
6781 gaccagatgc tgggtcatgt cctcctattg aattaaaatt aagctttata gaggatggag
6841 atatgatgga cataggattt ggggctgcga actttaaaga gttaaatgcc acaaagtcag
6901 atctacctct ggacattgcc aactccatct gtttgtatcc tgattactta aagatgactg
6961 aggaggcggc aggcaatagc atgttttttt ttgccaggaa ggaacaagtt tatgtaagac
7021 acatttggac cccgtggggt actgacaaag aactcccacc cgaggcctat tatctgaagc
7081 caccgggga gatggaactc aaaatgccaa gtgtttttctt tgcaagtcca agtgggagtt
7141 tagtatctac agatggccag ctattcaatc ggccatattg gatactgaga gctcagggaa
7201 tgaacaatgg tgtatgctgg aataataccc tatttgtgac agtgggagac aacacaaggg
7261 gcagcacact gaccatcacg gtccctaaca atgatgagcc tttgacggag tatgatacta
7321 gtaaatttaa tgtatatcaa agacatgtgg aggaatttaa gcttgcattc attctcgaac
7381 tatgctcagt ggagcttact cctgagactg tctctagtct ccaggctca atgccctcaa
7441 tcctggaaaa ctgggaaatt aacctgcaac ctccaacatc ctctgtgtta gaggatatct
7501 atcgctttat agattcccct gcaacaaagt gtgcagataa tgtatctccc agcaagcctg
7561 aggacccata ctctgctcat aagttttggg aggtaaactt aaaagagaaa ttatctttag
7621 atttggacca gtttcccta ggtcgcctcg tcctacagtt tgactgtcgt ctagacaggc
7681 ttttacctca aaaagaccac ttcacgtacc ctgaaaagcg gTATAAAcgg cacatgagga
                                          signal ->
7741 taacggggac ggtgagaaaa gtgcttctgt acatatgctt tagtttaaat tccTAATAAA
                                                          <- L1 end
                                              signal ->
7801 cctgagtttc ttgaatgtgt ctagtcatgt tccaattgag tgtgtcatgt cctgcttgta
7861 ttcctccgaa gtttgcacca cactcccggt gtcggaatga agctgataat acccttgaga
7921 cgcactgtca gcaactttta attcacagca ggcctcttgg ctcggcacat tgacgccgca
7981 gacagctgca tttgtctgga ttagggatac cgctggcgtt cAATAAAgca cttcgtctgc
                                              signal ->
8041 aactttcacg gtaccggcgc gcacgtcgag tccgaaccgc tgacggtggt tatggtaagt
8101 aggcggtcaa tagaggcgga actgaggcgg aacttataga ctataagccg gacatgtcct
8161 tgcaagttca atgcctatag aaccgtcctc ggtgctgtac aaggcgctaa accgttcgcg
8221 gtgtagtgtt taatgcatgc catcgcgagc ggtctcagct gatcttctta agttcccagg
8281 cgtcttgatt ggacggttcc aaatgtactg aagtactttg caggcatcgc tctcggtacg
8341 tgagaatctt gatttttcag tgtgaatgat tgtt
```

```
LOCUS       BPV1           7945 bp ds-DNA   Circular  VRL       30-SEP-1988
DEFINITION  Bovine papillomavirus type 1 (BPV-1), complete genome.
ACCESSION   J02044
KEYWORDS    complete genome; open reading frame.
SOURCE      Bovine papillomavirus type 1 DNA from cow, isolate 307.
REFERENCE   1  (bases 1 to 7945)
  AUTHORS   Chen,E.Y., Howley,P.M., Levinson,A.D. and Seeburg,P.H.
  TITLE     The primary structure and genetic organization of the bovine
            papillomavirus type 1 genome
  JOURNAL   Nature 299, 529-534 (1982)
REFERENCE   2  (bases 1 to 7945)
  AUTHORS   Danos,O., Engel,L.W., Chen,E.T., Taniv,M., Howley,P.M.
  TITLE     Comparative analysis of the human type 1a and bovine type
            1 papillomavirus genomes.
  JOURNAL   J. Virology 46, 557-566 (1983)
COMMENT     Full genomic sequences exist for BPV-1, BPV-2 and BPV-4,
            EEPV, DPV, and partial genomes for BPV-3, BPV-6 and RPV.
            The bovine papillomaviruses can be classified into two
            groups, subgroup A (BPV-1, BPV-2 and BPV-5) and subgroup
            B (BPV-3, BPV-4 and BPV-6), based on the tissues they
            infect (Jackson et al. Mol. Carc. 4: 382-387). The
            subgroup A viruses infect both dermal fibroblasts and
            squamous epithelial cells (Coggins et al. Vir. 143:
            603-611). BPV-1, isolated from a Swedish cow, has been
            linked specifically to frond-like fibropapillomas which
            occur on the teats, penis, and nose and equine sarcoida,
            a benign, naturally occurring fibroblastic tumor in horses
            (Coggins et al. Vir. 143: 603-611; Amtmann et
            al. J. Virol. 61: 3394-3400). In addition to differences
            in host tissue restriction, several other characteristics
            distinguish the subgroups of the bovine papillomaviruses.
            First, subgroup B viruses have smaller genomes (7.2 kB)
            than subgroup A viruses (7.9 kB).  Second, the analogous
            position of the subgroup A E6 ORF is occupied by the
            subgroup B E8 ORF (Jackson et al. Mol. Carc. 4: 382-387).
            This coding region encodes a protein which strongly
            resembles the E5 transforming protein of the subgroup A
            viruses (Jackson et al. Mol. Carc. 4: 382-387).

            This sequence has been corrected as stated in [2]: G
            changed to C at position 7305, and G deleted at position
            7585.
BASE COUNT     2270 a   1714 c   1886 g   2075 t
ORIGIN      3 bp upstream of unique HpaI site.
        1 gttaacaata atcacACCAT CACCGTTttt tcaagcggga aaaaaTAGcc agctaacTAT
                    E2 bind ->                      E6 orf start -> signal ->
       61 AAAaagctgc tgacagaccc cggttttcac ATGgacctga aacctttgc  aagaaccaat
                                          |-> mRNA start site from P(89) promoter
                                  E6 cds ->
      121 ccattctcag ggttggattg tctgtggtgc agagagcctc ttacagaagt tgatgctttt
      181 aggtgcatgg tcaaagactt tcatgttgta attcgggaag gctgtagata tggtgcatgt
      241 accatttgtc ttgaaaactg tttagctact gaaagaagac tttggcaagg tgttccagta
      301 acagGTgagg aagctgaatt attgcatggc aaaacacttg ataggctttg cataagatgc
        5' sj /\
      361 tgctactgtg ggggcaaact aacaaaaaat gaaaaacatc ggcatgtgct ttttaatgag
      421 cctttctgca aaaccagagc taacaTAAtt agaggacgct gctacgactg ctgcagacAT
                          E7 orf start ->                         E7 cds ->
      481 Ggttcaaggt ccaaataccc aTAGaaactt ggatgattca cctgcAGgac cgttgctgat
                                 <- E6 end                    /\ 3' sj
      541 tttaAGtcca tgtgcAGgca cacctaccag tctcctgca gcacctgatg cacctgattt
           /\ 3' sj    /\ 3' sj
      601 cagacttccg tgccatttcg gccgtcctac taggaagcga ggtcccacta ccctccgct
      661 ttcctctccc ggaaaactgt gtgcaacagg gccacgtctga gtgtgattct tgactgtctg
      721 ctgtggaaac tgcggaaaag agctgacttt tgctgtgaag accagctcga cgtccctgct
      781 tggatttgaa caccttttaa actcagattT AGacctcttg tgtccacgtt gtgaatctcg
```

```
                                  E1 orf start ->
      841 cgagcgtcAT GgcaaacgaT AAagGTagca attgggattc gggcttggga tgctcatatc
          E1 cds ->             <- E7 end                      |-> mRNA start
                        5' sj /\                                 site from
                                                              P(890) promoter
      901 tgctgactga ggcagaatgt gaaagtgaca aagagaatga ggaacccggg gcaggtgtag
      961 aactgtctgt ggaatctgat cggtatgata gccaggatga ggattttgtt gacaatgcat
     1021 cAGtctttcA Gggaaatcac ctggaggtct tccaggcatt agagaaaaag gcgggtgagg
          /\ 3' sj /\ 3' sj
     1081 agcagatttt aaattTGAaa agaaaagtat tggggagttc gcaaaacagc agcggttccg
               E8 orf start ->
     1141 aagcatctga aactccagtt aaaagacgga aatcaggagc aaagcgaaga ttatttgctg
     1201 aaaATGaagc taaccgtgtt cttacgcccc tccagGTaca ggggaggg gaggggaggc
       E8 cds ->                                5' sj /\
     1261 aagaacttaa tgaggagcag gcaattagtc atctacatct gcagcttgtt aaatctaaaa
     1321 atgctacagt ttttaagctg gggctcttta aatctttgtt cctttgtagc ttccatgata
     1381 ttacgaggtt gtttaagaat gataagacca ctaatcagca atgggtgctg gctgtgtttg
     1441 gccttgcaga ggtgtttttt gaggcgagtt tcgaactccT AAagaagcag tgtagtttc
                                                       <- E8 end
     1501 tgcagatgca aaaaagatct catgaaggag gaacttgtgc agtttactta atctgcttta
     1561 acacagctaa aagcagagaa acagtccgga atctgatggc aaacacgcta aatGTaagag
                                                                  5' sj /\
     1621 aagagtgttt gatgctgcag ccagctaaaa ttcgaggact cagcgcagct ctattctggt
     1681 ttaaaagtag tttgtcaccc gctacactta aacatggtgc tttacctgag tggatacggg
     1741 cgcaaactac tctgaacgag agcttgcaga ccgagaaatt cgacttcgga actatggtgc
     1801 aatgggccta tgatcacaaa tatgctgagg agtctaaaat agcctatgaa tatgctttgg
     1861 ctgcAGgatc tgatagcaat gcacgggctt ttttagcaac taacagccaa gctaagcatg
                /\ 3' sj
     1921 tgaaggactg tgcaactatg GTaagacact atctaagagc tgaaacacaa gcattaagca
                         5' sj /\
     1981 tgcctgcata tattaaagct aggtgcaagc tggcaactgg ggaaggaagc tggaagtcta
     2041 tcctaacttt ttttaactat cagaatattg aattaattac ctttattaat gctttaaagc
     2101 tctggctaaa aggaattcca aaaaaaaact gtttagcatt tattggccct ccaaacacag
     2161 gcaagtctat gctctgcaac tcattaattc atttttttgg tggtagtgtt ttatcttttg
     2221 ccaaccataa aagtcacttt tggcttgctt ccctagcaga tactagagct gctttagtag
     2281 atgatgctac tcatgcttgc tggaggtact ttgacacata cctcagaaat gcattggatg
     2341 gctaccctgt cagtattgat agaaaacaca aagcagcggt tcaaattaaa gctccACCCC
                                                                 E2 bind ->
     2401 TCCTGGTaac cagTAATATT gatgtgcagg cagaggacag atatttgtac ttgcatagtc
             signal ->                             |-> mRNA start site
                                                    from P(2443) promoter
     2461 gggtgcaaac ctttcgcttt gagcagccat gcacagatga atcggGTgag caaccttta
                                                            5' sj /\
     2521 atattactga tgcagattgg aaatcttttt ttgtaAGgtt atgggggcgt ttagaccTGA
                                                  /\ 3' sj
     2581 ttgacgagga ggaggatagt gaagaggATG gagacagcat gcgaacgttt acatgtagcg
       E2 orf ->                E2 cds ->
     2641 caagaaacac aaatgcagtt gatTGAgaaa agtagtgata agttgcaaga tcatatactg
                                  <- E1 end
     2701 tactggactg ctgttagaac tgagaacaca ctgctttatg ctgcaaggaa aaaagggtg
     2761 actgtcctag acactgcag agtaccacac tctgtagttt gtcaagagag agccaagcag
     2821 gccattgaaa tgcagttgtc tttgcaggag ttaagcaaaa ctgagtttgg ggatgaacca
     2881 tggtctttgc ttgacacaag ctgggaccga tatatgtcag aacctaaacg gtgctttaag
     2941 aaaggcgcca gggtggtaga ggtggagttt gatggaaatg caagcaatac aaactggtac
     3001 actgtctaca gcaatttgta catgcgcaca gaggacggct ggcagcttgc caaggctggg
     3061 gctgacggaa ctgggctcta ctactgcACC ATGGCCGGTg ctggacgcat ttactattct
                              |-> mRNA start site from P(3080) promoter
                                   -> E2 bind site
     3121 cgctttggtg acgaggcagc cagatttagt acaacagggc attactctgT AAgagatcag
                                                               E4 orf start ->
     3181 gacagagtgt ATGctggtgt ctcatccacc tcttctgatt ttAGagatcg cccagacgga
              E4 cds ->                               /\ 3' sj
     3241 gtctgggtcg catccgaagg accTGAagga gaccctgcag gaaaagaagc cgagccagcc
                   E3 orf start ->
```

```
                           NH₂ terminus unknown
3301 cagcctgtct cttctttgct cggctccccc gcctgcggtc ccatcagagc aggcctcggt
3361 tgggtacggg acggtcctcg ctcgcacccc tacaattttc ctgcaggctc gggggggctct
3421 attctccgct cttcctccac cccgtgcagg gcacggtacc ggtggacttg gcatcaaggc
3481 aggaagaaga ggagcagtcg cccgactcca cagaggaaga accagTGAct ctcccaaggc
                                                         <- E2 end
3541 gcaccaccaa TGAtggattc cacctgttaa aggcaggagg gtcatgcttc gctctaattt
                <- E4 end
3601 cAGgaactgc taaccaggta aagtgctatc gctttcgggt gaaaaagaac catagacatc
          /\ 3' sj
3661 gctacgagaa ctgcaccacc acctggttca cagttgctga caacggtgcT GAaagacaag
                                               E5 orf start ->
                                               NH₂ terminus unknown
3721 gacaagcaca aatactgatc acctttggat cgccaagtca aaggcaagac tttctgaaac
3781 atgtaccact acctcctgga atgaacattt ccggctttac agccagcttg gacttcTGAt
                                                              <- E3 end
3841 cactgccatt gcctttctt catctgactg gtgtactatg ccaaatctat ggtttctatt
3901 gttcttggga ctagttgctg caatgcaact gctgctatta ctgttcttac tcttgttttt
3961 tcttgtatac tgggatcatt ttgagtgctc ctgtacaggt ctgccctttT AAtgcctta
                                                         <- E5 end
4021 catcactggc tattggctgt gttttttactg ttgtgtggat ttgatttgtt tTATATActg
                                                     signal ->
4081 tatgaagttt tttcatttgt gcttgtattg ctgtttgtaa gttttttact agagtttgta
4141 ttccccctgc tcagatttta tatggttTAA gctgcagcAA TAAAaATGag tgcacgaaaa
                L2 orf start ->         L2 cds ->
                              signal ->
4201 agagtaaaac gtgccagtgc ctatgacctg tacaggacat gcaagcaagc gggcacatgt
4261 ccaccagatg tgatacgaaa ggtagaagga gatactatag cagataaaat tttgaaattt
4321 gggggtcttg caatctactt aggagggcta ggaataggaa catggtctac tggaaggggtt
4381 gctgcaggtg gatcaccaag gtacacacca ctccgaacag cagggtccac atcatcgctt
4441 gcatcaaatag gatccagagc tgtaacagca gggacccgcc ccagtatagg tgcgggcatt
4501 cctttagaca ccccttgaaac tcttggggcc ttgcgtccag gggtgtatga ggacactgtg
4561 ctaccagagg cccctgcaat agtcactcct gatgctgttc ctgcagattc agggcttgat
4621 gccctgtcca taggtacaga ctcgtccacg gagaccctca ttactctgct agagcctgag
4681 ggtcccgagg acatagcggt tcttgagctg caacccctgg accgtccaac ttggcaagta
4741 agcaatgctg ttcatcagtc ctctgcatac cacgcccctc tgcagctgca atcgtccatt
4801 gcagaaacat ctggtttaga aaatattttt gtaggaggct cgggtttagg ggatacagga
4861 ggagaaaaca ttgaactgac atacttcggg tccccacgaa caagcacgcc ccgcagtatt
4921 gcctctaaat cacgtggcat tttaaactgg ttcagtaaac ggtactacac acaggtgccc
4981 acggaagatc ctgaagtgtt ttcatcccaa acatttgcaa acccactgta tgaagcagaa
5041 ccagctgtgc ttaagggacc tagtggacgt gttggactca gtcaggttTA TAAAcctgat
                                                     signal ->
5101 acacttacaa cacgtagcgg gacagaggtg ggaccacagc tacatgtcag gtactcattg
5161 agtactatac atgaagatgt agaagcaatc ccctacacag ttgatgaaaa tacacaggga
5221 cttgcattcg tacccttgca tgaagagcaa gcaggttttg aggagataga attagatgat
5281 tttagtgaga cacatagact gctacctcag aacacctctt ctacacctgt tggtagtggt
5341 gtacgaagaa gcctcattcc aactcaggaa tttagtgcaa cacggcctac aggtgttgta
5401 acctatggct cacctgacac ttactctgct agcccagtta ctgaccctga ttctacctct
5461 cctagtctag ttatcgatga cactactact acaccaatca ttataattga tgggcacaca
5521 gttgatttgt acagcagtaa ctacaccttg catccctcct tgttgaggaa acgaaaaaaa
5581 cggaaacatg ccTAAtttt tttgcAGATG gcgttgtggc aacaaggcca gaagctgtat
             L1 orf start ->        /\ 3' sj
                           L1 cds ->
                     <- L2 end
5641 ctccctccaa cccctgtaag caaggtgctt tgcagtgaaa cctatgtgca aagaaaaagc
5701 attttttatc atgcagaaac ggagcgcctg ctaactatag acatccata ttacccagtg
5761 tctatcgggg ccaaaactgt tcctaaggtc tctgcaaatc agtatagggt atttaaaata
5821 caactacctg atcccaatca atttgcacta cctgacagga ctgttcacaa cccaagtaaa
5881 gagcggctgg tgtggccagt cataggtgtg caggtgtcca gagggcagcc tcttggaggt
5941 actgtaactg ggcaccccac ttttaatgct ttgcttgatg cagaaaatgt gaatagaaaa
6001 gtcaccaccc aaacaacaga tgacaggaaa caaacaggcc tagatgctaa gcaacaacag
6061 attctgttgc taggctgtac ccctgctgaa gggaatatt ggacaacaac ccgtccatgt
6121 gttactgatc gtctagaaaa tggcgcctgc cctcctcttg aattaaaaaa caagcacata
6181 gaagatgggg atatgatgga aattgggttt ggtgcagcca acttcaaaga aattaatgca
```

```
6241 agtaaatcag atctacctct tgacattcaa aatgagatct gcttgtaccc agactacctc
6301 aaaatggctg aggacgctgc tggtaatagc atgttctttt ttgcaaggaa agaacaggtg
6361 tatgttagac acatctggac cagaggggGc tcggagaaag aagcccctac cacagatttt
6421 tatttaaaga atAATAAAgg ggatgccacc cttaaaatac ccagtgtgca ttttggtagt
              signal ->
6481 cccagtggct cactagtctc aactgataat caaatttta atcggcccta ctggctattc
6541 cgtgcccagg gcatgaacaa tggaattgca tggaataatt tattgttttt aacagtgggg
6601 gacaatacac gtggtactaa tcttaccata agtgtagcct cagatggaac cccactaaca
6661 gagtatgata gctcaaaatt caatgtatac catagacata tggaagaata taagctagcc
6721 tttatattag agctatgctc tgtggaaatc acagctcaaa ctgtgtcaca tctgcaagga
6781 cttatgccct ctgtgcttga aaattgggaa ataggtgtgc agcctcctac ctcatcgata
6841 ttagaggaca cctatcgcTA TATAgagtct cctgcaacta aatgtgcaag caatgtaatt
                        signal ->
6901 cctgcaaaag aagacccta tgcagggttt aagtttttgga acatagatct taaagaaaag
6961 ctttctttgg acttagatca atttcccttg ggaagaagat ttttagcaca gcaaggggca
7021 ggatgttcaa ctgtgagaaa acgaagaatt agccaaaaaa cttccagtaa gcctgcaaaa
7081 aaaaaaaaaa AATAAAgct aagtttcTAT AAAtgttctg taaatgtaaa acagaaggta
                       <- L1 end
         signal ->          signal ->
7141 agtcaactgc acctAATAAA aatcacttaa tagcaatgtg ctgtgtcagt tgtttattgg
           signal ->                             |-> mRNA start site
                                                    from P(7185)
                                                    promoter
7201 aACCACACCC GGTacacatc ctgtccagca tttgcagtgc gtgcattgaa ttattgtgct
    -> E2 bind
7261 ggctagactt catggcgcct ggcaccgaat cctgccttct cagccaaaat gaataattgc
7321 tttgttggca agaaactaag catcaatggg acgcgtgcaa agcACCGGCG GCGGTagatg
                                                      -> E2 bind
7381 cgggGTaagt actgaatttt aattcgACCT ATCCCGGTaa agcgaaagcg acacgctttt
  5' sj /\                           -> E2 bind
7441 ttttcacaca tagcgggACC GAACACGTTa taagtatcga ttaggtctat ttttgtctct
                        -> E2 bind
7501 ctgtcggaAC CAGAACTGGT aaaagtttcc attcgtctg ggcttgtcta tcattgcgtc
              -> E2 bind
7561 tctatggttt ttggaggatt agacgggccA CCAGTAATGG Tgcatagcgg atgtctgtAC
                                 -> E2 bind                        ->
7621 CGCCATCGGT gcACCGATAT AGGTttgggg ctccccaagg gactgctggg atgacagctt
   E2 bind        -> E2 bind
7681 catattatat tgaatgggcg cataatcagc ttaattggtg aggacaagct acaagttgta
7741 acctgatctc cacaaagtac gttgccggtc ggggtcaAAC CGTCTTCGGT gctcgaAACC
                                                 -> E2 bind
                                                 ->    repeat region
7801 GCCTTAaact acagacaggt cccagccaag taggcggatc aaaacctcaa aaaggcggga
7861 gccaatcaaa atgcagcatt atattttaag ctcACCGAAA CCGGTaagta aagactatgt
                                      -> E2 bind
                                      5' sj /\
7921 attttttccc agtgaataat tgtt
```

```
LOCUS       BPV2            7937 bp ds-DNA           VRL        15-DEC-1988
DEFINITION  Bovine papillomavirus type 2 (BPV-2), complete genome.
ACCESSION   M20219 M19551
KEYWORDS    complete genome.
SOURCE      Bovine papillomavirus type 2 DNA.
REFERENCE   1  (bases 1 to 7937)
  AUTHORS   Groff,D.E., Mitra,R. and Lancaster,W.D.
  JOURNAL   Unpublished (1986) Georgetown Univ. Wasington D.C. 20007-2197
COMMENT     Full genomic sequences exist for BPV-1, BPV-2 and BPV-4,
            EEPV, DPV, and partial genomes for BPV-3, BPV-6 and RPV.
            The bovine papillomaviruses can be classified into two
            groups, subgroup A (BPV-1, BPV-2 and BPV-5) and subgroup
            B (BPV-3, BPV-4 and BPV-6), based on the tissues they
            infect (Jackson et al. Mol. Carc. 4: 382-387). The
            subgroup A viruses infect both dermal fibroblasts and
            squamous epithelial cells (Coggins et al. Vir. 143:
            603-11). BPV-2 is specifically associated with
            fibropapillomas of the head, neck, and alimentary canal
            (Coggins et al. Vir. 143: 603-11). In addition to
            differences in host tissue restriction, several other
            characteristics distinguish the subgroups of the bovine
            papillomaviruses.  First, subgroup B viruses have smaller
            genomes (7.2 kB) than subgroup A viruses (7.9 kB).
            Second, the analogous position of the subgroup A E6 ORF
            is occupied by the subgroup B E8 ORF (Jackson et al. Mol.
            Carc. 4: 382-387). This coding region encodes a protein
            which strongly resembles the E5 transforming protein of
            the subgroup A viruses (Jackson et al. Mol. Carc. 4:
            382-387).

            Draft entry and printed copy of sequence for [1] kindly
            provided by D.E.Groff and W.D.Lancaster, 30-MAY-1988).
BASE COUNT     2252 a   1711 c   1923 g   2051 t
ORIGIN
        1 gttaacaata atcacaccat caccgttttt tcaagcggga aaaaaaaagc cagctaacta
       61 taaaaagctg ctgacagacc cTGAttccac ATGgacctgc aaagtttttc cagaggcaat
                    E6 orf start ->       -> E6 cds
      121 cctttctcag gattggcctg tgtttggtgc agggagcctc tcacagaagt tgatgctttt
      181 aggtgcatga taaaagactt tcatgttgta taccgagatg gtgtgaaatt tggtgcatgt
      241 accacttgtc ttgagaactg cttagataaa gaaagaagac tgtggaaagg tgtgccagta
      301 acaggtgagg aagctcaatt attgcatggc aaatcccttg ataggctttg cataagatgc
      361 tgctactgtg ggggaaaact aaccaaaaac gagaagcagc ggcatgtgct ttataatgag
      421 ccttttttgca aaacgagatc taacataaTA Agaggacgct gctacgactg ctgcagacAT
                          E7 orf start ->                     E7 cds ->
      481 Ggttcaaggt ccaactaccc aTAGaaactt ggatgattca cctgcaggac cgttgctgat
                                <- E6 end
      541 tttaagtcca tgtgcaggca cacctaccag ggttcctgca gcacctgatg cacccgattt
      601 cagacttccg tgccatttcg gccgtcctac taggaagcga ggtccctcta cgcctccgct
      661 ttcctctccc ggaaaagtgt gtgcaacagg ccacgtcga gtgtactctg tgactgtctg
      721 ctgcggacac tgcggaaagg accttacatt tgctgtcaag actggctcta cgaccttgct
      781 gggcttcgaa cacctatTAA actcagattt ggacctgttg tgtccccgtt gcgaatctcg
                    E1 orf start ->
      841 cgagcgtcAT GgcaaacgaT AAaggtagca attgggattc ggcattggga tgctcatatc
                E1 cds ->              <- E7 end
      901 tgctgacaga ggcagaatgt gaaagtgaca agagaatga ggaacctggg gcaggtgtag
      961 aactgtctgt ggaatccgat cggtatgata gccaggatga ggatttcctt gacaatgcat
     1021 cagtatttca gggaaatcac ctggaggtct tccaggcttt agaaaaaaag gcgggtgagg
     1081 agcagctttt aaattTGAaa agaaaagtgt tggggagttc ggaaaacagc agcggttcag
                    E8 orf start ->
     1141 aggcatcaga gactccagct aagaggcaga aagccggagc aaaaagaaga ttattttcag
     1201 aaaATGaagc taaccgtgtt cttacgccct gcaggtaca gggaggggaa tggaggcaag
           E8 cds ->
     1261 gatttaatga ggaccaggca attagtcatc gacttctgca gctcgtcaaa tctaaaaatg
     1321 ctacagtttt taagctgggg cttttttaagt ctttatttct gtgcagcttt cacgatctta
     1381 ctagactgtt taagaatgat aaaaactacta atcagcagtg ggtgcTAGct gtgtttggga
```

```
                                                            <- E8 end
1441 ttgcagaggt gttttttgag gccagtctcg aactcctaaa gaagcaatgc agttttgtac
1501 aaatgcagaa aaggtcacat gaaggaggga cgtgtgcagt atatttatta tgctttaaca
1561 cagctaaaag cagagagaca gtgcggaacc tcatggcaaa catgctaaat gtgagagaag
1621 agtgcctact gatgcagccc cctaaaatta gaggcttaag tgcagcctta ttttggttca
1681 aaagtagttt gtcgcccgct actctgaagc atggtgcttt acctgagtgg atacgggcgc
1741 aaactacact gcacgacagc ctggcgacgg agaagtttga cttcggaacg atggtgcaat
1801 gggcgtatga tcacaaatat gctgaggaat caaaaatagc ttatgaatat gctttggcag
1861 caggttctga tagcaatgca cgagctttct tagcaactaa cagccaagct aagcatgtaa
1921 aggactgtgc aactatggtg aggcactatt taagagctga aacacaagcc ttaagtatgc
1981 ctgcatatat taaaactaga tgcaagctag caactgggga aggaagctgg aagtctattc
2041 tcacattctt taactatcag aacattgaat taatcacctt tattaatgcc ttgaaacttt
2101 ggctgaacgg tattccaaaa aaaaactgcc tggcattcat tggccctcca aagacaggca
2161 agtctatgct gtgcaactca ttaattcatt tcttaggggg aagtgtttta tcctttgcca
2221 atcataaaag tcacttttgg cttgcttcct tagcagatgc tagagcagct ttagtggatg
2281 atgctactca tgcctgctgg aggtactttg acacatacct cagaaatgca ttggatggct
2341 accccgtcag tattgataga aagcacaaag cagcggttca aattaaagct ccaccccttt
2401 tagtaaccag TAATATTgat gtgcaggcag aggaaagata cttgtacttg catagccggg
            signal ->
2461 tgcaaacctt tcgtttcgaa cagccatgca cagatgaatc gggtgagcaa ccttttacta
2521 ttactgatgc agattggaaa tcttttttttg taaggttatg ggggcgctTA Gacctggttg
                                              E2 orf start ->
2581 acgaggagga ggatagtgaa gaggATGgag acagcatgcg aacgtttaca tgtagcgcaa
                 E2 cds ->
2641 gaaacacaaa tgcagttgat TGAgaaagct agtgataagt tgcaagatca tatactgtat
                 <- E1 end
2701 tggactgctg ttagaactga gaacacactg ctctatgctg caaggaaaaa aggggtgact
2761 gtgcttggac actgtagagt accacactct gtagtttgtc aagaaagagc aaagcaggca
2821 attgaaatgc aattgtcttt gcaggagtta agcaaaactg agtttgggaa tgagccatgg
2881 tgtttgctag atacaagctg ggaccgatat atgtcagagc ctaaacggtg ctttaagaaa
2941 ggcgccaggg tggtagaggt ggagtttgat ggaaatgcaa gcaatacaaa ctggtacact
3001 gtctacagca aactgtacat gcgcacagag acggctggc agcttgcgaa ggctggggct
3061 gacggaactg ggctctacta ctgcaccatg gccggtgctg acgcattta ctattcgcgc
3121 tttggtgagg aggcagccag atttagtaca acagggcatt actctgTAAg agatcaggac
                                              E4 orf start ->
3181 agagtgtATG ctggtgtctc atccacctct tctgattta gagatcgccc agacggagtt
     E4 cds ->
3241 tccgcatcag aaggaccTGA aggagaccct gcaggaaaag aagccgagcc agcccagcct
                 E3 orf start ->
                 NH2 terminus unknown
3301 gtctcttctt tgctcggctc ccccgcctgc gtacccatca gagcaggcct cggttgggta
3361 cgggacggtc ctcgcccgca cccctaccat tttcctgcag gctcgggggg ctctcttctc
3421 cgctctgcct ccacccggt gcaggccccg gtaccggtgg acttggcacc aaggcaggaa
3481 gaagaagaaa tcaatcgcc cgactcaaca gaggaggaac ctgTGActgt accaaggcac
                                                        <- E4 end
3541 acctcTGatg ctgatggctt ccacctgctg aaggcgggac aatcatgctt tgctttaatt
         <- E3 end
3601 tcaggctctg ctaaccaggt aaagtgctat cgctttcgtg tgaaaaagaa ccatagacac
3661 cgctacgaga actgcaccac cacctccttc acagttgctg acaacggtgc TGAaagacaa
                                              E5 orf start ->
3721 ggtcaggcac agatcttgat caccttggga tcaccaggtc aaaggcaaga ctttctgaaa
3781 catgtcccac tacctcctgg aatgaacatt tccggcttta cggccagctt ggactttTAA
                                                           <- E2 end
3841 tcactgccat ttgtttttttt catatctcgt ctaggcatac tATGccgaat ctatggtttc
                 E5 cds ->
3901 tattgttctt gggactagtt gctgcaatgc aactgctgct gttactgttt ctgctgctat
3961 ttttcttggt atactgggat cattttgagt gctcctgtac aggtctgccc tttTAAtgca
                                                           <- E5 end
4021 tcactggtta ctggctttgt tgttactatt atgtggattt gatttgttta ctgtttttgta
4081 tgaagtgttc ttattttTGA ttgtgttgct gtttgtacat tttttactgg agtttgtatt
          L2 orf start ->
4141 cccctactc agattttttt tggtataaag ctgcagcaat aaacATGagt gcacgaaaaa
                                              L2 cds ->
4201 gggtgaaacg tgcaaatgtc tatgacctgt acaggacttg caagcaagcg ggcacctgtc
```

```
4261 caccagatgt gatacctaag gtagaaggtg acactatagc agacaagatt ttaaaattag
4321 gagggcttgc aatttatttg gggggcctag gtattggaac atggtctaca ggaagagtgg
4381 ctgcaggagg atcacctagg tatgtaccct taagaacctc tggatccact acaagcctgg
4441 catctgtagg atccagggct ggtgcagcca ctggcactcg cagcagcatc acaggaatcc
4501 cccttgacac cctagaaact attggggctc ttcgtcctgg agcatatgaa gacactgtgc
4561 tcccagaggc ccctgctatt gtcacccctg atgctgtacc tgcggacaca gggatagatg
4621 gcctttctat aggcactgac tcttccactg aaactctaat cacattatta gagcctgagg
4681 gtcctgaaga cgtggcagtc ttggagctgc aacctctaga ccatgcaaat tggcaagtta
4741 gcaatgctgt tcatcagggc tctgcatacc acgccctct gcagctgcag tcctccattg
4801 cagaaacatc tggactagaa aatatttttg taggaggggc tgggttaggg gatacaggcg
4861 gagagaacat agagctcaca tttttttggtt ccccacgcac aagtaccccc cgtaacctgc
4921 ctcaaactgc acggggcatc ttgaactggt ttagcaaaag atactcacaca caaatacccca
4981 cagaagaccc tgatgtattt tcatcacaga cattttcaaa cccattgtat gatcctgagc
5041 ctgcagtgct aaaaggtccc agtgccgtg tggggctaag ccaagtgtat aggcctgact
5101 ttattgaaac acggggtggg ggtcaggtgg gcccacagct gcatgtcagg tactccttaa
5161 gcactatcac agaagatgtg gaagccatac ctatagcagt tgatgaagac acacaagggc
5221 tagcatttat tcctttacat gaagaaccag gggactttga agaaattgag ctagatgatt
5281 taggtgaaga gcacgccttg ctccccaagt catatactgc acctattggt agtggagttc
5341 gtagggcgct cattccaggt caaggcttca gtgcaacacg gcccacaggt gtggtaacct
5401 atggctcacc tgacatgtac cctgctagcc cagttggccc tgactcgaca tcccctagcc
5461 tagttatcga tgacaacaca acaacaccaa taatcattat tgatggccac acagtggatc
5521 tgtatagcaa taactatagc ttgcatccct ccttgttgag gaaaagaaaa aaacggaaac
5581 atgccTAAtt ttttttgcag ATGgcgttgt ggcaacaagg ccaaaagctg tatctccctc
L1 orf start ->     L1 cds ->
          <- L2 end
5641 caaccccctgt aagcaaggtg ctatgcagtg aaacctatgt gcaaagaaaa agcatattct
5701 atcatgcaga aacggaacgc ctgttaactg taggacatcc atactaccaa gtcactgtgg
5761 gggacaaaac tgttcccaaa gtgtctgcta atcaatttag agttttttaaa atacagctcc
5821 ccgatcccaa tcagtttgca ttgcctgata ggactgtgca caatccaagc aaggagcgcc
5881 tggtttgggc tgtaataggg gttcaagtat ctcgtggcca accactagga ggcacagtta
5941 ctgggcaccc cacttttaat gctctgcttt atgcagaaaa tgttaataga aaagttactg
6001 cacaaacaac agatgacagg aagcaaacag gattagatgc taagcaacaa cagattctgt
6061 tgctgggctg tacccctgca gaagggggaat actggaccac agcccgtcca tgtgttactg
6121 atagactaga aaatggtgcg tgtcctcctt tagaattaaa gaacaaacac atagaagatg
6181 gagacatgat ggaaataggg tttggtgctg ctgactttaa aacactaaat gccagtaaat
6241 cagatctacc tcttgacatt caaaatgaaa tatgcctgta tccagactac ctcaaaatgg
6301 ctgaagatgc tgctggaaac agtatgttct tctttgcaag aaaagaacaa gtgtatgtaa
6361 ggcatatatg gactcggggg ggctctgaaa aagaagcacc cagtaaagac ttctacctca
6421 aaaatggtag aggtgaagaa actctaaaaa tacctagtgt gcactttggc agtcccagtg
6481 gatccttggt gtccactgat aatcaaatat ttaacaggcc ttattggcta ttcagggctc
6541 agggcatgaa caatgggatt gcatggaata atttattatt tttaactgta ggggataaca
6601 cacggggaac taaccttagt attagtgtag ctgcagatgg aaacgcattg tcagagtatg
6661 atactggcaa atttaaccta taccataggc atatggaaga atataagcta gcatttatat
6721 tggagctgtg ctctgttgag attactgcac aaacactgtc acatctgcaa ggactgatgc
6781 cctctgtgct acaaaactgg gaaatcgggg tgcaacctcc tgcttcttct attttagaag
6841 atacttatag gtacatagag tctcctgcaa ctaaatgtgc aagtaatgtt ataccaccca
6901 aagaagaccc ttatgcaggg cttaagtttt ggagcataga cttaaaagaa aagctgtctt
6961 tggacttaga ccaatttccc ttgggaagaa gattcttagc tcagcaaggg gcaggatgtt
7021 caactgtgag aaagagagct gttgcaacca gaaattccag taagcctgca aaaagaaaaa
7081 aaatcaaagc tTAAgttttt ataaatgtat aataaatgtg aaccataagg taagtcaact
               <- L1 end
7141 gcacctaata aaaatcactt aatagcaatg tgctgtgtca gttgtttatt acaACCACAC
                                                                -> E2 bind
7201 CCGGTacaca tcctgtccag catttgcagt gcatgcattg aattattgtg ctggccagac
7261 ttcatggcgc ctggcaccga tttctgctct ctcagccaaa atgaataatt gctttgttgg
7321 caagaaactg agcatcaatg ggacctgtgc aaagcaccgg aggcggatac ctgcggggta
7381 agcactgatt ttttattagA CCCATCTCGG Taaggtgaaa gcgacacgct ttttttttcgc
                       -> E2 bind
7441 ataacacggt ACCACAGTAG GTataagtat cgattatgtc tattttagtc tctctgtcgg
                 -> E2 bind
7501 aACCAGAACT GGTaaaagtt tccattacgc tttgggcttg tctagctctg cctctctatg
        -> E2 bind
7561 gtctttggag gattagacgg actACCGGTA TCGGTgtata gcggatgtct gcACCGCCAT
                          -> E2 bind                         -> E2 bind
```

```
7621 CGGTgcacga aatggattgg ggctccccaa gggattgacg ggatgacagc tttagtgtct
7681 agtttattga agcataatag gcttaattgg tgagtacaag tggcaagttg taacctgatc
7741 tccacaaagt ACCGCTCCCG GTtggggtca aACCGGCAGC GGTgctcaag gccgccttca
                -> E2 bind               -> E2 bind
7801 actagacaga cattcatgcc aagtaggcgt aaccgaaacc tgaaaaacgg cgggaaccaa
7861 tcaaaatgca ggattttatt tcaagacacg aaaccggtac gtaagtggtg tatacttttt
7921 cccatggaat aattgtt
//
```

```
LOCUS       BPV4            7265 bp ds-DNA          VRL        02-APR-1988
DEFINITION  Bovine papillomavirus type 4 (BPV-4), complete genome.
ACCESSION   X05817 D00146 X59063
SOURCE      BPV-4 DNA originally cloned in pAT153, separated from the plasmid
            vector by digestion with BamHI, recircularized by ligation.
REFERENCE   1
  AUTHORS   Patel K.R., Smith K.T., Campo S.
  TITLE     The Nucleotide Sequence and Genome Organization of Bovine
            Papillomavirus Type 4
  JOURNAL   J. Gen. Virol. 68, 2117-2128 (1987)
REFERENCE   2
  AUTHORS   Jackson M.E., Campo M.S.
  TITLE     Positive and negative E2-independent regulatory elements in the
            long control region of bovine papillomavirus type 4
  JOURNAL   J. Gen. Virol. 72, 877-883 (1991)
COMMENT     The Bovine papillomaviruses can be classified into two groups,
            subgroup A (BPV-1, BPV-2 and BPV-5) and subgroup B (BPV-3,
            BPV-4 and BPV-6), based on the tissues they infect (Jackson
            et al. Mol. Carc. 4: 382-387). In addition to differences in
            host tissue restriction, several other characteristics
            distinguish the subgroups of the bovine papillomaviruses.
            First, subgroup B viruses have smaller genomes (7.2 kB) than
            subgroup A viruses (7.9 kB).  Second, the analogous position
            of the subgroup A E6 ORF is occupied by the subgroup B E8 ORF
            (Jackson et al. Mol. Carc. 4: 382-387). This region encodes a
            protein which strongly resembles the E5 transforming protein of
            the subgroup A viruses (Jackson et al. Mol. Carc. 4: 382-387).
            BPV-4, cloned from alimentary epithelial papillomas, can
            progress to malignancy when infected cattle feed on bracken
            (Jackson et al. Mol. Carc. 4: 382-387). Because of its oncogenic
            potential, more research has focused on BPV-4 than any of the
            other subgroup B viruses.

            Replication of BPV-4 is controlled by proteins encoded in
            the E1 region.  Sequence similarity has been identified between
            the E1 C-terminals of BPV-4 and other HPVs including HPV-1
            (50.4%) (Patel et al. J. Gen. Virol. 68: 2117-2128).  The E2
            region of BPV-4 had the highest level of similarity with the E2
            C-terminus of CRPV (45.6%) (Patel et al. J. Gen. Virol. 68:
            2117-2128).  A putative metal-binding motif, common to many
            papillomaviruses, is located in the E8 ORF of BPV-4 (Patel et al.
            J. Gen. Virol. 68: 2117-2128). Areas of similarity have also been
            identified in the L1 ORF and the L2 N-terminal of BPV-4, CRPV,
            HPV-1, HPV-6, and HPV-16 (Patel et al. J. Gen. Virol. 68: 2117-
            2128).  These results indicate a closer evolutionary relationship
            between BPV-4, CRPV and epitheliotropic HPVs than between BPV-4
            and BPV-1 (Patel et al. J. Gen. Virol. 68: 2117-2128).
BASE COUNT    2245 a   1400 c   1651 g   1969 t
ORIGIN      Sp1 binding site.
        1 ACCCGCCTTG cacctggacg CTGTGTAAAG CAAAAACTAT GCAGGCACAC Ctaaaggctg
          -> Sp-1 bind            -> 5' enhancer
       61 tcaactttga agtttattg gccaCAAAat taatTGGCCG AGCCAAgttg gcaTATAAaa
                            signal ->          -> NF-1 bind        -> signal
```

```
 121 ctgttcttga atagtcacac TGGCTTTTGG CCAtaacaag ctcagacaaa gtgtACCGAT
                          -> NF-1 bind                         -> E2 bind
 181 TGCGGTCGaa actctcacgc tagGTaagtg tTGTACCTAA CAactattta ccTAGggaaa
                          5' sj /\                  E8 orf start ->
                                             -> similarity to SV-40
                                             large T antigen promoter
 241 atatcagttg catcccattc ggtcgtACCG AATCGGGTGC atatataagg agagcagtgc
                          -> E2 bind
 301 ggatttggtg catgaggcag tagctctcat cATGtctttg tggcttatct atgttttgtt
                                      E8 cds ->
 361 gctttttctgg tgtgctttta attttcttgc actgttattt gcaattattg tgtaccttct
 421 gttaatttct actattactc ggttagatgg atgggatTGA attttttattt attgtgtttta
                                      E8 end <-
 481 ttatttgctt gatttctgat ttgcttttgc tatatatagt ggagattatg cgtttggtgt
 541 tattgctgtc tttgctagat gaattagaac aggaggaaca gtaaatagaa tattgctgtg
 601 cttttacatt tgagaaggat tgtaggtgtg ggcatttaaa agcTGAcctt ccagtcttaa
                                           E7 orf start ->
 661 ttgcagtagg cgcctaagag ggtggtggtg gtataagttc aagttttttca agaaaATGaa
                                                              E7 cds ->
 721 ggggcagaac gtgacattac aggacattgc aatagaatta gaggatacaa ttagtccaat
 781 taacttgcat tgtgaagagg agattgaaac tgaggaggtg gacacccccta acccttttgc
 841 aataacagca acttgttatg cttgcgagca agtccttcgt ttagctgttg TAAcgtcaac
                                           E1 orf start ->
 901 agaaggaatt catcaactgc agcaactgct gttcgacaac ctctttctac tgtgtgcagc
 961 ttgctccaaa caagtgttct gtaaccgcag acccgagcga aATGgacccT AAagGTacta
                                                    5' sj /\
                                                E7 end <-
                                      E1 cds ->
1021 ccgtgcttga cttttattgag gaacaagcag aatgtagtca atctgatagc gagcaaggca
1081 gtgaagaaaa ttgttctgat atatctgacc taattgataa tgctgaatac gagcaaggga
1141 attctgcaga attgtttgca caacagcagg ctttcgattt ccataaagac atatgcacca
1201 caaagcgaaa cttaaaacgg tcactgcgaa acgctttttCA ATgcataacc agtcagagta
                                                  signal ->
1261 acacccctgc ttcTAGATCC GCGCCAAaaa gacgtttgct agacgacagt ggtTATAAtg
          NF-1 bind ->                                     signal ->
1321 aagatatttc tacagaagtg gtacagGTag atgaaaatgg cggctcagag ggttacggga
                          5' sj /\
1381 gcttaacctc gcaacatttg tctggtgtat gcctaaatca aggggataac ggaggggtag
1441 ataaagaaaa tgtagattgc acagcgctac tgcgagccgg tagtcgcagg gcagcgtatt
1501 tgggtgtttt taaagagaaa tttactatta gcttcactgc tttgacaAGa atctttaaaa
                                                  /\ 3' sj
1561 atgacaaaac ttgctgcaga aactgggtgg gaactgtata tagagccaga gaagaattgc
1621 tcgaggcttc aaaaacaatt ttgcaaaaat gctgtgattt tattctattg ttaacgcaca
1681 cttgtaaata tggtttcctg gcattgtttt tgctggaatt taagactgct aaaagtagag
1741 agactgtgca gcgcttgttt gagcatatat tgcagGTgga aaaagaagat atgctgttag
                          5' sj /\
1801 aacctcctaa gttaaaaagt ttacccgcgg caacctttttg gtggaaaatt cagcatagca
1861 acaactcttt taaatgggga actctacctg attggatagc tagacaaact atgatatctc
1921 atcaaatagc agatgatgag ccatttagtc tcagtgtaat ggtgcagtgg gcttatgatc
1981 ataattatac tgaagaatca acaattgcat atcattatgc aaaattagcc tcagaagatt
2041 caaatgctgc agcatttcta aaatgtaata atcaagtaaa gcatgtgaaa gaatgtgcac
2101 aaatgactag gtattataag actgcagaaa tgacagaaat gtcaatggga caatggatta
```

```
2161 aaaaatgcat tggtgagata gaagGTgtag gTGAttggaa gcaaatctgc aaatttttaa
                      5' sj /\           -> E3 orf start
                                         NH₂ terminus unknown
2221 agtttcaaaa tgtcaacttt ttatcattta tgtctgcttt aaaagatttg ttgcacagag
2281 tacccaaaag gaattgcatg gttatctgtg ggcctccaaa cacaggaaaa tctatgtttg
2341 taatgagttt tatgaaagct ttgcaaggaa aagtattatc atttgtaaat tcaaaaagcc
2401 attttttggct ccagccttta cgtggagcaa aagtagcgtg attagaTGAt gctacacgag
                                                    <- E3 end
2461 ccacatggac atactttgat acctacctca gaaatggcct agatggcacg cctgtgtcat
2521 tagatatgaa gcatagagcc cccttgcaaa tatgtttccc cccttttggta atcaccacta
2581 atgtcaatgt gatgcaggat cctgcatatt tttatttgca cagtagaatt gtatgtttg
2641 agtttccaaa tacatttcca ttagatgagg ctggaaatcc tcttttatta attgatgaac
2701 ttagttggaa atcttttttt gaaaggcttt ggacacagcT AGacctcaca gacgctgagg
                                      E2 orf start ->
2761 aagacgaagA TGgtgagcct cgaagcccgt ttcgatgctg tccaagatca gttgctacaa
        E2 cds ->
2821 gtttaTGAaa atgactctaa tacattagaa ctgtgtttac aatactgggc actcataagg
      E1 end <-
2881 agagagaatg cactttatta ttatgccagg caacaagGTa aaacaaggct agGTttgtac
                                            5' sj /\          5' sj /\
2941 acagtgcctc ccaccagagt atcagaacaa aaggctaagg atgcaatcaa gatgtactta
3001 tgtttggaaa gcctgcagaa atcagagttt gccaatcaaa gatggtcact tgtggacact
3061 agcatagaga catttaaggc gccacctgaa aacactttaa aaaaaagggg ccagcatgtg
3121 actgtcattt atgatcagaa tgcaatgaat tcaatggtat acactttgtg gaaagaagtg
3181 tattatgttg atgaaactga gacatggcat aaaaccagta gcgatttgga ttatgatgga
3241 atattttata ttgataatca ggggaacaag atatattatg TGAactttca ggacgATGca
                                      E4 orf start ->   E4 cds ->
3301 gcattgtatt ctaattctgg catggggcaa gtgcattttg aaagcaaagt tctttctccc
3361 tctgttacca GTtcgcttcg tgttgggagt accggaggac aacgcgggac ccaaaccggg
                /\ 3' sj
3421 gaccacgccc gaggacgTAG cagaccgtcc ccccgatctt ccagagacgc ccggggccgg
                E5 orf start ->
                NH₂ terminus unknown
3481 cagcagaggg cgcagtcgtc ttcgcgatcg cgatcacggt cacgatcacg atcgcctacg
3541 aaggggccgc actccagtgg acgagacacg cggctaccga gtcccgggag accccgggga
3601 ggaagacgaa ggggcacccc cgaacgggaa cgatgccctg gaacaccgac tccgccaact
3661 ccTGAccaag tgggaggacg atcttcaacg cctaagagac aagcttcgtc tagacttgct
        <- E5 end
3721 cagcttaTAG acgcggcata cgaccctcca gtactcctgc tacAAGGTGC TGcaaacact
                <- E4 end                             5' sj /\
3781 ttaaaatgct ttaggcgcag ggcaacgcag gctcatcctc acaaatttct gtgcatgagc
3841 acaagctgga catgggttag caaaacttcc cctttgaaat cgggacatag aatgctgatt
3901 gcattttcga actctgagca aaggaactgc tttttggctt ctgttcgatt accaaAAGGT
                                                              5' sj /\
3961 GTCAgtgctg tgaaaggggc tctTGAcggg ttaTAGcaaa tgctATTAAA ATGgttcgtg
                            E2 end <-  signal ->  <-
                L2 orf start ->                           -> L2 cds
4021 cAGcaagacg aaaacgtgcg tctgaggatg atttgtacag agggtgcaga atggggcaag
        /\ 3' sj
4081 actgccctat tgacatcaaa aacaaatatg aacacaacac attagctgac agaattttga
4141 agtgggtcag ctctttcttg tactttggac aattaggtat aagcagtggt aaaggtacag
4201 gggggtccac agGCTACACA CCattgggtg gcagaggagg aggtggagta acctctggga
                -> enhancer
4261 aaggagcaaa tgttgttaga cccactgtta ttgtggatgc attaggtcct actggggttc
4321 ctattgatcc tgcagtccca gatagtagta ttgtgccatt attggaaagc agtgggggca
4381 gcaccacatt agatgcaact cctggagctg aaattgaaat cattgcagag gtgcacccac
4441 cacctgtata tgaaggccct gaagttacta tagggacat agaagagccc ccaatattag
4501 aagTAGtgcc tgaaACACAT Cctacatctc gtgttagaag cacaacaagt aaacATGaca
L4 orf start ->        -> enhancer                       L4 cds ->
4561 acccagcttt cacagcatac gttgcaagtg cacaattacc aggagaaacc tcagcttcag
4621 acaatgttta cattttgcat ggtttcaatg gtgatttcgt ggggcaagct gatcctgaag
4681 gggatacaat atttgaagag attcctctgg aagaatttgg ggttcctgac atgccaccaa
4741 gcactagcac acctacaagc tcattcagaa gtgtttTAAa taaatttcag cgaaggctgt
                                      <- L4 end
```

```
4801 acaatagaaa gttagttcag caagtcaaaa tcacaaacag aaacactttt ttgaagcaac
4861 cgtcccaatt tgtgcagtgg gaatttgaca accctgccta tgttgatgac tctttgtcat
4921 taatctttca gcaagattta gatgaagtgt cagctgctcc tgatgctgat tttcaggata
4981 ttgttaagct aagcagacca gtattcacaa caaaagaagg cttagtcagg GTgagcagat
                                                          5' sj /\
5041 taggccaaag aggaactatt aaaacacgca gtggcttgca aataggtggc catgtgcatt
5101 attatacaga tttgagtcca attagaccat tagaagacat tgaaatgact tcctttgggg
5161 aggtgtctgg agaaagtgta attatgcagc cattaggtga aagcacattc attgatacag
5221 gtgcaagGTc agacaattta aatgaaggcg tcatagagta ctctgagagt gctttagagg
        5' sj /\
5281 atgctatgga ggaggatttt agtcaagtgc gtttggaaat aagcacttct gcaagaagta
5341 gaactagtat agtcactgtt caagaaggaa taccaccagg atcagtcaaa ctgtttataa
5401 atgatgctgc agcaacagtt gacactggtt atctgccagg aaagccagat tatgaggacc
5461 ctttcacccC AATagagcct gctgTCCCGC CTGATATAAT actcaatttt gaagatgaca
          signal ->      Sp-1 bind ->        <--> signal
5521 ctgcaacctt ttttctacat cccAGtcttt TGAaaaagca taaacataac aaacattggt
                              /\ 3' sj
                      L1 orf start ->
5581 ttcttTAAtg ttttacAGAT Gtctttctgg gttccaaatt ctgcaaagct gtacttgcca
    L2 end <-           /\ 3' sj
              L1 cds ->
5641 ccgcctacac ctgtcacaca atttcttgac acggatgagt ttgtcacacg cactgacatc
5701 ttttaccaca caagcacaga ccgactgctg tttgtaggcc atccatattt tgacctaaaa
5761 aaaggagaca ccgttgttgt tcccaaagtg tccggcagtc aatttagggt tttcaggcta
5821 aaattccctg atccaaataa gtttagcttt cctacacaag atatatacaa ccctgagaag
5881 caaagactag tgtgggctgt acgtggaata gaaatatgta gaggacagcc cttgggtgta
5941 ggagtaacag gtcacccaag ctttaataag ttcaaagatg cagaaaatct aaaccaaaat
6001 tcagatcaaa aagaagatga tagggttaat gtatgtatgg accccaagca agttcagctt
6061 tttattgttg gctgtgtacc ttgtgatgga gagcattgag ataaggcaca agcttgtgaa
6121 ccacaaggac caggggattg cccacctatt gagctcaaaa acacaaaaat acaggatggt
6181 gaaatgtgtg atacaggggtt tggtaatatg aattttgcag ctttgcaagc cagcaaatca
6241 ggagcccctt tggatattgt agaccaaata gtaaaatacc ctgattttct taaaatgggt
6301 aatgaccgt atggaaattc tatgttcttc tatgctaaaa gagaacagat gtatgtcagg
6361 catttatggg ctagagcagg gagagTAGga gATGatattc ctacaggaga gtctggtagc
                  L3 orf start ->
                          L3 cds ->
6421 ccatatttcc taccagccac tggacgagga cccctgccct cctcagtata tattggaagc
6481 cccagtggat ctctggtatc cagtgatcag cagatttata ataggcccctt ttggatacaa
6541 agggcacaag gcagtaataa tggcatgtgc tggaacaatg agcttttttgt tacagctgtt
6601 gatagtacta gaggcactaa ctttagtata tcagtacaca ctactgaccc agaggttaaa
6661 ccacaagaaa cttatacagc cactaagttt aagcactacc tcagacatgt ggaagaatgg
6721 gatttgtctt TAAttatgca gctttgcatt gtcaatcaa ctccagaatc aatagcttac
          L3 end <-
6781 ctgcacaata tgaatgaaag cattatagaa aactggaatt taggctttat tcagccaccg
6841 aatgatatag aggaccacta cagatttatt acctcattag ctactagatg ccccaaaaag
6901 acagacactc agGTtaaaga agacccctat aaagacctaa agttctggga cgtggaccta
        5' sj /\
6961 acagaaaagt ttaccacaaa cctgaaccag cactccctgg gtagaaaatt tttgttccag
7021 ataggtagaa aagctaccaa acgctctgcA CCGAAAACGG TCacatttga aaatactgaa
                                  -> E2 bind
7081 ggtaaaaagg cgccaaagcg taggcggaaa aatgtgTAGc tgccaacacc tgcagcagta
                                          <- L1 end
7141 tgagcaatat gataAATAAA cttttctttta tgtaagatat gaatgttggc AATAAAatgc
          signal ->                                signal ->
7201 agacaaacaa agcactgtta tctttggttt tttttcatct ccggcacttg tgtgtttgtg
7261 caagc
```

```
LOCUS       PCPV1       7902 bp ds-DNA   Circular   VRL       04-DEC-1993
DEFINITION  Pygmy Chimpanzee papillomavirus type 1 (PCPV-1), complete genome.
ACCESSION   X62844 S43934
KEYWORDS    genome.
SOURCE      Pygmy chimpanzee papillomavirus type 1 DNA.
REFERENCE   1  (bases 1 to 7902)
  AUTHORS   Opdenakker Mm G.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-1991) to the EMBL/GenBank/DDBJ databases.
            G. Opdenakker Mm, Rega Institute for Medical Research,
            Minderbroedersstraat 10, B 3000 Leuven, BELGIUM
REFERENCE   2  (bases 1 to 7902)
  AUTHORS   Van Ranst M., Fuse A., Fiten P., Beuken E., Pfister H., Burk R.D.,
            Opdenakker G.
  TITLE     Human papillomavirus type 13 and pygmy chimpanzee papillomavirus
            type 1: comparison of the genome organizations
  JOURNAL   Virology 190, 587-596 (1992)
COMMENT     To date, the complete genomes of two primate papillomaviruses have
            been sequenced: PCPV-1 and RhPV-1.  PCPV-1 hybridized to HPV-13
            under stringent conditions; subsequent analysis revealed a
            similarity of 85% (alignments of PCPV-1 are found with Group B
            viruses) (Van Ranst et al. Virol. 190: 587-596).  The fraction of
            nucleotide substitutions that represent nonsynonymous changes
            (amino-acid replacing substitutions) versus synonymous changes
            ("silent" substitutions) is remarkably low, suggesting that cross-
            species transmission may best account for the close similarity
            to HPV-13 (Part III).  Compared to the human genital
            papillomaviruses, the PCPV-1 genome does not seem to contain a
            homologous GRE or the homologous splice/acceptor pair potentially
            used to generate a truncated E6 protein product.(Van Ranst et al.
            Virol. 190: 587-596).
                Van Ranst et al. (Virol. 190: 587-596) isolated PCPV-1 from a
            pygmy chimpanzee with symptoms characteristic of human oral focal
            epthelial hyperplasia (FEH).  Pfister et al. (J. Virol. 47: 363-366)
            had earlier demonstrated the association of FEH and the
            papillomavirus HPV-13.  Subsequent restriction analysis and
            Southern blot hybridization characterized several analogous
            restriction sites in the two FEH-related viruses (PCPV-1 and HPV-13)
            but yielded divergent cleavage patterns.  Accordingly, HPV-13 and
            PCPV-1 had an overall sequence similarity of 85%.  The G + C content
            of PCPV-1 was 38%, the lowest figure for the sequenced nonhuman
            papillomaviruses and 1.5% lower than that of HPV-13.  Phylogenetic
            analysis of the E6 region conducted by Van Ranst et al. (Virol. 190:
            587-596) grouped HPV-13 and PCPV-1 with the HPV types connected to
            condyloma acuminata, orogenital lesions and low-grade cervical
            neoplasia: 6, 11, 43, and 44.
                Van Ranst et al. (Virol. 190: 587-596) identified two
            polyadenylation sites, potential signals for the termination of
            early and late gene transcription, in the upstream regulatory region
            (URR) of PCPV-1.  The absence of a glucocorticoid-responsive element
            (GRE) in the URR region distinguishes PCPV-1 and HPV-13 from other
            genital papillomaviruses.  Van Ranst et al. (Virol. 190: 587-596)
            notes that this feature of the two FEH- related viruses may inhibit
            steroidal sensitivity and thus preclude high incidence in genital
            mucosa.  PCPV-1 and HPV-13 lack a splice donor/acceptor pair in the
            E6 region which is also characteristic of HPV types 6, 11 and 44.
            HPV-13, PCPV-1 and HPV11 are further distinguished from the high-
            risk HPVs by the presence of a putative leucine zipper motif located
            in the carboxyterminal region of E2.  In both HPV-13 and PCPV-1 as
            well as HPV-11 and HPV-6, two short ORFs, E5a and E5b, lie between
            the E2 and L2 ORF's.  The late proteins encoded in L1 are highly
            conserved between HPV-13 and PCPV-1.  Van Ranst et al. (Virol. 190:
            587-596) propose that the late proteins, which constitute the
            structural coat of the virus, may be a crucial factor determining
```

```
          host viability.  Van Ranst et al. (Virol. 190: 587-596) further
          postulate that the high degree of similarity betwen HPV-13 and
          PCPV-1 suggests the possibility of cross-species transmission.
BASE COUNT     2509 a   1417 c   1582 g   2394 t
ORIGIN
        1 tttaataata atatcctgtt taaaaaatag gagggACCGA AAACGGTttc aACCGAAATC
                                                     -> E2 bind        -> E2 bind
       61 GGTgatatat aaaccagccc acaaattaag caagcggggc ataATGgaaa aagctaatgc
                                                              E6 cds ->
      121 ctccacgtcc gcaaaaacga tagaccagtt gtgcaaggag tgcaaccttt gtatgcacag
      181 tttgcaaatt ttatgcgtgt tttgcaggaa aaccctgtct actgcagagg tgtatgcatt
      241 tcagtataag gatttaaata ttgtgtggca gggcaatttt ccatttgcgg cctgtgcatg
      301 ttgtttagaa atacaaggaa aagttaatca atacaggcat tttgactttg ctgcatatgc
      361 tgtaacagtg gaagaagaaa taaacaagtc aattttgat gtgagaattc gatgctattt
      421 gtgccacaaa cctttgtgtg acgtggaaaa attacgacac atcttggaga aagcaagatt
      481 catTAAgtta aactgcgagt ggaaagggcg ctgcttccat tgctggacat cATGcatgga
  E7 orf start ->                                                  E7 cds ->
      541 aaatatacta cctTAAagga cattgtgtta gacctaagtc ctgaccctgt aggtctacat
                           <- E6 end
      601 tgcaatgagc aattagacag ctcagaagaa gatgaggtgg atgaacaagc cacgcaagcc
      661 acgcaagcca cgttcacaca acattaccaa atagtaacct gttgtggtca gtgTGAcagc
                                                                 E1 orf start ->
      721 aacgtgcgtt tggttgtgga ctgtacagga tcggacattc aacacctaca caagctgctg
      781 ctgggctcat aaatatagt gtgccccttg tgtgctccac aaaccTAAcc aggATGgccg
                                                                   E1 cds ->
                                                                   <- E7 end
      841 acaacacagg tacagacaac aagggtacgg ggtgctcagg atggttttta gtagaggcta
      901 tagtagacag gaaaactgga gaagaaatat cagatgatga ggatgaaaca gtggaagata
      961 gtgggttgga tatggtggat tttatagatg acaggtgtat tacacacaat tccttggaag
     1021 cacaggcatt gttaaacgag caggaggcgg atgctcatta tgcagctgtg caagacctaa
     1081 aacgaaagta tttaggtagt ccctatgtta gtccattagg ccatattgaa cagtcagtcg
     1141 agtgtgatat aagtcctcga ttgaacgcca tacaattaag tagaaaacct aaaaaagtaa
     1201 agcggcggct gttccaatca agggaaataa cggacagtgg atatggccat actgaagtgg
     1261 aagtggaagc agcaacgcag gtagaaagac atggcgaacc ggaaaatggc tgtggggggg
     1321 gtggacacgg aagggacaaa gaggggggagg acaggtgca tacggaagtg cacacagaaa
     1381 gcgagataga acaccacacg ggtactacgc gggtactaga actactaaaa tgtaaggata
     1441 taagggctac attgcatggt aagtttaaac aatgctatgg gctatcgttt acagatttaa
     1501 ttagacaatt taaaagtaat aaaacaacat gtgaggactg ggtggtggca gcatttggtg
     1561 tgcatcatag tgtgtctgag gcatttgaaa agttaataca gccattaact atatataggc
     1621 acatacaatg gctaacaaat gaatgggga tgctattgtt agtattacta agatttaaag
     1681 taaataaaaa tagatgtaca gtagcacgaa cactggcaac attgcttaac attccagaag
     1741 accacatgct aattgaacct ccaaaaatac aaagcagtgt ggcagcatta tactggttta
     1801 gaaccagtct atctaatgcc agtatagtaa caggagaaac acctgaatgg atagcaaggc
     1861 aaacaatagt agaacatggc cttgcagata gtcaatttaa attaactgaa atggtgcaat
     1921 gggcctatga caatgattat tgtgatgaat gcgacatagc atttgaatat gcaaaacgag
     1981 ctgattttga ttctaatgcc aaagcttttt taaatagtaa ttgtcaagca aaatatgtaa
     2041 aggattgtgc aacaatgtgt aagcattata aaaatgcaga aatgaaaaaa atgacaatga
     2101 atcaatggat aaaacataga agcaaaaaaa tagatgaaac aggtaattgg aaaccaatag
     2161 tgcaattttt aaggcaccaa aatatagaat ttatttcgtt tttaagtaaa ttaaagttgt
     2221 ggcttcaggg cacaccaaaa aaaaactgta ttgcaatagt ggggccacca gatacaggca
     2281 aatcaatgtt ttgcatgagt ttaataaaat ttttaggggg aactgtaatt agttatgtaa
     2341 attcaagcag ccatttttgg ctgcaaccgt tatgtaatac taaagtagct ttgctagatg
     2401 atgcaacaca ttcatgctgg ggatatatgg atacatatat gagaaattta ttagatggta
     2461 accctatgag tatagataga aaacataaat cgttagcatt aataaaatgt ccgccattat
     2521 tagtaacatc taatatagat attacaacag aggaaaaata taatatttg tatagtaggg
     2581 taacagtatt taaatttcca aatccattcc cttttgacag aaatgggaat gcagtatatg
     2641 agttgtgtga tgcaaactgg aaatgttttt ttgcaagatt atcagcaagt tTAGatatac
                                                            E2 orf start ->
     2701 aggactcaga ggacgaggac gATGgagaca ctagccaagc atttagatgc gtgccaggaa
                           E2 cds ->
     2761 cagttgttag aactgtaTGA agaaaatagt aatgaactta caaaacatat acaacattgg
                            <- E1 end
     2821 aaatgtgtac ggcacgaaaa tgtactgtta tacaaagcac gccaaatggg cctaagccat
     2881 attggactac aagttgtgcc accattaaaa gtatcacaaa ctaagggaca tgaggcaatt
```

```
2941 gaaatgcaaa tgacgttaga aacagtgcta aagtcagagt atggtacgga accatggacg
3001 ttacaagaga caagttttga aatgtggtta acaccaccaa aacattgttt taaaaaacag
3061 ggacaaactg tggaagtaag atatgactgc aatgcagaaa attcaatgca ttatgtattg
3121 tggaaataca tttatgtgtg tgaaaatgac agatggcaaa aggtaaaagg aatggtagac
3181 attaaaggat tatactatat ggttggacag tgtaaaacat attataTAGa ctttgaaaag
                                        E4 orf start ->
3241 gaggctaaac aatatagtaa aacattacaa tgggaagtat gttATGacag caaagttata
                                        E4 cds ->
3301 tgttctcctg catctgtatc tagtactgtg caagaagtat ccattgctgg gcctacttca
3361 cactccacaa ccacccttgc acaggccacc tgcgcagtgt catccatcgc cacagaagat
3421 agtgtgcaag cgccgcccta taacgactt cgaggaccct cccactgtgc tcgaaaactc
3481 caaaacacct ctaacattgt gtgtgccacg gaccgtggaa ccttggacag tgaaaacaac
3541 atcaacaata acaattacaa caacaacaac cagcaacgga caacagtaa cagtagtggt
3601 acacctaTAG tgcaattaca aggtgactca ataacttaa agtgtttcag atatagattg
             <- E4 end
3661 catgacaaat ataaacattt atttatgcta gcatcgtcta catggcattg gaccgcctct
3721 agcaattcaa caaaaaatgc aattgtaaca ttaacatatg tgaatgaaca acaaagacaa
3781 gatttttttaa atactgtaaa aatacctgcc actataaaac atacattagg gtttatgtcg
3841 tttcaattat tgTAAcacca tgtattgtac ataacatgta ttgtgtataa atgctgTAAA
             <- E2 end                               E5 orf and cds ->
3901 TGgaattaca ggttgtacct gtagatttta ctgcaaaagc aaccagtcaa tcattgttgc
3961 cactattaat tgctcttact gtatgttttc tcagtattat aatacttata tttgtatctg
4021 agtttctact atattcatct gttctagtac taaccttact tttatatctg ttgttgtggc
4081 ttttactaac tcccccctg cagttttttt tactaaccct gtctttgtgc tttttgcctg
4141 cgttttgtat acatcaatat attttgcaaa ctcagcaaca aTAActatac acaatgttaa
                                                       <- E5 end
4201 cgtgtacatt taatgatggt gatacatggt tgttattatg gcttttttgca tcagtgttag
4261 ttgcaattgt tgggttactt ttactgcata ttagaactgg aaaattgtaat tcacgttttt
4321 gttatagtaa ataagttttt ttaTAGgttg tgtgtagtgg tttccttATG gcacatagta
                L2 orf start ->                L2 cds ->
4381 ggcctcgcag acgcaaacgt gcgtctgcta cacagttata tcaaacctgc aaagcttctg
4441 gtacatgtcc tcctgatatt attgcaaagg tagaacaaaa tactcttgca gataaaatAT
                                                              signal ->
4501 TAAAgtgggg aagtttagga gtgttttttg gggggcttgg tattggcaca gggtctggta
4561 ctggcggtag aactggatat gttcctgtac agactgcccc acgccctgcc ataccctttg
4621 ggcctactgc acgtcctcct attattgttg atacagttgg ccctagtgac tcgtctattg
4681 tatccttagt agaagattca actattatta attcagcagc gtctgacttt gtgcctccta
4741 ttcgtgaggg atttgaaata agcacctctg aaactactac tccagccatt ttagatgtat
4801 ctgtaacaac acacaacact acctctacaa gtatatttaa aaatcctgcc tttgcagaac
4861 cttctattgt tcagtcacaa ccctcagttg aagcaagtgg acacgttctt acatctacat
4921 atacatctac tatttcctct cactctgtag aagacattcc attggatact tttattgtat
4981 cttcatcaga tagtaatcct gcatccagta ctcctgtgcc tacacctgtg gcacgtccac
5041 gacttggcct ttatagtaaa gcgttgcagc aagtgcaggt tactaatcct gccttttttat
5101 catcgccaca cgcctttata acttttgata accctgctta tgaaggtgaa gatataaagtt
5161 tacagtttca gcacaatact atacataacc ctcctgatga cgcttttatg gatattgtaa
5221 gattacatag gccggctata acgtctaggc gtggtattgt taggtttagt agaattggtc
5281 agcgagggtc catgtataca cgcagtggca aacatattgg tggacgggta catttttata
5341 cagacatttc tcctatatct gcagctgcag aggaacttga aatgcagcct cttgtggctg
5401 ctgcgcagga tgacagtgga ttatttgatg tttatgtaga ccctacccct ggtcctgtag
5461 cagtacaaaa tatgtcgtat ccctcctcca catcgtttgt acggtcctct atgtttacca
5521 ctaagtgggg taatactact gtcccttat cattaccaag taacatattt gcacagccag
5581 gccctgatat aatatttccc gccgcccctg gtgtaccacc gtataacct gttataccgt
5641 ctttacctaT AActcctatt tttattagtg ggtctcaatt ttatttacat ccttccttat
     L1 orf start ->
5701 atcttgcacg caaacgtcgt aaacgtgttt ccttgttttt tgcagATGtg gcggccTAGt
                                                 L1 cds ->              <- L2 end
5761 gacaacaaac tatatgtgcc tcctccccgcc cctgtatcaa aagtaattac tacggatgca
5821 tatgttacac gtacaaaaat attttatcat gctagcagtt ctagactact tgcagttggt
5881 aatccctatt ttcctattag aaagggtaac aaaactattg ttcccaaggt atctggattt
5941 cagtttagag tatttaaaat agtattacct gaccctaata aatttgcttt gcctgacaca
6001 tctatatttg attctactag tcaacgtttg gtgtgggcct gtataggttt agaggtaggt
6061 aggggtcagc cattaggtgt tggtattagt ggacatcctt tgttaaacaa attttgatgat
6121 gtagaaaatt ctgctagtta tgctgtaaac cctggccagg ataatagggt taatgtagca
6181 atggactata aacaaacaca gttatgtttg gtaggttgtg cacctccttt aggggaacat
```

```
6241 tggggtaaag gcacacaatg ttcaggtgta agtgtacaag atggcgattg tcctccattg
6301 gaattagtta ctagtgtaat tcaggatggt gatatggttg atacaggctt tggagccatg
6361 gattttgcac aattacaatc tAATAAAtca gatgtacctt tagatatatg cacagctact
                        signal ->
6421 tgcaaatatc ctgactattt acaaatggct gcagatcctt atggtgacag attatttttt
6481 tctttacgta aagaacaaat gtttgcaaga cattttttta atagggcagg gactgttggt
6541 gagcaaatac cagaagactt attggttaaa gggactacct cacgagctac tgtttctagt
6601 actatttatt ttaatactcc cagtggctct cttgtgtctt ctgaagctca gttgtttAAT
                                                            signal ->
6661 AAAccttatt ggttacataa ggctcaggga cacaataatg gcatatgttg gggcaatact
6721 ttgtttgtta ctgttgtaga tactacacga agtacaaaca tgactgtttg tgcttccaca
6781 acatcgtcgc cttctgccac atacacagct tcagaatata aacaatacat gcgacatgtg
6841 gaagaatttg atttgcaatt tattttttcaa ttatgcagta ttaaattaac tgctgaagta
6901 atggcatata ttcatactat gaatcctaca gtttttagaag aatggaactt tggattatct
6961 cccccctccca atggaacatt agaagacaca tatagatatg ttcagtctca ggctataaca
7021 tgtcaaaaac ctactcctga taaggagaaa caggatcctt atgcaggtct cagttttttgg
7081 gaagttaatc ttaaagaaaa gttttctagt gagctagatc aatatccgct tggtagaaag
7141 tttttattac aaacaggcgt tcaaactaca tcctttgctc gtgcaggaac aaaacgtgct
7201 gcaagtacat cttcatctac acctactaca cgtaaacggg ttaaacgaaa aTAGtttgtt
                                                            <- L1 end
7261 attttgtgtg tatgtttgtt atgtatgttt gttgtttata tgttgtatag tatatgtgta
7321 atgttatatg tatgttgttt ttatgtatta aagaatgtat gtgtgtgttt gtgtgtAATA
                                                            signal ->
7381 AActtaatct gtggtgtcct gttccaccct atgactaact attttgttgt ttatgttatg
7441 ttgtatatta tatagtattt atataactat ttaaaaacat atatattaaa ttatactcca
7501 tattgagtac gACCGAATTC GGTgtctac ttttacatcc tgtaaccaat ttgttgcagc
                    -> E2 bind
7561 acgttactTA TATAAtctta tctactgtct GCCAAAaata atttatccta ccacctacac
          signal ->              NF1 bind ->
7621 acctggcgtt accagggtgc ggttttttata tttacaaaat gttattaatc tctgtaacag
7681 tcattccact ttctaacaat cgttttgctt ttctagtaac ttttttcatac tttgttGCCA
                                                            NF1 bind ->
7741 ATgcctaaaa tcctttTTGG CTaccagcac tacatttttg taaacttact gttcatttta
              NF1 bind ->
7801 caaaatgagt aaccgtgttt gttttacaaa ctgagtaacc tacggtcaca cacctgcacA
                                                            E2 bind ->
7861 CCGGTATCGG Ttaaaacaca ccctgtcTAA tcctttatta ta
                        E6 orf start ->
//
```

```
LOCUS       MnPV         7687 bp    DNA    Circular  VRL      29-APR-1994
DEFINITION  Mastomys natalensis papillomavirus (MnPV), complete genome.
ACCESSION   U01834
SOURCE      Mastomys natalensis papillomavirus, isolated from a rat species
            native to southern Africa afflicted by a high incidence of skin
            tumors. Clone obtained from German Cancer Research Center from a
            population of rats established in 1969.
REFERENCE   1  (bases 1 to 7687)
  AUTHORS   Tan,C.H., Tachezy,R., van Ranst,M., Chan,S.Y., Bernard,H.U. and
            Burk,R.D.
  TITLE     The Mastomys natalensis papillomavirus: Nucleotide sequence, genome
            organization, and phylogenetic relationship of a rodent
            papillomavirus involved in tumorigenesis of cutaneous epithelia
  JOURNAL   Virology 198, 534-541 (1994)
REFERENCE   2  (bases 1 to 7687)
  AUTHORS   Chan,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-1993) Shih-Yen Chan, Institute of Molecular and
            Cell Biology, National University of Singapore, Lower Kent Ridge
            Road, Singapore, 0511, Republic of Singapore
COMMENT     In 1978, Muller et al. isolated and cloned the first rodent
            papillomavirus, MnPV.  (Only one other rodent PV has been cloned
            and characterized, MmPV.)  It was derived from Mastomys natalensis,
            the South African mouse species known to have a very high incidence
            of both keratoacanthomas and squamous carcinomas.  Subsequently,
            Tan et al. sequenced and published the complete genome in 1994.
            The size of the MnPV genome (7687 basepairs) is roughly equivalent
            to that of the other cutaneous papillomaviruses.  Tan et al.
            identified a protein motif in E7, LYCHEVLDEDEL, which has also been
            found in the adenovirus E1A, HPV-16 and HPV-18.  A distinguishing
            characteristic of MnPV is the presence of a relatively large hinge
            in the E2 transactivator and the absence of a homologous E5 coding
            region.  Tan et al. mapped several E1 binding sites to the long
            control region.  These sites have a high degree of similarity to
            analagous sites in BPV-1, CRPV and HPV-9.  A putative nuclear
            localization signal is located 200 basepairs downstream of the E1
            intitiation codon.  Phylogenetic analysis conducted for Tan et.
            al's publication places MnPV between the EV HPV's and animal
            and human papillomaviruses associated with cutaneous neoplasia.
            Tan et al. propose that MnPV may provide a model for cutaneous
            papillomavirus infection in humans.

            NCBI gi: 436472
BASE COUNT     2036 a   1821 c   2035 g   1795 t
ORIGIN
        1 GCAACAATct cctctccata ctttttttcca ctgcaccggt atcgtacaaa caTATATAAg
          E1 bind <-                                        signal ->
       61 aagaccccaT TGGCTatgga tttgttactg cggtagccgg ATGgatagga ccgtgcactc
          NFI bind ->                        E6 cds ->
      121 ctttgtggag cggctgggaa ttcctcggga ggacctcctg ctgccgtgca cattctgctc
      181 gaggtttctt acccaggagg aattaactgc atttgacttt agtgctttta accttgtttg
      241 gagaggaagG Tgtgcccatg gaatctgcac agcctgtgct cgtgtctgtg catccctaga
          5' sj /\
      301 cctgtttctg caccatcaga attcgcgacc attagcagat gttctgcggg acgaaaatct
      361 tacactccac ggactgaaag cacggtgtcg cgtgtgcatg aagatactgt cagtgacaga
      421 aaagcTAGag tgtgcagaaa gaggggaatc ctttgccaaa gtcAGgggcc agtggagggc
          E7 orf start ->                                    /\ 3' sj
      481 acggtgcaga atttgcaaac ccgtgTAAaA TGataggacc tgacaccacg cgctgtctca
                                <- E6 end
                                E7 cds ->
      541 ccggcgaaac tcctgactcg gtcagcctgt attgtcacga agttctcgac gaggacgaat
      601 taaaagagcc aacagaggcg gctccgccac cggaacaata caccttgtac caggtactca
      661 ttgagtgtcc tgagtgtaat aagacaattc ggctgacgtg cgcggcacaa gcacaccaga
      721 tccgtgggct agaacatcta ctgcttgacg ggcTAAgagt gatctgtccg cggtgtaacc
                                E1 orf start ->
```

```
 781 agaagaATGg aagatctTGA agaaggtact ggcgaaggtt gcagtggctg gtttgataga
        E1 cds ->          <- E7 end
 841 gaagctattt gtagtgacgg gtctagtgat gaggagccaa atgagtcctt tgaatctatt
 901 gcggacatgt tcgatgacgg aacacaaaca cagggcaatt ccctagagtt gttccatacc
 961 caggaaaagg aggagactag gacacagata caagctcTAA AGCGAAAGTa cattcccagt
                  nuclear localization signal ->
1021 ccagaggcag gtggggatct ctcaccacgg ctgagggcta tctctattac ccccaaaaaa
1081 aagaaaccta gcagacggtt gtttgagacc ccagaggata gcggcaacgg gagtcttggg
1141 aatgagacta cagatacttc ttcgggggtt caggtagtag gggactcagc tgtggatgta
1201 tgcgatgcgg ggcggctgct taatctgaat ctgcttcaaa gccataatag ggtggcgagg
1261 ttgcttgctg tcttcaagga agcttatggg gtgtcataca aagagcttac acgggagtac
1321 aagagcgata aaacctgcaa tccagattgg gttatcgcat tgtactcctt gagtgagccc
1381 atcctaaatg cggcgcggac aacactacag ggaatttgtg agtatgtgtt tatgcaaagc
1441 cgccctacag cggcagccac agttgcttta ctaactgttc gctttaaatg cagtaaaagc
1501 agggagacag taagaaaaca aatgtgcggc atgttccact cagatccgct actctgcctg
1561 tgtgatcccc ccaaggtcca aagtgtgcct gcagctctat actggtataa gagcagcatg
1621 tatagtggga cattcacaca cggagaggcg cctgagtgga tcaagagaca gaccatgatt
1681 acctgtgcaa tggaagagac taaatttgac ctttcagaaa tggtgcagtg ggcatatgac
1741 aataactatg aggacgaatc ccaaatagca tttgaatatg ctagaacagc cactgagagc
1801 cctaatgcga atgcctggct ggcttccaat gcacaagcta aacatgtgag ggactgtgct
1861 acaatggtga ggcattataa acgggcggag atgaaggcta tgagtatgtc acagtgggta
1921 tggaagtgct gtagagagga acctgaggag ggcacttgga cacctatttc cctatatctc
1981 gcgtccgaag gggtggaagt gataagattt ctatctgcta tgaagagttg gttgcggggg
2041 attccaaaga aaaattgtct ggtattttac ggccctccaa atacaggaa gagtctgttt
2101 actatgagcc ttattaagtt tttgagaggg cgagttatat catttgccaa tagcaaaagc
2161 catttctgga tgcagccact ggctgaggca aaggtagtgc ttttagatga tgccacaagg
2221 gccacatggg actatgtaga tacatatatg aggaatgcca tggatggaaa tccattatca
2281 attgattgca agtatagaac acctgtgcag gtaaaatgcc cccccatgct tgtcacaaca
2341 aatgaggatg tgcacttgaa tgataggtgg cgctaccttc atagcagaat acaagtcttt
2401 cacttaaagg aacctatgcc tatagacact gccggtaacc cagagtattc cttctctaat
2461 agacattgga aggcgttttt cgaaaagtta cagaagccac tagatcTAAg cgaggacgag
                                          E2 orf start ->
2521 ggtgacccca aggacaATGg agagcataca cagccgttta gctgctgtgc aagaggaact
                  E2 cds ->
2581 gatgtgcatg taTGAggatg gggaggagac actggaggcc cagcttaaac attggggctt
                  <- E1 end
2641 gttgaggaaa gagcaagtct tgttacatgc agcacgccag catggacata acaaaatagg
2701 actgcaggcc gtgccccctc tttcagtgac ccagcagaat gccaagaatg ctattgaaat
2761 gcatttgctg ttgcaaagtc ttgcagagac accatatgct agggaagcat ggacactaag
2821 ccagaccagc agggaaatgt atatggcagg tccatccggc accttcaaga aagacggcac
2881 cattgtggag gttatatttg atggtgacaa gactaatatg atgacatata caaagtgggg
2941 gaagatatac tttgctgatc caaatggcaa ttggagcaga acaacctccc atacggacat
3001 taatggcata tattttaata agtctgggga taaggagtac tatgtgcggt tcaaagagga
3061 agcaaagagg tactcattaa caggaacttg ggaagtacat gatggactag agacacattc
3121 ccttcttatt cctgtcacca gctctacacc gcagaccgga tttccTAGag gggatccggt
                                          E4 orf start ->
3181 acgccttcac gggaatacca ccacAGgact gcccataccg cttcggaaca gcagcagcaa
                          /\ 3' sj
3241 ccagatatta ctacgagagg gaagaggaga ctatccgac ggcgcacgcc gcgagacgag
3301 gaggtactac caggggccaa caccgacgcc caggtctcta tctcccccca tctaccgtcc
3361 cccgccaagc tacgaagagt cgaggaggag gaggaagcta aggcgccgcc aagacgggcg
3421 agtcaaatac cccgcgtctc cctacaggac aaaaccaccg ggggaaacca gcagcgacga
3481 cgaagacgag gggagagggg ggcacgaacc ccgtccccag agacgactgc ccagaggcct
3541 aagagaccgc ggagagcgtg cacccgaaag gaggagaccc ccagttcagg aggggtagga
3601 ggacgtggac ggcgtagggg ccttgctgga cgacctgaag ctgtaccaag aaccacctgg
3661 agacccagtg aggactcgg actccccagg cagtcgtctt accccgccc cgccagacct
3721 atctcggtac gactctaccc ggttacaggt ggacgcggag agcagccctc cTAGgacacc
                                                      <- E4 end
3781 cagaccggcc cccactctcg tggcagagtg cactcctggg agaccttctc cgcagactgg
3841 aagcggacag caagcactgg gagaaccgcc ttctcggcct tcacgcggac attgccgcga
3901 ccctcggact gcctgccttt tgatcatcaa aggatcatca atcaggtta agtgcttgcg
3961 attttagactt aaatcctggc atcacagcct gttttcctac atcagcacca catggcagtg
4021 ggttccttca gtaggaagta ataggattgg acggtcacgc attctggtga tgtgtgagga
4081 ctcagcgcag atggacagat tcctatgtac tgttaagatc cctgctggta tgacagttga
```

```
4141 acagtgcagc atggcgtctg tcTGAtgccc cccccccctcg cataacatac taacgcacac
                                <- E2 end
4201 tgcAATAAAg tttttccttt acacagtacT AAcctactaa tattagcATG tctagaagga
  signal ->                   L2 orf start ->       L2 cds ->
4261 gaaagcgaca tacacgagtc cctcgtgact cggccactca catatatcaa acatgtaagc
4321 aggcaggcac atgtccgcct gatgtagttA ATAAgttga aggcacaacc acagctgata
                                signal ->
4381 agattcttca atatggcggg gcggctgtat tcctcggtgg ccttggtatt ggtacaggta
4441 ggggaagtgg tggtgcaaca gggtatgtac cggtcggcga gacacctggt atttccgtgg
4501 gtgcaagacc agttcctcga cctaatgtgc ccttagaaac tgttggtccc caggacctgt
4561 ttcctgtgga tgccattagg cctactgatc cttcggtgat tgatgtcgcc agtgtgccta
4621 ctccccactga cacctctatt aatgtacccg aggtggaggt cattgctgag atccaccctg
4681 tacctcctga cggtccctcc aacacaccaa caaccacaat taacacatca ggctcagggg
4741 atgcagccat attagaggta gctcctgaac catccccagc cgtcaggact cggtggagag
4801 ctagcaagac aaccttccat aatcctgcct ttcacagctt ctcctctact ggttcaactg
4861 taggcgaggc cacaggtatg gacaatattg ttgtttacag cggtagtggg gggaggacga
4921 taggtgggga cagcatagag cttatgccct ttactagcag tgataccccta gatttaagta
4981 ttgtggagga gacctccttt ggaggtagga ccagcacacc acgaaccaag cccctccctt
5041 ctcggttgcc ttcccggagg tattatgaat atagagaaag cagtcttggt gagttatggt
5101 cacctaggag ggctatgggt cccacgtata taaatcctgc ctttgaagct gaggatagta
5161 tcctttttcc tgaatgtagc atgcaggccg ctaatccaga ttacacaggc attaccaggc
5221 ttggtcatct cttttggtact gagcagggtg gccgtgtccg tattggtcgt ctgggacaaa
5281 agacatccct gcacacacgc agcggtatgg caataggccc taaggcatac ttttataagg
5341 acatttctag catttctgtt gtcccagagg agagtataga actcagcacc tatacctcag
5401 ctgccccttt gggtgaggat gcaggtataa tagtggagga ctctatggag ggttcttttg
5461 acaatatcac cctcagttct tggagtcatg gatccatgga cgggcttctt gaggatgatg
5521 ctagttatga ttttcacggc cacctggtgt ggggaacacg ccgtagctct aagcaaataa
5581 gcatgccatt ccgccggtcg tggtatcctg aaaactgctgt gtacgtgcag gagggtgggt
5641 ctgTAAtgga tcctgaggct tctgcagagc tggttcccag tagggacagt gctcgtcccc
L1 orf start ->
5701 ATGtcatata tagggggctat aatgggacgg actattatct acacccgtca ttgtccagac
L1 cds ->
```

```
5761 gcaggcgtaa cagcaggcat atctattttt cagatggcgt actggctgcc TAAtaaccag
                                                                <- L2 end
5821 aagttgtacc tgcccccggc cccggtgcag cgcatactgt ctacagatga atttactaca
5881 cgaacagaca tatattacta tgctagtagt gacaggttat taactgttgg taatccatat
5941 tatcctatac tggatgggga tactgttact gttcctaagg tcagtcctaa tcaatacagg
6001 gtgttccgtt gtaaattacc ggaccctaac cggtttgcat ttggtgagaa gtcggtttac
6061 gaccctgaga agcaacggct tgcatggtgt atacggggag tggagatagc tcgtggccaa
6121 cctctgggaa tagggattac tgggcatccc ctatataaca ggctagagga tgtggagaac
6181 cctggaaagt atccatctgc tccgggcacg gacaatagac aaaatgtagg ccttgatccg
6241 aagcagactc agatgtttat tgtcggttgt gtacctgcac agggtgagca ctggagtaga
6301 gcacttacct gcagcaatca ggtggttaag aagggtgact gtccacctat tcagcggatg
6361 tctgggatga ttgaggatgg tgacatgggg gacataggtt atggcaactt agacttccga
6421 gtgttgcagg aaaacaagtc agaggttccc ctcgaggtag ttgactctat ctgtaagtac
6481 cccgattatt taggaatgtc caaggaaacc cacggcaact catgcttctt ctatgctagg
6541 caggcgagat tatacagcag gcacttcttt aaccgtgcag gtgttcaggt tgagactgtg
6601 ccggagtcat tatacaagaa gggcaaggat ggacaggcac agagcacact ggcactagct
6661 acatactcag ggactccgtc agggtcacta gtgtcatctg atgctgtact gttcaaccgt
6721 ccatactggc ttgagagggc acaaggacaa aacaatggca ttctgtggaa taatgatttg
6781 ttcgtgaccg tgctggacaa cactcgtggg acccattct ccatcagcat tgctacacag
6841 gatgaaaatg attacaccgc ctcaaactac aagcaatata ctcgacatgt tgaagaattt
6901 gagcttgaat ttattttcca actggttaag atcaaccttt ctactgaggt gctagcatac
6961 ctgcatggga tggacccatc tatactggat aactggaact tgactctggg acccccaat
7021 gatggtagcc ttgctgataa gtacagattt atagaatccc ttgctacaaa atgccctgac
7081 aatgtggaag tcactaagcc tgatccctac aaaggacgga tattctggaa cattgacctg
7141 actgaaagac tgacagctga tctggaccaa ttctcacttg gacggaagtt cctctaccag
7201 cacgcgcgaa tttcaaaccg taaacggtcc cttcctgctt ccagaaacgg cggcggaacc
7261 tcctcctctt ccaccaagcg gagaaaaaaa TAGttggAAT AAAgactgct gacactgcac
                                         <- L1 end
                                         signal ->
7321 ttgtgtcccg ccttttctta atcccgcctt tgctggggct gcagtacagc acgctgccaa
7381 gtttatGGGA GGtgctggaa cactGGGCGG tgcttggatc cggaatgcgc cgccttggaa
     Sp1 bind ->                    -> Sp1 bind
7441 gccagcgcca gtcttgttca gacaccgaga cgccaggtgt gcagcttcat tggtgcACCG
                                                                      ->
7501 TGCCAGGTat ACCTCTTTCG GTgcagttct tatgccaagt ctattgttgc tttttgccaa
     E2 bind        -> E2 bind
7561 ctcggTGAGT AAcatcctgc TTGGCActgt ctgcgacgta cctgctgcag agacttgtAC
E6 orf start ->   NFI bind ->                                      ->
AP-1 bind ->      <-
7621 CGGGTGCGGT acTTGGCAgt acaaacacaa ttaggtttgg acaagACCGA TATGGGTgTG
     E2 bind        -> NFI bind                              -> E2 bind     ->
7681 AATGTTG
     E1 bind
//
```

RhPV1

```
LOCUS       RhPV1        8026 bp ds-DNA           VRL        26-JAN-1993
DEFINITION  Rhesus papillomavirus type 1 (RhPV-1), complete genome.
ACCESSION   M60184 M37718
KEYWORDS    complete genome.
SOURCE      Rhesus papillomavirus type 1 DNA, isolated from a metastasized
            tumor from a Macaca mulatta.
REFERENCE   1  (bases 1 to 8026)
  AUTHORS   Ostrow,R.S., LaBresh,K.V. and Faras,A.J.
   TITLE    Characterzation of the complete RhPV 1 genomic sequence and an
            integration locus from a metastatic tumor
  JOURNAL   Virology 181, 424-429 (1991)
COMMENT     To date, the complete genomes of two primate papillomaviruses have
            been sequenced: PCPV-1 and RhPV-1.  RhPV-1 was isolated and cloned
            from a metastatic tumor of a rhesus monkey.  Ostrow et al. (Virol.
            181: 424-429) reported a 71\% prevalence of RhPV1 DNA within a
            mating cohort of monkeys including a male with penile carcinoma and
            subsequent lymph node metastasis (Ostrow et al. Virol. 181:
            424-429).  These data indicate that the most likely mode of RhPV
            transmission is sexual activity.  Notable sequence similarity to the
            sexually transmitted Group A Human anogenital papillomaviruses
            (alignments of RhPV DNA are found with Group A viruses) further
            substantiates this conclusion.  RhPV's oncogenic potential is
            indicated by its ability to cooperatively transform primary
            epithelial cells with activated \it Ha-ras, independent of
            dexamethasone treatment (Ostrow et al. Virol. 181: 424-429).  The
            physical state of RhPV DNA is integrated with high copy number
            (roughly 100 copies per cell) (Ostrow et al. Virol. 181: 424-429).
            Regulation of the viral genome is likely to be dependent on the
            glucocorticoid hormones, as the GRE sequences of HPV-6, HPV-16,
            HPV-11 and RhPV differ by only a single nucleotide (Ostrow et al.
            Virol. 181: 424-429).  The RhPV-1 E5 ORF does not have a methionine
            start and therefore likely undergoes post-transcriptional
            processing.  Ostrow et al. (Virol. 181: 424-429) have determined
            that the integration of RhPV-1, at at least one locus, does not
            directly inactivate the E1 or E2 ORF.  An additional unique feature
            of the RhPV-1 genome consists of two 18 bp direct repeats occuring
            at the end of the E5 ORF and 6 bp upstream of the E2 ORF.  Ostrow
            et al. (Virol. 181: 424-429) suggest that these repeats may be the
            artifact of an insertion in an ancestor of the rhesus virus),
            Ostrow et al. (Virol. 181: 424-429) further note that unregulated
            early gene expression is a critical factor in oncogenic
            transformation.
BASE COUNT     2118 a   1789 c   2067 g   2052 t
ORIGIN   First nt of E6 orf
        1 tacttaacta tactcctgag tATGaaaaag ggtgtaACCG AAAACGGTgc aACCGAAAGC
  E6 orf ->              E6 cds ->              -> E2 bind       -> E2 bind
       61 GGTgcaTATA Aaaagctcct gaaactttgg tttttgtgg caatggtaga ctgccctggc
          signal ->
      121 gagccaaacg aattgcccag gaccattcac gaactatgcg agcagcgtga ggagaccctg
      181 cacgagcttc aattggagtg cgtgtattgc ctgaaggaac taacacgcat tgaggtatat
      241 gattttgcac ggtgggattt aagattggtg catagacaag gcaagccata cggggtatgt
      301 cccatatgct tgaggtttta ctcaaaaatt cgaaaatata ggcgatacga gtattcaata
      361 tatgggtgta ctttagagcg tagaaactaga aaacagttag tggaggtatt aataaggtgt
      421 tattgttgtc agaagcccct gtgtcccatt gaaaagcaaa gacacgtgga ccaaggacaa
      481 aggttccaca gaatagcggg acagtggacc ggaaggtgct tgatgtgctg agaccaaca
      541 gtacctgaga cccagccaga cactgatcaa cagggcagTA Gtttcttgca agcATGAttg
                                             E7 orf start ->    E7 cds ->
                                                                     <- E6 end
      601 ggcctaaacc taccctcgag gacattgtcc tagatttgca accatttcca caACCGCAAC
                                                                  -> E2 bind
      661 CGGTcgacct tatgtgttat gagcaattat ctgacagctc agaggatgag gatgaagtag
      721 accatcatca caataatcag cagcagcatc atcagcacgc cagacctgaa gtaccagagg
      781 atggtgattg ttatagaatt gtgagcgatt gttacagctg tggcaagcca ctgaggctgg
      841 ttgtggttag tagccacgaa gagttacgtg tgctagagga cctgctgatg ggcacgctTG
```

```
 901 Acattgtgtg tcccagctgt gccagcagag tgTAActgca ATGgaccctg aaggtacacc
E1 orf ->                                       E1 cds ->
                                             <- E7 end
 961 aggggaaggg gtggggtgta cggggtggtt taatgtggag gctatagtag aacgtaaaac
1021 gggggatgtg gtgtcagagg acgaagacga cacagaggat acaggatag atttggtaga
1081 ctttatagat gacacatgtg gaagtgtgca gacaggggac gaggcacctg gggcgttgtt
1141 gcacgcacag gaaacacaag cgcatgcaga ggcagtgcag gtttttaaaac gaaagtttgt
1201 aggcagtccg gcagttagtc cgttgggaaa ctacaatccc tgtgtagaca gggatttaag
1261 tcccagatta aatgaaataa gtttaaacca aggcagcgga caggcaaaac ggagactgtt
1321 tttgccggac agcggttatg gcaatactga agtggaaacg tcgctattgc aggtagcagg
1381 gggggggcggc caggatgtac aggcagggggg gaaggaaaac acacggccag atgacggggg
1441 gggggatgcc acgcagctgc tccgttgcag caacttaaaa gccactttgc tgAGTAAAtt
                                                   signal ->
1501 taaatctgtg tatggagtta gctttttcaga gttggtgcga agctttaaaa gcgacaggac
1561 cacgtgcgct gactgggtgg tggggggcagc gggggtccat catagcgtgg cagaggggtt
1621 aaagcagctc attcagcctt tttgcagtta tgcacacatc cagtgcctta catgcgactg
1681 gggggtgtac ctgctactgc tggcacggtt taagtgtggc aaaaacagac taacagtttc
1741 taaatgcatg agcacgctgt taaatgtgca agaaacgcac atgctaattg aaccaccgaa
1801 gctgcgtagc gcagcagcag ctctatactg gtacaggaca ggtatatcaa atgtaagtga
1861 agtaataggg gaaacacctg agtggattac aagacagaca atgtttcaac atggcctgga
1921 ggacagtata tttgatttgt ctgaaatggg gcagtgggca tacgaccacg actttacaga
1981 tgacagtgtg atagcgtacg agtatgcaca gctggcaggg atagacagca acgctgctgc
2041 attttttaaaa agtaatgcac aggccaaata tgtgaaggat tgtgccacta tgtgtaggca
2101 ctacaaaaga gccgaaaggc aacagatgac tatgtcacag tggataaaac aaaggtgtga
2161 aaaaactgat gatggagggg actggaggcc aatagtgcag ttttttaaggt accaagggggt
2221 ggagtttata gcatttttag cagctttaaa gctgtttttg aagggcattc caaaaaaaaa
2281 ctgcatagtg ttatttggac cgccaaatac aggtaaatcc tactttggca tgagcttaat
2341 acatttcttg caagggtcta tcatttcata tgtaaattcc aacagtcact tttggttgca
2401 gcctctggca gatgctaagg tggccatgtt ggatgatgca actcctcagt gctggtccta
2461 tatagataat tatttaagga acgcactgga cgggaacccc attagtgttg atagaaaaca
2521 taaaaatctt gtacagatga agtgcccccc attgcttatt acctcaaacA CCAATGCAGG
                                                     -> E2 bind
2581 Tcaggatgac aggtggatgt atttgcacag tagaatggtt gtgtttacat ttgaacagcc
2641 atttccattt gatcagaacg gtaatccagt ttatgagtta aatgataaaa actggaaatc
2701 ctttttctca aggacatggt ccagatTAGa tttacaagag gaagaggaga cggaaaATGa
                       E2 orf start ->                      E2 cds ->
2761 tggaagcact tgcagagcgt ttaagtgcgt tgcaggacag aatcttagaa ctgtaTGAag
                                                           <- E1 end
2821 ctgatagcaa ggacttaaaa gaccaaatag agcactggaa atgtgtgcgc caagaatgtg
2881 cagtgttgTA TAAggcacgg gaagtagggt tttcccacct gaaccatcag gtggtgccat
       signal ->
2941 cattaactgt gtcacgggct aaagcccaca aagcaattga agtgcagctg cattagaga
3001 gtttacaaaa ttcggagTAT AAcaatgagg agtggacgct gcaagatgcc agcttggaga
              signal ->
3061 tgtggcacac agaacctaag ggatgcttca aaaaaacagg tgttccagta acagttttgt
3121 ttgactgtga caaagacaat accatggagt atgtgctgtg gggacacata tatgtgtggg
3181 gggacaatgg atgggtgaag acattcggtg aggcggacaa ctgggggtctg cactataccg
3241 ttgctgggga aaaggtgtac tatgtgcagt tttatgagga tgctaaaaaa tatggacatg
3301 gaaatggaaa tggagatggc tatgagtggg aggtgcatgt tggtgggacg gTAAtgcatt
                                            E4 orf start ->
                                            NH2 terminus unknown
3361 attctgactc tgtgtctagc gctacccact gcgacaaact acccactgtt gaaattgtta
3421 gcggactgca acacatcaac ccatcacccc cccccgccaa ccccagcgcc aaggaaaacg
3481 tgtggtcatc gcctgcaaag cgagtgcgtc ggtcagactc aggtggagat ccagtgcggg
3541 ccttggacgg taaaagccgg tcagtccttt gtggatctgc acacaacaac gctacaggga
3601 gttccggtga cagtgactat acgcctaTAG tgcacctaaa aggtgaatct aactgtttga
                                        <- E4 end
3661 agtgtttgcg gttcagactg ggaaagcata agcacctgta tattaatata tcgtccacct
3721 ggaggtgggc aaaccatgca agtgagaaag caattgtaac tgtgacattt gcaaatgagc
3781 ttcaaagaca acagtttttta aacactgtaa aaataccttc tactgtaact ctgtcacaag
3841 gagtaatgac tgtgtAGTGT GCATTGGCAC ACAGggtttt gtatttttttt ttttacaagt
                 <- E2 end
                 -> 18 bp repeat <-
3901 actgtttgta attaattttg tatattgact gtatatTGAa ttgtgGTGTG CATTGGCACA
```

```
                                          E5 orf start ->      -> 18 bp repeat
                                          NH2 terminus unknown
3961 CAGtggtctc atttcaagcc tgtacataca ttgaacagta tccaggtact gtgtaaagcc
        <-
4021 aattgttgct gctacgcttg taaaccgcca ccattctgct gtttctggtt gtgttttgc
4081 tgctgttttt gcttggcctt gtgttttgtg cacttgttga tcgctgctt ctgtgttttt
4141 ccggtatgcc tcagtgttgc tgcttatgct gttgttctgg gtgtccatag tgaacccgtt
4201 tgcagctttt ggtctgtgtt tgttttgttt tttaaccccg ttgcttttga tacacctgca
4261 tgccctcagt gtggtttaca gcagaatgat gtaaatactg cacatagaca tgttattatc
4321 agttattttg ctattgttgc tgttaatatt tactttgttt tggcactact tgttggtgct
4381 gcgtttaaag ccaccagcag ggcgcgcacg TAAatgtaaa cagctgagac ggcggcgcag
                                          <- E5 end
4441 atAATAAAcg tcacacAATA AAgcgtcATG aagcatgcac acttgtcgcg gcgcaagcga
                L2 orf start ->      -> L2 cds
signal ->       signal ->
4501 gcagccccgc gcccacctgg tggggcggcaa aagcgtgcat ctgccacgca gctgtaccaa
4561 acctgcaagg cggcaggcac atgcccccc gatgttatcc ctaaggtgga aggcacaacc
4621 gtagcagatc aaattttaaa gtatggcagc atgggtgtat actttggggg tttgggcatt
4681 ggctctggtg ctggcacggg cggaagaagc ggctacgtgc ccctaggttc acgtcccgca
4741 tccattcccg agccgttgcc acgaccacca gtaacaattg agcctgtggg cccttccgat
4801 ccctccattg tgtcattgct ggaagagtcc agactaaatag aggcaggtgt tccagccccc
4861 acattcccca ctcatggggg gtttgaaatt agcacatctg aagttagcac acccgctgtc
4921 ctggatgtgt ctagcggtgg ctctgatgtg cacgttagtg tgacctcctt tacaaaccct
4981 acctttactg agccatctgt gctgcgaccc ccgcccccg tagaggcgtc tggacgcctg
5041 gtaatctctg catcctctgt cagcacgcat agctacgaag aaataccccat ggacacattt
5101 gtaataactg gagaccacaa cTATAAcaca accagcacac ccattcctgg ttcacgtgcc
                signal ->
5161 cctgcacgac ttggtctata tggacgtgct acccagcaag tgcgggtggt ggatcctgca
5221 tttataacca cccctgcgcg actggtgaca tatgacaacc ctgcatatga gggtgtggac
5281 gatgccaccc tgcaattttc ccactctgac attcaccagc cgccagatcc tgacttcctt
5341 gacattgtgg cattgcacag gcccgccttg acctcacgta agggcaccgt gcgctttagc
5401 cgattaggcc agcgggcaac actaaccacg cgcagtggta agcgtattgg ggccaaggtg
5461 catttctatc atgacctcag tcccattgcc cctgcagaaa gcatcgagtt gcagcccctg
5521 tcatctcagg gagagctgta tgacatatat gcagatgtag acgggcaaga ggacgctgca
5581 gctgtggcta cacccccatt aaacagcaac agcagtggca ttgcaagccc ctggaacacc
5641 acagtgccac tcagtgcagg ggcggacgtg acgctgcagt ccggccccga cgtgtccctg
5701 gatgcaccag tggctgaatc gcctggtgac ctctcaaggcc ttctgcacat
5761 attattctgt acgggggaga ctttttatttg caccctagct acctcggtat cgcaggaaaa
5821 cgtaaacgca tgcacaattt cttttcagat gtctatgtgg cggccTAGTG Actccaaggt
                                          L1 orf start ->
                                                    <- L2 end
5881 ctacctacca cctgtcctgt gtctaaggtg gtcagcacgg atgaatatgt ctctcgcaca
5941 agcatatact atcacgctgg cagttccaga cttctggctg ttggacatcc ctactatgct
6001 gtaaagaagg gcaacaacaa agtgtcagtg cccaaggttt ctggtttaca ataccgagtg
6061 tttcgagtgc gtttgcctga ccccaataag tttggccttc cagatgctaa cttttatgac
6121 cctaacacac agcgccttgt gtgggcctgt ttaggcgtgg aggtggggcg tggacagcca
6181 ctgggagtgg gcaccagtgg tcatccactg ctgaacaaac tagatgacac ggaaaatggc
6241 cctaaagtgg ccggggggaca aggagcagat aacagggaat gcgtgtcaAT Ggactacaag
                                                    L1 cds ->
6301 caaacacagc tgtgcatgct aggatgcaag ccccctgtgg gtgagcattg gggaaaagga
6361 aatccttgca ccactggcgc tgcaggtgac tgccctgcac ttgagcttgt taactcagtt
6421 atacaggatg gggacatggt tgatacaggg tatggcgcta tggacttaa tgcactgcag
6481 gccaacaaat cagatgttcc tatagacata tgcactagcg tgtgcaaata ccctgactat
6541 ttaaaaatgg catcagatcc ctatggcgac agcttgtttt tttacctgcg aagggagcaa
6601 atgtttgtca gacacctgtt taacagagct ggcacaatgg gtgacagtgt ccctgatgac
6661 ttgtatatta aaggcagtgg aagcaatgtc aagcttgcca gccacgtgtt ttaccccaca
6721 cctagtggct caatggtgac atctgatgcc caattattta caagccata ctggttacag
6781 aaagcccagg ccataacaa tggcatctgt tggggcaacc aagtgttcct tactgtagtt
6841 gacaccacta ggagcacaaa catgacactg tgtgcatcca ctgcctccac agttactaca
6901 ccaTATAATa atgagagttt taaagagtac ctgcgacatg tggaggagtt tgatttgcaa
 signal ->
6961 tttatatttc agctgtgcaa ggtaaccctg aacactgaag taatggccta catacacagc
7021 atggatgcca gcatactgga ggactggaac tttggtttgc agcctcctcc gtctggctcc
7081 ttgcaggaca cctataggtt tgtgacgtct gccgccatca cctgtcaaaa acctgcaccc
```

```
7141 cccaaagaaa aggaagaccc gttggctaag tatacctttt gggaggtgga tttaaaggaa
7201 aagttttctg cagatttaga ccaatttccc ttaggccgca aatttttgct gcaagctggc
7261 atgcgtgcac ccctaccct gcgcgccccc aaacgcacag cctcatctac ctcatcttcc
7321 agcccccgca aacgcaaacg caccaaacgc TAAcgttgct tatatttatg ttgttgtacc
                                     <- L1 end
7381 cagtgtgcat gatttatgta tgtgtgcatg ttgtacgtga ttttgtattt cctgtgttgt
7441 gcgtgtcact gtttgtgtt gttgcgtgag tgtgttgcac ttatgtgttt attaaagtat
7501 gcgtggtcgc acccgagtga gtaactgtgt gtgtccggcg tgtagtttct gtcacatgca
7561 tgcatgcaca cccaaacact gttgccactg cctttaacag cttgcctgct gcacttccat
7621 tttgaaccct tctccatttt ccctgcaaac cctccatttt atggtctcgA CCGGTTTCGG
                                                             -> E2 bind
7681 Tcgcgcttgg cacgcatttt gggcaaacaa aaccacaaca ctgctaatcc tctggcttcc
7741 tgcctctcct actgctgcat acctgtggtt gtgctttggc gctccctggt gactcactgt
7801 ctctgcaaac aaaaatttgc acacacactt aatccaaccc tcttTGTACA AAATGCTTTt
                 glucocorticoid responsive element ->
7861 ggcagtacat ttctaagagt tactcatgct aattgcatag ttggccacaa tttcagggtt
7921 ggattgccaa tactatgtcc ttttaaatgt gattaatttt caaaatgttc ttgcaggtgt
7981 gtgtgACCGG GATCGGTcaa actttcacaa gcatttttta tagtaa
        -> E2 bind
```