

Groups A12, A13, A14

A12	
RhPV-1	RhPV-a
RhPV-b	RhPV-d
RhPV-e	MfPV
A13	
RhPV-c	RhPV-f
RhPV-g	RhPV-j
RhPV-i	RhPV-m
A14	
RhPV-h	RhPV-k

INTRODUCTION

Groups A12, A13, and A14 comprise the genital/mucosal viruses isolated from 286 Rhesus monkeys (*Macaca mulatta*) and a long-tailed macaque (*Macaca fascicularis*) living in three primate research laboratories in the United States [1]. Many animals showed pathological evidence of genital PV infection, including RhPV-1. "Gross lesions have included increased incidence of post-coital bleeding, vaginal leukoplasia, and vaginal polypoid masses that physically obstructed breeding. Histologic lesions included vaginal epithelial hyperplasia, mild-to-moderate lymphocytic vaginitis, and one case of severe vaginal epithelial dysplasia." [1]

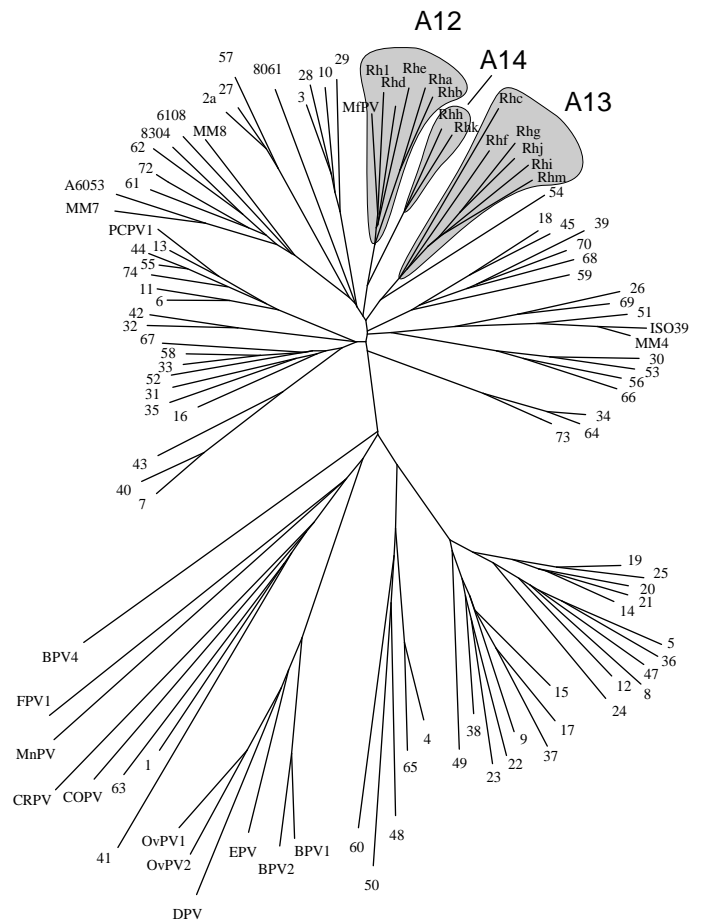
The twelve novel RhPV types were discovered from cervical smears amplified by the MY09-MY11 consensus primers. The MY09-MY11 region of each virus differed by more than 10% from each other and from all other viruses, thus qualifying them for "type" status. The three clades, A12, A13, and A14, contain all the Rhesus types, and, significantly, exclude all human types. HPV54, which appears to fall close to the A13 clade, actually diverges from the stem "below" the A13 branch and is considered the sole representative of an as-yet-unnamed A clade.

The sequences of these novel Rhesus PVs are presented on the following pages in alphabetical order, RhPV-a, RhPV-b ... RhPVm, MfPV. Note there is no RhPV-l ('el') to avoid confusion with RhPV-1 (one).

We also republish a revised sequence of the complete genome of the first Rhesus PV to be sequenced, RhPV-1, which differs from the original in two ways, a T has been found to occur between the two Gs (of the original sequence) at nt position 4144, and the C shown in this sequence at position 5897 was missing in the original [2]. The effect of the first insertion is to frameshift the E5 ORF and break it into two shorter ORFs, E5a, and E5b. The second insertion occurs within the L1 ORF and significantly lengthens it in a 5' direction. The corrected protein translations are given in the features table.

References

- [1] Chan, S-Y, Bernard, H-U, Ratterree, R. R., Birkebak, T. A., Faras, A. J., and Ostrow, R. S., Genomic Diversity and Evolution of Papillomaviruses in Rhesus Monkeys, *Journal of Virology* **71**:7 (1997).
- [2] Ostrow, R.S., Liu, Z., Schneider, J.F., McGlennen, R.C., Forslund, K. and Faras, A.J., The products of the E5, E6, or E7 open reading frames of RhPV 1 can individually transform NIH 3T3 cells or in cotransfections with activated ras can transform primary rodent epithelial cells, *Virology* **196**:2 861-867 (1993).



LOCUS RHPV1R 8028 bp DNA VRL 16-OCT-1995
DEFINITION Rhesus papillomavirus type 1, complete genome.
ACCESSION M60184 M37718
NID g1019918
KEYWORDS complete genome.
SOURCE Rhesus papillomavirus type 1 DNA, isolated from a metastasized tumor from a Macaca mulatta.
ORGANISM Rhesus papillomavirus type 1
Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
REFERENCE 1 (bases 1 to 8028)
AUTHORS Ostrow,R.S., LaBresh,K.V. and Faras,A.J.
TITLE Characterization of the complete RhPV 1 genomic sequence and an integration locus from a metastatic tumor
JOURNAL Virology 181 (1), 424-429 (1991)
MEDLINE 91135018
REFERENCE 2 (bases 1 to 8028)
AUTHORS Ostrow,R.S., Liu,Z., Schneider,J.F., McGlennen,R.C., Forslund,K. and Faras,A.J.
TITLE The products of the E5, E6, or E7 open reading frames of RhPV 1 can individually transform NIH 3T3 cells or in cotransfections with activated ras can transform primary rodent epithelial cells
JOURNAL Virology 196 (2), 861-867 (1993)
MEDLINE 93383416
REFERENCE 3 (bases 1 to 8028)
AUTHORS Ostrow,R.S.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-1991) Ronald S. Ostrow, Inst. of Human Genetics, University of Minnesota, Minneapolis, MN 55455, USA
COMMENT RhPV-1 was isolated and cloned from a metastatic tumor of a rhesus monkey. Ostrow et al. [1] reported a 71% prevalence of RhPV1 DNA within a mating cohort of monkeys including a male with penile carcinoma and subsequent lymph node metastasis. These data indicate that the most likely mode of RhPV transmission is sexual activity. Notable sequence similarity is seen in alignments of RhPV-1 DNA to Group A9 viruses, the sexually transmitted "high and intermediate risk" group of human anogenital papillomaviruses that includes HPV-16 and HPV-18. Recently, twelve new RhPV clones have been partially sequenced and found to fall in the "A", or "genital" supergroup (Chan, S-Y et al., Genomic Diversity and Evolution of Papillomaviruses in Rhesus Monkeys, Journal of Virology, July 1997). However, the new RhPVs are not associated with any of the existing clades within the A supergroup but cluster in three new clades. It is believed that each new RhPV is a new type based on L1 sequence dissimilarity of 10% with all other known papillomaviruses. Papillomavirus diversity among rhesus monkeys may be similar to that among humans.

The oncogenic potential of RhPV-1 is indicated by its ability to cooperatively transform primary epithelial cells with activated Ha-ras, independent of dexamethasone treatment (Ostrow et al. [1]). Ostrow et al. [2] have further shown that E5, E6, and E7 gene products can individually transform rodent epithelial cells. 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 GRE sequences of HPV-6, HPV-16, HPV-11 and RhPV differ by only a single nucleotide [1]. The integration of RhPV-1, at at least one locus, does not directly inactivate the E1 or E2 ORF [1]. An additional unique feature of the RhPV-1 genome consists of two 18 bp direct repeats occurring at the end of the E5 ORF and 6 bp upstream of the E2 ORF. These repeats may be the artifact of an insertion in an ancestor of the rhesus virus). Unregulated early gene expression is a critical factor

RhPV-1R

in oncogenic transformation [1].

SEQUENCE CORRECTION:

The sequence published here differs from the original in two ways, a T has been found to occur between the two Gs (of the original sequence) at nt position 4144, and the C shown in this sequence at position 5897 was missing in the original [2]. The effect of the first insertion is to frameshift the E5 ORF and break it into two shorter ORFs, E5a, and E5b. The second insertion occurs within the L1 ORF and significantly lengthens it in a 5' direction.

FEATURES

Location/Qualifiers

CDS	22..597 /note="E6 ORF from bp 1 to 597" /product="transforming protein" /gene="E6" /note="putative" /codon_start=1 /db_xref="PID:g333907" /translation="MKKGVTENGATESGAYKLLKLVFFVAMVDCPGEPNELPRTIHELCEQREETLHELQLECVYCLKELTRIEVYDFARWDLRLVHRQKPYGVCPICLRFYSKIRKYRRYEYSIYGCTLERRTRKQLVEVLIRCYCCQKPLCPIEKQRHVDQGRFHRIAGQWTGRCLMCWRPTVPETQPDDTQQGSSFLQA"
protein_bind	37..48 /bound_moiety="E2 protein" /note="putative"
protein_bind	52..63 /bound_moiety="E2 protein" /note="putative"
TATA_signal	67..71 /note="putative"
CDS	594..935 /note="E7 ORF from bp 582 to 935" /product="transforming protein" /gene="E7" /note="putative" /codon_start=1 /db_xref="PID:g333908" /translation="MIGPKPTLEDIVLDLQPPFPQPVDLMCYEQLSDSSEDEDEVDHHNNQQQHHQHARPEVPEDGDCYRIVSDCYSCGKPLRLVVVSSHEELRVLEDLLMGTLDIVCPSCASRV"
protein_bind	653..664 /bound_moiety="E2 protein" /note="putative"
CDS	941..2818 /note="E1 ORF from bp 902 to 2818" /product="replication protein" /gene="E1" /note="putative" /codon_start=1 /db_xref="PID:g333909" /translation="MDPEGTPGEGVGCTGWFNVEAIVERKTGDVVSEDEDDTETDGTGIDLVDLIDTTCGSVQTGDEAPGALLHAQETQAHAEAVQVLKRKFVGS PAVSPLGYNPCVDRDLS PRLNEISLNQSGQAKRRLFLPDSGYGNTEVETSLLQVAGGGGQDVQAGGKENTRPDDGGGDATQLLRCSNLKATLLSKFKSVYGVVSFSELVRSFKSDRTTCADWVVGAAGVHHSVAEGLKQLIQPFCSYAHIQCLTCDWGVYLLLLARFKCGKNRLTVSKCMSTLLNVQETHMLIEPPKLRSAALYWYRTGISNVSEVIGETPEWITRQTMFQHGLEDSIFDLS EMVQWAYDHDFTDSSVIAEYEAQLAGIDSNAAAF LKSNAQAKYVKDCATMCRHYKRAERQQMTMSQWIKQRCEKTDGGDWRPIVQFLRYQGVFEFIAFLAALKLFLKGI PKKNCIVLFGPPNTGKSYFGMSLIHFLQGSII SYVNSNSHFWLQPLADAKVAMLDDATPQCWSYIDNYLRNALDGNPISVDRKHKNLVQMKCPPLLITSNAGQDDRWMLHSRMVVFTEFQPPFPDQNGNPVYELNDKNWKSFFSRTWSRLDLQEEETENDGSTCRAFCKCVAGQNLRTV"
polyA_signal	1493..1498

protein_bind /note="putative"
 2570..2581
 /bound_moiety="E2 protein"
 /note="putative"
 CDS 2757..3857
 /note="E2 ORF from bp 2730 to 3857"
 /product="regulatory protein"
 /gene="E2"
 /note="putative"
 /codon_start=1
 /db_xref="PID:g333910"
 /translation="MMEALAERLSALQDRILELYEADSKDLKDQIEHWKCVRQECAVL
 YKAREVGFSHLNHQVPSLTVSRAKAHKAIEVQLALESLQNSEYNNEEWTLDASLEM
 WHTEPKGCFKKTGVPVTVLFDCKDNTMEYVLWGHYVWGDNGWVKTFGEADNWLHY
 TVAGEKVYYVQFYEDAKKYGHGNGNGDGYEWEVHVGGTVMHYSDSVSSATHCDKLPV
 EIVSGLQHINSPPPANPSAKENVWSSPAKRVRRSDSGGDPVRALD GKRSRVLCSAH
 NNATGSSGSDSYTPIVHLKGESNCLKLFRFLGKHKHLYINISSTWRWANHASEKAIV
 TVTFANELQRQQFLNTVKIPSTVTLTSGVMTV"
 TATA_signal 2889..2893
 /note="putative"
 TATA_signal 3018..3022
 /note="putative"
 CDS <3355..3630
 /note="E4 ORF from bp 3355 to 3630"
 /gene="E4"
 /note="putative"
 /codon_start=1
 /translation="CIILTCLALPTATNYPLLKLLADCNTSTHHPPPPTPAPRKTCG
 HRLQSECVGQTQVEIQCGPWTVKAGQSFVDLHTTTTLQGVVTVTIRL"
 repeat_region 3857..3963
 /rpt_unit=3857..3874, 3946..3963
 /standard_name="Direct repeat"
 /note="putative"
 CDS 4167..4295
 /gene="E5a"
 /note="ORF is thought to begin at nucleotide position
 4044; first methionine codon is found at nucleotide
 position 4167."
 /codon_start=1
 /db_xref="PID:g1019110"
 /translation="MLLFWVSI VNPFAAFGLCLFCFLTPLLLIHLHALSVVYSRMM"
 gene 4311..4445
 /gene="E5b"
 CDS 4311..4445
 /gene="E5b"
 /codon_start=1
 /db_xref="PID:g1019111"
 /translation="MLLSVILLLLLLIFTLFWHYLLVLRKPPAGRARKCKQLRRRRR
 "
 polyA_signal 4444..4449
 /note="putative"
 polyA_signal 4458..4463
 /note="putative"
 CDS 4469..5869
 /note="L2 ORF from bp 4463 to 5869"
 /product="minor capsid protein"
 /gene="L2"
 /note="putative"
 /codon_start=1
 /db_xref="PID:g333912"
 /translation="MKHAHLRRKRAAPRPPGGRQKRASATQLYQTCKAAGTCPDVI
 PKVEGTTVADQILKYGSMGVYFGGLGIGSGAGTGRSGYVPLGSRPASIPPELPRPPV
 TIEPVGSPDPSIVSLLEESRLIEAGVPAPTFFPTHGGFEISTSEVSTPAVLVDVSSGGSD
 VHVSVTSFTNPTFTPEPSVLRPPPPVEASGRLVISASSVSTHSYEEIPMDTFVITGDHN

RhPV-1R

```

YNTTSTPIPGSRAPARLGLYGRATQQVRVVDPAFITTPARLVTYDNPAYEGVDDATLQ
FSHSDIHQPPDPDFLDIVALHRPALTSRKGTVRF SRLGQRATLTTRSGKRIGAKVHFY
HDLSPIAPAESI E L Q P L S S Q G E L Y D I Y A D V D G Q E D A A A V A N T P L N S N S S G I A S P W N T T
V P L S A G A D V T L Q S G P D V S L D A P V A E S P V H P G V P L R P S A H I I L Y G G D F Y L H P S Y L G I R R
K R K R M H N F F S D V Y V A A "
TATA_signal 5123..5127
              /note="putative"
CDS          5850..7355
              /note="L1 ORF from bp 5748 to 7355"
              /product="major capsid protein"
              /gene="L1"
              /note="putative"
              /codon_start=1
              /db_xref="PID:g1019919"
              /translation="MSMWRPSDSKVYLPVSKVSTDEYVSRTSIYYHAGSSRLLAV
GHPYYAVKKGNNKVSVPKVSGLQYRVFRVRLPDPNKFGLPDPANFYDPNTQRLVWACLG
VEVGRGQPLGVGTS GHPLLNKLDDETFENGPKVAGGQADNRECVSMDYKQTQLCMLGCK
PPVGEHWGKGNPCTTGAAGDCPALELVNSVIQDGMVDTGYGAMDFNALQANKSDVPI
DICTSVCKYPDY LK M A S D P Y G D S L F F Y L R R E Q M F V R H L F N R A G T M G D S V P D D L Y I K G S
G S N V K L A S H V F Y P T P S G S M V T S D A Q L F N K P Y W L Q K A Q G H N N G I C W G N Q V F L T V V D T T R
S T N M T L C A S T A S T V T P Y N N E S F K E Y L R H V E E F D L Q F I F Q L C K V T L N T E V M A Y I H S M D
A S I L E D W N F L Q P P P S G S L Q D T Y R F V T S A A I T C Q K P A P P K E K E D P L A K Y T F W E V D L K E
K F S A D L D Q F P L G R K F L L Q A G M R A R P T L R A P K R T A S S T S S S S P R K R K R T K R
TATA_signal 6906..6910
              /note="putative"
protein_bind 7672..7683
              /bound_moiety="E2 protein"
              /note="putative"
protein_bind 7847..7861
              /bound_moiety="hormone receptor"
              /standard_name="glucocorticoid responsive element"
              /note="putative"
protein_bind 7988..7999
              /bound_moiety="E2 protein"
              /note="putative"
source       1..8028
              /organism="Rhesus papillomavirus type 1"
              /sequenced_mol="DNA"
BASE COUNT   2118 a   1790 c   2067 g   2053 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 ttttttgtag caatggtaga ctgccctggc
signal ->
              121 gagccaaaag aattgcccag gaccattcac gaactatgag agcagcgtga ggagaccctg
              181 cagcagcttc aattggagtg cgtgtattgc ctgaaggaac taacacgcat tgaggtatat
              241 gattttgcac ggtgggattt aagattggtg catagacaag gcaagccata cggggatagt
              301 cccatatgct tgaggtttta ctcaaaaatt cgaaaatata ggcgatacga gtattcaata
              361 tatgggtgta ctttagagcg tagaactaga aaacagttag tggaggtatt aataaggtgt
              421 tattgttgtc agaagcccct gtgtcccatt gaaaagcaaa gacacgtgga ccaaggacaa
              481 aggttccaca gaatagcggg acagtggacc ggaaggtgct tgatgtgctg gagaccaaca
              541 gtacctgaga cccagccaga cactgatcaa cagggcagTA Gtttcttgca agcATGAttg
              E7 orf start ->      E7 cds ->
              <- E6 end
              601 ggcctaaacc taccctcag gacattgtcc tagatttgca accatttcca caACCGCAAC
              -> E2 bind
              661 CCGTcgacct tatgtgttat gagcaattat ctgacagctc agaggatgag gatgaagtag
              721 accatcatca caataatcag cagcagcatc atcagcacgc cagacctgaa gtaccagagg
              781 atgggtattg ttatagaatt gtgagcgatt gttacagctg tggcaagcca ctgaggctgg
              841 ttgtggttag tagccacgaa gaggttacgtg tgctagagga cctgctgatg ggcacgctTG
              901 Acattgtgtg tcccagctgt gccagcagag tgTAActgca ATGgaccctg aaggtacacc
E1 orf ->      E1 cds ->
              <- E7 end
              961 aggggaaggg gtgggggtgta cgggggtggt taatgtggag gctatagtag aacgtaaac

```

```

1021 ggggggatgtg gtgtcagagg acgaagacga cacagaggat acagggatag atttggtaga
1081 ctttatagat gacacatgtg gaagtgtgca gacaggggac gaggcacctg gggcggttgtt
1141 gcacgcacag gaaacacaag cgcattgcaga ggcagtgacg gttttaaaac gaaagtttgt
1201 aggcagtcctg gcagtttagtc cgttgggaaa ctacaatccc tgtgtagaca gggattttaag
1261 tcccagatta aatgaaataa gttaaaccga aggcagcggg caggcaaac ggagactgtt
1321 tttgcccggac agcggttatg gcaatactga agtggaaacg tcgctattgc aggtagcagg
1381 gggggcgcc caggatgtac aggcaggggg gaaggaaaac acacggccag atgacggggg
1441 ggggggatgcc acgcagctgc tccgttgacg caacttaaaa gccactttgc tgAGTAAAtt
                                     signal ->
1501 taaatctgtg tatggagtta gcttttcaga gttggtgcca agctttaaaa gcgacaggac
1561 cacgtgcgct gactgggtgg tgggggcagc ggggggccat catagcgtgg cagagggggtt
1621 aaagcagctc attcagcctt tttgcagtta tgcacacatc cagtgcctta catgcccagt
1681 gggggtgtac ctgctactgc tggcaggggt taagtgtggc aaaaacagac taacagtttc
1741 taaatgcctg 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 acgctgtgc
2041 atttttaaaa agtaaatgcac aggcacaata tgtgaaggat tgtgccacta tgtgtaggca
2101 ctacaaaaga gccgaaaggc aacagatgac tatgtcacag tggataaaac aaagggtgtga
2161 aaaaactgat gatggagggg actggaggcc aatagtgcag tttttaaggt accaaggggt
2221 ggagtttata gcatttttag cagctttaaa gctgtttttg aagggcattc caaaaaaaaa
2281 ctgcatagtg ttatttggac cgccaaatac aggtaaatcc tactttggca tgagcttaat
2341 acatttcttg caaggttcta tcatttcata tgtaaattcc aacagtcact tttggttga
2401 gcctctggca gatgctaagg tggcattgtt ggatgatgca actcctcagt gctggtccta
2461 tatagataat tatttaagga acgcactgga cgggaacccc attagtgttg atagaaaaca
2521 taaaaatctt gtacagatga agtgccccc attgtttatt acctcaaacA CCAATGCAGG
                                     -> E2 bind
2581 Tcaggatgac aggtggatgt atttgcacag tagaatgggt gtgtttacat ttgaacagcc
2641 atttccattt gatcagaacg gtaatccagt ttatgagtta aatgataaaa actggaatc
2701 cttttctca aggacatggg ccagatTAGa tttacaagag gaagaggaga cggaaaATGA
                                     E2 orf start ->
2761 tggaagcact tgcagagcgt ttaagtgcgt tgcaggacag aatcttagaa ctgtaTGAag
                                     <- E1 end
2821 ctgatagcaa ggacttaaaa gaccaaatag agcactggaa atgtgtgccc caagaatgtg
2881 cagtgttgTA TAAggcacgg gaagtgggt tttcccacct gaaccatcag gtggtgccat
                                     signal ->
2941 cattaactgt gtcacgggct aaagcccaca aagcaattga agtgacgctg gcattagaga
3001 gtttacaaaa ttcggagTAT AAcaatgagg agtggacgct gcaagatgcc agcttgaga
                                     signal ->
3061 tgtggcacac agaacctaaag ggatgcttca aaaaaacagg tgttccagta acagtttgt
3121 ttgactgtga caaagacaat accatggagt atgtgctgtg gggacacata tatgtgtggg
3181 gggacaatgg atgggtgaag acattcgggt aggcggacaa ctggggctctg cactataccg
3241 ttgctgggga aaagggtgtac tatgtgcagt tttatgagga tgctaaaaaa tatggacatg
3301 gaaatggaaa tggagatggc tatgagtggg aggtgcatgt ttggtgggacg gTAAatgcatt
                                     E4 orf start ->
                                     NH2 terminus unknown
3361 attctgactc tgtgtctagc gctaccacct gcgacaaact acccactgtt gaaattgtta
3421 gcggactgca acacatcaac ccatcaccac cccccgcaa cccagcgcg aaggaaaacg
3481 tgtgtcctc gcctgcaaag cgagtgcgtc ggtcagactc aggtggagat ccagtgcggg
3541 ccttggacgg taaaagccgg tcagtcctt gtggatctgc acacaacaac gctacagggg
3601 gttccggtga cagtactat acgcctaTAG tgcacctaaa aggtgaatct aactgtttga
                                     <- E4 end
3661 agtgtttgag gttcagactg ggaaagcata agcacctgta tattaatata tcgtccacct
3721 ggaggtgggc aaaccatgca agtgagaag caattgtaac tgtgacattt gcaaatgagc
3781 ttcaaagaca acagttttta aacactgtaa aaataccttc tactgtaact ctgtcacaag
3841 gagtaatgac tgtgTAGTGT GCATTGGCAC ACAGggtttt gtattttttt ttttacaagt
                                     <- E2 end
                                     -> 18 bp repeat <-
3901 actgtttgta attaattttg tatattgact gtatattgaa ttgtgGTGTG CATTGGCACA
                                     -> 18 bp repeat
                                     NH2 terminus unknown
3961 CAGTggtctc atttcaagcc tgtacataca ttgaacagta tccaggctact gtgtaaagcc
<-

```

RhPV-1R

```

4021 aattgttgct gctacgcttg TAAaccgcca ccattctgct gtttctgggt gtgtttttgc
      E5a orf start ->
4081 tgctgttttt gcttggcctt gtgttttggg cacttggtga gtcgctgctt ctgtgttttt
4141 ccgtgtatgc ctcagtggtg ctgcttATGc tgttgttctg ggtgtccata gtgaaccctg
      E5a cds ->
4201 ttgcagcttt tggctctgtg ttgttttggg ttttaacccc gttgcttttg atacacctgc
4261 atgccctcag tgtggtttac agcagaatga tgTAAaact gcacatagac ATGttattat
      <- E5a end -> E5b cds
      E5b orf start ->
4321 cagttatfff gctattgttg ctgttaatat ttactttggt ttggcactac ttgttggtgc
4381 tgcgttttaa gccaccagca gggcgcgcac gTAAatgtaa acagctgaga cggcgccgca
4441 gaTAATAAAC gtcacacAAT AAAGcgtcAT Gaagcatgca cacttgctgc ggcgcaagcg
E5b end <- L2 orf start -> -> L2 cds
      signal -> signal ->
4501 agcagccccg cgcccacctg gtgggcgcca aaagcgtgca tctgccacgc agctgtacca
4561 aacctgcaag gcggcaggca catgcccccc cgatgttatc cctaaggtgg aaggcacaac
4621 cgtagcagat caaatfntaa agtatggcag catgggtgta tactttgggg gtttgggcat
4681 tggctctggg gctggcagg gcggaagaag cggctacgtg cccctagggt cagctcccg
4741 atccattccc gagccgttgc cagcaccacc agtaacaatt gagcctgtgg gcccttccga
4801 tccctccatt gtgtcattgc tggaaagagtc cagactaata gaggcagggtg ttccagcccc
4861 cacattcccc actcatgggg ggtttgaaat tagcacatct gaagttagca cccccgctgt
4921 cctggatgtg tctagcggtg gctctgatgt gcacgttagt gtgacctcct ttacaaacct
4981 tacctttact gagccatctg tgcctgcgacc cccgcccccc gtagaggcgt ctggacgcct
5041 ggtaatctct gcacccctcg tcagcagcga tagctacgaa gaaataccca tggacacatt
5101 tgtaataact ggagaccaca acTATAAcac aaccagcaca cccattcctg gttcacgtgc
      signal ->
5161 ccctgcacga cttggtctat atggacgtgc taccagcaa gtgccccgtg tggatcctgc
5221 atttataaacc acccctgcgc gactgggtgac atatgacaac cctgcatatg aggggtgagg
5281 cgatgccacc ctgcaatfff cccactctga cttaccagc ccgccagatc ctgactcttg
5341 tgacatttgg gcattgcaca gggccgcctt gacctcacgt aagggcaccg tgcgctttag
5401 ccgattagge cagcgggcaa cactaacacc gcgcagtggt aagcgtattg gggccaaggt
5461 gcatttctat catgacctca gtcccattgc ccctgcagaa agcatcgagt tgcagccct
5521 gtcactctag ggagagctgt atgacatata tgcagatgta gacgggcaag aggacgctgc
5581 agctgtggct aacaccccat taaacagcaa cagcagtggt attgcaagcc cctggaacac
5641 cacagtgcca ctcagtgagc gggcggagct gacgctgcag tccggccccg acgtgtccct
5701 ggatgcacca gtggctgaat cgctgtgca ccctggagtg cctcTAAGgc cttctgcaca
      L1 orf start ->
5761 tattattctg tacggggggag acttttattt gcaccctagc tacctcggta ttgcaggaa
5821 acgtaaaccg atgcacaatt tcttttcagA TGtctatgtg gcggccTAGt gactccaagg
      L1 cds -> <- L2 end
5881 tctacctacc acctgtccct gtgtctaagg tggtcagcac ggatgaatat gtctctcgca
5941 caagcatata ctatcacgct ggcagtcca gacttctggc tgttggacat ccctactatg
6001 ctgtaaagaa gggcaacaac aaagtgtcag tgcccagggt ttctggttta caataccag
6061 tgtttcgagt gcgtttgcct gaccccaata agtttggcct tccagatgct aacttttatg
6121 accetaaacac acagcgcctt gtgtgggcct gtttaggcgt ggagggtgggg cgtggacagc
6181 cactgggagt gggcaccagt ggtcatccac tgcctgaacaa actagatgac acggaaaatg
6241 gccctaaagt gggcggggga caaggagcag ataacagggga atgcgtgtca atggactaca
6301 agcaaacaca gctgtgcatg ctaggatgca agccccctgt ggggtgagcat tggggaaaaag
6361 gaaatccttg caccactggc gctgcaggtg actgcctgct acttgagctt gttactcag
6421 ttatacagga tggggacatg gttgatacag ggtatggcgc tatggacttt aatgcaactg
6481 aggccaacaa atcagatggt cctatagaca tatgcaactag cgtgtgcaaa taccctgact
6541 atttaaaaat ggcacagat ccctatggcg acagcttggg tttttacctg cgaagggagc
6601 aatgtttgt cagacacctg ttaacagag ctggcacaat ggggtgacagt gtcctctgatg
6661 acttgtatat taaaggcagt ggaagcaatg tcaagcttgc cagccacgtg ttttaccaca
6721 cacctagtgg ctcaatggg acatctgatg cccaattatt taacaagcca tactggttac
6781 agaaagccca gggccataac aatggcactc gttggggcaa ccaagtgctt cttactgtag
6841 ttgacaccac taggagcaca aacatgacac tgtgtgcatc cactgcctcc acagttacta
6901 caccaTATAA Taatgagagt ttaaaagagt acctgcgaca tgtggaggag tttgatttgc
      signal ->
6961 aatttatatt tcagctgtgc aaggtaacct tgaacactga agtaatggcc tacatacaca
7021 gcatggatg cagcactatg gaggactgga actttgggtt gcagcctcct ccgctctggct
7081 ccttgcaagg cacctatagg tttgtgacgt ctgccgccat cacctgtcaa aaacctgcac
7141 cccccaaaga aaaggaagac ccgttggcta agtatacctt ttgggagggtg gatttaaagg
7201 aaaagttttc tgcagattta gaccaatffc ccttaggccg caaatfnttg ctgcaagctg

```

```

7261 gcatgctgac acgccctacc ctgctgctcc ccaaacgcac agcctcatct acctcatctt
7321 ccagcccccg caaacgcaaa cgcaccaaac gcTAAcgttg cttatattta tgttggtgta
      <- L1 end
7381 cccagtgtgc atgatttatg tatgtgtgca tgttgtagct gattttgtat ttctgtgttt
7441 gtgctgtgca ctgttttctg ttgttgctgt agtgtgttgc acttatgtgt ttattaaagt
7501 atgctgtggtc gcacccgagt gagtaactgt gtgtgtccgg cgtgtagttt ctgtcacatg
7561 catgcatgca cacccaaaca ctgttgccac tgcctttaac agcttgctct ctgcacttcc
7621 attttgaacc cttctccatt ttccctgcaa accctccatt ttatggctct gACCGGTTTC
      -> E2 bind
7681 GGtcgctgctt ggcacgcatt ttgggcaaac aaaaccacaa cactgctaata cctctggctt
7741 cctgcctctc ctactgctgc atacctgtgg ttgtgctttg gcgctccctg gtgactcact
7801 gtctctgcaa acaaaaatth gcacacacac ttaatccaac cctcttTGTA CAAAATGCTT
      glucocorticoid responsive element ->
7861 ttggcagtac atttctaaga gttactcatg ctaattgcat agttggccac aatttcaggg
7921 ttggattgcc aatactatgt ctttttaaat gtgattaatt ttcaaaatgt tcttgcaggt
7981 gtgtgtgACC GGGATCGGTc aaactttcac aagcattttt tatagtaa
      -> E2 bind

```


RhPV-a

LOCUS RHPVA 449 bp DNA VRL 25-MAY-1997
DEFINITION Rhesus monkey papillomavirus strain RhPV-a L1 gene, partial cds.
ACCESSION U89656
NID g2108178
KEYWORDS .
SOURCE rhesus monkey papillomavirus, L1 MY09/MY11 region DNA
ORGANISM rhesus monkey papillomavirus
Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
Papillomavirus.
REFERENCE 1 (bases 1 to 449)
AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
and Ostrow,R.S.
TITLE Genomic diversity and evolution of papillomaviruses in rhesus
monkeys
JOURNAL J. Virol. (1997) In press
REFERENCE 2 (bases 1 to 449)
AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
and Ostrow,R.S.
TITLE Direct Submission
JOURNAL Submitted (14-FEB-1997) Institute of Molecular & Cell Biology,
National University of Singapore 119260, Republic of Singapore
COMMENT Original GenBank locus name is RMU89656.
In the first systematic study of diversity of papillomaviruses
in a nonhuman host (the Rhesus monkey) Chan et al. (1) isolated
twelve different and novel PV genomes which were given the temporary
names of RhPV-a through RhPV-m. The MY09/MY11 consensus primer
region of the L1 gene was sequenced in each case. Each of these
new L1 fragments differs by more than 10% from all other known
papillomaviruses, including RhPV1, thus qualifying them as new
"types." They fall into three new branches of the "genital," or
A-supergroup of the papillomavirus phylogentic tree and are distinct
from the eleven clades already found there. The frequency with which
new genomes were discovered in this study suggests that the ultimate
diversity of papillomaviruses in Rhesus monkeys may be comparable
to that of HPVs in Homo sapiens.
FEATURES Location/Qualifiers
source 1..449
/organism="rhesus monkey papillomavirus"
/note="name is provisional, pending complete isolation of
the genome whereupon it will be assigned a type number;"
/strain="rhesus monkey papillomavirus type a (RhPV-a)"
gene <1..>449
/gene="L1"
CDS <1..>449
/gene="L1"
/note="MY09/11 segment"
/codon_start=1
/db_xref="PID:g2108179"
/translation="AQGHNNGICWGNQVFLTVVDTRSTNITLCATATTEGTYKNDNF
KEYLRHVVEYDLQFVFLCKITLTTEVMSYIHNMDANILEDWNFGVQPPPTGTLQDTY
RFVQSEAIRCQKTAAPKQKEDPLSKYTFWDVLDLRDKFSADLDQFPLGR"
BASE COUNT 134 a 104 c 104 g 107 t
ORIGIN
1 gcacagggac ataataatgg catttggtg ggcaaccaag tgttcttgac agtagttgac
L1 cds ->
-> start MY09/MY11 region
61 accaccgca gactaacat aactatgt gccactgcca caactgaggg cacgtataag
121 aatgacaact tcaaagagta cttgcgtcac gttgaggaat atgacctgca gtttgtgtc
181 cagctatgca aaataacgct tactactgag gttatgagtt acatacacia catggacgcc
241 aatatattag aggattgaa tttgggggt caaccctc cactggcag gttgcaagc
301 acatatcggg ttgtacagtc agaggcaatt aggtgccaag agactgccg cccaaaacaa
361 aaagaagacc ccctcagca gtacacctt tgggatgtgg accttaggga taagttttc
421 gcagatttgg atcagtttcc cctgggagc
L1 cds ->
<- end MY09/MY11 region

LOCUS RHPVB 449 bp DNA VRL 25-MAY-1997
 DEFINITION Rhesus monkey papillomavirus strain RhPV-b L1 gene, partial cds.
 ACCESSION U89657
 NID g2108180
 KEYWORDS .
 SOURCE rhesus monkey papillomavirus, L1 MY09/MY11 region DNA
 ORGANISM rhesus monkey papillomavirus
 Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
 Papillomavirus.
 REFERENCE 1 (bases 1 to 449)
 AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
 and Ostrow,R.S.
 TITLE Genomic diversity and evolution of papillomaviruses in rhesus
 monkeys
 JOURNAL J. Virol. (1997) In press
 REFERENCE 2 (bases 1 to 449)
 AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
 and Ostrow,R.S.
 TITLE Direct Submission
 JOURNAL Submitted (14-FEB-1997) Institute of Molecular & Cell Biology,
 National University of Singapore 119260, Republic of Singapore
 COMMENT Original GenBank locus name is RMU89657.
 In the first systematic study of diversity of papillomaviruses
 in a nonhuman host (the Rhesus monkey) Chan et al. (1) isolated
 twelve different and novel PV genomes which were given the temporary
 names of RhPV-a through RhPV-m. The MY09/MY11 consensus primer
 region of the L1 gene was sequenced in each case. Each of these
 new L1 fragments differs by more than 10% from all other known
 papillomaviruses, including RhPV-1, thus qualifying them as new
 "types." They fall into three new branches of the "genital," or
 A-supergroup of the papillomavirus phylogentic tree and are distinct
 from the eleven clades already found there. The frequency with which
 new genomes were discovered in this study suggests that the ultimate
 diversity of papillomaviruses in Rhesus monkeys may be comparable
 to that of HPVs in Homo sapiens.
 FEATURES Location/Qualifiers
 source 1..449
 /organism="rhesus monkey papillomavirus"
 /note="provisional nomenclature until complete genome is
 isolated"
 /strain="rhesus monkey papillomavirus type b (RhPV-b)"
 gene <1..>449
 /gene="L1"
 CDS <1..>449
 /gene="L1"
 /note="MY09/11 segment"
 /codon_start=1
 /db_xref="PID:g2108181"
 /translation="AQGHNNGICWGNQVFLTVVDTRSTNITLCATKTS EDTYKNDNF
 REYLRHMEEFDLQFVFLCKITLTTEVMAYIHNMDPSILEDWNFGVQPPPSGTLQDQTY
 RFVQSEAIRCQKTAAPKVKEDPLSKYTFWDVLDLRDKFSADLDQYPLGR"
 BASE COUNT 133 a 104 c 104 g 108 t
 ORIGIN
 1 gcacagggac ataataatgg catatgctgg ggcaatcaag tgttcttgac agtagttgac
 L1 cds ->
 -> start MY09/MY11 region
 61 acaacacgga gactaacaat cactctttgt gccacaaaa catctgagga cacctacaag
 121 aatgacaact ttagggagta cttgcgtcac atggaggaat ttgacttgca gtttgtgttc
 181 cagctgtgca aatcacgct acaactgag gttatggcat acatacaca catggaccct
 241 agcatactag aggactggaa tttggtgtg caaccocctc cgtctggtac actgcaggat
 301 acttataggt ttgtacaatc agaggccatt agatgccaga aaaccgctgc cccaagggt
 361 aaggaagacc ccctcagca gtacacattt tgggatgtgg acctaaggga caaatatttc
 421 gcagatttgg atcagtatcc ccttggacg
 L1 cds ->
 <- end MY09/MY11 region

RhPV-c

LOCUS RHPVC 445 bp DNA VRL 25-MAY-1997
DEFINITION Rhesus monkey papillomavirus strain RhPV-c L1 gene, partial cds.
ACCESSION U89658
NID g2114413
KEYWORDS .
SOURCE rhesus monkey papillomavirus, L1 MY09/MY11 region DNA
ORGANISM rhesus monkey papillomavirus
Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
Papillomavirus.
REFERENCE 1 (bases 1 to 445)
AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
and Ostrow,R.S.
TITLE Genomic diversity and evolution of papillomaviruses in rhesus
monkeys
JOURNAL J. Virol. (1997) In press
REFERENCE 2 (bases 1 to 445)
AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
and Ostrow,R.S.
TITLE Direct Submission
JOURNAL Submitted (14-FEB-1997) Institute of Molecular & Cell Biology,
National University of Singapore 119260, Republic of Singapore
COMMENT Original GenBank locus name is RMU89658.
In the first systematic study of diversity of papillomaviruses
in a nonhuman host (the Rhesus monkey) Chan et al. (1) isolated
twelve different and novel PV genomes which were given the temporary
names of RhPV-a through RhPV-m. The MY09/MY11 consensus primer
region of the L1 gene was sequenced in each case. Each of these
new L1 fragments differs by more than 10% from all other known
papillomaviruses, including RhPV-1, thus qualifying them as new
"types." They fall into three new branches of the "genital," or
A-supergroup of the papillomavirus phylogentic tree and are distinct
from the eleven clades already found there. The frequency with which
new genomes were discovered in this study suggests that the ultimate
diversity of papillomaviruses in Rhesus monkeys may be comparable
to that of HPVs in Homo sapiens.
FEATURES Location/Qualifiers
source 1..445
/organism="rhesus monkey papillomavirus"
/note="provisional nomenclature until complete genome is
isolated"
/strain="rhesus monkey papillomavirus type c (RhPV-c)"
gene <1..>445
/gene="L1"
CDS <1..>445
/gene="L1"
/note="MY09/11 segment"
/codon_start=1
/db_xref="PID:g2114414"
/translation="AQGHNNIGCWNNQLFLTVDTRTTNLTVCATATQSGTFKAADF
KEYVRHVEEFDLQFIFQLCTITLTSVDMAYIHGMDPSILEDWNFGIQPPSSSLEDKY
RFIQSQAITCQKPDPKAPKEDPLSQFNFEVDLKERFSADLDQYPL"
BASE COUNT 113 a 110 c 104 g 118 t
ORIGIN
1 gccacgggac ataataatgg catatgctgg aacaatcagc tgttccttac tgttggtggat
L1 cds ->
-> start MY09/MY11 region
61 actacgcgga cgactaacct tactgtgtgt gcaacagcca cgcaatctgg aacctttaag
121 gctgcagatt ttaaggaata tgtaggcac gtggaggaat tgacactgca gttcattttc
181 cagttgtgca ccatcacctt gacctctgat gttatggcat acatacatgg catggaccct
241 tccattcttg aggactggaa tttcggcata cagcccctc cctctagcag ttagaggat
301 aagtataggt ttatccagtc ccaggccatt acatgccaga agcccgacc taaagccct
361 aaggaggatc ccctgtcaca atttaatttc tgggaagtag atttgaagga gcggttttct
421 gcagacctgg atcagtatcc cctgg
L1 cds ->
<- end MY09/MY11 region

LOCUS RHPVD 438 bp DNA VRL 25-MAY-1997
 DEFINITION Rhesus monkey papillomavirus strain RhPV-d L1 gene, partial cds.
 ACCESSION U89659
 NID g2108184
 KEYWORDS .
 SOURCE rhesus monkey papillomavirus, L1 MY09/MY11 region DNA
 ORGANISM rhesus monkey papillomavirus
 Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
 Papillomavirus.
 REFERENCE 1 (bases 1 to 438)
 AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
 and Ostrow,R.S.
 TITLE Genomic diversity and evolution of papillomaviruses in rhesus
 monkeys
 JOURNAL J. Virol. (1997) In press
 REFERENCE 2 (bases 1 to 438)
 AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
 and Ostrow,R.S.
 TITLE Direct Submission
 JOURNAL Submitted (14-FEB-1997) Institute of Molecular & Cell Biology,
 National University of Singapore 119260, Republic of Singapore
 COMMENT Original GenBank locus name is RMU89659.
 In the first systematic study of diversity of papillomaviruses
 in a nonhuman host (the Rhesus monkey) Chan et al. (1) isolated
 twelve different and novel PV genomes which were given the temporary
 names of RhPV-a through RhPV-m. The MY09/MY11 consensus primer
 region of the L1 gene was sequenced in each case. Each of these
 new L1 fragments differs by more than 10% from all other known
 papillomaviruses, including RhPV-1, thus qualifying them as new
 "types." They fall into three new branches of the "genital," or
 A-supergroup of the papillomavirus phylogentic tree and are distinct
 from the eleven clades already found there. The frequency with which
 new genomes were discovered in this study suggests that the ultimate
 diversity of papillomaviruses in Rhesus monkeys may be comparable
 to that of HPVs in Homo sapiens.
 FEATURES Location/Qualifiers
 source 1..438
 /organism="rhesus monkey papillomavirus"
 /note="provisional nomenclature until complete genome is
 isolated"
 /strain="rhesus monkey papillomavirus type d (RhPV-d)"
 gene <1..>438
 /gene="L1"
 CDS <1..>438
 /gene="L1"
 /note="MY09/11 segment"
 /codon_start=1
 /db_xref="PID:g2108185"
 /translation="AQGHNNGICWGNQVFLTVVDTRSTNMTLCAATANDATYNNDSF
 KEYLRHVEEYDIQFIFQLCKITLTDVMAYIHGMDAGILEDWNFGLQPPPSGSLQDTY
 RFVTSSAIACQKTTTPKEDPLAKYTFWEVDLKEKFSADLDQF"
 BASE COUNT 129 a 98 c 98 g 113 t
 ORIGIN
 1 gcacagggtc ataacaatgg catttgctgg ggtaaccagg tatttctcac tgttggtggat
 L1 cds ->
 -> start MY09/MY11 region
 61 accacaagaa gactaacaat gacactgtgt gcagctactg ctaatgatgc aacatataac
 121 aacgatagtt ttaaagagta cttgcgccac gtggaggaat atgacataca atttatattt
 181 caattgtgca aatcactct cactactgat gttatggcct acattcacgg catggatgca
 241 ggcatccttg aggactggaa tttggggttg cagcccctc cgtccggcag tttgcaagac
 301 acgtataggt ttgtaacgtc ctccgccatt gcatgccaaa aaaccacccc gccaaaggaa
 361 aaggaagacc ccttagcaaa atacacattt tgggaggtgg atcttaagga aaagttctct
 421 gcagacttgg atcagttt
 L1 cds ->
 <- end MY09/MY11 region

RhPV-e

LOCUS RHPVE 449 bp DNA VRL 25-MAY-1997
DEFINITION Rhesus monkey papillomavirus strain RhPV-e L1 gene, partial cds.
ACCESSION U89660
NID g2108186
KEYWORDS .
SOURCE rhesus monkey papillomavirus, L1 MY09/MY11 region DNA
ORGANISM rhesus monkey papillomavirus
Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
Papillomavirus.
REFERENCE 1 (bases 1 to 449)
AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
and Ostrow,R.S.
TITLE Genomic diversity and evolution of papillomaviruses in rhesus
monkeys
JOURNAL J. Virol. (1997) In press
REFERENCE 2 (bases 1 to 449)
AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
and Ostrow,R.S.
TITLE Direct Submission
JOURNAL Submitted (14-FEB-1997) Institute of Molecular & Cell Biology,
National University of Singapore 119260, Republic of Singapore
COMMENT Original GenBank locus name is RMU89660.
In the first systematic study of diversity of papillomaviruses
in a nonhuman host (the Rhesus monkey) Chan et al. (1) isolated
twelve different and novel PV genomes which were given the temporary
names of RhPV-a through RhPV-m. The MY09/MY11 consensus primer
region of the L1 gene was sequenced in each case. Each of these
new L1 fragments differs by more than 10% from all other known
papillomaviruses, including RhPV-1, thus qualifying them as new
"types." They fall into three new branches of the "genital," or
A-supergroup of the papillomavirus phylogentic tree and are distinct
from the eleven clades already found there. The frequency with which
new genomes were discovered in this study suggests that the ultimate
diversity of papillomaviruses in Rhesus monkeys may be comparable
to that of HPVs in Homo sapiens.
FEATURES Location/Qualifiers
source 1..449
/organism="rhesus monkey papillomavirus"
/note="provisional nomenclature until complete genome is
isolated"
/strain="rhesus monkey papillomavirus type e (RhPV-e)"
gene <1..>449
/gene="L1"
CDS <1..>449
/gene="L1"
/note="MY09/11 segment"
/codon_start=1
/db_xref="PID:g2108187"
/translation="AQGHNNGICWGNQVFLTVVDTRSTNMTLCSTGTDATYKNDNF
KEYMRHVVEFDLQFIFQLCKITLTTEVMAYIHNMDASILEDWNFGLQAPPTGSLQDQTY
RFVTSAAITCQKTAPPKEKEDPLAKYAFWDVNLKEKFSADLDQFPLGR"
BASE COUNT 135 a 101 c 98 g 115 t
ORIGIN
1 gccacgggtc ataataatgg catttgctgg gaaaccaag tatttttaac agttgtggat
L1 cds ->
-> start MY09/MY11 region
61 accaccagga gcacaaatat gacgctgtgt gcatccactg ggacagatgc cacgtacaaa
121 aatgacaact ttaaagagta catgctcat gttgaggagt ttgacttgca gtttatatt
181 caactgtgta aaattaccct caccactgag gtgatggcct acatccacaa catggatgcc
241 agcatattag aggactggaa ctttggtttg caggccccc caacaggtag tttgcaggac
301 acatacagat ttgtaacctc tgctgctata acttgccaga aaaccgccc cccaaaagaa
361 aaagaggacc ctctagctaa atatgcattc tgggatgta acttaaagga aaaattctct
421 gctgatcttg atcagtttcc ccttgagc
L1 cds ->
<- end MY09/MY11 region

```

LOCUS      RHPVF          445 bp    DNA          VRL          25-MAY-1997
DEFINITION Rhesus monkey papillomavirus strain RhPV-f L1 gene, partial cds.
ACCESSION  U89661
NID        g2108188
KEYWORDS   .
SOURCE     rhesus monkey papillomavirus, L1 MY09/MY11 region DNA
ORGANISM   rhesus monkey papillomavirus
           Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
           Papillomavirus.
REFERENCE  1 (bases 1 to 445)
AUTHORS    Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
           and Ostrow,R.S.
TITLE      Genomic diversity and evolution of papillomaviruses in rhesus
           monkeys
JOURNAL    J. Virol. (1997) In press
REFERENCE  2 (bases 1 to 445)
AUTHORS    Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
           and Ostrow,R.S.
TITLE      Direct Submission
JOURNAL    Submitted (14-FEB-1997) Institute of Molecular & Cell Biology,
           National University of Singapore 119260, Republic of Singapore
COMMENT    Original GenBank locus name is RMU89661.
           In the first systematic study of diversity of papillomaviruses
           in a nonhuman host (the Rhesus monkey) Chan et al. (1) isolated
           twelve different and novel PV genomes which were given the temporary
           names of RhPV-a through RhPV-m. The MY09/MY11 consensus primer
           region of the L1 gene was sequenced in each case. Each of these
           new L1 fragments differs by more than 10% from all other known
           papillomaviruses, including RhPV-1, thus qualifying them as new
           "types." They fall into three new branches of the "genital," or
           A-supergroup of the papillomavirus phylogentic tree and are distinct
           from the eleven clades already found there. The frequency with which
           new genomes were discovered in this study suggests that the ultimate
           diversity of papillomaviruses in Rhesus monkeys may be comparable
           to that of HPVs in Homo sapiens.
FEATURES   Location/Qualifiers
   source   1..445
            /organism="rhesus monkey papillomavirus"
            /note="provisional nomenclature until complete genome is
            isolated"
            /strain="rhesus monkey papillomavirus type f (RhPV-f)"
   gene     <1..>445
            /gene="L1"
   CDS      <1..>445
            /gene="L1"
            /note="MY09/11 segment"
            /codon_start=2
            /db_xref="PID:g2108189"
            /translation="QGHNNGICWNNQLFVTVVDTRSTNLTVCATEKSEETFKASNFK
            EYVRHVEEFDLQFIFQLCTITLTAEIMQYIHTMDPNILEAWEFVTPPPSSSLEDKYR
            FVQSQAITCQKDAPAKQKEDPYANLNFVWVVDLKERFSADLDQFPLG"
BASE COUNT 124 a    103 c    101 g    117 t
ORIGIN
1 acagggacat aacaatggta tatgctggaa caaccagctg tttggtactg tggtagatac
L1 cds ->
-> start MY09/MY11 region
61 caccgtagc accaatctta ctgtctgcmc cacagaaaaa tctgaggaaa cattcaaagc
121 ctctaacttt aaggaatagc tgcgccatgt ggaggagttt gatctgcagt ttatatttca
181 actgtgcacc atcacactca cagctgaaa atgcagtagc atacacacta tggatcctaa
241 cattttagag gcctgggagt ttggtgtaac accccccct tccagtagtt tggaggataa
301 gtatcgcttc gtgcagtccc aggctatcac gtgtcagaag gacgcccctg caaagcaaaa
361 ggaggacccc tatgctaacc ttaacttttg ggtggtagat ttgaaggaac ggttttctgc
421 agatttgat cagtttcccc ttgga
L1 cds ->
<- end MY09/MY11 region

```

RhPV-g

LOCUS RHPVG 433 bp DNA VRL 25-MAY-1997
DEFINITION Rhesus monkey papillomavirus strain RhPV-g L1 gene, partial cds.
ACCESSION U89662
NID g2108190
KEYWORDS .
SOURCE rhesus monkey papillomavirus, L1 MY09/MY11 region DNA
ORGANISM rhesus monkey papillomavirus
Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
Papillomavirus.
REFERENCE 1 (bases 1 to 433)
AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
and Ostrow,R.S.
TITLE Genomic diversity and evolution of papillomaviruses in rhesus
monkeys
JOURNAL J. Virol. (1997) In press
REFERENCE 2 (bases 1 to 433)
AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
and Ostrow,R.S.
TITLE Direct Submission
JOURNAL Submitted (14-FEB-1997) Institute of Molecular & Cell Biology,
National University of Singapore 119260, Republic of Singapore
COMMENT Original GenBank locus name is RMU89662.
In the first systematic study of diversity of papillomaviruses
in a nonhuman host (the Rhesus monkey) Chan et al. (1) isolated
twelve different and novel PV genomes which were given the temporary
names of RhPV-a through RhPV-m. The MY09/MY11 consensus primer
region of the L1 gene was sequenced in each case. Each of these
new L1 fragments differs by more than 10% from all other known
papillomaviruses, including RhPV-1, thus qualifying them as new
"types." They fall into three new branches of the "genital," or
A-supergroup of the papillomavirus phylogentic tree and are distinct
from the eleven clades already found there. The frequency with which
new genomes were discovered in this study suggests that the ultimate
diversity of papillomaviruses in Rhesus monkeys may be comparable
to that of HPVs in Homo sapiens.
FEATURES Location/Qualifiers
source 1..433
/organism="rhesus monkey papillomavirus"
/note="provisional nomenclature until complete genome is
isolated"
/strain="rhesus monkey papillomavirus type g (RhPV-g)"
gene <1..>433
/gene="L1"
CDS <1..>433
/gene="L1"
/note="MY09/11 segment"
/codon_start=1
/db_xref="PID:g2108191"
/translation="AQGHNNGICWGNEVFVTVVDTRSTNLTVCTTESEATTFQASNF
KEYTRHVEEYDLQFIFQLCTITLTAEVMQYIHTMDPAILEDWKFVTPPPSSLENKY
RFITSQAITCQKDAPPKEKEDPYARLNFVVDLKD RFSADLD"
BASE COUNT 112 a 105 c 103 g 113 t
ORIGIN
1 gcacagggtc ataacaatgg catatgctgg ggcaatgagg tgtttgttac agttgtggac
L1 cds ->
-> start MY09/MY11 region
61 accactagaa gcaccaacct tactgtgtgc actactgagt ccgaggctac cacatttcag
121 gcctccaatt ttaaggaata cacacgacac gtggaggagt atgacctgca gtttattttt
181 cagctgtgca ccattactct cactgcagag gttatgcagt acatccatac tatggatcct
241 gcgattttga aggactggaa gtttggggtc actcctccgc cgtccagcag cttagagaat
301 aagtacaggt ttataacttc tcaggcaatt acatgccaga aggacgcccc cccaaggag
361 aaggaagatc cttatgcccg ccttaacttc tgggttgtgg accttaagga ccgcttttct
421 gcagatttag atc
L1 cds ->
<- end MY09/MY11 region

LOCUS RHPVH 449 bp DNA VRL 25-MAY-1997
 DEFINITION Rhesus monkey papillomavirus strain RhPV-h L1 gene, partial cds.
 ACCESSION U89663
 NID g2108192
 KEYWORDS .
 SOURCE rhesus monkey papillomavirus, L1 MY09/MY11 region DNA
 ORGANISM rhesus monkey papillomavirus
 Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
 Papillomavirus.
 REFERENCE 1 (bases 1 to 449)
 AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
 and Ostrow,R.S.
 TITLE Genomic diversity and evolution of papillomaviruses in rhesus
 monkeys
 JOURNAL J. Virol. (1997) In press
 REFERENCE 2 (bases 1 to 449)
 AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
 and Ostrow,R.S.
 TITLE Direct Submission
 JOURNAL Submitted (14-FEB-1997) Institute of Molecular & Cell Biology,
 National University of Singapore 119260, Republic of Singapore
 COMMENT Original GenBank locus name is RMU89663.
 In the first systematic study of diversity of papillomaviruses
 in a nonhuman host (the Rhesus monkey) Chan et al. (1) isolated
 twelve different and novel PV genomes which were given the temporary
 names of RhPV-a through RhPV-m. The MY09/MY11 consensus primer
 region of the L1 gene was sequenced in each case. Each of these
 new L1 fragments differs by more than 10% from all other known
 papillomaviruses, including RhPV-1, thus qualifying them as new
 "types." They fall into three new branches of the "genital," or
 A-supergroup of the papillomavirus phylogentic tree and are distinct
 from the eleven clades already found there. The frequency with which
 new genomes were discovered in this study suggests that the ultimate
 diversity of papillomaviruses in Rhesus monkeys may be comparable
 to that of HPVs in Homo sapiens.
 FEATURES Location/Qualifiers
 source 1..449
 /organism="rhesus monkey papillomavirus"
 /note="provisional nomenclature until complete genome is
 isolated"
 /strain="rhesus monkey papillomavirus type h (RhPV-h)"
 gene <1..>449
 /gene="L1"
 CDS <1..>449
 /gene="L1"
 /note="MY09/11 segment"
 /codon_start=1
 /db_xref="PID:g2108193"
 /translation="AQGHNNGICWGNQVFVTVVDTRSTNMTLCPATSSSETTYDASKF
 KEYLRHVVEYDLQFIFQLCKITLNAEVMSYIHTMNAGLLDDWNFGLSPPPSSLEDY
 RFIQSAAIRCQKDTPPPEKKDPFAQYTFWDVLDLKEKFSLLDLQFPLGR"
 BASE COUNT 116 a 114 c 105 g 114 t
 ORIGIN
 1 gcacagggac ataataatgg catatgctgg ggtaaccagg tgtttgtcac ggtggtggac
 L1 cds ->
 -> start MY09/MY11 region
 61 accacacgca gactaacaat gaccctgtgt cgggccacgt cctcagagac cacgtatgac
 121 gcttccaagt ttaaggagta cttgcgacat gtggaggaat atgatttgca gtttattttc
 181 caactgtgca aatcaccct taatgctgag gttatgtcct acattcacac aatgaatgca
 241 ggcctcctag atgactggaa tttcggcctt agtcccctgc cgtccagcag tctggaggat
 301 acctataggt ttattcagtc tgctgccata cggtgccaaa aagacactcc acccctgaa
 361 aaaaaggacc cctttgcaca atacacctt tgggacgtgg atctcaagga aaagttttct
 421 ctggatttgg atcagtttcc cctgggacg
 L1 cds ->
 <- end MY09/MY11 region

RhPV-i

LOCUS RHPVI 449 bp DNA VRL 25-MAY-1997
DEFINITION Rhesus monkey papillomavirus strain RhPV-i L1 gene, partial cds.
ACCESSION U89664
NID g2108194
KEYWORDS .
SOURCE rhesus monkey papillomavirus, L1 MY09/MY11 region DNA
ORGANISM rhesus monkey papillomavirus
Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
Papillomavirus.
REFERENCE 1 (bases 1 to 449)
AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
and Ostrow,R.S.
TITLE Genomic diversity and evolution of papillomaviruses in rhesus
monkeys
JOURNAL J. Virol. (1997) In press
REFERENCE 2 (bases 1 to 449)
AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
and Ostrow,R.S.
TITLE Direct Submission
JOURNAL Submitted (14-FEB-1997) Institute of Molecular & Cell Biology,
National University of Singapore 119260, Republic of Singapore
COMMENT Original GenBank locus name is RMU89664.
In the first systematic study of diversity of papillomaviruses
in a nonhuman host (the Rhesus monkey) Chan et al. (1) isolated
twelve different and novel PV genomes which were given the temporary
names of RhPV-a through RhPV-m. The MY09/MY11 consensus primer
region of the L1 gene was sequenced in each case. Each of these
new L1 fragments differs by more than 10% from all other known
papillomaviruses, including RhPV-1, thus qualifying them as new
"types." They fall into three new branches of the "genital," or
A-supergroup of the papillomavirus phylogentic tree and are distinct
from the eleven clades already found there. The frequency with which
new genomes were discovered in this study suggests that the ultimate
diversity of papillomaviruses in Rhesus monkeys may be comparable
to that of HPVs in Homo sapiens.
FEATURES Location/Qualifiers
source 1..449
/organism="rhesus monkey papillomavirus"
/note="provisional nomenclature until complete genome is
isolated"
/strain="rhesus monkey papillomavirus type i (RhPV-i)"
gene <1..>449
/gene="L1"
CDS <1..>449
/gene="L1"
/note="MY09/11 segment"
/codon_start=1
/db_xref="PID:g2108195"
/translation="AQGHNNGICWNNQLFVTVVDTTRSTNMTVCATATQANNFQAGNF
KQYIRHVEEYDLQFVFLCSITLTAEVQMVMHTMDPSILEEWKFGVTPPPSSLEDKY
RFIQSRAISCQKDAAPQAKEDPYDKLNFVWVVDLKD RFSADLDQFPLGR"
BASE COUNT 117 a 103 c 110 g 119 t
ORIGIN
1 gcacagggtc ataataatgg catctgttgg aataatcagt tgtttgttac tgttgtagac
L1 cds ->
-> start MY09/MY11 region
61 accacgcgta gtaccaatat gacagtgtgc gctacagcca cgcaggccaa taatttccag
121 gcaggaact ttaagcagta catacgccat gtggaggaat atgacctgca gtttgtgtt
181 cagttgtgct ccataacact cactgctgaa gttatgcagt acatgcatac tatggatccc
241 tctatattgg aggaatggaa atttggcgtt actccccac cctccagctc tttggaggac
301 aagtataggt ttatacagtc gcgggccatt agttgccaga aggacgaggc cccccaggcg
361 aaggaagacc cctatgataa gttaaacttt tgggttgtgg atttaaagga ccgcttctct
421 gcagacctag atcagtttcc cctgggagc
L1 cds ->
<- end MY09/MY11 region

LOCUS RHPVJ 449 bp DNA VRL 25-MAY-1997
 DEFINITION Rhesus monkey papillomavirus strain RhPV-j L1 gene, partial cds.
 ACCESSION U89665
 NID g2108196
 KEYWORDS .
 SOURCE rhesus monkey papillomavirus, L1 MY09/MY11 region DNA
 ORGANISM rhesus monkey papillomavirus
 Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
 Papillomavirus.
 REFERENCE 1 (bases 1 to 449)
 AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
 and Ostrow,R.S.
 TITLE Genomic diversity and evolution of papillomaviruses in rhesus
 monkeys
 JOURNAL J. Virol. (1997) In press
 REFERENCE 2 (bases 1 to 449)
 AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
 and Ostrow,R.S.
 TITLE Direct Submission
 JOURNAL Submitted (14-FEB-1997) Institute of Molecular & Cell Biology,
 National University of Singapore 119260, Republic of Singapore
 COMMENT Original GenBank locus name is RMU89665.
 In the first systematic study of diversity of papillomaviruses
 in a nonhuman host (the Rhesus monkey) Chan et al. (1) isolated
 twelve different and novel PV genomes which were given the temporary
 names of RhPV-a through RhPV-m. The MY09/MY11 consensus primer
 region of the L1 gene was sequenced in each case. Each of these
 new L1 fragments differs by more than 10% from all other known
 papillomaviruses, including RhPV-1, thus qualifying them as new
 "types." They fall into three new branches of the "genital," or
 A-supergroup of the papillomavirus phylogentic tree and are distinct
 from the eleven clades already found there. The frequency with which
 new genomes were discovered in this study suggests that the ultimate
 diversity of papillomaviruses in Rhesus monkeys may be comparable
 to that of HPVs in Homo sapiens.
 FEATURES Location/Qualifiers
 source 1..449
 /organism="rhesus monkey papillomavirus"
 /note="provisional nomenclature until complete genome is
 isolated"
 /strain="rhesus monkey papillomavirus type j (RhPV-j)"
 gene <1..>449
 /gene="L1"
 CDS <1..>449
 /gene="L1"
 /note="MY09/11 segment"
 /codon_start=1
 /db_xref="PID:g2108197"
 /translation="AQGHNNGICWGNQLFVTVVDTRSTNLTVCCTESEATNFQASNF
 KEYTRHVEEFDLQFIFQLCTITLSAEVMQYIHTMDPAILEDWKFVTPPPSSLEDKY
 RFITSQAITCQKDTPPKEKEDPYARLNFVVDLKD RFSADLDQFPLGR"
 BASE COUNT 116 a 107 c 101 g 125 t
 ORIGIN
 1 gccacgggac ataataatgg catctgctgg gaaatcagt tgtttgttac tgttggtggac
 L1 cds ->
 -> start MY09/MY11 region
 61 accaccagaa gtacaaactt aacagtgtgc actactgagt ccgaggccac caatttccag
 121 gcctccaatt ttaaagaata cacgcgtcac gtagaagaat ttgacttgca gtttatcttc
 181 cagttgtgca ctattacctt gtctgcagaa gtgatgcagt acatccatac tatggaccct
 241 gctatattgg aggactggaa gtttgggggtt actcccccc cttccagcag tttggaggat
 301 aagtatagggt ttatcacgtc ccaggctatt acatgtcaaa aagatacccc cccgaaagaa
 361 aaggaagatc cctatgcccg ccttaacttt tgggttggg atcttaagga ccgcttctct
 421 gcagatttgg atcagtttcc ccttgggacg
 L1 cds ->
 <- end MY09/MY11 region

RhPV-k

LOCUS RHPVK 446 bp DNA VRL 25-MAY-1997
DEFINITION Rhesus monkey papillomavirus strain RhPV-k L1 gene, partial cds.
ACCESSION U89666
NID g2108198
KEYWORDS .
SOURCE rhesus monkey papillomavirus, L1 MY09/MY11 region DNA
ORGANISM rhesus monkey papillomavirus
Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
Papillomavirus.
REFERENCE 1 (bases 1 to 446)
AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
and Ostrow,R.S.
TITLE Genomic diversity and evolution of papillomaviruses in rhesus
monkeys
JOURNAL J. Virol. (1997) In press
REFERENCE 2 (bases 1 to 446)
AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
and Ostrow,R.S.
TITLE Direct Submission
JOURNAL Submitted (14-FEB-1997) Institute of Molecular & Cell Biology,
National University of Singapore 119260, Republic of Singapore
COMMENT Original GenBank locus name is RMU89666.
In the first systematic study of diversity of papillomaviruses
in a nonhuman host (the Rhesus monkey) Chan et al. (1) isolated
twelve different and novel PV genomes which were given the temporary
names of RhPV-a through RhPV-m. The MY09/MY11 consensus primer
region of the L1 gene was sequenced in each case. Each of these
new L1 fragments differs by more than 10% from all other known
papillomaviruses, including RhPV-1, thus qualifying them as new
"types." They fall into three new branches of the "genital," or
A-supergroup of the papillomavirus phylogentic tree and are distinct
from the eleven clades already found there. The frequency with which
new genomes were discovered in this study suggests that the ultimate
diversity of papillomaviruses in Rhesus monkeys may be comparable
to that of HPVs in Homo sapiens.
FEATURES Location/Qualifiers
source 1..446
/organism="rhesus monkey papillomavirus"
/note="provisional nomenclature until complete genome is
isolated"
/strain="rhesus monkey papillomavirus type k (RhPV-k)"
gene <1..>446
/gene="L1"
CDS <1..>446
/gene="L1"
/note="MY09/11 segment"
/codon_start=1
/db_xref="PID:g2108199"
/translation="AQGHNNGICWGNQLFVTVVDTRSTNMTVCAATSKETTYDASKF
KEYLRHVVEYDLQFIFQLCKIALNAEVMSYIHTMNASLLDDWNFGLAPPVQSLEDTYR
FIQSAAIRCQKDSPPPEKQDPYAQYTFWDVDLKEKFSLLDLQYPLGR"
BASE COUNT 122 a 107 c 104 g 113 t
ORIGIN
1 gccacgggac ataataatgg catatgctgg ggtaatcaat tgtttgtcac agtgggtggac
L1 cds ->
-> start MY09/MY11 region
61 actacacgga gcaccaacat gaccgtgtgc gccgccacat ccaaggaaac cacatatgat
121 gcttccaaat ttaaggaata ttgcccggac gtggaggaat atgatttgca atttatattc
181 caactgtgca aaatcgccct taatgctgag gtgatgtcct atattcacac catgaatgcc
241 agcctgctgg atgattggaa ctttggctta gcccccccgc tccagagttt ggaggatacc
301 tacagattta tacagtctgc tgccattagg tgccaaaagg acagcccccc gcctgaaaag
361 caggaccctc atgcacaata cacatcttgg gatgtggatc ttaaggaaaa gttctccctt
421 gatttggatc agtatacctc tgggacg
L1 cds ->
<- end MY09/MY11 region

LOCUS RHPVM 444 bp DNA VRL 25-MAY-1997
 DEFINITION Rhesus monkey papillomavirus strain RhPV-m L1 gene, partial cds.
 ACCESSION U89667
 NID g2108200
 KEYWORDS .
 SOURCE rhesus monkey papillomavirus, L1 MY09/MY11 region DNA
 ORGANISM rhesus monkey papillomavirus
 Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
 Papillomavirus.
 REFERENCE 1 (bases 1 to 444)
 AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
 and Ostrow,R.S.
 TITLE Genomic diversity and evolution of papillomaviruses in rhesus
 monkeys
 JOURNAL J. Virol. (1997) In press
 REFERENCE 2 (bases 1 to 444)
 AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
 and Ostrow,R.S.
 TITLE Direct Submission
 JOURNAL Submitted (14-FEB-1997) Institute of Molecular & Cell Biology,
 National University of Singapore 119260, Republic of Singapore
 COMMENT Original GenBank locus name is RMU89667.
 In the first systematic study of diversity of papillomaviruses
 in a nonhuman host (the Rhesus monkey) Chan et al. (1) isolated
 twelve different and novel PV genomes which were given the temporary
 names of RhPV-a through RhPV-m. The MY09/MY11 consensus primer
 region of the L1 gene was sequenced in each case. Each of these
 new L1 fragments differs by more than 10% from all other known
 papillomaviruses, including RhPV-1, thus qualifying them as new
 "types." They fall into three new branches of the "genital," or
 A-supergroup of the papillomavirus phylogentic tree and are distinct
 from the eleven clades already found there. The frequency with which
 new genomes were discovered in this study suggests that the ultimate
 diversity of papillomaviruses in Rhesus monkeys may be comparable
 to that of HPVs in Homo sapiens.
 FEATURES Location/Qualifiers
 source 1..444
 /organism="rhesus monkey papillomavirus"
 /note="provisional nomenclature until complete genome is
 isolated"
 /strain="rhesus monkey papillomavirus type m (RhPV-m)"
 gene <1..>444
 /gene="L1"
 CDS <1..>444
 /gene="L1"
 /note="MY09/11 segment"
 /codon_start=1
 /db_xref="PID:g2108201"
 /translation="AQGHNNIGICWSNELFVTVVDTRSTNLTVCATSSEAATYQASNF
 KEYTRHVEEYDLQFIFQLCTITLTREVMQYIHTMNPALIEDWKFGVTPPPSSSLEDKY
 RFIQSQAITCQRDGPPEKEDPYAKLNFVVDLKD RFSADLDQYPL"
 BASE COUNT 110 a 115 c 107 g 112 t
 ORIGIN
 1 gccacgggac ataacaatgg catctgctgg agtaatgagc tgtttgttac tgtagtggac
 L1 cds ->
 -> start MY09/MY11 region
 61 accaccgca gtaccaact cactgtgtgt gccacgtcct ccgaggccgc cacgtaccag
 121 gcgtccaact ttaaggaata caccagggcat gtggaggaat atgacctgca gtttattttt
 181 cagctgtgca ccatcacgct tactcgagaa gttatgcagt acatacatac catgaatccc
 241 gctattttgg aggattggaa gtttggggct acccctccac cttccagcag tttggaggac
 301 aaatataggt ttatacagtc ccaggccatt acatgccagc gtgacgggcc tccaaggaa
 361 aaggaggacc cctatgctaa gctgaatttt tgggtttag atcttaagga ccgcttttct
 421 gcagacctgg atcagtatcc cctt
 L1 cds ->
 <- end MY09/MY11 region

MfPV

```

LOCUS           MFPV             449 bp    DNA             VRL             25-MAY-1997
DEFINITION     Macaca fascicularis (long-tailed macaque) papillomavirus strain
                MfPV-a L1 gene, partial cds.
ACCESSION     U89668
NID           g2108176
KEYWORDS       .
SOURCE        Macaca fascicularis papillomavirus.
  ORGANISM     Macaca fascicularis papillomavirus
                Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
                Papillomavirus.
REFERENCE     1 (bases 1 to 449)
  AUTHORS     Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
                and Ostrow,R.S.
  TITLE       Genomic diversity and evolution of papillomaviruses in rhesus monkeys
  JOURNAL     J. Virol. (1997) In press
REFERENCE     2 (bases 1 to 449)
  AUTHORS     Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J.
                and Ostrow,R.S.
  TITLE       Direct Submission
  JOURNAL     Submitted (14-FEB-1997) Institute of Molecular & Cell Biology,
                National University of Singapore 119260, Republic of Singapore
COMMENT       In a study of new papillomaviruses from rhesus monkeys, Chan et al.
                [1] sequenced samples from the congeneric long-tailed macaque
                (Macaca fascicularis). Using the MY09/11 system they isolated this
                sequence whose protein translation is closely related (89%) to
                RHPV1. It is 82% similar over MY09/11 to HPV16. Phylogenetically
                the sequence falls in the "A", or genital supergroup and clusters
                with RHPV1, RHPVa, RHPVb, RHPVd, and RHPVe.
FEATURES       Location/Qualifiers
  source       1..449
                /organism="Macaca fascicularis papillomavirus"
                /note="provisional nomenclature until complete genome is
                isolated"
                /strain="Macaca fascicularis papillomavirus type a
                (MfPV-a)"
  gene         <1..>449
                /gene="L1"
  CDS          <1..>449
                /gene="L1"
                /note="MY09/11 segment"
                /codon_start=1
                /db_xref="PID:g2108177"
                /translation="AQGHNNGICWGNQVFLTVVDTRSTNMTLCASTASEPTYKDNF
                KEYLRHVEEYDLQFIFQLCKITLTTDVMSYIHSMDASILEDWNFGLQPPPSGSLEDTY
                RFVTSAAITCQKNAPPKEKEDPLDKYTFWVDVNLKEKFSADLDQYPLGR"
BASE COUNT     130 a      101 c      95 g      123 t
ORIGIN
  1 gccacgggtc ataataatgg catatgctgg ggtaaccagg tgtttttgac tgttgtagat
L1 cds ->
  -> start MY09/11 region
  61 acaactagaa gcactaacat gacgctctgt gcatctactg cctcagaacc tacctataaa
  121 aatgacaact ttaaggagta ttgcgctcac gtggaagaat atgacctgca atttattttt
  181 cagctgtgca aaataaccct aaccactgat gtaatgtcct acattcacag tatggatgct
  241 agtattttgg aggactggaa ctttggggtg cagcctcccc catctggctc tctggaggac
  301 acctatcgct ttgttacctc cgcggccatc acatgtcaaa agaatgcacc cccaaggaa
  361 aaggaagacc ccttggataa gtatacattt tgggatgtta acctaaaaga aaagttttca
  421 gcagatttag atcagtatcc ccttggacg
                L1 cds ->
                <- end MY09/11 region

```

Other A Groups

HPV-74 HPVGA6053

Two new viral sequences that fall within the A supergroup have been released this year.

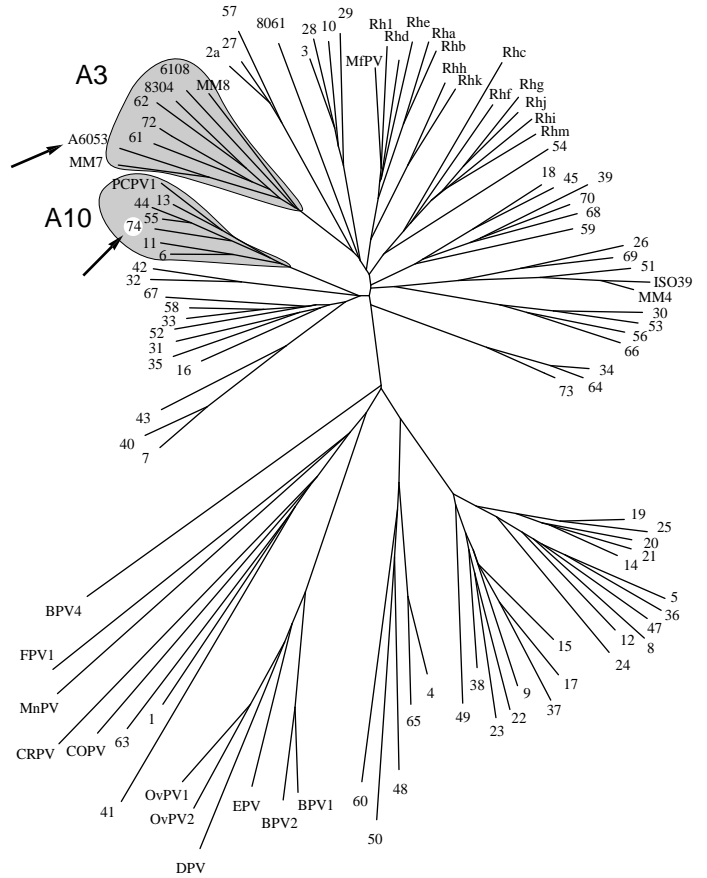
HPV-74 is a group A-10 virus which has been sequenced from roughly 700 bp from the start of the L2 cds to ~ 225 bp past the end of the E7 cds. The central part of the genome, including the genes E1, E2, E4, and E5, remains unsequenced as of this writing [1].

HPVGA6053 is known only from the L1 consensus primer region. The virus appears to be loosely linked to group A3, its closest known relative being HPVMM7 [2].

References

[1] Longuet, M., Cassonnet, P. and Orth, G., A novel genital human papillomavirus (HPV), HPV type 74, found in immunosuppressed patients, *J. Clin. Microbiol.* **34**:7, 1859–1862 (1996).

[2] Astori, G., Arzese, A., Pipan, C., de Villiers, E.M. and Botta, G.A., Characterization of a putative new genomic sequence from a cervical lesion using L1 consensus primers and restriction fragment length polymorphism, *Virus Research* **50**:1, 57–63 (1997).



HPV74

LOCUS HPV74 3891 bp DNA VRL 16-AUG-1996
DEFINITION Human papillomavirus type 74 E6 protein (E6), E7 protein (E7),
partial E1 protein (E1), and L1 protein (L1) genes.
ACCESSION U40822
NID gl491796
KEYWORDS .
SOURCE Human papillomavirus type 74.
ORGANISM Human papillomavirus type 74
Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
Papillomavirus.
REFERENCE 1 (bases 1 to 3891)
AUTHORS Longuet,M., Cassonnet,P. and Orth,G.
TITLE A novel genital human papillomavirus (HPV), HPV type 74, found in
immunosuppressed patients
JOURNAL J. Clin. Microbiol. 34 (7), 1859-1862 (1996)
MEDLINE 96379050
REFERENCE 2 (bases 1 to 3891)
AUTHORS Longuet,M., Cassonnet,P. and Orth,G.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-1995) Michele Longuet, Laboratoire des
Papillomavirus, Institut Pasteur, 25 Rue du Dr Roux, Paris 75015,
France
COMMENT Original GenBank locus name for HPV74 is HPU40822.
The sequence begins roughly 700 bp from the start of the L2 cds and
runs without a break to 225 bp past the end of the E7 cds, ending
in the 5' portion of the E1 orf. The central part of the genome,
including the genes E1, E2, E4, and E5, is unsequenced. Compared
over the L1 protein, HPV-74 shows 86% similarity to HPV-44 and
HPV-55, making it a member of the genital group A10. The A10 viruses
are not significantly involved in the development of invasive
cervical carcinomas, being instead associated with genital warts,
anogenital warts, laryngeal papillomatosis, and low-grade cervical
lesions.
HPV-74 was isolated from persistent vaginal lesions of an
immunosuppressed woman who had undergone renal transplant.
Subsequently, the virus was also found in approximately 5% of other
renal transplant patients studied, but, significantly was never
observed in about 3,000 anogenital specimens from the general
population. This suggests HPV-74 is, under normal circumstances,
efficiently controlled by the immune system.
FEATURES Location/Qualifiers
source 1..3891
/organism="Human papillomavirus type 74"
CDS 2..673
/note="partial"
/note="putative"
/note="ORF L2 from bp 2 to 673"
/product="minor capsid protein"
/gene="L2"
/codon_start=1
/translation="DPAFLSSPQRLITFDNPVYEGEDVSLQFQHDTIHNPPDDAFMDI
IRLHRPAITSRRLVRFVSRIGQRGSMHTRSVMKIGARVHFFQDISPISAAEEIEIHLPL
VAHAQDSSGLFDVYAEPDLEVMEEPVSLSFPTSTPFQRSVAVSATWGNNTTVPLSLPGDM
FIQPGPDIMFPTASSTTPYNPVTALPTGPPVVIHGSTFYIYPSWYFARKRRKRVPVPLFF
TDVAA"
gene 582..2165
/gene="L1"

CDS 582..2165
 /note="ORF L1 from bp 507 to 2165"
 /gene="L1"
 /function="putative major capsid protein"
 /codon_start=1
 /product="L1 protein"
 /db_xref="PID:g1491797"
 /translation="MVLHFIYIHHGILHANAVNVFPCFLQMWRPSDNQVYVPPAPVS
 KVISTDAYVTRTNIFYHASSRLLAVGNPYFPIRQSNKTVVPKVSQYQFRVFKVVLDP
 PNKFALPDTISIFDPSSQRLVWACTGLEVGRGQPLGIGISGHPLLNKLDDEVENSASYAA
 NPGQDNRVNVAMDYKQTQLCMVGCAPPLGEHWGKQCSNVNVKPGDCPALELITSVI
 QDGMVDTGFGAMNFADLQPIKSDVPLDICNTTCKYPDYLQMAADPYGDRLFFYLKKE
 QMFARHFFNRAGTVGEDIPQDLYIKGTTRSTLANAIYFNTPSGSLVSSSETQLFNKPFW
 LQRAQGHNNGICWGNQLFVTVVDTRSTNMTVCAPTSQSPSATYNSSDYKQYMRHVEE
 FDLQFIFQLCSIKLTAEVMAYIHTMNPTVLEEWNFGLTPPPNGTLEDTYRYVQSQAIT
 CQKPTPDKAKPNPYANLSFWEVNLKEKFSSELDQYPLGRKFLQTVQARSSVRVSKK
 RSAPTAPSSATKQKRSRKR"

5'UTR 2166..2947
 /function="regulatory region"
 /standard_name="LCR"

protein_bind 2476..2486
 /function="gene transcription"
 /bound_moiety="E2"
 /note="putative"

protein_bind 2796..2807
 /function="gene transcription"
 /bound_moiety="E2"
 /note="putative"

protein_bind 2837..2853
 /function="gene transcription"
 /bound_moiety="E1"
 /note="putative"

protein_bind 2880..2891
 /function="gene transcription"
 /bound_moiety="E2"
 /note="putative"

protein_bind 2896..2907
 /function="gene transcription"
 /bound_moiety="E2"
 /note="putative"

CDS 2948..3400
 /gene="E6"
 /function="putative transforming protein"
 /codon_start=1
 /product="E6 protein"
 /db_xref="PID:g1491798"
 /translation="MESANASTSAKSIDQLCKDCNIPMHNLIQSCVFCRKTLSVPEVY
 SFAYKNLYIVYRGNFPFAACAICLEVQGVNQFRHFNYAGYALTVEEETKKSILDVLI
 RCYLCHKPLCHVEKVRHIVERARFIKLGNTWKGRCFHCWTTTCMENILP"

gene 2948..3400
 /gene="E6"

CDS 3376..3666
 /gene="E7"
 /function="putative transforming protein"
 /codon_start=1
 /product="E7 protein"
 /db_xref="PID:g1491799"
 /translation="MHGKYSTLKEIVLELQDPVGLHCNEQLDSSEEEVDELATQATQ
 QLTQAYQIVTCCGVCNRSRLRLVQCTGPDINNLLHTLLLGLTLNLVCPKAPKT"

HPV74

```

CDS                3672..3891
                   /partial
                   /gene="E1"
                   /function="putative replication protein"
                   /codon_start=1
                   /product="E1 protein"
                   /translation="MAENSGTEGTGCSGWFLVEAIVEHTTGQQISEDEDEAVEDNGSDM
                                VDFIDDRPITHNSLEAQALLNEQEAISSL"
BASE COUNT        1083 a    786 c    819 g    1203 t
ORIGIN
    1  ggatcctgct tttttatcca gtctcaacg cttataacc tttgataatc cagtatatga
      -> partial L2 orf start
    61 aggtgaggat gttagtttgc aatttcagca cgatacaata cataatcctc ctgatgatgc
   121 ctttatggat attattcgct tacacagacc tgctataacg tcacggcgtg gtcttgtacg
   181 gtttagtagg attgggcaac ggggtctat gcatacacgc agtgtaaagc atataggtgc
   241 acgggtgcat tttttcaag acatatctcc tataatctgct gcagaagaaa tagaactgca
   301 ccctcttggt gctcatgcac aggatagcag tggcttattt gatgtttatg cagaacctga
   361 ccttgagggt atggaagaac ctgtttcatt gtcttttcca acatccacac cctttcagcg
   421 gtcggctgtg tcagccacat ggggcaatac tactgtccct ctttcattgc caggtgacat
   481 gtttatacag cctggtcctg acaTAAgttt tccaactgca tcttctacaa caccctataa
      L1 orf start ->
   541 tcctgttact cctgctttac ctacaggctc tgttgttata cATGgttcta cattttatat
      L1 cds ->
   601 atatccatca tggatTTTTG cacgcaaacg ccgtaaactg gttcccttgt tttttacaga
   661 tgtggcggcc TAGtgacaac caggtttatg tgctcctcct cgcccctgta tccaaagtca
      <- L2 end
   721 tttccacgga tgcctatggt acacgcacca acatctttta tcatgctagc agttctagac
   781 tacttgctgt aggaaatccc tattttccca taaggcaaag caacaaaaca gttgttccta
   841 aggtgtctgg atatcaatth agagtgttth aggtggthth gccagatccc aataaatttg
   901 ctttacctga cacttctata ttgaccctcc ccagtcaacg cctgggatgg gcctgtacgg
   961 gtttgagggt aggtagggga cagcctttag gtattggcat tagtggtcac cctttattaa
  1021 ataaattaga tgatgtggag aattcagcta gttatgcagc taatcctggg caggataaca
  1081 ggggtaagtgt tgctatggac tataagcaga cacagttatg catggtaggc tgtgctcccc
  1141 cgtttaggtga acactgggga aaaggcaagc aatggtctaa tgttaatgta aaaccaggcg
  1201 attgcctgca attagaatta attactagtg taattcagga tgggtgatag gtatgatacag
  1261 gctttgggtg catgaattht gctgacttgc agcctattaa atctgatggt ccattggaca
  1321 tatgcaacac tacatgcaaa tctctgatt atttacaat ggctgaggat ccatatgggtg
  1381 atagattggt tttttactta cgcaaggaac aaatgthtgc aagacactth tttaataggg
  1441 ctggaacagt ggggtgaggac attccccagg acctatata taaaggtagt acacgcagta
  1501 cgcttgccaa tgctattht ttaatactc ccagtggctc tcttgtatct tctgaaacac
  1561 agttgtthaa taagccgtht tggctgcaaa gggcgcaagg ccacaataat ggtatthtgt
  1621 ggggtaatca attatthtgt acagttgtgg ataccacacg cagtactaac atgactgtgt
  1681 gtgctcctac ctcaaatcg ccttctgcta catataatag ttcagactac aaacaatata
  1741 tgcgacatgt ggaggaatth gatttgcact ttattthtca attatgtagt attaatgtaa
  1801 ctgctgagggt tatggcctat attcatacta tgaatcctac agthtttagaa gagtggaaact
  1861 ttgggctaac gcctcccccc aatggtactt tagaagacac ctacagatat gtgcagtccc
  1921 aggctattac atgtcaaaaa cctacgcctg ataaagcaaa gcccaatccc tatgcaaat
  1981 taagthtttg ggaagttaat cttaaggaaa agthttctag tgaattagat caatacccc
  2041 ttggtagaaa gthtttattg cagacaggcg tgcaggcacg ctcttccgth cgtgtaagta
  2101 aaaaacggtc ggcccctaca gcccctcct cagccactaa acaaaaacgc tctagaaaac
  2161 gaTAGcatgc tgttgttgg tgtgtaatat gttgttgcac ggtatatgtg tgttgtatta
      <- L1 end
  2221 tattgtgtgt catatgttgt tgttgtggta tatatgttgt atgttctgtg gtgtcatatt
  2281 gtatgtgtca tgctgtgtaa tgttgtatgt actgcatgta tgtgtthtatt gttgtatgta
  2341 ctgcatgtat gtgttthtgg ttttcaataa actthtthtt tgtgtthtgg gtaccctgta
  2401 cgthttgcatg gtatthtata cthttatata cccatthtacc cthtttataa cctthtcccc
  2461 tccatthttt gtgcaACCGT TTTCGTtatt actcatctct ccacatcctg taaccaatth
      -> E2 bind
  2521 gttgcagcaa gcattthtatt taatcctata tactgtthtgc caaggctata ctgtgttgcc

```

```

2581 aagtacacac caggtgcaca aataggcgcg gtttggcatc cacataatac attcctaatac
2641 tatataccac tcctgtctgc ctgtttaaca attggttgct tttotaatac atttttgtag
2701 ttgggtggcca actcttaaaa gcatttttgg ctgccagtac aacatttttg tacagttact
2761 gttaggttta tataatgagt aacctagggt tacacACCGC AACCGGTatc gttaacacac
                -> E2 bind
2821 accctgtgcc cttcttATTA TACATAATAA CAAtttaaac tttgtaaaaa ataggaggaA
                -> E1 bind                E2 bind ->
2881 CCGAAAACGG TtccaACCGA AAACGGTtat ataTAAacca gcccaaaaat taagcaagcg
                E2 bind ->                E6 orf start ->
2941 gggcataATG gaaagtgcaa atgcctccac gtctgcaaaa agtatagacc agttgtgcaa
                E6 cds ->
3001 ggactgcaac attcctatgc acaatttgca aatttcatgc gtgttttgca ggaaaacctt
3061 gtctgttcca gaggtatatt catttgcta taaaaatttg tatatagtat accgaggaaa
3121 ctttccattt gcagcgtgtg ccatttgttt agaagtacaa ggtaaagtaa atcagtttag
3181 acactttaac tacgcgggat atgcgttaac agtgaagaa gaaacaaaaa agtcaatttt
3241 ggatgtgcta atacgctgct acctgtgcca taagccggtg tgccacgtgg aaaaagtacg
3301 gcacattgtg gaaagggcaa gatttattaa acTAGGcaac acctggaaaag gacgctgctt
                E7 orf start ->
3361 tcattgctgg acaacATGca tggaaaatat tctacctTAA aggaaattgt gttagagcta
                E7 cds ->                <- E6 end
3421 caacctgacc ctgtaggcct acattgcaat gagcaattag acagctcaga agaggaggtg
3481 gatgaactag ccacgcaagc cacgcaacaa ctgacacagg cgtaccaaat agtaacctgc
3541 tgtggtgtat gtaatcgcag ccttcggctg gttgtgcagt gtacaggacc TGAcatcaac
                E1 orf start ->
3601 aatctacata cgcttctact gggtagactg aacctagtgt gtcctctgtg tgcgccc aaa
3661 acgTAAcgac aATGgcggaa aattcaggta cagagggaaac ggggtgctca ggggtggttt
                E7 end <-                -> E1 cds
3721 tagtagaagc aatagtagaa catacaacag ggcagcagat ttcagaggat gaggatgaag
3781 cagtagagga taatgggtct gacatggtag actttataga tgacaggccg attacacaca
3841 attccttggg agcacaggca ttattaaacg agcaggaggc gatctcatta t
                E1 cds ->

```

HPVGA6053

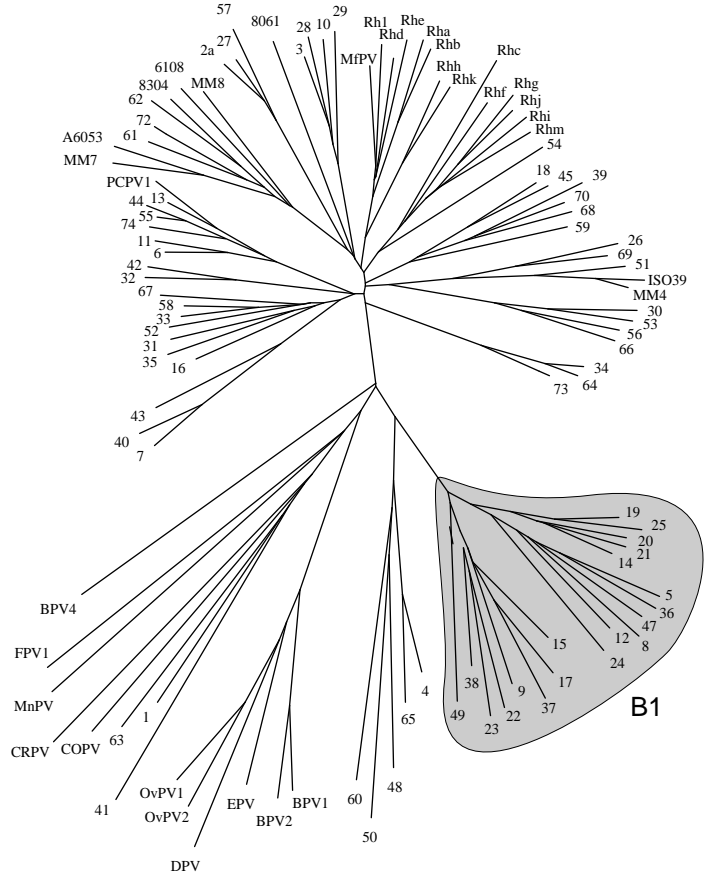
LOCUS HPVGA6053 452 bp DNA VRL 21-APR-1997
DEFINITION Human papillomavirus gene encoding major capsid protein, partial.
ACCESSION Y11911
NID g1945753
KEYWORDS major capsid protein.
SOURCE Human papillomavirus.
ORGANISM Human papillomavirus
Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
Papillomavirus.
REFERENCE 1 (bases 1 to 452)
AUTHORS Astori,G., Arzese,A., Pipan,C., de Villiers,E.M. and Botta,G.A.
TITLE Characterization of a putative new genomic sequence from a cervical
lesion using L1 consensus primers and restriction fragment length
polymorphism
JOURNAL Virus Research 50 (1), 57-63 (1997)
REFERENCE 2 (bases 1 to 452)
AUTHORS Astori,G.
TITLE Direct Submission
JOURNAL Submitted (17-MAR-1997) G. Astori, Udine Medical School, Institute
of Microbiology, P.le S.M.della Misericordia - Udine, 33100, ITALY
COMMENT This novel HPV type was detected in a study of 323 patients with
genital complaints or abnormal PAP smears. Using PCR and RFLP methods
on the MY09/MY11 region of L1, HPV DNA was detected in 20% of
patients with normal cervical cytology, and over 70% in those with
LSIL or HSIL. HPV-16 and 18 were observed most frequently.
HPVGA6053, which was isolated from a patient with an HSIL lesion,
comprises the L1 consensus primer region of a putative new HPV
type. The most similar sequence -- 88% similarity in the aligned
protein -- is HPVMM7 (GB Accession No. U12489). HPVGA6053 should
probably be classified in the genital group A3, a group that
also contains HPVs 61, 62, and 72.
FEATURES Location/Qualifiers
source 1..452
/organism="Human papillomavirus"
/isolate="GA6053"
/specific_host="human"
/note="fragment was isolated from cervical lavage of a
woman with an HSIL lesion"
primer_bind 1..20
/note="primer MY11"
CDS <1..>452
/note="putative"
/codon_start=1
/product="major capsid protein"
/db_xref="PID:e308258"
/db_xref="PID:g1945754"
/translation="AQGHNNGICWFNEVFVTVVDTRSTNITISAATSQSGEYQASNF
KEYLRHTEEYDLQFIFQLCKIRITPEIMAYLHGTNESLLDDWNFGVLPSTSLDDTY
RYLQSRAITCQKGV SAPDPKKDPYDGLAFWEVDLKD KLSMDLDQYPLGR"
primer_bind 433..452
/note="primer MY09"
BASE COUNT 122 a 103 c 93 g 134 t
ORIGIN
1 gcacagggtc ataataatgg catttggttg ttaatgagg tatttgtaac agttgtagat
L1 cds ->
61 actaccgcta gcaccaatat tactatttct gcagccacat cacagtccgg tgaataccag
121 gcctctaact ttaaggaata cctacgccac acagaagaat atgatttaca atttatcttt
181 cagctgtgca aaatacgtat tacacccgaa attatggcct acctccatgg tacgaatgaa
241 tctttgctag atgactggaa ttttggtgta ttgccccctc cttccaccag ccttgatgat

```
301 acatataggt attgcagtc ccgtgctatt acctgccaaa aggggggttc tgctcctgac
361 cccaagaagg atccatatga tggccttgct ttttgggagg tggatctcaa agacaaactg
421 tctatggatc tggatcagta tccccttgga cg
L1 cds ->
```

//

Group B1 Sequences

HPV5	HPV8
HPV9	HPV12
HPV14d	HPV15
HPV17	HPV19
HPV20	HPV21
HPV22	HPV23
HPV24	HPV25
HPV36	HPV37
HPV38	HPV47
HPV49	HPV75 (VS40)
HPV76 (CR148)	HPVICPX1
HPVRTRX1	HPVRTRX2
HPVRTRX3	HPVRTRX4
HPVRTRX5	HPVRTRX6
HPVRTRX7	HPVRTRX8
HPVRTRX9	HPVRTRX10
HPVUWSA	HPVUWSD
HPVVS20	HPVVS42
HPVVS73	HPVVS75
HPVVS92	HPVVS102
HPVVS200	HPVTogawa
CgPV2	



What's new?

The type 2 Colobus monkey PV, CgPV-2, which is the first animal PV member of this group, is given after the introductory material below [33]. The sequences of four new "RTRX" viruses [34] and HPVUWSA and HPVUWSD are also new for 1997.

Group B1 viruses are primarily associated with the multifactorial disease Epidermodysplasia Verruciformis (EV). Recent work suggests they are also commonly associated with immunosuppressed renal transplant recipients [1,2]. Several isolates which appear to constitute new types have been found in skin lesions of renal transplant patients [5,6]. Association of EV-related HPV types with squamous cell carcinomas (SCC) of the skin, and with SCCs of the esophagus has recently been suggested [3,4,5,7,27]. Two potential new types have been isolated from immunocompetent patients [5,5a].

Patients with EV tend to have depressed cell-mediated immunity [8]. In roughly one-third of EV-associated HPV infection, the sun-exposed flat wart-like or macular lesions transform into malignant squamous cell carcinomas [9]. Benign wart scrapings tend to be multiply-infected, with as many as six different viral types. However, in contrast, EV carcinomas tend to harbor only a few types, specifically HPV-5 and HPV-8, and less frequently HPV-14, HPV-17, HPV-20 and HPV-47 [9,10]. These types are rarely detected in lesions afflicting the general population. A key to host restriction of these viruses may be in part due to the unusual organization of the LCR in these viruses. The LCR of these viruses is short compared to the viruses in other groups; most of the types in this group also contain two EV-specific regulatory regions: M33 and M29, both shown to be involved in protein binding [10,11,12]. Study of variants of HPV5 and HPV8 has revealed a higher level of sequence diversity within these types [13,14,15] than has been observed in primarily mucosal types such as HPVs 6, 11, 16 and 18 [16,17,18], suggesting that the EV-related types may accumulate mutations at a higher rate. One possible explanation for this is that the cutaneous tropism of the EV types could lead to additional

mutations through UV-induced DNA damage; however, sequencing of variants of primarily cutaneous HPV2 show levels of variability comparable with that seen in the mucosal types [19].

This group forms two major branches based on phylogenetic analysis, each of which can be subdivided into two minor branches. These clusters have been designated as a_1 , a_2 , b_1 , and b_2 . This phylogenetic classification is compatible with other classifications based on hybridization [21], transforming activity of the E6 gene [20], and conservation of the M33 and M29 regions and E2 binding sites in the LCR [11]. In addition, HPV-24, HPV-49, and several of the new sequences seem to form isolated branches, which may be related with their detection in immunosuppressed, non-EV patients.

Cluster a_1 consists of HPV-5, HPV-8, HPV-12, HPV-36 and HPV-47; the available sequence of HPVICPX1 suggests that it too is a member of this cluster. Both HPV-5 and HPV-8 are associated with macular lesions which frequently progress to malignancy [22,23,24]. Yabe et al. studied the characteristics of HPV-5 in lesions of differing severity. In a primary carcinoma, HPV-5 was present in an episomal state with a 40% subgenomic segment amplified. In the metastatic tumor, only the 40% subgenomic region was present, but integrated into the host genome [24]. The segment was determined to be the entire sequences of E6, E7, and the noncoding region and portions of E1 and L1, with no mutations present [25]. In addition, amplifications of the LCR have been reported in HPV-5 associated carcinomas [26]. HPV-5 and HPV-8 have also been found in significant numbers in squamous cell carcinomas of renal allograft patients. Barr et al. detected either HPV-5 or HPV-8 in nearly 60% of the cases surveyed in Scotland [27]. HPV-47 is primarily associated with benign lesions; however, it has also been detected in cases of malignancy [20]. HPV-12 induces benign macular and flat wart-like lesions [28]. HPV36 was isolated from two patients with actinic keratosis. HPVICPX1 was isolated from an immunocompetent patient [5]; other information is not currently available.

Cluster a_2 consists of HPV-14, HPV-19, HPV-20, HPV-21 and HPV-25. HPV types forming this cluster produce benign macular or flat wart-like lesions and malignant lesions in isolated cases. Both HPV-19 and HPV-25 induce macular lesions, which are benign in character [21,20,29]. HPV-14, HPV-20 and HPV-21 induce flat-wartlike lesions; HPV-20 and HPV-14 have been detected in carcinomas [20,29].

The new Colobus monkey virus CgPV-2 is the first animal virus to be classified with the Supergroup B viruses [33]. Isolated from pedunculate cutaneous papillomas on the hands and feet of an Abyssinian monkey, the virus is most closely related to the a_2 cluster, but may well be the first representative of a new cluster of monkey cutaneous viruses.

Cluster b_1 includes of HPV-9, HPV-15, HPV-17, and HPV-37; available sequence indicates that HPVVRTRX3, HPVVS92 and HPVVS102 are also members of this cluster [5,6]. HPV-15 was isolated from a benign flat wart-like lesion [29]. HPV-17 was isolated from benign macules and subsequently from squamous cell carcinomas and the malignant melanoma of an immunosuppressed patient [29,30]. HPV 9 DNA induces both macular and flat wart-like lesions, however it has also been identified in a keratoacanthoma [28,31]; HPV-37 was found in the same keratoacanthoma. HPVVRTRX3, HPVVS92, and HPVVS102 were isolated from a squamous cell carcinoma, skin wart, and dysplastic wart respectively in renal transplant patients [5,6]. Also possibly belonging to this cluster is HPVVRTRX6, although its position in phylogenetic trees is rather unstable. HPVVRTRX6 was isolated from an SCC in one renal transplant patient [5].

Cluster b_2 includes HPV-22, HPV-23, and HPV-38; available sequence indicates that HPVVRTRX1, the Togawa isolate, HPVVS42 and HPVVS73 are also members of this cluster. HPV-22 and HPV-23 were isolated from macules of EV patients [29]. HPV-38 was isolated, along with HPV17a, from a superficial spreading melanoma in an immunocompromised patient [31]. The Togawa isolate was found in multiple SCCs of the esophagus in nonimmunocompromised patients [7]. HPVVRTRX1, HPVVS42 and HPVVS73 were isolated from an SCC, a verrucous biopsy and a skin wart biopsy in renal transplant patients [5,6].

Isolated types Several EV-related types or potential new types seem to be relatively unrelated to the clusters defined above, and, for the most part, to each other. HPV-49 was isolated from the flat warts of a Polish renal transplant patient. Favre et al. screened benign and malignant lesions from the general population, EV patients and transplant patients for the presence of HPV-49. In the survey,

Introduction

HPV-49 was not detected in any of the patients with EV but was detected in two additional cases of flat warts in renal transplant patients [32]. Related to HPV49 are HPV-75 (VS40) and HPV-76 (CR148), from a dysplastic wart biopsy and a skin wart biopsy, respectively, from renal transplant patients [6].

HPV-24 was isolated from macules in an EV patient [29]. HPVVS75 and HPVVS20 appear to be relatively closely related to HPV24, and were isolated from skin wart biopsies of renal transplant patients [6]. HPVVS200, isolated from a basal cell carcinoma of a nonimmunosuppressed patient [5a], is most similar to HPVVS75 and HPV-24.

HPVRTRX2, HPVRTRX4, and HPVRTRX5 appear to form their own cluster, and were all isolated from cutaneous SCCs of renal transplant patients [5]. HPVRTRX2 and HPVRTRX5 were each isolated from one SCC in each of two patients, and HPVRTRX4 was isolated once, out of 53 SCCs from 26 renal transplant patients.

HPV-5, HPV36 and HPV-47 are close enough to each other to be considered “close types”—sequences that qualify to be distinct types under the criterion of ten percent dissimilarity at the nucleotide level, but between which most of these changes are “silent”, causing no difference at the amino acid level (Part III). Also qualifying as close types are HPV-19 and HPV-25, and HPV-14d, HPV-20f and HPV-21.

References

- [1] Gassenmaier A; Fuchs P; Schell H; Pfister H. Papillomavirus DNA in warts of immunosuppressed renal allograft recipients. *Archives for Dermatological Research. Archiv fur Dermatologische Forschung*, 1986, **278**(3):219–23.
- [2] Blessing K; McLaren KM; Morris R; Barr BB; Benton EC; Alloub M; Bunney MH; Smith IW; Smart GE; Bird CC. Detection of human papillomavirus in skin and genital lesions of renal allograft recipients by in situ hybridization. *Histopathology*, 1990 Feb, **16**(2):181–5.
- [3] Bouwes Bavinck JN. Epidemiological aspects of immunosuppression: role of exposure to sunlight and human papillomavirus on the development of skin cancer. *Human and Experimental Toxicology*, 1995 Jan, **14**(1):98
- [4] Hartevelt MM; Bavinck JN; Kootte AM; Vermeer BJ; Vandenbroucke JP. Incidence of skin cancer after renal transplantation in The Netherlands. *Transplantation*, 1990 Mar, **49**(3):506–9.
- [5] Berkhout RJ; Tieben LM; Smits HL; Bavinck JN; Vermeer BJ; ter Schegget J. Nested PCR approach for detection and typing of epidermodysplasia verruciformis-associated human papillomavirus types in cutaneous cancers from renal transplant recipients. *Journal of Clinical Microbiology*, 1995 Mar, **33**(3):690–5.
- [5a] Shamanin V; zur Hausen H; Lavergne D; Proby CM; Leigh IM; Neumann C; Hamm H; Goos M; Hausteil UF; Jung E; Plewig G; Wolff H; de Villiers EM. Human papillomavirus infections in nonmelanoma skin cancers from renal transplant recipients and nonimmunosuppressed patients. *Journal of the National Cancer Institute*, 1996, **88**(12):802–11.
- [6] Shamanin V; Glover M; Rausch C; Proby C; Leigh IM; zur Hausen H; de Villiers EM. Specific types of human papillomavirus found in benign proliferations and carcinomas of the skin in immunosuppressed patients. *Cancer Research*, 1994 Sep 1, **54**(17):4610–3.
- [7] Togawa K; Rustgi AK. A novel human papillomavirus sequence based on L1 general primers. *Virus Research*, 1995 MAY, **36**(2-3):293–297.
- [8] Lutzner, M.A. Epidermodysplasia verruciformis. An autosomal recessive disease characterized by viral warts and skin cancer. A model for viral oncogenesis. *Bull Cancer (Paris)* **65**: 169–82 (1978)
- [9] Gassenmaier, A., Lammel, M., and Pfister, H. Molecular cloning and characterization of the DNAs of human papillomaviruses 19, 20, and 25 from a patient with epidermodysplasia verruciformis. *J Virol* **52**: 1019–23 (1984)
- [10] May, M., Helbl, V., Pfister, H., Fuchs, P.G. Unique topography of DNA-protein interactions in the non-coding region of epidermodysplasia verruciformis-associated human papillomaviruses. *J Gen Virol* **72** (Pt 12): 2987–97 (1991)

- [11] Ensser,A. and Pfister,H. Epidermodysplasia verruciformis associated human papillomaviruses present a subgenus-specific organization of the regulatory genome region. *Nucleic Acids Res* **18**:3919–22 (1990)
- [12] Krubke,J., Kraus,J., Delius,H., Chow,L., Broker,T., Iftner,T., and Pfister, H. Genetic relationship among human papillomaviruses associated with benign and malignant tumours of patients with epidermodysplasia verruciformis. *J Gen Virol* **68**(12): 3091–103 (1987)
- [13] Deau MC; Favre M; Jablonska S; Rueda LA; Orth G. Genetic heterogeneity of oncogenic human papillomavirus type 5 (HPV5) and phylogeny of HPV5 variants associated with epidermodysplasia verruciformis. *Journal of Clinical Microbiology*, 1993 Nov, **31**(11):2918–26.
- [14] Deau MC; Favre M; Orth G. Genetic heterogeneity among human papillomaviruses (HPV) associated with epidermodysplasia verruciformis: evidence for multiple allelic forms of HPV5 and HPV8 E6 genes. *Virology*, 1991 Oct, **184**(2):492–503.
- [15] Kawase,M., Favre,M., Pehau-Arnaudet,G., Durand,E., Majewski,S., Jablonska,S., and Orth,G. Unpublished data presented at the 14th International Papillomavirus Conference, Quebec City, July 23–28, 1995. Abstract title: Phylogeny and antigenic properties of Human Papillomavirus type 5 (HPV5).
- [16] Heinzl,PA., Chan,SY., Ho,L., O’Connor,M., and others. Variation of Human Papillomavirus type 6 (HPV-6) and HPV-11 genomes sampled throughout the world. *Journal of Clinical Microbiology*, 1995 Jul, **33**(7):1746–1754.
- [17] Ho L; Chan SY; Burk RD; Das BC; Fujinaga K; Icenogle JP; Kahn T; Kiviat N; Lancaster W; Mavromara-Nazos P; et al. The genetic drift of human papillomavirus type 16 is a means of reconstructing prehistoric viral spread and the movement of ancient human populations. *Journal of Virology*, 1993 Nov, **67**(11):6413–23.
- [18] Ong CK; Chan SY; Campo MS; Fujinaga K; Mavromara-Nazos P; Labropoulou V; Pfister H; Tay SK; ter Meulen J; Villa LL; et al. Evolution of human papillomavirus type 18: an ancient phylogenetic root in Africa and intratype diversity reflect coevolution with human ethnic groups. *Journal of Virology*, 1993 Nov, **67**(11):6424–31.
- [19] Chan,SY., Rübber,A., Grussendorf-Conen,E., Chew,SH., and Bernard,HU. Unpublished data presented at the 14th International Papillomavirus Conference, Quebec City, July 23-28, 1995. Abstract title: Molecular variants of Human Papillomavirus types 2, 24 and 57 found in cutaneous and genital lesions.
- [20] Kiyono,T., Hiraiwa,A., and Ishibashi,M. Differences in transforming activity and coded amino acid sequence among E6 genes of several papillomaviruses associated with epidermodysplasia verruciformis. *Virology* **186**: 628–39 (1992)
- [21] Pfister, H. Papilomaviruses: general description, taxonomy, and classification. In *The Papovaviridae 2* (N. P. Salzman and P.M. Howley, Eds.), pp. 1–38. Plenum, N.Y.
- [22] Zachow,K.R., Ostrow,R.S. and Faras,A.J. Nucleotide sequence and genome organization of human papillomavirus type 5. *Virology* **158**: 251–4 (1987)
- [23] Orth,G. Epidermodysplasia verruciformis: a model for understanding the oncogenicity of human papillomaviruses. *Ciba Found Symp* **120**: 157–74 (1986)
- [24] Yabe, Y., Tanimura, Y., Sakai, A., Hitsumoto, T., and Nohara, N. Molecular characteristics and physical state of human papillomavirus DNA change with progressing malignancy: studies in a patient with epidermodysplasia verruciformis. *Int J Cancer* **43**: 1022–8 (1989)
- [25] Yabe, Y., Sakai, A., Hitsumoto, T., Kato, H., and Ogura, H. A subtype of human papillomavirus 5 (HPV-5b) and its subgenomic segment amplified in a carcinoma: nucleotide sequences and genomic organizations. *Virology* **183**: 793–8 (1991)
- [26] Deau, M.C., Favre, M., and Orth, G. Genetic heterogeneity among human papillomaviruses (HPV) associated with epidermodysplasia verruciformis: evidence for multiple allelic forms of HPV5 and HPV8 E6 genes. *Virology* **184**: 492–503 (1991)

Introduction

- [27] Barr,B.B., Benton,E.C., McLaren,K., Bunney,M.H., Smith,I.W., Blessing,K., Hunter,J.A. Human papilloma virus infection and skin cancer in renal allograft recipients. *Lancet* **1**: 124–9 (1989)
- [28] Kremsdorf,D., Jablonska,S., Favre,M., Orth,G. Human papillomaviruses associated with epidermodysplasia verruciformis. II. Molecular cloning and biochemical characterization of human papillomavirus 3a, 8, 10, and 12 genomes. *J Virol* **48**: 340–51 (1983)
- [29] Kremsdorf,D., Favre,M., Jablonska,S., Obalek,S., Rueda,L.A., Lutzner,M.A., Blanchet-Bardon,C., Van Voorst Vader,P.C. and Orth,G. Molecular cloning and characterization of the genomes of nine newly recognized human papillomavirus types associated with epidermodysplasia verruciformis. *J Virol* **52**: 1013–8 (1984)
- [30] Yutsudo,M and Hakura,A. Human papillomavirus type 17 transcripts expressed in skin carcinoma tissue of a patient with epidermodysplasia verruciformis. *Int J Cancer* **39**: 586–9 (1987)
- [31] Scheurlen,W., Gissmann,L., Gross,G., and zur Hausen,H. Molecular cloning of two new HPV types (HPV 37 and HPV 38) from a keratoacanthoma and a malignant melanoma. *Int J Cancer* **37**: 505–10 (1986)
- [32] Favre,M., Obalek,S., Jablonska,S., and Orth,G. Human papillomavirus type 49, a type isolated from flat warts of renal transplant patients. *J Virol* **63**: 4909 (1989)
- [33] Chan,S.-Y., Ostrow,R.S., Faras,A.J. and Bernard,H.U. Genital Papillomaviruses (PVs) and Epidermodysplasia Verruciformis PVs occur in the same Monkey species: Implications for PV Evolution *Virology* **228**:213–217 (1997)
- [34] Hopfl,R., Bens,G., Wieland,U., Petter,A., Zelger,B., Fritsch,P. and Pfister,H. Human papillomavirus DNA in non-melanoma skin cancers of a renal transplant recipient: detection of a new sequence related to epidermodysplasia verruciformis associated types *J. Invest. Dermatol.* **108** (1), 53–56 (1997)

LOCUS CgPV2E6L1 2932 bp DNA VRL 07-JAN-1997
DEFINITION Colobus monkey papillomavirus type 2 major capsid protein (L1) and oncoprotein (E6) genes, complete cds.
ACCESSION U72630
NID gl764167
KEYWORDS .
SOURCE Colobus monkey papillomavirus type 2.
ORGANISM Colobus monkey papillomavirus type 2
Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
REFERENCE 1 (bases 1 to 2932)
AUTHORS Kloster,B.E., Manias,D.A., Ostrow,R.S., Shaver,M.K., McPherson,S.W., Rangen,S.R., Uno,H. and Faras,A.J.
TITLE Molecular cloning and characterization of the DNA of two papillomaviruses from monkeys
JOURNAL Virology 166 (1), 30-40 (1988)
MEDLINE 88322887
REFERENCE 2 (bases 1 to 2932)
AUTHORS Chan,S.-Y., Ostrow,R.S., Faras,A.J. and Bernard,H.U.
TITLE Genital Papillomaviruses (PVs) and Epidermodysplasia Verruciformis PVs occur in the same Monkey species: Implications for PV Evolution
JOURNAL Virology 228 (1997) In press
REFERENCE 3 (bases 1 to 2932)
AUTHORS Chan,S.-Y., Ostrow,R.S., Faras,A.J. and Bernard,H.U.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-1996) Institute of Molecular & Cell Biology, National University of Singapore, Lower Kent Ridge Road, Singapore 119260, Republic of Singapore
COMMENT CgPV2, whose original GenBank Locus Name is CMU72630 was cloned from an Abyssinian Colobus monkey. Tissue was derived from pedunculate cutaneous papillomas on the hands and feet. A 2.9 kb segment that includes the L1, LCR, E6, and E7 part of the genome was inserted into pUC19.
CgPV2, a clear member of the B supergroup, or EV (cutaneous) papillomavirus family, is the first nonhuman PV to be found in that supergroup. The implication of this finding is that the evolutionary divergence of the genital and EV supergroups predated the divergence of the monkeys and humans roughly 22 million years ago [2].
The end of this sequence contains the E7 gene, but was not identified as such in the original GenBank entry. Perhaps this was because the E7 cds contains what is undoubtedly a sequencing error; a missing nucleotide at position 2858. We have indicated this missing nucleotide with an 'x' in the sequence below.
FEATURES Location/Qualifiers
source 1..2932
/organism="Colobus monkey papillomavirus type 2"
/note="CPV2; sequence derived from 2.9 kb fragment of subcloned as a SacI-BamHI fragment into pUC19"
CDS 118..1668
/gene="L1"
/codon_start=1
/product="major capsid protein"
/db_xref="PID:gl764168"
/translation="MSVWQAATGKVYLPPSTPVARVQSTDEYIQRTOIYYHAYSDRLLT
TVGHPYFNVYNNNGTVLEVPKVSNGQHRVFRRLRPLDPNRFALADMSVYNPDKERLVWA
CRGIEIGRGQPLGVGSGTGTPIFNKVGDSTENSNTYRAQANSTDDRQNI SFDPKQIQMFI
IGCAPCIGEHWDRALPCVEQNPPAYACPPPIELRNTVIEDGDMADIGYGNLNFKALQQN

CgPV2

RSDVSLDIVNETCKYPDFLRMQNDVYGDSCFFYARREQCYARHFFVRRGGKTGDDIPGA
 QIDEGSMTNAFYIPAANAQAQNNIGNAMYFPTVSGSLVSSDAQLFNRPFWLQRAQGHN
 NGICWFNQLFVTVDNTRNTNFSIAVSTDEKDISKIANYDSQKFAEYLRHVEEYEVSL
 ILQLCKVPLKAEVLAQINAMNSGILEDWQLGFVPAPDNPIHDTYRYIDSLATRCPDKE
 PPKEKEDPYKQYTFWNVDLTERLSLDDLQYSLGRKFLFQAGLQRSTVNGTKNAPVTSR
 GVKRRKRS"

gene 118..1668
 /gene="L1"
 misc_feature 1669..2145
 /note="long control region (LCR)"
 gene 2185..2634
 /gene="E6"
 CDS 2185..2634
 /gene="E6"
 /codon_start=1
 /product="oncoprotein"
 /db_xref="PID:g1764169"
 /translation="MASDNAEESQIPELTKTAVGLAHLDDIPLDDLLLCTFCGRFLS
 FSEVCEFDKSLIWHNYTVSACCRCCCVATATYEFNEFYEQTVNGREIEQVTGLSI
 LHLDVRCQNCLRFLDNIEKLDICGRNRPFHKVRNWWKVICRHCKYLL"
 CDS 2631..2933
 /gene="E7"
 /codon_start=1
 /product="transforming protein"
 /translation="MIGKDADLQDIILELSEIQPEGQPVDLLCEEELPAEQELEEEPT
 TARTTFKVVAPCGCCEANLRLFVRATTFGIRTLQDILTEELQLLCPGCRGNCKHGGS"

BASE COUNT 829 a 617 c 690 g 796 t

ORIGIN

```

1 gagctccctg ttgtaattat ttaccctcat gacaatactg gggatTTTTa cttacatcct
61 agtctgcaca aacgaaaacg gcgtaacggg aagtatttct aatgttttct tttacagATG
                                     -> L1 cds
121 tctgtctggc aagccgctac tggttaagggt tacctgccac catccacacc ggtggccaga
181 gttcaaagca cggatgagta catacaaaaga acacaaatat actaccatgc atatagcga
241 cgtctgctta ctgtcgggca cccctacttt aatgtctata acaataatgg cacagtgtct
301 gaggtcccta aagtatcggg aaaccaacac agagtgttcc gtttacgttt gccggacccc
361 aatcgttttg ctttggtgga tatgtccgta tacaaccgga acaaggagcg tttagtgtgg
421 gcctgtagag gcatagaaat tggacgaggg cagcctttag gcgttggcag tactggcacc
481 cctatTTTTa ataaggtggg ggatactgaa aattccaaca catatagggc ccaggcaaat
541 tccactgatg acagacaaaa catacttttt gatcctaagc aaatacaaat gtttatcatt
601 ggctgtgccc cctgtatagg tgagcattgg gatagggcct tgccatgcgt ggaacaaaa
661 ccacctgcat atgcatgccc tcctatagag cttagaaaca cagtaatcga ggatggcgat
721 atggctgata taggatatgg caacttaaat ttttaaggcct tgcaacaaaa cagatctgat
781 gttagttagg acattgtaaa tgaaacatgc aagtaccggg atTTTTtaag aatgcaaaa
841 gacgtgatg gggattcctg ttttttttgc gcacgtaggg agcaatgcta tgcaagacat
901 ttttctgccc gtggaggcaa aacaggcgat gatattcctg gggcacaaat tgatgaaggc
961 agtatgacca atgctTTTTa catacctgct gctaattgccc aagctcagaa taatataggc
1021 aacgcaatgt atttcccaac tgtagtgagg tcattgggtt ctagtgatgc acagctgttc
1081 aacaggccat tttggttga gcgtgcacaa ggccacaaca atggcatatg ttggtttaat
1141 cagctatttg tcacagtggg ggacaacacc cgaataacca acttttagcat tgctgttagc
1201 acagatgaaa aggatatctc aaaaatagca aactacgatt cccagaagtt tgcggagtat
1261 cttagacacg tagaggagta tgaagtttcc ttaatattac agctctgtaa ggttccccta
1321 aaagctgagg ttctggcgca aatcaatgcc atgaactcag gtattctgga ggactggcaa
1381 ttgggctttg taccgcccc tgataacccc atccatgata cttatcgata tatagactcc
1441 cttgcaactc gttgccccga taaagagccc cccaaagaaa aagaggacc ttataaaca
1501 tacacatTTT ggaatgtgga tctcacagag cgtctatcct tagattttag tcaatactcg
1561 ctgggtcgta agtttctctt tcaggcaggc ctacaacggg caaccgttaa cgttacaAAA
1621 aacgcgccag ttacatcgcg gggggtcaag cgcaaacgta agagcTAAct gtaccgTTTT
                                     L1 end <-
                                     -> LCR start

```

```

1681 cggtagctgg cacatcaagg ggtgggtgta taaacctaaa cagtctcttg taccgttttc
1741 ggtaccacat gtgaactttt gcacagtatt taaaaaatgc ctgagtaagc ttgctgtgta
1801 gccgggaggc cgcaccgggt acataaattg ctcgctgtg tcattttatg ccaaagtcgg
1861 cgcgccagct ggcatagacc cttggctctc tacatgcccg gataagacgt ctgccaaacg
1921 gattatacat ttgccaact gcttcgggcg tgttctttca ggcacatag ctgtaccagt
1981 aacggtaagt acctcaaatt gtaccgggtg cggctcaaat taatcacgtc aatgggtaat
2041 tggtgtgccc aacaagcata gtttacaagt acttgcctgt aacggattgg ataatacaac
2101 ggtcataaat caatatatat atatatatat atatatatat aaTAAaata tatatgcaac
                                end LCR <-
                                -> E6 orf
2161 ttggtgtgga gatggtgtgc ttcaATGgcg tctgacaacg cggaggaatc ccaaattcct
                                -> E6 cds
2221 gagcttccaa aaactgctgt agggtagca catttattag acattccttt agacgacctg
2281 ctgctgcctt gcacattttg cggcaggttt ttatcgtttt ctgaggtctg tgagtttgat
2341 gataagaaac tgtctttaat ttggcacaat tatactgtgt ccgctgtgtg tcgctgctgc
2401 tgtgtggcaa cagcgactta tgaattcaat gagttttatg agcaaacagt aaacggacga
2461 gagatagagc aggtgacagg actcagcatc ctgcatctgg atggttagatg tcaaaaactgt
2521 ttgaggttcc tagataacat agaaaagcta gatatatgtg gtaggaaccg tccgttccac
2581 aaagtaagaa actggtggaa gggcatctgc aggcactgTA Agtatttatt ATGAttggta
                                E6 end <-
                                -> E7 orf -> E7 cds
2641 aggacgcgga cttgcaagac attattttag agttgagtga gattcagcct gaagggcagc
2701 cagttgacct gctttgtgaa gaggagttac cagctgagca ggagttggag gaggagccaa
2761 ctactgcaag gaccacattt aaagttgtcg ctccgtgtgg atgctgtgaa gcaaacctac
2821 gtctgtttgt tcgggcaaca acatttgtaa ttagaacxtt gcaggacata ctgacggagg
                                ^ missing nucleotide
2881 agctgcagct gttgtgcccg gagtgtcgcg gaaactgcaa gcatggcgga tcc
                                -> E7 cds

```

//

HPVTRX7

LOCUS HPVTRX7 1043 bp DNA VRL 10-FEB-1997
DEFINITION Human papillomavirus strain RTRX7 major capsid protein L1 gene, partial cds.
ACCESSION U85660
NID g1835605
KEYWORDS .
SOURCE Human papillomavirus.
ORGANISM Human papillomavirus
Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
REFERENCE 1 (bases 1 to 1043)
AUTHORS Hopfl,R., Bens,G., Wieland,U., Petter,A., Zelger,B., Fritsch,P. and Pfister,H.
TITLE Human papillomavirus DNA in non-melanoma skin cancers of a renal transplant recipient: detection of a new sequence related to epidermodysplasia verruciformis associated types
JOURNAL J. Invest. Dermatol. 108 (1), 53-56 (1997)
MEDLINE 97134727
REFERENCE 2 (bases 1 to 1043)
AUTHORS Bens,G., Pfister,H. and Wieland,U.
TITLE Direct Submission
JOURNAL Submitted (17-JAN-1997) Institute for Virology, University Koeln, Fuerst-Pueckler-Strasse 56, Koeln 50935, Germany
COMMENT GenBank locus name HPU85660. In this study 17 biopsies taken from malignant and nonmalignant skin tumors of a renal transplant patient were subjected to nested PCR using broad spectrum primers capable of amplifying all known EV-type L1 gene fragments. Nine HPV L1 sequences were obtained, three of which were identical to previously identified types -- HPV38, RTRX1, and ICPX1. Identical L1 sequences of the present type, HPVTRX7, were found in the other six samples. RTRX7 is a member of the B1 group of EV-associated viruses, a group that contains the other nine known RTRX viruses. It is most closely related (83%) to HPV-12. Hopfl et al. [1] note that the "extraordinary widespread non-melanoma skin cancers in the patient examined in this study may increase speculation that RTRX7 has a high oncogenic potential."
FEATURES Location/Qualifiers
source 1..1043
/organism="Human papillomavirus"
/strain="RTRX7"
CDS <1..>1043
/codon_start=3
/product="major capsid protein L1"
/db_xref="PID:g1835606"
/translation="GVGSTGHPYFNKVKDTENTNNTTGGSTDDRQNTSFDPKQIQMFI
VGCTPCVGEHWEKALPCGDAPADNGICPPIELKNTFIEDGDMADIGFGNMNFKTLQON
RADVSLDIVNETCKYPDFLKMQNDVYGDACFFYARREQCYARHFFVVRGRKTGDNIPDA
YIDDGNMKNSFYIPAASNQSQKDIGNAMYFPTVSGSLVSSDAQLFNRPFWLQRAQGHN
NGILWANQMFVTVVDNTRNTNFSISISSENQDIQQIQSYDSQKFREYLRHVEYEYIS
ILQLCKIPLQAEVLAQINAMNPSLLEDWQLGFVPTPDNPIQDTYRFIDSLATRCPKDN
PPKEKPDPEKLT"
BASE COUNT 349 a 188 c 207 g 299 t
ORIGIN
1 aaggagttgg cagtactgga catccatatt ttaataaggt aaaagacacg gaaaacacta
L1 cds ->
61 acaattatac cacaggctct acagatgata ggcagaacac atcttttgac cccaacaa
121 ttcagatggt tatagtaggt tgtacacctt gtgtgggaga gcattgggaa aaggccttac
181 cttgtggaga tgcgcctgca gataatggta tttgcctcc aattgaatta aaaaacactt
241 ttattgaaga tggagacatg gcagatattg gttttggcaa tatgaatttc aaaactttac
301 agcagaatag ggctgatgtc agtctggata tagttaatga aacttgcaaa tatccagact
361 ttttggaaat gcaaaatgat gtctatggcg atgcgtggtt tttctatgct cgcagggagc

```
421 agtgttatgc cagacatfff tttgttaggg ggcgcaagac aggtgacaat attcctgacg
481 catatattga tgatggcaat atgaaaaatt ctttttacat accagcagct tcaaatcaat
541 ctcaaaagga tattggtaat gctatgtatt tcccaacagt cagtggctca ctagtctcta
601 gtgatgctca attgtttaac aggcccttct ggctgcaaag agcacagggt cataataatg
661 gcattctgtg ggctaatacag atgtttgtca cagttgtaga caacacgcga aacaccaatt
721 tcagtatafc tatatctagt gaaaatcaag atatacagca aatacaatca tatgactcac
781 aaaagtttag ggaatattta aggcacgtag aagaatatga aatttctatt attttacagt
841 tgtgtaagat tccactacaa gcagaagttt tagcacaaat aaatgcaatg aaccctcct
901 tactagagga ttggcagtta ggatttgtgc caactcccga taatcctatc caggacacat
961 acagatttat tgattcctta gctaccaggt gtcccgataa aaatccacca aaggaaaaac
1021 ctgatcctta tgaaaaatta aca
```

//

L1 cds ->

HPVRTRX8

LOCUS HPVTRX8 341 bp DNA VRL 10-FEB-1997
DEFINITION Human papillomavirus strain RTRX8 major capsid protein L1 gene,
partial cds.
ACCESSION U85661
NID g1835607
KEYWORDS .
SOURCE Human papillomavirus.
ORGANISM Human papillomavirus
Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
Papillomavirus.
REFERENCE 1 (bases 1 to 341)
AUTHORS Bens,G., Pfister,H. and Wieland,U.
TITLE Human papillomavirus DNA in non-melanoma skin cancers of a renal
transplant recipient
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 341)
AUTHORS Bens,G., Pfister,H. and Wieland,U.
TITLE Direct Submission
JOURNAL Submitted (17-JAN-1997) Institute for Virology, University Koeln,
Fuerst-Pueckler-Strasse 56, Koeln 50935, Germany
COMMENT GenBank locus name HPU85661.
FEATURES Location/Qualifiers
source 1..341
/organism="Human papillomavirus"
/strain="RTRX8"
CDS <1..>341
/codon_start=3
/product="major capsid protein L1"
/db_xref="PID:g1835608"
/translation="ADNTRNTNFTNSVSSENAVAQEYNANNIREFLRHVGEYQLSLIL
QLCKVPLISEVLSQINAMNSSILENWQLGFVPTPDNSVHDMYRYIHSKATKCPDAVEP
TEKEDPFNKYT"
BASE COUNT 133 a 62 c 56 g 90 t
ORIGIN
1 tagcagataa tactagaaac acaaacttca ctaatagtgt cagttcagag aatgcagtag
L1 cds ->
61 cacaggaata taatgctaat aacataagag agttottaag acatgtagga gaatatcagt
121 tgtccttaat attacaacta tgcaaagtcc ccttaatttc agaggтата tcacaaatca
181 atgcaatgaa ctctagtatt ttggaaaact ggcagttagg atttgttcca acaccagaca
241 attcagttca tgacatgtac agatatattc attccaaggc cactaaatgt ccagatgctg
301 tagaaccaac agaaaaggaa gatccattta acaaatcac a
// L1 cds ->

```

LOCUS      HPVTRX9      349 bp      DNA              VRL              10-FEB-1997
DEFINITION Human papillomavirus strain RTRX9 major capsid protein L1 gene,
partial cds.
ACCESSION  U85662
NID        g1835609
KEYWORDS   .
SOURCE     Human papillomavirus.
  ORGANISM Human papillomavirus
            Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
            Papillomavirus.
REFERENCE  1 (bases 1 to 349)
  AUTHORS  Bens,G., Pfister,H. and Wieland,U.
  TITLE    Human papillomavirus DNA in non-melanoma skin cancers of a renal
            transplant recipient
  JOURNAL  Unpublished (1997)
REFERENCE  2 (bases 1 to 349)
  AUTHORS  Bens,G., Pfister,H. and Wieland,U.
  TITLE    Direct Submission
  JOURNAL  Submitted (17-JAN-1997) Institute for Virology, University Koeln,
            Fuerst-Pueckler-Strasse 56, Koeln 50935, Germany
COMMENT    GenBank locus name HPU85662.
FEATURES   Location/Qualifiers
  source    1..349
            /organism="Human papillomavirus"
            /strain="RTRX9"
  CDS       <1..>349
            /codon_start=2
            /product="major capsid protein L1"
            /db_xref="PID:g1835610"
            /translation="LDNTRNTNFSIAVHQEQKQVKEIQNYDSAKFNEFQRHVVEEYEVS
            LILQLCKIPLKAEVLAQINAMNSDILENWQLGFVPTPDNPIHDTYRYLDSLATRCPEK
            VPAKEKVDPYAKYV"
BASE COUNT 130 a      56 c      61 g      102 t
ORIGIN
  1 attagataat actagaaata caaacttttag tattgctgta catcaagagc agaagcaggt
L1 cds ->
  61 taaagaaata caaaattatg attctgcaaa gtttaatgaa tttcaaagac atgttgagga
 121 atatgaagtg tctctcattt tacaattgtg taaaattcca ttgaaagctg aggttcttgc
 181 acagattaat gcaatgaact ctgatatttt agaaaactgg cagttagggt ttgtaccaac
 241 accggacaat cctattcatg atacttacag atacttagac tcattagcaa cacgctgtcc
 301 agaaaaagtt cccgcaaagg aaaaggtaga cccttatgct aaatatgta
//
                                                    L1 cds ->

```


HPVRTRX10

LOCUS HPVTRX10 337 bp DNA VRL 10-FEB-1997
DEFINITION Human papillomavirus strain RTRX10 major capsid protein L1 gene,
partial cds.
ACCESSION U85663
NID g1835611
KEYWORDS .
SOURCE Human papillomavirus.
ORGANISM Human papillomavirus
Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
Papillomavirus.
REFERENCE 1 (bases 1 to 337)
AUTHORS Bens,G., Pfister,H. and Wieland,U.
TITLE Human papillomavirus DNA in non-melanoma skin cancers of a renal
transplant recipient
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 337)
AUTHORS Bens,G., Pfister,H. and Wieland,U.
TITLE Direct Submission
JOURNAL Submitted (17-JAN-1997) Institute for Virology, University Koeln,
Fuerst-Pueckler-Strasse 56, Koeln 50935, Germany
COMMENT GenBank locus name HPU85663.
FEATURES Location/Qualifiers
source 1..337
/organism="Human papillomavirus"
/strain="RTRX10"
CDS <1..>337
/codon_start=2
/product="major capsid protein L1"
/db_xref="PID:g1835612"
/translation="ADNTRNTNFTISVASDSSTVNYDAGKIREYMRHVEEYQLSFIQ
LCRIPLEAEVLTQLNAMNHGILENWQLGFVPTPDNAVHDTYRCISSKATKCPDAVPET
QKEDPFCQYT"
BASE COUNT 117 a 59 c 64 g 97 t
ORIGIN
1 ggctgataac acacgtaaca ctaattttac tattagtgtt gccagtgata gtagcacagt
L1 cds ->
61 gaattatgat gctggaaaaa tcagagaata catgcgcat gttgaagaat atcagttatc
121 atttatttta caattatgta gaataccttt agaggcagaa gtgttaacac agcttaatgc
181 tatgaatcat gggatattag aaaattggca attgggggttt gtacctacac cagacaatgc
241 tgtgcacgat acctacaggt gtatatcttc caaagctaca aaatgtcctg atgcagttacc
301 cgaaacacaa aaagaagacc ctttttgtca atacact
// L1 cds ->

```

LOCUS      HPVUWSA      337 bp      DNA            VRL            23-JUL-1997
DEFINITION Human papillomavirus strain uwS-A major capsid protein L1 gene,
partial cds.
ACCESSION  AF012460
NID        g2271511
KEYWORDS   .
SOURCE     Human papillomavirus.
  ORGANISM Human papillomavirus
            Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
            Papillomavirus.
REFERENCE  1 (bases 1 to 337)
  AUTHORS  Wieland,U.
  TITLE    Direct Submission
  JOURNAL  Submitted (20-JUN-1997) Virology, Universitaet Koeln,
            Fuerst-Pueckler-Street 56, Koeln, NRW D-50935, Germany
FEATURES   Location/Qualifiers
  source   1..337
            /organism="Human papillomavirus"
            /strain="uwS-A"
  CDS      <1..>337
            /codon_start=2
            /product="major capsid protein L1"
            /db_xref="PID:g2271512"
            /translation="ADNTRNTNFTISVSTEANAQQYNAGNIREYLRHVEEYQLSLILQ
            LCKVSLVPEVLSQINAMNSGILEDWQLGFVPTPDNSVHDTYRFINSTATKCPDKVAPK
            EKEDPFAQYF"
BASE COUNT      119 a      55 c      52 g      111 t
ORIGIN
  1 agcagataat accagaaaca ctaatthttac tataagcgta tccacagaag ctaatgcaca
L1 cds ->
  61 gcaatataat gctggtaata ttagagaata ttttaagacat gtagaagaat atcaattgtc
  121 tttgattctt caactatgta aagtttcttt agttocagaa gttttatctc aaattaatgc
  181 aatgaattca ggcattttgg aagactggca attaggtttt gtaccaacac cagataattc
  241 agttcatgac acttatagat ttattaattc tactgctact aaatgtcctg acaaggttgc
  301 tcctaaagaa aaggaagatc cttttgctca atacttt
                                     L1 cds ->
//

```

HPVUWSD

LOCUS HPVUWSD 340 bp DNA VRL 23-JUL-1997
DEFINITION Human papillomavirus strain uwS-D major capsid protein L1 gene,
partial cds.
ACCESSION AF012461
NID g2271513
KEYWORDS .
SOURCE Human papillomavirus.
ORGANISM Human papillomavirus
Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
Papillomavirus.
REFERENCE 1 (bases 1 to 340)
AUTHORS Wieland,U.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-1997) Virology, Universitaet Koeln,
Fuerst-Pueckler-Street 56, Koeln, NRW D-50935, Germany
FEATURES Location/Qualifiers
source 1..340
/organism="Human papillomavirus"
/strain="uwS-D"
CDS <1..>340
/codon_start=2
/product="major capsid protein L1"
/db_xref="PID:g2271514"
/translation="ADNTRNTNFTISVSTENGGAEYDATNIREYLRHVEEYQLSFIL
QLCKVPLTAEVLTQINAMNSNILEEWQLGFVPAPDNPIHDTYRYINSAATRCPKNPP
KEREDPYKDLN"
BASE COUNT 123 a 56 c 62 g 99 t
ORIGIN
1 cgctgataat actcgaaata caaatTTTtac aattagtgta tccactgaaa acggaggtgc
61 tcaggaatat gatgctacaa atattagaga atatttaaga catggtgagg aataccaatt
121 gtcatttata ttgcaattgt gtaaggTtcc tTtaactgct gaagtgctga cacagataaa
181 tgctatgaat tcaaataTat tagaggagtG gcagttagga tTtgTtctg caccggataa
241 tcctatccac gatacataca gatataTtaa tTctgcagct actagatgTc ctgataaaaa
301 tcctcaaaaa gaaagagaag atccttTcaa ggatcTaaac

//

Animal Sequences

BPV5E1E2	BPV5E7E8
CgPV1	FPV1L1
MfPV	OvPV1
OvPV2	RPV1L1

INTRODUCTION

New sequences presented in this section are a mixed bag of phylogenetically isolated animal viruses.

BPV-5 was previously known from an L1 sequence (GenBank accession number U21863, published in the 1995 compendium page I-C-246). Two new sequences now extend the known genome to the E1, E2, E6, E7, and E8 regions. The sequences are preliminary and may contain errors.

CgPV-1, a genital/mucosal supergroup A virus is probably the first known member of a new group-level clade as can be seen in the phylogenetic tree on page I-3. The sequence presented here covers the MY09/MY11 region and was obtained by extending the original CgPV-1 L1 sequence (GenBank Accession M64366). As part of the same study [1], the existing sequences of both the Chaffinch papillomavirus sequence, **FPV1L1**, and the reindeer papillomavirus, **RPV1L1** were extended into the MY09/11 region in order to allow phylogenetic comparisons. The long-tailed macaque sequence, **MfPV**, also part of the Chan et al. study, is closely related to RhPV-1.

Two complete ovine papillomavirus genomes, **OvPV-1** and **OvPV-2** were released this year. They cluster in the C2 group of viruses with the deer and elk papillomaviruses. The major genomic sites are annotated in the sequences presented here, but in the absence of a publication on these viruses little is known of their biology.

References

- [1] Chan, S.Y., Bernard, H.U., Ratterree, M., Birkebak, T. A., Faras, A. J., Ostrow, R. S. Genomic diversity and evolution of papillomaviruses in Rhesus monkeys. *J. Virol.*, **71**, 4938–4943 (1997).

BPV5 E1 E2

LOCUS BPV5E1E2 1657 bp DNA VRL 17-APR-1996
 DEFINITION Bovine papillomavirus type 5 E1 ORF and E2 ORF genes, partial cds.
 ACCESSION U43367
 NID g1263203
 KEYWORDS .
 SOURCE Bovine papillomavirus strain=Type 5.
 ORGANISM Bovine papillomavirus
 Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
 Papillomavirus.
 REFERENCE 1 (bases 1 to 1657)
 AUTHORS Bloch,N., Breen,M. and Spradbrow,P.B.
 TITLE Direct Submission
 JOURNAL Submitted (14-DEC-1995) Natalie Bloch, Veterinary Pathology, Univ.
 of Queensland, St Lucia Campus, Brisbane, QLD 4072, Australia
 COMMENT This is a preliminary sequence which may contain some sequencing
 errors. The E1 and E2 ORFs of this sequence are most similar to
 the homologous ORFs of the deer papillomavirus (DPV) and the European
 elk papillomavirus (EEPV), thus placing BPV5 in group C2. The
 beginning of this sequence lies roughly 500 nt in a 3' direction
 from the actual start of the E1 CDS as estimated by alignment against
 DPV and EEPV. Approximately 878 nt are absent from the 3' end of
 the E2 ORF in this sequence.

FEATURES Location/Qualifiers
 source 1..1657
 /organism="Bovine papillomavirus"
 /strain="Type 5"
 /note="BamH1-HindIII fragment of a BPV5 genomic clone"
 CDS <1..1343
 /note="E1 ORF"
 /codon_start=3
 /product="unknown"
 /db_xref="PID:g1263204"
 /translation="LAVFKFVYAASFCDLTRPFKNDKTTNYQWVAAVFGVSEELFEAS
 KQLLGRSCTYLHATCRAHENASVALLLSFHVAKSTKTVPNLLKNNLNLRAEHMLQP
 PKLRGVTSAMFWYKMTLSPNTYTWGQLPRWIEHQILITENSSEVLKFDFFSHMVQWALD
 NEMMDESSIAFHYAQMADHDSNARAWLGLSNQAKIVKDVCPYGTSLSESICVADNVI
 CAPKCVKEKNCDWVLVSDHAVFEVSWIEPIKICECLAPMASRSAQKKLSCIYRPPDTG
 KSLFTNSLMSFLLKGVNLFANSASHFWLPLTEAKVALIDDATHACLKYCDTYLRNFF
 DGYSVCIDRKHKNVQIKAPPMLLTSNIDI PGRKKSILTLKSRVTCFYFNDKCPLNED
 GKPLFQITDPDWKSFFERLWQRLELSDQEEEEEGDENGSRGTFICSTRNSNDFT"
 CDS 1288..>1657
 /note="E2 ORF"
 /codon_start=1
 /product="unknown"
 /db_xref="PID:g1263205"
 /translation="MAAAERLSAAQETQMTLLEKPSFDLKDHSYYGPLRTEENTIFYA
 ARKKGLTSLGHCPVPTLATAAANAKAAIEMQLLLKDLLRSPFAKNDWSPNDVSHERYK
 APPSDTLKRKPRIVEVIFDKD"

BASE COUNT 511 a 317 c 357 g 472 t
 ORIGIN
 1 agcttgctgt atttaagttt gtgtatgctg cgagtttttg tgacttaact agacctttta
 -> partial E1 orf start
 61 aaaatgataa aacaacaaac tatcagtggg tggcggcggt ctttgggggt tgggaggagt
 121 tgtttgaagc tagtaagcag ttgctaggta gaagctgcac atatttgcac gcgacctgca
 181 gagcccatga aatgcctca gttgctttgc ttttattatc ctttcacgtg gcaaaatcta
 241 caaaaacagt cccaaactcg ttaaaaaatt tgctcaattt aagagctgag cacatgatgc
 301 tgcagcccc caaacttaga ggggtaacat ctgcaatggt ttggataaaa atgacattaa
 361 gcccgaaatac ttatacatgg gggcagttac ctagggtgat agaacaccaa atattaatta

```

421 cagaaaatag ttcagaagtt ttaaaatttg atttctctca catggtccaa tgggccccttg
481 ataatgagat gatggatgag tcctccatag cttttcatta tgcgcagatg gctgatcatg
541 actctaatagc cagagcatgg ctaggtttaa gtaatcaagc taagatagtt aaagatgtct
601 gcccctatgg tacatcatta tcagagagca tatgocgtagc tgacaatgtc agcatatgtg
661 ccccaaatg tgtgaaagag aaaaattgtg actgggtcctt ggttagtgat catgcagttt
721 ttgaagtttc atggattgag ccaataaaga tttgtgaatg ccttgcgccc atggcctcaa
781 ggagtgccca aaaaaactg tcttgcatth ataggccacc tgatactggc aaatccttat
841 tcactaatag cctgatgagt tttctaaaag gcaaagttht aaatthtgca aatagtgcaa
901 gtcactthtg gctgcccccc ctgactgaag ccaaggtagc tthaatagat gatgccacgc
961 atgcctgctt aaaatactgt gatacttacc ttagaaatth tthtgatggt tattctgtgt
1021 gcattgatag gaagcataaa aatgcagtcc aaataaaagc acctccaatg cththtaacta
1081 gcaatataga tataaccaggc aggaagaaaa gtattcttac cctcaaaagc agggtgacct
1141 gcttctatth taatgataaa tgtcctctaa atgaagatgg aaaaccactg thccaaataa
1201 ctgacccccg ttggaaatct thththgaaa ggctthggca gcgthTAGag ctcagtgacc
      E2 orf start ->
1261 aggaggagga ggaggagggg gacgaaaATG gcagccgcgg aacgthtatc tgcagcacia
      E2 cds ->
1321 gaaactcaaa tgactthtact TGAgaagcct agctthgatt taaaagatca tatatcatat
      <- E1 end
1381 tatgggcctc tgcgaacaga aaacactatt ththtatgcag ctgcacaaaa aggtctgacc
1441 tcacttgagc actgtccagt tcctaccctg gcaactgcag cagccaatgc aaaagcagca
1501 attgaaatgc agctgctgct aaaagacttg ttacgthcac cththtgcaa aatgattgg
1561 tcaccaacg atgttagcca tgagcgctac aaggcccctc ctagtgcac tthgaaaaga
1621 aagcctagaa thgtggagggt tatatthgat aaggatc
      E2 cds ->

```

BPV5 E7 E8

LOCUS BPV5E7E8 1229 bp DNA VRL 04-APR-1996
DEFINITION Bovine papillomavirus HindIII-HindIII fragment corresponding to the 5' end of the E1 ORF and the E7 and E8 ORF of BPV1.
ACCESSION U50841
NID g1255224
KEYWORDS .
SOURCE Bovine papillomavirus strain=type 5.
ORGANISM Bovine papillomavirus
Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
REFERENCE 1 (bases 1 to 1229)
AUTHORS Bloch,N and Breen,M.B.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-1996) Natalie Bloch, Veterinary Pathology, University of Queensland, Brisbane, QLD 4072, Australia
COMMENT This is a preliminary sequence which may contain some sequencing errors. When this sequence is translated to protein in all three reading frames and the resulting proteins used in BLAST searches against GenBank, significant matches are found with bovine papillomavirus E6, E7, and E1 genes. However, the translations contain stop codons, and it is difficult to obtain good alignments with these genes. For this reason we have been unable to determine the beginning and end of genes in the sequence with any certainty, and so the sequence is presented below with no annotation.
FEATURES Location/Qualifiers
source 1..1229
/organism="Bovine papillomavirus"
/strain="type 5"
/note="HindIII-HindIII fragment; similar to 5' end of the E1 ORF and the E7 and E8 ORF of BPV1"
BASE COUNT 322 a 274 c 308 g 325 t
ORIGIN

```

1 agctttaggg tccctggatg ctctaaaatg caagaatcat aaatatagga ggggtgcatag
61 aggggaaag ccttatggca tgtgtcaaat ttgcttagag gctttgctgc aattagaaag
121 gcaagaatth ccttgacat tgctactgca aaggactttg ttaaagttg ggggagactg
181 cctggaggac tattgtgtac gctgctatta ctgtggctgc gtgttgtcag acagtgaaaa
241 aagatcgcca cgccttggc ccacgaagt acctgtacgt ccgtggaaga gccagaggcc
301 gctgctactc ttgctctagt gatggtcgcc ggccctgcgt gttctaaatt tctgccacaa
361 gaccctccac caccctcagt gacactggtt ctcatgatt tgactcaaga agaggatgaa
421 caggatthttg taacgctgca cgcacaatat agaccaactt ttaaagataa aactcctaga
481 cgacctggct acaatccttc gtgctgctcc ctgccacatt cagaggctct ctgcgagggtg
541 ctgagthttt tttgtgcaag tcggatgccc ctgtggctcag cctttgaaga ttgctgtgca
601 aagtacccca gactgtatct ctgagthttg acaacttctg cgaggacctt tagatcttct
661 gtgtcctcac tgcgattcc cgctgctatg gcagataaat caggtagatt gctggggggc
721 tgctctthttg tattagatga ggctgactgt agtgatcttg aaataagtag tgatgatgag
781 tctgataagg aaaaatgtgcc aaatggacag gatatgtgca atagcttca tgctgaatth
841 atagacaatg cgcctthtag acagggaaat acccctgccc cthttccaga gccaggtagc
901 ccaggcggga aaacagaaag taaactatct caaaagaaac tgcacctca gtcgagtgac
961 tgggcgggtt gtagagcagt gctgcagcct gtgaaccaca gcacccaac agctgaaagg
1021 cgcctthttg agtgctcaag tagtgaatg gaagttagtt atgcttcttc gcccgccgcc
1081 ccaaacacac aggtatthtag aatcaaat agtgggtctg tgggaggaag tagcgggtt
1141 gggtcacagg ctgcgtaag tcagtctcaa ccaaacagta atthacatth tcagatthta
1201 aagtctatac atagtctcc ttgcaagct

```

LOCUS CgPV1MY911 453 bp DNA VRL 07-JAN-1997

DEFINITION Colobus monkey papillomavirus type 1 major capsid protein (L1) gene, partial cds.

ACCESSION U72629

NID gl764165

KEYWORDS .

SOURCE Colobus monkey papillomavirus type 1.

ORGANISM Colobus monkey papillomavirus type 1

Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.

REFERENCE 1 (bases 1 to 453)

AUTHORS Reszka,A.A., Sundberg,J.P. and Reichmann,M.E.

TITLE In vitro transformation and molecular characterization of Colobus monkey venereal papillomavirus DNA

JOURNAL Virology 181 (2), 787-792 (1991)

MEDLINE 91196277

REFERENCE 2 (bases 1 to 453)

AUTHORS Chan,S.-Y., Bernard,H.U., Ong,C.K., Chan,S.P., Hofmann,B. and Delius,H.

TITLE Phylogenetic analysis of 48 papillomavirus types and 28 subtypes and variants: a showcase for the molecular evolution of DNA viruses

JOURNAL J. Virol. 66 (10), 5714-5725 (1992)

MEDLINE 92407963

REFERENCE 3 (bases 1 to 453)

AUTHORS Chan,S.-Y., Delius,H., Halpern,A.L. and Bernard,H.U.

TITLE Analysis of genomic sequences of 95 papillomavirus types: uniting typing, phylogeny, and taxonomy

JOURNAL J. Virol. 69 (5), 3074-3083 (1995)

MEDLINE 95222760

REFERENCE 4 (bases 1 to 453)

AUTHORS Chan,S.-Y., Ostrow,R.S., Faras,A.J. and Bernard,H.U.

TITLE Genital Papillomaviruses (PVs) and Epidermodysplasia Verruciformis PVs occur in the same Monkey species: Implications for PV Evolution

JOURNAL Virology 228 (1997) In press

REFERENCE 5 (bases 1 to 453)

AUTHORS Chan,S.-Y., Ostrow,R.S., Faras,A.J. and Bernard,H.U.

TITLE Direct Submission

JOURNAL Submitted (25-SEP-1996) Institute of Molecular & Cell Biology, National University of Singapore, Lower Kent Ridge Road, Singapore 119260, Republic of Singapore

COMMENT The first L1 genetic sequence obtained from a Colobus monkey [1] was published in GenBank in 1991 under the Accession Number M64366. The virus was isolated from a penile lesion pathologically similar to venereal lesions caused by genital HPV. The present sequence, also L1, was obtained by extending the same (original) clone in a 3' direction by primer walking so as to cover the phylogenetically informative MY09/MY11 consensus primer region. The sequence reported below covers only the MY09/11 region; the sequence between the 3' end of the original sequence and the 5' end of this sequence has not been reported. Thus a gap of approximately 621 unknown nucleotides exists between the two L1 sequences of CgPV-1.

CgPV1 is a genital/mucosal supergroup A virus, loosely related to HPVCP8061 (Acc. No. U12479). It is probably the first known virus of a new group-level clade. The only other known Colobus monkey PV, CgPV-2, is only very distantly related to CgPV-1. It is a member of the B, or EV/cutaneous, supergroup. The implication of this finding is that the evolutionary divergence of the genital and EV supergroups predated the divergence of the monkeys and humans roughly 22 million

CgPV1 MY911

```
years ago.
FEATURES             Location/Qualifiers
  source             1..453
                    /organism="Colobus monkey papillomavirus type 1"
                    /note="CPV1; sequence derived from a pUC18 clone of the
                    complete genome linearized at the unique EcoRI site in the
                    L1 ORF"
  gene               1..453
                    /gene="L1"
  CDS                <1..>453
                    /gene="L1"
                    /note="MY09/11 segment"
                    /codon_start=1
                    /product="major capsid protein"
                    /db_xref="PID:g1764166"
                    /translation="AQGRNNGICWGNQVFLTVVDTRSTNLTVCATATSETTYKASNF
                    KEYLRHGEEFDLEFIFQLCVVNLTRVMTYIHGMDPSLLEDWNFGTLPPPSASLGDTY
                    RFQQSQAITCQRPPAPEKDKQDPYAGLTFWEVDLTERFSVDLDQFPLGR"
BASE COUNT          116 a    121 c    105 g    111 t
ORIGIN
    1 gctcagggac gcaacaatgg catctgctgg ggcaatcagg tgtttttaac agtagtagac
L1 CDS ->
    -> start MY09/11 region
    61 actaccgta gcaccaacct cactgtgtgt gcaactgcca cgtctgaaac cacatataag
    121 gcctccaact ttaaagaata cttaaggcat ggagaggaat ttgatctcga gtttattttt
    181 cagctatgtg ttgtaaacct tactcgagag gtcatgacat acatacatgg catggaccgc
    241 tctctgctag aggactggaa ctttgggtacc ttgcctccgc cgtctgctag tttggggggc
    301 acctacaggt tccagcagtc ccaggccata acatgccaac gccccccagc ccctgaaaag
    361 gacaagcaag acccctatgc aggcctaacc ttttggggagg tggaccttac agagcgtttc
    421 tcagttgatt tagatcaatt tcccctggga cgc
                                L1 CDS ->
                                <- end MY09/11 region
//
```

LOCUS FPV1L1R 609 bp ds-DNA VRL 30-MAY-1997
 DEFINITION Chaffinch papillomavirus strain FPV-1 L1 gene, partial cds.
 ACCESSION K02020 U89669
 SEGMENT 2 of 2
 SOURCE FPV-1 DNA from chaffinch epithelial warts.
 ORGANISM Avian papillomavirus
 Viridae; ds-DNA nonenveloped viruses; Papovaviridae.

REFERENCE 1 (bases 1 to 609)
 AUTHORS Moreno-Lopez,J., Ahola,H., Stenlund,A., Osterhaus,A. and Pettersson,U.
 TITLE Genome of an avian papillomavirus
 JOURNAL J. Virol. 51, 872-875 (1984)
 MEDLINE 84292470

REFERENCE 2 (bases 1 to 609)
 AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J. and Ostrow,R.S.
 TITLE Genomic diversity and evolution of papillomaviruses in rhesus monkeys
 JOURNAL J. Virol. (1997) In press

REFERENCE 3 (bases 1 to 609)
 AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J. and Ostrow,R.S.
 TITLE Direct Submission
 JOURNAL Submitted (14-FEB-1997) Institute of Molecular & Cell Biology, National University of Singapore 119260, Republic of Singapore

COMMENT FPV1 and FPV1a were isolated from chaffinches in the Netherlands and Sweden, respectively; FPV1 was isolated from a skin papilloma on the leg, FPV1a from a wart on the foot. The isolates give identical restriction fragment patterns for several restriction enzymes, although the patterns for ClaI are distinct. The genomes have been cloned and their genomes partially characterized and sequenced; the genome is approximately 7.8 kb. Low stringency hybridization to BPV1 revealed some sequence homology. The organization of the genome appears to be similar to that of mammalian PVs. Partial sequencing of the E1 and L1 ORFs revealed greatest homology to BPV1 and related types within L1; within E1, high similarity to the same sequences is observed, although HPV65 is more similar in this region. FPV DNA (crude and purified) failed to raise tumors in the chaffinch and canary foot or tarsus, and also did not demonstrably lead to transformation of C127 mouse cells.

Corrections: Chan et al. [2] extended the original FPV L1 sequence (Acc. No. K02020) by primer walking in order to obtain sequence of the entire MY09-MY11 region. In so doing they discovered an 18-bp in-frame deletion between positions 129 and 130 of the original sequence. They also extended the original known sequence in a 3' direction to the end of the MY09-MY11 region. The sequence as recorded below is thus composite; the first 129 nucleotides are from the original sequence (K02020), then come the 18 nt missing from the original sequence, followed by the remaining sequence as reported by Chan et al. [2] in GenBank Acc. No. U89669.

NCBI gi: 332992

FEATURES Location/Qualifiers
 source 1..609
 /organism="chaffinch papillomavirus"
 /strain="chaffinch papillomavirus type 1 (FPV-1)"
 gene <1..>609
 /gene="L1"
 CDS <1..>609

FPV1 L1

```
/gene="L1"  
/note="composite of 2 GB sequences: K02020 and U89669"  
/codon_start=1  
/translation="EPVPETVPIASREQIEKNN SAYMACPSGSVITSDTNLFNRSYWL  
SRAQGTNNGILWENL FVTVLDNSRNVIMKISSLAEGAQENNA TVYDWKNYYECVRHV  
EEYGISAIVRLCRVTLTAENL GSIYRMNPDILKKWGIQEAPLGPQSAEDKYRFTSSQA  
ITCQLPQNPPNAPDDPYKTENYWTVDCRERLSDDLRLRYPLGR"
```

BASE COUNT 174 a 133 c 159 g 143 t

ORIGIN About 4.3 kb after <fpv11>.

1 gaacctgtac cagagacagt tcccatcgct tctagggaac agattgaaa gaacaatagt
L1 cds ->

```
61 gcctacatgg cctgcccgtc tggctccggt atcacgagtg atacgaatct ttttaacagg  
121 tcatactggc tttcccgggc gcagggcacg aacaatggca tattgtggaa cgaaaactta  
181 ttcgtgacag tgctggataa tagcaggaat gtcattatga aaataagcag cttagctgaa  
241 ggtgctcagg agaataatgc cacagtctat gactggaaaa attactacga gtgtgtcagg  
301 catgtagagg agtatggcat atctgcaata gtaaggcttt gcagagtcac tttgactgcc  
361 gagaatctag ggtccatcta taggatgaac cccgacattc tgaaaaagtg gggatttcag  
421 gaggcaccct tggggccgca atctgctgag gataagtaca ggtttacaag tagccaggct  
481 ataacttgcc agttgcctca aaaccccccc aacaatgctc cggatgatcc gtacaagaca  
541 gagaattatt ggacggtgga ctgcagggag cggctttcgg acgaccttct acggtatcct  
601 ttgggcaga
```

L1 cds ->

LOCUS OVPV1 7761 bp DNA circular VRL 02-FEB-1997
 DEFINITION Ovine papillomavirus 1 complete genome.
 ACCESSION U83594
 NID g1813785
 KEYWORDS .
 SOURCE ovine papillomavirus 1.
 ORGANISM ovine papillomavirus 1
 Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
 Papillomavirus.
 REFERENCE 1 (bases 1 to 7761)
 AUTHORS Karlis,J., Delius,H., Baird,P.J., Meischke,H.R.C., Burrell,C.J. and
 Higgins,G.D.
 TITLE Cloning and Sequencing of Two Ovine Papillomaviruses
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 7761)
 AUTHORS Delius,H., Karlis,J., Baird,P.J., Meischke,H.R.C. and Higgins,G.D.
 TITLE Direct Submission
 JOURNAL Submitted (02-JAN-1997) Infectious Diseases Laboratories, Institute
 of Medical and Veterinary Science, Frome Road, Adelaide, South
 Australia 5000, Australia
 FEATURES Location/Qualifiers
 source 1..7761
 /organism="ovine papillomavirus 1"
 CDS 102..473
 /note="E6 ORF from 54 to 473"
 /codon_start=1
 /product="transforming protein"
 /gene="E6"
 /translation="MLSSKFLGKQCLWCRVELTSLDKYRCTSKGINVVQRGKGLYGV
 CLACLETALTLESLYPAQSIPATVEHLERTIRCCYCGGKLNIDEKRRHFLENECYA
 LVRGCLRGRCYECTKD GARTKYP"
 CDS 448..750
 /note="E7 ORF from 421 to 750"
 /codon_start=1
 /product="transforming protein"
 /gene="E7"
 /translation="MVHGPNTHKDL PQDESPETVTLHLRPLIQPSEHGSLPSLKPLK
 IQKKS RPLLRTYYVTVACPCSTRLNFAVSTSSKSILIFEELLTADFQILCPTCAKRP"
 CDS 747..2624
 /note="E1 ORF from 675 to 2624"
 /codon_start=1
 /product="replication protein"
 /gene="E1"
 /translation="MSDEPGSSGIGKGFILLEAECSSDSEADSQVDS DAGHDLED
 FVDNATVFQGNHRELFQIQEKEAGDKAIQKLRKRLALSPSSDASPEIDQLSPGLAAIT
 LQPRRNPLVKRRLFDNIGPKAQDEANSPSQVSLQVHGKENGSRQEDSPHGSLQTDHSV
 SEQKDSQRMVLDILKSKNSAACKLKFKTIFACSYSDLTRVFSNKTTNLQWVIAAYG
 PSETMFEASFELLKACSYLLSVRRSHETGTVALFLACFNNAKSRDTRVTKLFASILNV
 HPEQLMQPPKIRGVCAALFWFRLTFSPATLTYGTL PQWIRTQTIAAEYTDQALKFDF
 GTMVQWAYDNSYCEESKIAEYAMLANCDSNAKAF LASNNOAKMVKDCATMVRHYKRA
 EVQAMSISEYIKRRC EQAPEGGSWLPIMNLFKFGIEPIRFVNSMRQWLRGVPKKNCI
 CIVGPPNSGKSL LCNLSIFLGGRVLTFAMHKSHFWLAPLSEARVALIDDATYACWKY
 FDTYLRNALDGYPICIDRKHKTA VQMKAPLLVTSNIDVHADEKLYYLHSRIVSFYFK
 ETCTTANGEP MFSITNADWKIFFERLWGRLELSDQEEEGEEDVSQRPF TCSARAADAA
 D"
 CDS 2569..3780
 /note="E2 ORF from 2536 to 3780"
 /codon_start=1
 /product="regulatory protein"

OvPV1

```
/gene="E2"  
/translation="MSASDRLRAAQEQMQLLIEKDSTKLSDHVLFWEAVRTEENGLLYA  
ARQKGITSMNGFPVPALAVSSTKAKQAIEMQLLLELMHTTWRDAWTLDDVSYERYL  
MEPKCTLKKGPRVVEVMYDNDPANRTWYTAWSLLYMRTPAGWLFATGGADECGLYYTD  
MTAVRHYYDFGRDAARYGSTGTWSVRDQDRVFGPLSLRPESEWSEHNTERPEDSP  
DRAVPPSPGTPSGFGPIRSSQYRSRTYTRPSPYGADQCQGVLLRCDSTSSVSSPVSL  
GLERREEGEETPPSDSTDIVSAPATPPDPNQFSLFKSSGGQQCVLLSGSGNQAKCFRF  
RLKKGHRDKYLHCTTTWWTVGDEGSERRGQASLLVTFDSPTQELFLKTVPIPDSMTVR  
GYTIQPDF"  
CDS 3128..3460  
/codon_start=1  
/product=""  
/gene="E4"  
/translation="EIRIGCFTLDLYLLDPNQDVNTTQRDDRRTRPTEPSLLLLQGH  
PPDSVPSEANIGAGPILARLPTVQTSKAGFYAAIPPPRCRARYRWVWNEGRKGKRR  
RRRQTQQT"  
CDS 3942..4214  
/codon_start=1  
/product="transforming protein"  
/gene="E5"  
/translation="MQHPFLLLFLGLLWGVQLLLMFFLLFFFIFWDKYGCRCDKLPV"  
CDS 4130..5623  
/note="L2 ORF from 4127 to 5623"  
/codon_start=1  
/product="minor capsid protein"  
/gene="L2"  
/translation="MANKRVKRANPYDLYRTCKQAGTCPPDVIKVEGSTIADKILQY  
GSMGVFLGGLGIGTQQKPVSGGYVPLRTFGSTASISSTAGRGTAAIKPFAGGIPLET  
LETIGAFRPSAVEESSLGTGVGIHSEAPSILIPDSAPSGNVLGGLDVSTDSADTLIT  
LLEPEGPDDIAVLEVRPTEHSRAHLSSSSMHPNPLFQQLEPAAIGDTSTFENVLVGGS  
GIGDNYSESIELTLFSEPRSTPEPGFKKPRLRGIFNFYSRRYYTQLPTSDPDGAAA  
GSYVFENPVFDSKAFEPPELPTDAPPTHADAEESPFISAAKLLQGFSGRVGVSRIARPT  
AMGTRSGVRVGPLFHLRHSFSTIAPSAESVELVPTVLEEVEVITVAESHLGDNAFE  
VDLDSVTSDTPLLETEHLSHSGKKNLPAAGGMSRPVIAVDVDAGQSEGLQPPRIRDI  
STGHDHYNPDSTSTIVIDGNIIIIYSTYFKHYLLHPSLYRRKRKRLLD"  
CDS 4965..5162  
/note="L3 ORF from 4815 to 5162"  
/codon_start=1  
/product=""  
/gene="L3"  
/db_xref="PID:g1813786"  
/translation="MSLRILYSIQRRNLNSCPRMHHLHMQLRNPPLFLQONCKGPL  
GGLVLAELRVPQLWERVVASG"  
CDS 5635..7128  
/codon_start=1  
/evidence=not_experimental  
/product="major capsid protein"  
/gene="L1"  
/db_xref="PID:g1813786"  
/translation="MAFWHAGQKLYLPPTPVSKTLCSESYVYRKDIYYHAETERLLTV  
GHPFYPIQVKDKTVPKVSPNQYRAFMVQLPDPNQFALPDRTIYDPTKEKLVWGLIGLQ  
VSRGQPLGGAISGHTSFNALLDAENVGRRNTNTQSTDDRKAAGMDAQQQVLLVGCTPA  
LGEYWDTARPCVSDSPEKGSPPALELKHKPIEDGMDMIGFGAANFKELNSSKSDLPL  
DITNEICLYPDYLRMTEETAGNSMFFARKEQVYVRHVWSRGGTDEKAPPNGYFLKPK  
SSAQTISGVFIGAPSGSLLSTDGLIFNRPYWLRYAQGMNNGICWNNTVFVTVGDNTR  
GTNLTITVPSDESVDYDSAKINVYHRHVEEYKLAIFQLCSVQLTPTVSSLQGLM  
PSILQNWVNVQPPASSILEDTYRYLES PATKCADNVSPTKPDYDGLKFKWIDLKEK  
FSLDIDQFPLGRRFLAQQGLGCTTIVRRRASRSTSSKVPKRRRRGT"
```

```

BASE COUNT      2171 a   1696 c   1851 g   2043 t
ORIGIN
      1 taacaccata gccgttatag aaaaagcggg aaagtgtttt tcaagcccTA TAAAAGctga
                                          E6 orf start ->
                                          signal->
      61 accgcttagg ttgttgacc caagcacttc taggactcag gATGctgagc agtaaattct
                                          E6 CDs start ->
     121 tgggaaaaaca gtgcttatgg tgcagggtag aattaactag cttagataaa tatcgctgca
     181 cctcaaaagg tattaatgta gtgcagagag gaaaaggact atatggggtc tgcttagcat
     241 gtctagaaac tgctttaacc ctagaagggt cctatatcc  tgcacagagt attcctgcaa
     301 cagtagagca tcttgagcgc actattagat gctgttactg tgggggaaag ttaaataatag
     361 atgaaaaaag gagacacttt ttggagaatg aatgctatgc acttgtaaga ggctgcttga
     421 GAGgacggtg ctacgagtgc actaaggATG gtgcacggac caaatacca TAAggaccta
E7 orf start->          E7 CDs start ->          <- E6 end
     481 ccacaagatg aatcacctga gactgtcaca ttgcacctcc ggccgctgat acagccaagc
     541 gaacatggaa gtctgccctc cctgaaacct taaaaaatac agaagaaaag ccgcccactt
     601 cttegcacgt attatgtgac tgtcgcttgt cctgctcca  ctcggtgaa cttcgctgtg
     661 agcacaagct ctaaATCaat cttatcttc  gaggagctcc taactgcaga ttttcaaatc
          E1 orf start ->
     721 ctgtgtccta cctgtgcaaa gcggccATGA gtgacgaacc aggtagctcg gggattggga
                                          <- E7 ends
          E1 CDs start->
     781 aagggtctga gtttattttg ctggaagccg aatgtgatag ttccgattcg gaggctgact
     841 cacaggtaga cagcgatgct gggcatgac tagaggattt tgtcgataat gcaactgtgt
     901 ttcagggaaa ccaccgggaa ttattccaaa tacaggaaaa agaggcggga gacaaagcta
     961 ttcaaaaact caaaagaaaa ctagctttaa gtcccagttc tgacgcctca cccgaaattg
    1021 atcagctcag cccaggcttg gctgctataa cgttgcagcc gaggagaaat ccgttggtaa
    1081 agagaaggct ttttgacaat attggaccga aagcacaaga tgaagctaat agtccttctc
    1141 aagtcagcct ccaggtaacat gggaaagaaa atgggaggag ccaggaagac agtccccatg
    1201 gctcactgca aactgaccac agcgtcagtg agcagaaaga cagtcaaagg atgggtgctgg
    1261 atataacttaa gtctaagaat agtgctgcct gcaagctgaa gctgtttaa actatttttg
    1321 cgtgcagtta ctctgactta actagagtgt ttcagagcaa taagaccact aatcttcagt
    1381 ggggttatgc tgcttatgga ccttctgaaa ctatgtttga agccagcttt gaactctca
    1441 aaaaagcatg cagctacctg ctaagtgtta gaaggctcca tgaactgga actgtagcat
    1501 tgtttttagc ttgctttaac aacgcgaaaa gcagggatac tgttcgaaag ctgtttgcaa
    1561 gcattctcaa tgtccatcct gagcagctgc tcatgcaacc tccaaaaatt aggggtgtgt
    1621 gtgcagcttt attctggttt agactcactt tctcgctgc aacgctaaca tacgggactt
    1681 taccacagtg gatcaggacg caaacaattg ctgcagagta tacagaccaa gcaactaagt
    1741 ttgactttgg gactatggtt cagtgggcct atgataatag ttattgtgaa gaatctaaaa
    1801 tagcctatga gtatgctatg cttgctaact gtgatagtaa tgcaaaggcc tttcttgcta
    1861 gcaataatca agctaaaatg gtaaaagact gtgcaactat ggttaggcat tataagcgtg
    1921 cagaagtaca agctatgagt atttctgagt atattaaaag acgttgtgaa caagctccag
    1981 agggggggag ctggctcccc attatgaatt tgtttaaatt tcaaggcatt gagcctataa
    2041 ggtttgttaa cagtatgctg cagtggttac gtggcgttcc caaaaaaac tgtatatgta
    2101 tagtaggccc gcctaatagt ggaaaatcat tgcgtgtgaa ctccctaatc tcatttttgg
    2161 gcgggagagt gctaacgttc gctatgcata aaagccattt ttggctcgca cccttaagtg
    2221 aagcaagggt ggctctcata gatgatgcca cgtatgcatg ttggaaatac tttgacacat
    2281 acctcagaaa tgcctggat gggtaacctt tttgtattga taggaaacac aaaactgctg
    2341 tgcaaatgaa agctccacc cttttagtga ctagtaacat cgatgtgcat gcagatgaaa
    2401 agctgtatta tttgcacagc aggattgtga gcttttactt caaggagaca tgcacaacag
    2461 ctaacggtga gcctatgttt tctattacta atgctgattg gaaaattttt tttgaaaggc
    2521 tatggggacg cttagAGCtc agtgaccagg aggaagaggg tgaagaagAT Gtcagccagc
          E2 orf start ->          E2 CDs start ->
    2581 gaccgtttac gtgcagcgca agagcagcag atgctgctga tTGAaaaaga cagtactaaa
                                          <-E1 end
    2641 ctatctgacc atgttctatt ctgggaagct gtgaggactg aaaatggact gttgtatgct
    2701 gctagacaga agggatttac ctctatgaat ggatttccag tgcctgcatt agcagtatct
    2761 agtactaaag caaagcaagc tattgaaatg cagcttttat taactgaatt aatgcacact
    2821 acatggggtc gtgatgcatg gacattgctt gatgtcagct atgaaagata cctcatggag

```

OvPV1

```
2881 cctaaatgca ccttaaaaaa gggaccccg gtagtagaag tgatgatga taatgatcct
2941 gcaaatagga cgtggtatac agcctggagc cttttgtaca tgcgactcc agcaggctgg
3001 ctttttgcaa ctggtggggc tgatgagtgt ggctctact acacagacat gacggcagtt
3061 cgccattatt atgacttctt tggaaagagac gctgcaagat acggcagtac tgggacatgg
3121 tcagtaaGAG atcaggatag ggtgttttac tctggacctt tatctcttag acccgaatca
E4 cds ->
3181 ggatggagtg aacacaacac agagagacga ccggaggact cgcccagacag agccgtccct
3241 cctccttctc cagggacacc ctccggattc ggtcccattc gaagcagcca atatcggagc
3301 aggacctata ctgcccgtc tcctacggt gcagaccagt gccaaaggggt tttactacgc
3361 tgcgattcca cctcctcggc gtcgagcccg gtatcgttgg gtttggaaacg aagggagcaa
3421 ggggaaagaga cgccgcccgtc gccagactca acagacaTAG tttctgcacc tgcaacgcct
<- E4 end
3481 ccggacccca atcagttttc actcttcaaa agctcagggtg ggcagcagtg tgtacttttg
3541 tctggttctg gaaaccaggc taagtgtctt aggtttcggg tgaaaaaggg acatagagac
3601 aagtacctac attgcacaac cacctgggtg actgttgggtg acgagggatc agagcggccg
3661 ggtcaagcaa gtctactggc cacctttgac agtcctacac agcagagaact gtttcttaag
3721 actgttctca tacctgattc tatgactgta agaggctata ctatacagcc agacttcTGA
<- E2 ends
3781 tgggactctg ctgttgcaag cttggcatgc agcatccctt tctgctggtg tttcttggac
3841 ttctgtgggg agttcaattg ctgttgatgt tttttctgct ctttttcttc tttatctttt
3901 gggataaata tggctgtcgc tgcgacaaac tcctgtgtga aataatgcca ctgtggctgt
3961 ggctcctgat atttttgata cttaccaggt ttaatgtgtt catttatctg tttatttttc
4021 ttatactgtt tattagttac attttctggt tattgttggc cgaggagtgt ggtggttccg
4081 tagacgcgatg tcccgggtgt gcagacgtta accccacaaa gaataaAGAA TGgcaaacaa
L2 orf start ->
L2 CDs ->
4141 aagagttaag cgtgccaacc cctatgatct gtacaggact tgcaagcaag caggaacctg
4201 tccacctgat gtgatcccca aagtggaggg aagcactatt gcagacaaaa tcctgcagta
4261 tgggagcatg ggggtattcc tgggagggct aggcattggg acaggtaacg gaaagccagt
4321 ctacagtggtg tatgttcttc tgcggacctt tggatccact gcatccatca gcagcaccgc
4381 aggtaggggc acagcagcca taaagccttt tgcaggagga ataccgctgg agaccttggc
4441 aacaattggt gcattccgct cctctgctgt cgaggaaagc tctttgggaa caggtgtagg
4501 tattcacagt gaagctccct caattcttat tccagactct gccccaagtg ggaatgtgtt
4561 aggggggcta gacgtcagca cagacagctc tgcagacact ttgatcactt tacttgaacc
4621 tgagggctct gatgatattg ctgtattaga ggtgagacca acagaacata gcagagccca
4681 cctctcttcc agcagtatgc atcctaacc cttatttcaa cagttggaac ctgcagctat
4741 aggtgacact tcaacattcg agaatgtgct agtgggtggg agcggcatag gggataatta
4801 cagtgaagt atagAACTta ccttattttc agagccaagg accagcacgc ctgaacctgg
L3 orf start ->
4861 atttaagaag ccaaggcttc gtggaatatt taactatttc agtcgcagat actacactca
4921 gctacctacg tctgatcctg atcctggcgc tgcctgcaggc tcctATGtct ttgagaatcc
L3 CDs start ->
4981 tgtattcgat tcaaaggcgt ttgaacctga gctgcccacg gatgaccac ctacacatgc
5041 agatgctgag gaatccccct ttatttctgc agcaaaactg ctgcaagggc cctctgggag
5101 ggttgggtgt agcagaattg cgcgtcccac agctatggga acgcgtagtg gcgtcagggT
5161 AGGaccttta tttcacctta gacactcctt cagtacaatt gcaccgtcag cagaatctgt
<- L3 ends
5221 ggagctagtc ccaacagtgc tggaggaagg ggaagtgatt acaacagtgg cagagtcaca
5281 tttaggcgac aatgcatttg aggaagttga cctagacagt gttacaagtg acacgccact
5341 acttgaacaa gagcacctct ctactctgg aaaaaaaagc aacctccctg cagcaggtgg
5401 tatgtcccgg cctgtcatag ctgtggatgt ggatgctggc caatctgagg ggttacagcc
5461 tccacgcata cgtgacataa gcactgggtc tgatcattat aatcccagca gcacatcaac
5521 aatagttatt gatggcaaca tcataattta cagtacatac ttttaagcact attatctgca
5581 tcccagcctg tataggcgca agcgcgaaacg cctccttgat TGAttttttt gcagATGgcg
<- L2 ends
L1 CDs ->
5641 ttctggcacg ctggacagaa gctctaccta cctcctacac ctgtgagcaa aacactgtgc
5701 tcggaatcct atgtgtacag gaaagacata tactaccatg cagaaactga acggctgctt
```

```

5761 actgtaggcc acccatttta cctatttcaa gtaaaagaca aaacagttcc taaggatatcc
5821 cctaaccaat atagagcatt catgggtgcag cttccagatc ctaatcagtt tgcattgcct
5881 gacagaacaa tatatgatcc tactaaagag aagctagtat ggggccttat agggttgcag
5941 gtatctcgtg gacagcctct aggcggggca atcagtggtc atacatcatt caatgcactt
6001 ttagatgcag aaaacgtagg aagacgtacc aatacacaat ctactgatga tagaaaagct
6061 gcaggtatgg atgcaaagca gcagcaagtg ctgcttgtgg ggtgcacacc tgcattaggg
6121 gagtattggg atacagctag gcctgtgtc tcagatagcc cagagaaagg ctctgtcct
6181 gcactagagc taaagcataa acctatagag gatggagata tgatggacat aggccttggg
6241 gctgcaaact ttaaagagct taatagctct aagtctgatc tacctcttga tattaccaat
6301 gagatttgtc tatatcctga ttatttgcgt atgacagaag aaacagctgg aaatagcatg
6361 ttcttttttg caagaaagga gcaagtttat gtgcgacacg tatgggtctag aggcgggaca
6421 gataaggaag cccccctaa cgggtacttt ttaaagccta aaagctccgc gcagcctact
6481 atttcagggg tttttattgg tgcacctct gggctctctgc tttcaacaga tggtttaata
6541 tttaatcgtc catattggct atatcgtgct cagggaatga acaatgggat atgctggaat
6601 aatactgtat ttgtcactgt tggagataac acacgtggta caaatttaac tattactgtg
6661 ccctctctcg acgagtcagt aacagattat gacagcgcaa agatcaatgt atatcataga
6721 cacgtggaag aatataaact agcattcatc tttcaactgt gctctgtgca gttaaccct
6781 gaaacagtga gtagtctgca ggggttaatg cccagcattt tgcaaaactg ggaagtaaat
6841 gtacaacctc ctgcctcttc aattctggaa gatacctacc gttatctaga atcgccagcc
6901 actaagtgtg cagataatgt gtctcctact aagccagatc cctatgatgg gttaaaattc
6961 tggaagattg atctaaaaga gaagtttct ttggatattg atcagttccc tctagggaga
7021 cggttcctag ctcagcaagg ttaggatgc acaacaatag tgcgcaggcg agcttctaga
7081 tctacctcaa gtaaggtacc ctctaagcgc cgaagaaggg gcaccTAAGg acactgtcct
                                     <-L1 cds end
7141 agcatattaa gcactttggt aataaataaa aactgcattt aattataaat gtgtcatgtc
7201 cgctgtctac accacaccg gtctgggatg cacaaaatag ctctgtcagc agctataaac
7261 tccctgcagc tggcgctca aaataaaatc cagacgtctt gacgccagct gtggaaaaaa
7321 agggctcgtg gcatagcacc gctggcgggt taagagttca acccacgcag aaccgtctcc
7381 ggttctgtct gctaaaagca atcccacgca caccgggagc ggtctccgtc ggggtcttaa
7441 aattgaagca ccggattcgg ttcataggag taagatttag ttttttttc cggaaatcaa
7501 taggaaaagc aataggaaaa gcaaacactt attttgtaa atgctcaagt tcacaaccgt
7561 ctgctgttca ttaaaaccac agtaccttcc acggttaggt ttcaatatcg ctgacgggtc
7621 ttgctacttt tctaaactgc tccccattg gtgggttcag tgaattacta actgtctcct
7681 gctctgattg ctagatcttc tacgggtcgt tactactctg cacagtaatt ttttctaga
7741 gatgattggt gttaacaata a

```

//

OvPV2

LOCUS OVPV2 7758 bp DNA circular VRL 02-FEB-1997
DEFINITION ovine papillomavirus 2 complete genome.
ACCESSION U83595
NID g1813787
KEYWORDS .
SOURCE ovine papillomavirus 2.
ORGANISM ovine papillomavirus 2
Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
Papillomavirus.
REFERENCE 1 (bases 1 to 7758)
AUTHORS Karlis,J., Delius,H., Baird,P.J., Meischke,H.R.C., Higgins,G.D. and
Burrell,C.J.
TITLE Cloning and Sequencing of Two Ovine Papillomaviruses
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 7758)
AUTHORS Delius,H., Karlis,J., Baird,P.J., Meischke,H.R.C. and Higgins,G.D.
TITLE Direct Submission
JOURNAL Submitted (02-JAN-1997) Infectious Diseases Laboratories, Institute
of Medical and Veterinary Science, Frome Road, Adelaide, South
Australia 5000, Australia
FEATURES Location/Qualifiers
source 1..7758
CDS 100..471
/organism="ovine papillomavirus 2"
/note="E6 ORF from bp 1 to 471"
/codon_start=1
/product="transforming protein"
/gene="E6"
/translation="MLSSKFLGRQCLWCRVELTNLDKYRCTSKGINVVQRGKSFYGVCLACLEVALTLERSLYPATPITATAEHIERTIRCCYCGGKLNLDKRRHLIEGECYAYV
RGCLRGRCYECTKDGARTKYP"
CDS 446..748
/note="E7 ORF from bp 419 to 748"
/codon_start=1
/product="transforming protein"
/gene="E7"
/translation="MVHGPNTHKDLPODESPGTVTLHLRPLKTPKEHGSLSPLKPLKI
QKNCRPLLRTYYVTVACTCSTRLNFAVSTSSKSILLFEELLTTDFITILCPTCAKRP"
CDS 745..2628
/note="E1 ORF from bp 673 to 2628"
/codon_start=1
/product="replication protein"
/gene="E1"
/translation="MSDEPGSSGLGKGSEFILLEAECSSDSEADSQVDSAGHDLED
FVDNATVSQGNHRELFQTQEKEAGEKAIQKLRKRLALSPGSDASPEIDQLSPGLAAIS
LQPRRNPLVKRRLFENIGPKAQDEANSPSQVSFQVHGEDNGRSQEDSGHGSQTQDHSV
SEQQKDSQRMVLDILKSKNSAACKLKLKFTIFACSFCDLTRVFQSNKTTNLQWVIAA
YGPSETMYEASFELLKACSYVLAVRRSHETGTIALILACFNNAKSRDTVQKLFNLI
NVHHEQLLMQPPKIRGVCAALFWFRLTFSPATLTHGTLQPWIRTQTIAAEHTDQALKF
DFGTMVQWAYDNSYCEESKIAYEYAMLNCDTNAKAFNASNNQAKMVKDCATMVRHYK
RAEVQAMTMESEYIKRRCEQAPEGGSWLPIMNLFKFGQIEPIRFVNSMRQWLRGVPKKN
CICIVGPPNSGKSLCNSLISFLGGRVLTFFAMHKSHFWLAPLSEARVALIDDATYACW
KYFDTYLRNALDGYPICIDRKHKTAVQMKAPPLLVTSNIDVHADEKLFYLHSRIVSFY
FKETCTTANGPEPMFSITNADWKIFFERLWGRLELSDQEEEGEEDVSRQTFACRSARAAD
AAY"
CDS 2573..3778
/note="E2 ORF from bp 2540 to 3778"
/codon_start=1
/product="regulatory protein"

```

/gene="E2"
/translation="MSASERLRAAQEQMQLLIEKDSTKLCDHILFWEAVRTENGLLYA
ARQKGITTLNGFVFPALAVSSSKAKQAIEMQLLLTELLNTTWGRDAWTLTDTSYERYL
LEPKCTLKKGPRVVEVMYDNDPANRMICYTAWSLLYMRTPAGWLFATGGADECGLYYTD
MTAVRHYYDFGRDAARYGSTGTWTVRDQDRVFYSGPLSLRPESEGWSEPAERPESPD
RAVPYPSAGTPSGFGPIRSSQYRSRTHSRSPYGSNQCGVLLRCDSTSSLSSEPVSVG
LERGEEGETPPSPDSTDVVTTPSTPPDPNKFSLFKTSGGQQCLLLSGSGNQAKCFRF
RMKKGHRHKYVHCTTTWWTVGDEGAERRGQASLLVTFDSPTQRELFKTVPIPESMTV
QGLTVQADF"
CDS
3132..3458
/note="E4 ORF from bp 3132 to 3458"
/codon_start=1
/product=""
/gene="E4"
/translation="EIRIGCFILDLYLLDPNQDGVSLQNDRRTRPTEPSLTLQGHPP
DLVPSEAASTGAGPILARLPTAQTSKAGFYAAIPPPRCRARVRWVWNEGRKGRKRRR
RQTQQT"
CDS
3796..3930
/note="E5 ORF from bp 3736 to 3930"
/codon_start=1
/product="transforming protein"
/gene="E5"
/translation="MQHPFLLFLGLLWGVQLLLMFFLLFFFFVFDKYGCRCEKLPV"
CDS
4122..5615
/note="L2 ORF from bp 4119 to 5615"
/codon_start=1
/product="minor capsid protein"
/gene="L2"
/translation="MANKRVKRANPYDLYRTCKQAGTCPPDVIPKVEGSTVADKILQY
GSMGVFLGGLGIGTGQKPVSGGYVPLRTFGSTSSLASTASRGATVSKPFAGGIPLT
LETIGAFRPSAVEESSLGTGVGLSNEAPAILIPDSAPSGDTLGGLDVSTDSSADTLIT
LLEPEGPDIAVLEVRPTEHSRAHLSFSSTHPNPLYHYVEPAAIGETSTFENVLVGGS
GIGDNYSESIELNLFSEPKTSTPEPAFKRPRLRGIFNYFSRRYYTQLPTSDPDGAAA
GSYVFENPVFDSKAFEPPELPTDNAPAAAPNEEAPHVSAAKLLQGPSGRVGISRIVRPT
SMGTRSGVRVGPLYHLRHSFSTIAPSAESVE LVPTVLEEGETVITVAESHLGDNAFE
EVDLDSVTSDTPLLDTEHLSHSGKSNLPAASGTSRPLIAVEVDSGQSEGLHPARVRD
ISTNHTHYDPDTTSTVVIDGNIVIYSTYLNKHYLHPSLYRRRRKRLLD"
CDS
5017..5097
/note="L3 ORF from bp 4807 to 5097"
/codon_start=1
/product=""
/gene="L3"
/translation="MHLLLHQMKKPLMFLQLNCCRVPLEE"
CDS
5627..7120
/note="L1 ORF from bp 5612 to 7120"
/codon_start=1
/evidence=not_experimental
/product="major capsid protein"
/db_xref="PID:g1813788"
/translation="MAFWQAGQKLYLPPTPVTKTLCSEYVSRDDIYYHAETER
LLTVGHFPFYSVKVNEKTVPKVSPNQYRAFMIQLPDPNQFALPDRTLYDPGKEKL
VWGLIGLQVSRGQPLGGAISGHTTFNALLDAENVSRRTTAQTDDRKPAGMDAK
QQQVLLVGCTPAVGEYWAKARPCVSDRVDAGSCPPLELKHKPIEDGDMMDIGFG
AANFKELNSSRSDLPDIVNEICLYPDYLRMTEETAGNSMFFARKEQVYVRHV
WAGGSDKEPPPEDYLLKPKGSDQPKIAGVFIGAPSGSLLSTDGLIFNRPYWL
YRQGMNNGICWNNTVFVTVDNTRGTNLTITVSSDSDPLKEYETGKINVYHRHV
EEYKLAIFIFQLCSVQLTPETVSSQLGLMPSILENWEINVQPPTSSILEDTYRYI
ESPATKCADNVPASKPDPEGLRFWKIDLKEKFLDIDQFPLGRRFLAQQGLGC
TAAVRRRVPKSTSDKAPSKRRRRGT"

```

OvPV2

```

BASE COUNT      2140 a   1719 c   1884 g   2015 t
ORIGIN
  1 tcacaccata gccgttatag aaagagcggg aaagtgtttt caggccctat aaaagctaaa
    -> E6 orf start
 61 ccgcgtacgt tgttgacccc aagcctcttt gaagacatcA TGttgagcag taaattcctg
      E6 cds ->
121 ggaagacagt gcttatggtg cagggttgaa ttaaccaact tagacaaata tcgctgcacc
181 tcaaaaaggta ttaatgtagt gcaaagagga aaatcttttt atggagtgtg tttagcttgc
241 ttagaggttg ctctaacatt agaaagatct ctttaccctg caacacctat tactgcaact
301 gcagagcata ttgaacgcac tattagggtc tgttactgtg gtgggaaatt gaatttagat
361 gaaaagagga gacatttaat agagggtgag tgttatgcat atgtgagagg ctgccTGAga
      E7 orf start ->
421 ggacgggtgct acgagtgtac taaggATGgt gcacggacca aatacccaTA Aggacctacc
      E7 cds ->                                <- E6 end
481 acaagatgaa tcacctgga ctgtcacatt gcacctccgg ccgcttaaga cgccaaagga
541 acatggaagc ttgccctcac tgaaccctt aaaaatacag aagaattgcc gccactcct
601 tcgcacgtat tatgtgactg tggcttgtag ttgctccact cggctgaact tcgctgtgag
661 cacaagctcT AAatcaatac ttctcttcga ggagctgcta actacagact ttaccatcct
      E1 orf start ->
721 gtgtcctacc tgtgcaaagc ggccATGAgT gacgaaccag gtagctcggg gctggggaaa
      E1 cds ->
                                <- E7 end
781 gggctctgagt ttattttgct ggaagccgaa tgtgatagtt ccgattctga ggctgactca
841 caggtagaca gcgatgctgg gcatgatcta gaggattttg tcgataatgc aactgtctct
901 cagggaaacc accgggaatt attccaaaca caggaaaaag aggggggaga aaaagctatt
961 caaaaactca aaagaaact agctttaagt cccggttctg acgctcacc cgaaattgat
1021 cagctcagcc caggcttggc tgctataagt ttgcagccga ggagaaatcc gttggtaaag
1081 agaagcctt ttgaaaatat cggaccgaaa gcacaagatg aagctaatag tccttctcaa
1141 gtcagcttcc aggtacatgg ggaagacaat gggaggagcc aggaagacag tgggcatggc
1201 tcaatatac ctgaccacag cgtcagtgtg cagcagcaga aagacagtca aaggatgggtg
1261 ctggatatac ttaaatctaa gaatagtgtg gcgtgcaaac tgaagctgtt taaaacaatt
1321 tttgctgca gcttttgtga cttaaactaga gtgtttcaaa gcaataagac cactaactta
1381 cagtgggtga tagctgctta tggaccctcc gaaactatgt atgaagccag ctttgaactg
1441 ctaaaaaagg catgcagcta cgtgttagct gttagaaggt ctcagagac tggcactata
1501 gccttaatat tagcatgctt taacaacgcc aaaagcaggg atactgttca aaagcttttc
1561 tcaatatac tgaatgttca ccacgagcag ctacttatgc agcccccaa aataaggggc
1621 gtgtgtgcag ctttattctg gtttaggctt accttttcgc ctgcaacgct tacacatggg
1681 actctgccac agtggatcag aacgcagacc attgctgcag aacatacaga tcaagcacta
1741 aaatttgact ttgggacaat ggtgcagtgg gcgtatgaca atagttattg tgaagaatct
1801 aagattgctt atgaatatgc tatgcttgct aactgtgata ctaatgctaa agcatttctt
1861 gccagcaaca atcaggctaa aatgggtgaag gactgtgcaa ctatgggtgag aactataaaa
1921 cgtgcagaag tgcaagctat gactatgtca gaatatataa aaagacgctg tgaacaggcc
1981 ccagaggggtg ggagctggct cctattatg aatttattta aatttcaagg tattgagcct
2041 attcggtttg ttaatagcat gaggcaatg ttacgaggtg ttcctaaaaa gaattgtatt
2101 tgtattgtgg gcccgccaaa cagtgggaag tctttgctgt gcaattcatt aatttcattt
2161 ttaggcggga gagtctaac gtttgcaatg cataagagcc atttttggct tgcaccttta
2221 agtgaagcga ggggtgctct aatagatgat gccacgtatg catggtggaa gtactttgac
2281 acatacctca gaaatgcatt ggacgggtac cctatttgta ttgatagaaa acacaaaaca
2341 gcagtgcaaa tgaagctcc gccctctta gtgactagta acatagatgt gcatgcagat
2401 gaaaagctgt tctatttgca cagcaggatt gtgagctttt acttcaagga gacatgcaca
2461 acagctaacg gtgagcctat gttttctatt actaatgccg attggaaaat ttttttgaa
2521 aggctatggg gacgctTAGa gctcagtgtc caggaggaag aggggtgaaga agATGtcagc
      E2 orf start ->                                E2 cds ->
2581 cagcgaacgt ttgcgtgcag cgcaagagca gcagatgctg cttatTGAAA aagacagtac
      <- E1 end
2641 taagctatgt gaccatatcc tgttttggga agctgtgagg actgaaaatg gactgctgta
2701 tgctgctaga caaaaaggta taaccacttt gaatggattc cctgtgcctg cactagcagt
2761 atctagcagt aaagcaaacg aagcaattga aatgcagctt ttactgacag aattacttaa
2821 cactacatgg ggtcgtgatg catggacact gacagatacc agctatgaga gatacttatt

```

```

2881 ggaacctaaa tgcaccttaa aaaaaggacc ccgagttagtg gaagttatgt atgacaatga
2941 tcctgcgaat aggatgtggt atacagcctg gagccttttg tacatgcgca ctccagctgg
3001 ctgggtgttt gcaactgggtg gggctgatga gtgtggcctt tactacacag acatgacggc
3061 ggttcgccat tactatgatt tctttggaag agatgctgca agatacggca gcaactgggac
3121 atggacagTG Agagatcagg atagggtggt ttattctgga cctttatctc ttagaccoga
    E4 orf start ->
3181 atcaggatgg agtgagcctg cagaacgacc ggaggactcg cccgacagag ccgtccctta
3241 cccttctgca gggacacccct cggatttgg tcccatcaga agcagccagt accggagccg
3301 gaccattctc cgcccgtctc cctacggctc aaaccagtgc caagggggtt tactacagctg
3361 cgattccacc tcctcgttgt cgagcccggg atcgggtgggt ttggaacgag gggaggaagg
3421 ggaagagacg ccgcgctcgc cagactcaac agacgTAGtg actccacctt caacacccc
                                <- E4 end
3481 ggaccctaac aaatthtgc tcttcaagac ttcaggtggg cagcagtgtc ttcttttctc
3541 tggctctggg aaccaggcaa agtgctttag gtttgcgatg aaaaaaggtc acagacacaa
3601 atacgtccat tgcacaacta cctggtggac tgttggagac gagggagcag agcgcctggtg
3661 ccaagcaagt ctgctagtga cctttgatag ccctacacag cgtgaactgt tcttaaaaaac
3721 tgttccata ccTGAgtcta tgactgtgca aggtttaaact gtgcaagcag acttcTGAtg
    E5 orf start ->                                <- E2 end
3781 ggacactgcc taggaATGca acatcccttt ctgctgctgt tccttggact tctgtggggg
    E5 cds ->
3841 gtgcagcttc tgttgatggt ctttttgctt ttcttttctt ttgtattttg ggataaatat
3901 ggctgcccgt gcgagaaatt acctgtgTAA ataatgtctg tctggttgtg gcttttactg
                                <- E5 end
3961 tttcttatto ttactcagtt taatgtgttc atatatthtgc ttattttttt tatactgttt
4021 attagttacc ttctgctggt attgctggac gaggagtctg ggggctctct aggacgcctg
4081 acaagtggtg gcagcggtta agcctccgtc ctaaaTAAag gATGgcaaac aaaagagtta
                                L2 orf start -> -> L2 cds
4141 agcgtgccaa tccctatgat ttatacagaa cctgcaagca agcaggaacc tgtccccctg
4201 atgtgattcc taaggtagaa ggagcacggt ttgcagataa aatacttcag tatgggagca
4261 tgggggtctt cctgggagga ctgggcatcg gtactggaca ggggaagcca gtatctggag
4321 ggtatgtacc attgcgacc tttgggtcca cctcttcttt ggccagcact gcaagcaggg
4381 gagccacagt aagtaaacct tttgcaggag ggattccact tgaaacctg gaaacaatag
4441 gtgcattctg gccttctgct gtcgaagaaa gctctctagg cacaggtgta gggcttagca
4501 atgaagcccc tgctatacta attcctgact ctgctcctag tggggatacc ttaggtggtc
4561 ttgatgttag cacagacagc tctgcagaca cgctgatcac tcttctggaa cctgagggac
4621 cggatgatat tgctgtacta gaggtgagac caacagagca tagcagagcc cacctctctt
4681 tcagcagtac tcaccctaac cccttatatc attacgtgga gcctgctgct ataggtgaaa
4741 cctcaacggt cgagaatgta ctggtagggt ggagcggcat aggggataat tacagtgaaa
4801 gtaTAGaact taacctgttt tcagagccaa aaacaagcac gcctgaacct gcatttaaga
L3 orf start ->
4861 gaccgaggct tcgtggaata ttaactatt tcagtcgag atactacact cagctgcca
4921 cttcagatcc tgatcctggc gctgctgag gctcttacgt ctttgagaat cccgtgtttg
4981 attcaaaggc gtttgaacct gagctgcccc cggataATGc acctgctgct gcaccaaag
                                L3 cds ->
5041 aagaagcccc tcatgtttct gcagctaaac tgttgcaggg tccctctgga agagTAGgca
                                <- L3 end
5101 taagcagaat agttaggctt acatccatgg gaaccagaag cggcgtcagg gtaggacctt
5161 tgtatcacct tagacactcc ttcagacta ttgcaccgtc agcagaatct gtggagctag
5221 tgccaacagt gctggaggaa ggggaagtga ttacaacagt agcagagtca catttgggctg
5281 acaatgcatt tgaggaagtt gacctagaca gtgttacaag tgacacgcca ctacttgaca
5341 cagagacct gtctcattct ggaaaaaaaa gcaacctccc tgctgcaagt ggtacgtccc
5401 ggctctcat agctgtggag gtggactctg gtcaatctga ggggttgcac cctgcacggg
5461 tacgtgacat aagcacaat catactcatt atgaccccga cactacctct acagtagtta
5521 ttgatggcaa catcgttatt tacagtacat acctaaagca ctattacctg catcccagcc
5581 tgtacaggcg caggcgcaaa cgcctcctTG AtTGAtttct ttacagATGg cgttctggca
                                L1 orf start ->                                L1 cds ->
                                <- L2 end
5641 agctggacag aagctgtacc taccaccgac acctgtgacc aaaacactgt gctctgaaac
5701 ctatgtaagc aggagagaca tatactatca tgcagaaaca gaacgggttc tcaactgtag

```

OvPV2

```
5761 acatcccttc tattctgtca aagtcaatga gaaaacagtc cctaaagttt ctccaaatca
5821 atatagagcc tttatgatac aactccccga tcctaaccag tttgcaactcc ctgataggac
5881 tttgtatgat ccaggtaaag aaaagctagt ctggggcctt attgggctac aggtctctcg
5941 tgggcagcca ttaggagggg ctatcagtgg ccacacaacc tttaatgctc tgctagatgc
6001 agaaaatgta tcacgccgca ccacagccca aggcactgat gataggaaac ctgctgggat
6061 ggatgctaaa cagcagcaag tgctgcttgt aggggtgcaca ccagctgttg gtgagtactg
6121 ggctaaagca agaccctgtg tatcagaccg tgtggacgct ggatcatgcc ctccactaga
6181 gcttaagcat aaacctatag aagacgggga tatgatggat atagggtttg gtgctgcaaa
6241 tttcaaagag ctaaacagtt cccgctctga tttaccttta gatatagtta atgaaatctg
6301 tctctatcct gattatthaa gaatgactga ggagactgct ggcaatagta tgtttttctt
6361 tgcaagaaag gaacaagtat acgtccgcca tgtatgggct aggggggggct cagataaaga
6421 acctccccct gaagactatt acttaaagcc taaaggcagt gaccagccta aaatagcagg
6481 cgtgtttatt ggtgcacctt caggctccct tttatctact gatggattaa tctttaatag
6541 gccatactgg ctataccgtg ctcagggtat gaacaatggg atatgctgga ataatactgt
6601 gtttgttact gtaggtgata atactcgtgg cactaacctc acaattactg tgtcctctga
6661 tgattcccca ctaaaggaat atgaaacagg taaaatcaat gtgtaccaca gacacgttga
6721 agaataaag ctagcattta tatttcagct ttgctctgtg cagttaaccct ctgaaacagt
6781 tagcagcttg caggggttaa tgcccagcat tttagaaaac tgggaaatca acgtgcaacc
6841 tcctacctcc tctatactgg aagacactta ccgctacata gagtctcctg ccactaaatg
6901 tgcagataat gtcctgcta gtaagccaga cccctatgag gggttgctgct tctggaaaat
6961 agatttaaaa gaaaagtctt ctcttgatat tgatcagttt cccctcggga gacggtttct
7021 agctcagcag ggcttaggat gcacagctgc agtgccgagc aggggtaccta aatctacctc
7081 tgacaaaagcc ccctcaaagc ggcgtagaag gggcaccTAA ggacactggt ttggcatatt
                                     <- L1 end
7141 aagcactttg ttaatgaata aaaactgttg atttaattgc aaatgtgtca tgtccgctgt
7201 ctacaccaca cccggtctgg tatgcacaaa atagccccgt cagcagctat aaacagcctg
7261 cagctggcgc cttaaataa aatccagacg ccttgacgcc agctgtggaa aaaagggctg
7321 tgtgcacagc accgtaagcg gtgtgaagaa ctcagcccac gcagaaccgt cttcgggtcca
7381 gacaaataga gtctcgtecc acgcgcaccg ggagcgggat cgtcggggtc ttgggattga
7441 agcaccggct gcggttcatt gaggtaaaaat tcagcttttt tctttgcgga aatcaatagg
7501 aaaagaata ggaagctaaa cacttagttt gttaaaggtc ctgggtcacg accgtctgcg
7561 attttgcttc aaccacagta ctttgcgca ttaggtttca ttaccgctca cggctactcg
7621 tgcttttcta aactgctccc ccattggttg gtgctgggaa ttactaactg ctgcttactc
7681 tgattgcttg aatcctctgc ggtgttttat tactctgcac agtaattttt ttctagagat
7741 gattgttgtt aacaaTAA
```

-> E6 orf start

LOCUS RPVL1R 528 bp ds-DNA VRL 2-JUN-1997

DEFINITION Reindeer papillomavirus genomic L1 region, partial cds.

ACCESSION M18175 U89670

SOURCE Reindeer papillomavirus (from epithelial layer of a single fibropapilloma) DNA.

ORGANISM Reindeer papillomavirus
Viridae; ds-DNA nonenveloped viruses; Papovaviridae; Polyomaviruses.

REFERENCE 1 (bases 1 to 264)

AUTHORS Moreno-Lopez,J., Ahola,H., Eriksson,A., Bergman,P. and Pettersson,U.

TITLE Reindeer papillomavirus transforming properties correlate with a highly conserved E5 region

JOURNAL J. Virol. 61, 3394-3400 (1987)

REFERENCE 2 (bases 1 to 444)

AUTHORS Chan,S.-Y., Bernard,H.U., Ratteree,M., Birkeback,T.A., Faras,A.J. and Ostrow,R.S.

TITLE Direct Submission

JOURNAL Submitted (14-FEB-1997) Institute of Molecular & Cell Biology, National University of Singapore 119260, Republic of Singapore

COMMENT The Reindeer Papillomavirus was isolated from a reindeer, specifically from the epithelial layer of a fibropapilloma. The isolate was cloned and a restriction map was determined. The isolated genome was unintegrated and circular, as indicated by restriction fragment patterns; total length was approximately 8.1 kb. Segments of E1, E5 and L1 were sequenced. The RPV sequences were most similar to the homologous segments of DPV and EEPV genomes, and more distantly related to BPV1. Syrian hamsters were inoculated with purified RPV and subsequently developed fibrosarcomas. Mouse C127 cells were transformed by RPV, although more slowly (2 to 4 weeks) than by BPV1 or EEPV (10 to 14 days). Transformed mouse cells produced several mRNA species in a pattern similar to that of EEPV-transformed cells. No antibody crossreactivity was detected between anti-EEPV nor anti-BPV1 serum, although there was reactivity with anti-RPV serum. A highly hydrophobic E5 protein of 44 amino acids is predicted; both the length and the sequence of the protein are highly conserved among those PVs which induce both fibromas and fibropapillomas, namely the group of PVs related to BPV1, including BPV2, BPV5, EEPV and DPV. There may be a distant relationship between the E5 of these viruses and an E5 ORF of HPV6b which has a predicted hydrophobic protein, although it has not been shown that this HPV6B ORF encodes a functional protein.

This sequence, a segment of L1, is a composite of two sequences, Accession Numbers M18175 and U89670. The former sequence was originally submitted to GenBank concatenated onto a fragment of the E1 gene. This E1/L1 chimera was assigned the accession number M18175. We have retained that accession number for both this sequence and for the sequence RPVE1 which comprises the E1 portion of the original E1/L1 submission. Sequence M18175 was 264 bp long and ended part way through the MY09-MY11 consensus primer region. Chan et al. [2] extended the sequence by primer walking to the end of the MY09-MY11 segment. They reported the sequence as Accession Number U89670. The two L1 sequences are here combined to make one of 528 bp length.

FEATURES Location/Qualifiers

source 1..528
/organism="reindeer papillomavirus"
/note="RPV"

Introduction

```
gene          <1..>528
              /gene="L1"
CDS           <1..>528
              /gene="L1"
              /note="contains MY09/11 segment"
              /codon_start=1
              /db_xref="PID:g2108203"
              /translation="PSVFFPVPSPGSLVSTDGQLFNRPYWLFRAQGMNNGICWNNQLFV
TVGDNTRGTTTLTITVPSPGGKKSPLTEYDTSKFNVYQRHVVEYKLAFFVQLCSVELTAD
TISHLQGLMPSILETWDINLQPPQSSILEDTYRYIESPATKCKDNVSPSPKPEDPYSGL
KFEVNLKGLSLDLD"

BASE COUNT   150 a    112 c    110 g    156 t
ORIGIN
      1 ccaagtgctt ttttcccagt acccagtgagg agccttgctt ctacggatgg tcagcttttc
L1 cds ->
      61 aatagacctt attggctatt tagagctcag gccatgaata atggatatatg ttggaataac
          start MY09/11 region ->
     121 cagctgtttg tcacagttgg ggacaacact cgtggtacca cactgacctt tactgtacca
     181 agtgggtggaa agaagtcccc cctcactgaa tatgacacaa gcaagtttaa tgtttatcag
     241 agacacgtag aagagtataa gcttgctttt gtatttcagc tttgctctgt ggagctcaca
     301 gcagatacca tctcacatct gcaaggggta atgccttcta ttttagaaac ttgggatatt
     361 aacctgcagc ctctcaatc atctattcta gaggatactt accggtatat agaatctcct
     421 gcaactaaat gcaaagacaa tgtatcccct agtaagccag aagaccctta ctcagggctt
     481 aaatthttggg aggtgaatth aaaaggaaaa ctgtcccttg atttggat
                                          L1 cds ->
                                          <- end MY09/11
                                          region

//
```

Revised HPV Sequences

In Part III of this compendium, John Meissner discusses the causes of and corrections to sequencing errors in reference HPV clones. In this section we republish annotated GenBank-style files of some of the revised (R) sequences described in his article: HPV-1aR, HPV-5R, HPV-6bR, and HPV-18R. The COMMENTS portion of each of these files tabulates the changes that have been performed and their consequences, if any, to the related amino acid sequences. Other revised sequences, RhPV-1R, RPVL1R, and FPVL1R, were presented earlier in this chapter.

Revised HPV1a

LOCUS HPV1aR 7816 bp ds-DNA circular VRL 06-JUL-1989
 DEFINITION Human papillomavirus 1a (HPV-1a), complete genome, revised.
 ACCESSION V01116 X03321
 KEYWORDS circular; genome; origin of replication.
 SOURCE Human papillomavirus type 1a DNA.
 ORGANISM Human papillomavirus type 1a
 Unclassified.
 REFERENCE 1 (bases 1 to 7811)
 AUTHORS Danos,O., Katinka,M. and Yaniv,M.
 TITLE Human papillomavirus 1a complete DNA sequence: a novel type of
 genome organization among papovaviridae
 JOURNAL EMBO J. 1, 231-236 (1982)
 STANDARD full staff_review
 REFERENCE 2 (bases 1 to 7816)
 AUTHORS Danos,O.
 TITLE Direct Submission
 JOURNAL Submitted (23-JAN-1985) Danos O.
 STANDARD full staff_review
 REFERENCE 3 (bases 7692, 7693 and sites; revision)
 AUTHORS Palermo-dilts,D., Broker,T. and Chow,L.
 TITLE Human papillomavirus type 1 produces redundant as well as
 polycistronic mRNAs in plantar warts
 JOURNAL J. Virol. 64, 3144-3149 (1990)
 STANDARD full staff_review
 COMMENT HPV-1a has a strong preferential association with deep plantar
 warts, characterized by a highly thickened corneal layer
 (hyperkeratosis). Generally, they are present in a single
 location. Skin warts are transmitted by direct contact with
 infected tissue or with contaminated objects. A majority of warts
 regress, spontaneously within two years. This is thought to be the
 result of a cell-mediated immune response. HPV-1 DNA has also been
 rarely detected in anogenital tissue.

Palermo-Dilts et al. (J. Virology 64, 3144-3149) report that the HPV1a mRNAs are highly analagous to those produced by HPV11 and HPV6 in splice site usage and major promoter locations. The exception to this similarity is the existence of a URR promoter in HPV1a, not found in either HPV11 or HPV6. This promoter, located at nt 7490, is active in plantar warts but not in primary keratinocytes. Thus, the HPV1a genome contains three putative promoters: one within the E7 ORF, one preceeding the E6 ORF, and one in the URR. The splice locations and their coding potentials are shown in the table below. All splice junctions shown annotated in the sequence are experimentally determined [3].

mRNA species potentials	Splice donor/acceptor pair	Coding
a	827/3200	E1^E4, E5a
b	1231/3200	E1M, E2C, E5a
c	827/2545	E2, E5a
e	827/3200	E1^E4, L1
f	3592/5432	
	7711/3200	L1
	3592/5432	
g	7711/5432	L1
h	7711/3200	E5a, E5, L2

SEQUENCE CORRECTIONS:

SITE	CHANGE	EFFECTS	SOURCE
1283	A->G	E1 aa:N->D	[1]
2301	T->C	E1 aa: I->T	[1,2]
3884	ins .->A (add base)	E5 frameshift reunites E5a and E5b orfs	[1]
4331	A->T	L2 (silent)	[1]
4375-	GAGGGAA->	L2 aa:RGN-> QGS,	[1]
4381	AGGGAAG	"loss" of BSM1 site	
7691-	CC->GG	"gain" of NlaIV site	[1,3]
7692			

A "correction" to the reference sequence which we made in 1994 needs to be "unmade." Base 7574 should be a C not a T.

7574	T->C	LCR	[1]
------	------	-----	-----

NOTES ON CORRECTIONS:

a) The "HPV1a" sequence currently available from GenBank corresponds to a "corrected" version of the original GenBank entry. The original seems to have been 7811 nt based on annotations on the GenBank entry. Corrections were submitted directly to GenBank by O. Danos, apparently with no associated publication; We have no documentation on what was corrected here, although Meissner [1] reports that it included a mistaken "correction" to the region 4374-4382.

b) A sequence which had been further corrected according to Ref. [2], with changes at nt. 7574 and 7691-7692 was published in "Human Papillomaviruses 1994." However, according to Meissner [1], the change at 7574 is present in a variant clone, but not the reference clone, and thus this "correction" is wrong. That is, for this site, it appears that the reference clone truly has "C".

REFERENCES:

1. Meissner, John, Virus Genes 9(2):189-191, 1994.
2. Clad, A et al, Virology 118:254-259, 1982. (J02208)
3. Palermo-Dilts, D.A. et al, J Virol 64:3144-3149, 1990.

FEATURES	Location/Qualifiers
5'UTR	join(6941..7816,1..44) /function="regulatory region" /standard_name="LCR" /note="putative"
TATA_signal	68..74 /note="putative"
CDS	104..526 /note="putative" /note="ORF E6 from bp 86 to 526" /product="transforming protein" /gene="E6" /note="putative" /codon_start=1 /translation="MATPIRTVTRQLSESLCIPYIDVLLPCNFCNYFLSNAEKLLFDHF DLHLVWRDNLVFGCCQGCARTVSLLEFVLYYQESYEVPEIEEILDRPLLQIELRCVTC IKKLSVAEKLEVVSNGERVHRVRNRLKAKCSLCRLYAI "
CDS	165..356 /note="putative" /note="ORF E8 from bp 45 to 356" /gene="E8"

Revised HPV1a

```

/note="putative"
/codon_start=1
/translation="MFYCLVIFVIFCLMLRSCFLIILICILSGETIWCLDAVKGVLE
LLAYWSLFYIIRSLMRYRK"
TATA_signal 520..526
/note="putative"
CDS 529..810
/note="putative"
/note="ORF E7 from bp 502 to 810"
/product="transforming protein"
/gene="E7"
/note="putative"
/codon_start=1
/translation="MVGEMPALKDLVLQLEPSVLDLDLYCYEEVPPDDIEEELVSPQQ
PYAVVASCAYCEKLVRLTVLADHSAIRQLEELLRLSLNIVCPLCTLQRQ"
CDS 812..2650
/note="putative"
/note="ORF E1 from bp 773 to 2650"
/product="replication protein"
/gene="E1"
/note="putative"
/codon_start=1
/translation="MADNKGTEWDFLVEATDCEETLEETS LGDLDNVSCVSDLSDLL
DEAPQSQGNSLELFHKQESLESEQLNALKRKLLYSPQARSADETDIASISPRLETIS
ITKQDKKRYRRQLFSQDDSGLELSLLQDETENIDESTQVDQQKEHTGEVGAAGVDIL
KASNIRAALLSRFKDTAGVSFTDLTRS YKSNKTC CGDWVLAVWGVREN LIDS VKELLQ
THCVYIQLEHAVTEKNRFLFLVRFKAQKSRETVIKLITLIPVDASYILSEPPKSRS
VAAALFWYKRSMSTVFTWGTLEWIAQQTLINHQLDSESPFELCKMVQWAYDNGHTE
ECKIAYYYAVLADEENARAFLLSSNSQAKYVKDCAQMRHYLRAEMAQMSMSEWIFRK
LDNVEGSGNWKEIVRFLRFQEVFISFMIAFKDLLCGKPKKNCLLIFGPPNTGKSMFC
TSLKLLGGKVISYCNSKSQFWLQPLADAKIGLLDDATKPCWDYMDTYMRNALDGNTI
CIDLKHRAPQQIKCPPLLITSNIDVKS DTCWMYLHSRISAFKFAHEFPFKDNGDPGFS
LTDENWKSFFERFWQQLELSDQEDEGNDGKPPQQLRLTARAANEPI"
polyA_signal 821..826
/note="putative"
misc_feature 827..831
/standard_name="Splice donor"
protein_bind 887..898
/function="gene transcription"
/bound_moiety="E2"
/note="putative"
misc_feature 1228..1235
/standard_name="Splice donor"
protein_bind 1359..1370
/function="gene transcription"
/bound_moiety="E2"
/note="putative"
TATA_signal 1661..1666
/note="putative"
misc_feature 2541..2548
/standard_name="Splice acceptor"
CDS 2592..3797
/note="putative"
/note="ORF E2 from bp 2568 to 3797"
/product="regulatory protein"
/gene="E2"
/note="putative"
/codon_start=1
/translation="MENLSSRLDLLQEQLMNLVEQDSKLI EDQIKQWNLRQEQVLFH
```

```

FARKNGVMRIGLQAVPSLASSQEAKAKTAIEMVLHLESKDSPTYGTEDWSLQDTSRELF
LAPPAGTFKKSGSTLEVITYDNNPDNQTRHTIWNHVYYQNGDDVWRKVSSGVDVAVGVYY
LEHDGYKNYYVLF AE EASKYSTTGQYAVNRYGKRFTNVMSSTSSPRAAGAPAVHSDYP
TLSESDTAQQSTS IDYTELPGQGETSQVRQRQKTPVRRRPPYGRRRSRSPRGGGRREG
ESTPSRTPGSVPSARDVGSIH TTPQKGHSSRLRLLQEAWDPVVCVKGGANQLKCLR
YRLKASTQVDFDSISTTWHWTD RKNTERIGSARMLVKFIDEAQREKFLERVALPRSVS
VFLGQFNGS"
TATA_signal 3087..3092
              /note="putative"
CDS          3181..3561
              /note="putative"
              /note="ORF E4 from bp 3157 to 3561"
              /gene="E4"
              /note="putative"
              /codon_start=1
              /translation="MLCLPLAPQGLLGLLQYTPPTQPYPRVTPPSNRRPSTTPNSQDR
GRPRRSDKDSRKHLYADGLTDGEDPEVPEVEDEEKENQRPLGHPDLSLLRETLLEVYTO
RLKRDILQDLDDFCRKLGIHPWSV"
misc_feature 3196..3203
              /standard_name="Splice acceptor"
CDS          3263..3433
              /note="putative"
              /note="ORF E3 from bp 3263 to 3433"
              /gene="E3"
              /note="putative"
              /codon_start=1
              /translation="HRPAIDVHRLHRTPTGGDLAGPTKTAENTCTQTALRTAKIQKS
QRWRTKRRRINAL"
misc_feature 3589..3606
              /standard_name="Splice donor"
TATA_signal 3844..3849
              /note="putative"
CDS          3803..4156
              /note="putative"
              /note="ORF E5 from bp 3788 to 4156"
              /gene="E5"
              /note="putative"
              /codon_start=1
              /translation="MEVDFAWTCVHSPCIYSPPTPTYLEACKHCNKCIAYVENALPPKI
YTPHAKYQTPAHLTFKIKLSIQQLLIKVCNMAVWEFFWEVWALEQPEALEEELVILPSV
RVVGLELLLVQLQ"
CDS          3898..5421
              /note="putative"
              /note="ORF L2 from bp 3895 to 5421"
              /product="minor capsid protein"
              /gene="L2"
              /note="putative"
              /codon_start=1
              /translation="MYRLRRKRAAPKDIYPSCKISNTCPPDIQNKIEHTTIADKILQY
GSLGVFLGGLGIGTARGSGGRIGYTPLGEGGGVRVATRPTPVRPTIPVETVGPSEIFP
IDVVDPTGPAVILPQLGRDFPIPTVQVIAEIHPI SDIPNIVASSTNEGESAILDVLQ
GSATIRTVSRTQYNNPSFTVASTSNISAGEASTSDIVFVSNSSGDRVVGEDIPLVELN
LGLETDTSSVVQETAFSSSTPIAERPSFRPSRFYNRRLYEQVQVQDPRFVEQPQSMVT
FDNPAFEPPELDEVSIIFQRDLDALAQTPVPEFRDVVYLSKPTFSREPGGRLRVSRLGK
SSTIRTRLGTAIGARTHFFYDLSSIAPEDSIELLPLGEHSQTTVISSNLGDTAFIQGE
TAEDDLEVISLETPQLYSEEELDTNESVGENLQLTITNSEGEVSI LIDLTSRVRPPF
GTEDTSLHVYYPNSSKGTPIINPEESFTPLVI IALNNSTGDFELHPSLRKRKRAYV"
TATA_signal 3936..3941
              /note="putative"

```

Revised HPV1a

polyA_signal 3985..3990
 /note="putative"
 CAAT_signal 4612..4616
 /note="putative"
 TATA_signal 4651..4659
 /note="putative"
 CDS 5414..6940
 /note="putative"
 /note="ORF L1 from bp 5348 to 6940"
 /product="major capsid protein"
 /gene="L1"
 /note="putative"
 /codon_start=1
 /translation="MYNVFQMAVWLP AQNKFYLP P QPITRILSTDEYVTRTNLFYHAT
 SERLLL VGHP LFEISSNQTVTIPK VSPNAFRVFRVRFADPNRFAFGDKAIFNPETERL
 VWGLRGIEIGRQPLGIGITGHPLLNKLD AENPTNYINTHANGDSRQNTAFDAKQTQ
 MFLVGCTPASGEHWTSSRCPGEQVKLGDCPRVQMIESVIEDGDMMDIGFGAMDF AALQ
 QDKSDVPLDVVQATCKYPDYIRMNHEAYGNSMFF FARREQMYTRHFFTRGG SVGDKEA
 VPQSLYL TADAEPRTTLATTNYVGT P SGSMVSSDVQLFNRSYWLQRCQGNNGICWRN
 QLFITVGDNTRGTSLSISMKNNA STYSNANFNDFLRHTEEFDL SFIVQLCKVKLTPE
 NLAYIHTMDPNILEDWQLSVSQPPTNPLEDQYRFLGSSSLAAKCEQAPPEPQ TDPYSQ
 YKFW EVDLTERMSEQLDQFPLGRKFLYQSGMTQRTATSSTTKRKT V RVSTSAKR RKA
 "

misc_feature 5428..5435
 /standard_name="Splice acceptor"
 TATA_signal 6764..6769
 /note="putative"
 TATA_signal 6941..6946
 /note="putative"
 TATA_signal 6948..6953
 /note="putative"
 TATA_signal 6950..6955
 /note="putative"
 TATA_signal 6982..6987
 /note="putative"
 polyA_signal 7381..7386
 /note="putative"
 polyA_signal 7427..7432
 /note="putative"
 TATA_signal 7488..7493
 /note="putative"
 protein_bind 7509..7520
 /function="gene transcription"
 /bound_moiety="E2"
 /note="putative"
 repeat_region 7598..7627
 /rpt_type=Direct
 /rpt_unit=7598..7612, 7613..7627
 protein_bind 7702..7713
 /function="gene transcription"
 /bound_moiety="E2"
 /note="putative"
 misc_feature 7708..7715
 /standard_name="Splice donor"
 protein_bind 7776..7787
 /function="gene transcription"
 /bound_moiety="E2"
 /note="putative"
 source 1..7816

```

/organism="Human papillomavirus type 1a"
/sequenced_mol="DNA"
BASE COUNT      2395 a   1481 c   1669 g   2271 t
ORIGIN
   1 gttaactacc atcattcatt attctagtta caacaagaac cTAGgagtta tatgccagaa
                                     E8 orf start ->
   61 gtaagccTAT AAAAtacaca ggTAAgactc tgcacaggac cagATGgcga caccaatccg
      signal ->      E6 orf start ->      E6 cds ->
  121 gaccgtcaga cagctttccg aaagcctctg tatcccatat attgATGttt tattgccttg
                                     E8 cds ->
  181 taatttttgt aattattttt tgtctaagtc tgagaagctg ctttttgatc attttgattt
  241 gcatcttgtc tggagagaca atttggtggt tggatgctgt caaggggtgtg ctagaactgt
  301 tagcctattg gagtttgttt tatattatca ggagtcttat gaggtaccgg aaaTAGaaga
                                               <- E8 end
  361 aattttggac agacctttat tgcaaattga actccgttgt gttacatgca taaaaaaact
  421 gagtgttgct gaaaaattgg aggttgtgtc aaacggagaa agagtgcata gagttagaaa
  481 cagacttaaa gcaaagtgTA Gtttgtgtcg cttgtatgcT ATATAAcaAT Ggtgggcgaa
      E7 orf start ->      E7 cds ->
                                               <- E6 end
                                     signal ->
  541 atgccagcac taaaggacct ggttcttcaa ctgaaaccaa gcgtcctaga tttagatctt
  601 tattgttacg aggaggtgcc tcctgatgac atagaggagg agttagtgtc gcctcagcaa
  661 ccttatgctg tcgttgcttc ctgtgcctat tgcgagaaac tggttcgatt gaccgtcctc
  721 gcggatcaca gcgccattag acagctggag gaactccttc tgcgatcttT GAacatcgtg
                                               E1 orf start ->
  781 tgcccactgt gcaccctaca gcgacagTAA aATGgcagat AATAAAGGTa ctgaaaacga
      E1 cds ->      5 sj /\
      <- E7 end      -> signal
  841 ttggtttttg gtggaggcga cagattgtga ggaaacgtta gaggaaACCT CACTTGGTga
                                               -> E2 bind
  901 cctagataat gtttcttggt ttagcgactt atctgattta ttagacgagg cgccgcaaag
  961 ccagggaat tcctggaat tgttccacaa gcaagaatcg ctggaaagcg aacaggaact
 1021 taatgtttaa aaacgaaagt tactttacag tcctcaggcg agaagcgcgg acgaaacaga
 1081 cattgctagc attagtcccta gattagaaac tatttctatt acaaagcaag acaaaaaaag
 1141 gtatcgaagg caactgtttt ctcaggatga tagtggttta gagctatcgc tgcttcagga
 1201 tgaactgaa aatattgatg aatcgacaca gGTagatcaa cagcagaaag aacatactgg
      5 sj /\
 1261 ggaagtggg gccgctggg tggacatttt gaaagctagt aatatccgcg cgcattatt
 1321 aagcagattt aaagatacgg ctggcgtcag ttttacagAC CTGACGCGGT cgtacaagag
      -> E2 bind
 1381 caacaaaacc tgttgtggag attgggtttt ggcagtttgg ggtgtccgtg aaaatttaat
 1441 tgacagtgta aaagaattat tgcaaaccca ttgtgtgtat attcaattgg aacatgcagt
 1501 aactgaaaaa aatagatttt tatttttatt ggtacgattt aaagcccaga aaagtagaga
 1561 gactgtgata aaactataaa ccacaatttc tccagttgat gctagctata ttttgtctga
 1621 gcctccaaaa tcaagaagtg tggctgctgc attattttgg TATAAAagat ctatgtcttc
      signal ->
 1681 aactgttttt acatggggta caactttgga gtggattgca cagcaaacc ttattaatca
 1741 tcagttagat tccgaaagtc ctttgagct ttgtaaaatg gttcagtggt cctatgataa
 1801 tggacataca gaagagtgta aaattgcata ttattatgct gttttagcag atgaggatga
 1861 aaatgcaagg gcatttctaa gctctaattc acaggcaaaa tatgtgaaag actgtgcaca
 1921 aatggttaaga cactatttac gtgctgagat gcacaaaatg tctatgtcag agtggatttt
 1981 tagaaaacta gataatgtag aaggttctgg taattggaaa gaaattgtaa gatttttaag
 2041 atttcaagaa gttgaattta taagctttat gattgcattt aaagatttgt tatgtggtaa
 2101 gccaaagaaa aactgtttgt taatatttgg acctccaaat acaggaaaat caatgttttg
 2161 tacaagttaa ttaaagtgtg taggagggaa agtgatttca tactgtaaca gtaaaagtca
 2221 gttttggttg cagcctctgg ctgatgctaa gatagggcta ttagatgatg caacaagacc
 2281 atgttgggat tatatggaca cttatatgag aaatgcattg gatggtaaca ctatttgtat
 2341 tgatttaaaa catagagctc ctcaacaaat taaatgccca cctttactta ttactagtaa
 2401 tattgatggt aatcagata cctgttggat gtatttgcac agtagaatat cagcttttaa

```

Revised HPV1a

```

2461 atttgctcat gagtttccat ttaaagacaa tggatgatcca ggattttcct taacagacga
2521 aaattggaaa tctttctttg aaAGgttttg gcaacagtta gaatTAAgtg accaagaaga
      /\ 3 sj      E2 orf start ->
2581 cgagggaaac gATGgaaaac ctcagcagtc gcttagactt actgcaagag cagctaataga
      E2 cds ->
2641 acctataTGA acaggacagt aaattgatag aagatcaaat taagcagtgg aatctaatta
      <- E1 end
2701 gacaagaaca agttcttttc catttcgcca gaaaaaatgg ggtaatgaga attggattgc
2761 aggcagttcc atcttttagcg tcctcacagg agaaggcaaa gacagctatt gaaatgggtgt
2821 tacatttaga gtctttaaag gactcacctt atggcacaga ggattgggtca cttcaagaca
2881 ctagcagaga gctgtttttg gcacccccag ctggcacctt caagaagagt ggcagcacac
2941 ttgaggttac ctatgacaat aaccctgata atcagacaag gcacacaatt tggaatcatg
3001 tgtattatca aaatggggac gatgtatgga gaaaagtatc cagtgggtgt gatgctgtag
3061 gagtgtacta tttagaacac gatggcTATA AAaattatta tgtgttattt gctgaggagg
      signal ->
3121 cctctaagta cagcacaaca ggacaatatg ctgTAAatta caggggtaaa aggtttacaa
      E4 orf start ->
3181 ATGttatgtc ttccactAGc tcccgaaggg ctgctggggc tcctgcagta cactccgact
E4 cds ->      /\ 3 sj
3241 acccaaccct atccgagagT GAcaccgccc agcaatcgac gtccatcgac tacaccgaac
      E3 orf start ->
3301 tcccaggaca gggggagacc tcgcaggtcc gacaaagaca gcagaaaaca cctgtacgca
3361 gacggcctta cggacggcga agatccagaa gtcccagagg tggaggacga agagaaggag
3421 aatcaacgcc ctctAGgaca cccggatctg tcccttctgc gcgagacgtt ggaagtatac
      <- E3 orf end
3481 acacaacgcc tcaaaagga cattcttcaa gacttagacg acttctgcag gaagcttggg
3541 atccaccgct ggtctgtgTA Aaagggggtg ccaatcagct taagtgtctc agGTacagac
      <- E4 end      5 sj /\
3601 ttaaagcatc tactcaagtt gactttgaca gcataagcac cacatggcat tggacagata
3661 gaaaaaacac cgagaggata ggtagtgcta gaatgttagt aaagtttatt gatgaggctc
3721 aacgagagaa gtttcttgag agagtgtctt tgcccagatc agtgtctgtg tttttgggac
3781 agttTAAatg gtctTAAaat taATGgaagt tgattttgct tggacgtgtg tacatagtcc
E5 orf start ->      <- E2 end
      E5 cds ->
3841 ctgTATATAT tcccctccta cccccacata cttgaagct tgcaaacatt gTAAcaaATG
      signal ->      L2 orf start ->
      L2 cds ->
3901 tatgccttac gtagaaaacg cgctgcccc aaagaTATAT Acccctcatg caaaatatca
      signal ->
3961 aacacctgcc cacctgacat tcaaAATAAA attgagcata caacaattgc tgataaaata
      signal ->
4021 ttgcaatatg gcagtctggg agtttttttg ggaggtttg gcattggaac agccagaggc
4081 tctggaggaa gaattgggta tactcccctc ggtgagggtg gtggggttag agttgctact
4141 cgtccaactc cagTAAggcc tacaatacct gtggaaacag taggccccag tgaattttc
      <- E5 orf end
4201 cccatagatg ttgtagatcc tacaggcctt gctgttattc ccctacaaga tttaggtaga
4261 gacttcccaa taccaactgt gcaggttatt gcagaaattc accctatttc tgacatacca
4321 aacattgttg cttcttcaac aaatgaagga gaatctgcca tattagatgt gttacagggg
4381 agtgcaacca tacgactgt ttcaagaaca caatacaata acccctcttt cactgttgca
4441 tctacatcta atataagtgc tggagaagca tcaacatcag atattgtatt tgttagcaat
4501 ggttcaggtg acaggggtgg ggcgaggat atccccttgg tagaattaa cttaggcctt
4561 gaaacagaca catcttctgt tgtacaagaa acagcatttt ccagcagcac aCCAATtgct
      signal ->
4621 gaaagaccct cttttaggcc ctcaagattc TATAATAGgc gtctatatga acaggtgcaa
      signal ->
4681 gtacaagacc ctaggttcgt tgagcagcca cagtcaatgg tcaacttttg taatccagca
4741 tttgagccag agcttgatga ggtgtctatt atcttccaaa gagacttaga tgctctgtct
4801 cagacaccag tgctgaatt tagagatgta gtttatctga gcaagcccac attttcgagg
4861 gaaccagggg gacgggtaag ggtagccgc cttggcaaaa gttcaactat tcgtacacgc

```

```

4921 ctggggcacag caattggcgc cagaaccac tttttctatg atttaagttc tattgctcca
4981 gaagactcaa ttgaattatt gccttttaggt gagcatagtc aaacaacagt ctaggttcc
5041 aacttaggtg acacagcatt tatacaaggt gagacagcag aggatgactt agaagttatc
5101 tctttagaaa caccacaatt atattcagaa gaagagcttt tagacacaaa cgaagtggtg
5161 ggcgaaaatt tgcaacttac tattactaac tcagaggggtg aggttttctat actagattta
5221 acacaaagca gagtcaggcc accttttggc actgaagata ctagcttgca tgtatattac
5281 ccaaattctt ctaaagggac tccaataatt aatcctgaag aatcatttac acctttgggt
5341 attatAGctc ttaacaactc aacaggggat tttgagttac atcctagtct tagaaagcgt
L1 orf start ->
5401 cgtaaaagag cttATGtaTA AtgtttttcA Gatggctgtc tggttaccag cgcagaataa
      L1 cds ->      <- L2 end      /\ 3 sj
5461 gttctatctt cctccccagc ccatcactag aatcctgtcc actgatgaat atgtaaccag
5521 aaccaatctc ttctaccatg caacatctga acgtctactg ctggtcggac atcctttggt
5581 tgagatctcc agtaatcaaa ctgtaactat accaaaagtg tcaccaaagc catttagagt
5641 ttttaggggtg cgttttgctg atccaaatag atttgcattt ggggataagg caatttttaa
5701 tccagaaaca gaaagattag tttggggcct aagagggata gagataggta gaggccagcc
5761 tttagggtata ggaataacgg gccaccctct tttaaataag ttagatgatg cagaaaatcc
5821 aacaaattat attaatactc atgcaaatgg agattctaga caaaatactg cttttgatgc
5881 aaaacagaca caaatgttcc tcgtcggctg tactcctgct tcagggtgaac actggacaag
5941 tagtcgttgc ccaggggaaac aagtgaaact tggggactgc cccaggggtg aaatgataga
6001 gtctgtcata gaagatgggt acatgatgga tattggtttt ggggctatgg attttgctgc
6061 tttacagcaa gacaagtctg atgtcccttt agatgtttgt caagcaacat gcaaatatcc
6121 tgattatata agaatgaacc atgaagccta tggcaactct atgttttttt ttgcacgtcg
6181 cgagcaaatg tataccaggc acttttttac tcgcggggggt tcggtgggtg ataaggaggc
6241 agtcccacaa agcctgtatt taacagcaga tgctgaacca agaacaactt tagcaacaac
6301 aaattatgta ggcacaccaa gtggctctat ggtttcatct gatgtccaat tgtttaatag
6361 atcttactgg cttcagcgat gtcaaggcca gaataatggc atttgctgga gaaaccagtt
6421 atttattaca gttggagata ataccagagg aacaagttta tctatcagta tgaaaaacaa
6481 tgcaagtact acatattcca atgctaattt taatgatttt ctaagacata ctgaagaatt
6541 tgatctttct tttatagttc agctttgtaa agtaaagtta actcccgaag atctagccta
6601 cattcatata atggacccta atattttaga ggattggcaa ctatctgtat ctcaaccacc
6661 taccaatcct ctagaagatc aatataggtt tttagggctc tccttggcag caaaatgtcc
6721 agaacagggcg cctcctgagc cccagactga tccttatagt caaTATAAAT tctgggaagt
      signal ->
6781 cgatctcaca gaaaggatgt ccgaacaatt agaccaatth ccactaggaa ggaaatthct
6841 atatcaaagt ggcatgacac aacgtactgc tactagttcc accacaaagc gcaaaacagt
6901 gcgtgtatct acgtcagcca agcgcagggcg taaggctTAG TATATAtTAT ATATAactat
      <- L1 end
      signal ->
6961 atttattagt agattattta tTATATAttt ttatattttt atactttttta tacttgttta
      signal ->
7021 gttctaaata gacatgtaag atttacatta gtataagtag gcatgtattt acataaaata
7081 gtcttgaaa ccttttatta gtgaaccatc atttacaata gtgacatcat agttcatctg
7141 caattgctat tccatcgttc ttcacatatt ctacagtagt gttctctaga ttgtattgct
7201 attttctctg taggcaaaca acaacatctg tacatggacc aaacaacca ctttcatttt
7261 attgtgctgc atataattoca gattgttgag gatttatttg tttagactcc ggtgcattat
7321 acacaagtgt gcattttttg tgttctctga ttgattgtgt gttattttcc tgcaaatatgc
7381 AATAAAagtg agctgtcctt tctttttggt aatccctccc tactccAATA AAaaatccct
signal ->      signal ->
7441 acccctaaaa tctgtttggt ctggttttat taataattgc gctctttTAT ATAAaagta
      signal ->
      |-> mRNA start site
      from P(7490)
      promoter
7501 ctattaacAC CGCACCCGTT gtggctaadc ccttatggta tttaaaagac tacacctaca
      -> E2 bind
7561 ggatgtattg tcttcattgt ttatggttta ccgctgctCCA AAGACGGTTT GCCCAAAGAC
      -> repeat region start

```


Revised HPV1a

```
7621 GGTTCGcCaa ccgcggttag gacttgtttc aatttgctgc caaacttata tggtcgtgct
      <- repeat region end
7681 ccaacggggtt tggcgccaag cacctaaaac gTAggtgtg tactcttttc aagaattaac
      5 sj /\
      -> E2 bind
7741 aaaggagatt tctcccgcca aattagtttc gagcgACCGA ATTCGGTcgt aaaaatctaa
      -> E2 bind
7801 agtgatgatt gttggt
```

LOCUS HPV5R 7746 bp ds-DNA VRL 30-SEP-1988
 DEFINITION Human papillomavirus type 5 (HPV-5), complete genome.
 ACCESSION M17463
 KEYWORDS complete genome.
 SOURCE Human papillomavirus type 5 DNA recovered from a benign flat
 wart from an EV patient.
 ORGANISM Human papillomavirus
 Viridae; ds-DNA nonenveloped viruses; Papovaviridae;
 Papillomavirus.
 REFERENCE 1 (bases 1 to 7746)
 AUTHORS Zachow,K.R., Ostrow,R.S. and Faras,A.J.
 TITLE Nucleotide sequence and genome organization of human papillomavirus
 type 5
 JOURNAL Virology 158, 251-254 (1987)
 STANDARD full staff_review
 COMMENT Draft entry and printed copy of sequence for [1] kindly provided by
 R.S.Ostrow, 10/23/87.

HPV-5 has been associated with macular lesions which frequently progress to malignancy. Yabe et al. (Int J Cancer 43: 1022-8) studied the characteristics of HPV-5 in lesions of differing severity. In a primary carcinoma, HPV-5 was present in an episomal state with a 40% subgenomic segment amplified. In the metastatic tumor, only the 40% subgenomic region was present, but integrated into the host genome. The segment was determined to be the entire sequences of E6, E7, and the noncoding region and portions of E1 and L1, with no mutations present (Yabe et al. Virology 183: 793-8). In addition, amplifications of the LCR have been reported in HPV-5 associated carcinomas (Deau et al. Virology 184: 492-503). HPV-5 and HPV-8 have also been found in significant numbers in squamous cell carcinomas of renal allograft patients. Barr et al. (Lancet 1: 124-9) detected either HPV-5 or HPV-8 in nearly 60% of the cases surveyed in the Scotland area. HPV-5 is considered to be part of the a1 cluster based on phylogenetic analysis. This cluster includes HPV-5, HPV-8, HPV-47, and HPV-12. Patients with EV tend to have depressed cell mediated immunity. In roughly one-third of EV-associated HPV infection, the sun-exposed flat wart-like or macular lesions transform into malignant squamous cell carcinomas. Benign wart scrapings tend to be multiply-infected, with as many as six different viral types. However, in contrast, EV carcinomas tend to harbor only a few types, specifically HPV-5 and HPV-8, and less frequently HPV-14, HPV-17, HPV-20 and HPV-47. These types are rarely detected in lesions afflicting the general population. A key to host restriction of these viruses may be in part due to the unusual organization of the LCR in these viruses. The LCR of these viruses is short compared to the viruses in other groups and contains two EV-specific regulatory regions: M33 and M29, both shown to be involved in protein binding.

SEQUENCE CORRECTIONS: Four corrections to the original GenBank sequence have been made to the present sequence.

POSITION	CHANGE	EFFECT	REFERENCE
4055-4058	AATG -> TGCT	NV -> CF in E2 aa	[1]
6175	C -> G	P -> A in L1 aa	[2]
6265	C -> G	R -> G in L1 aa	[2]
6502	C -> G	P -> E in L1 aa	[2]

[1] Deau & Favre, J Clin Micro, Nov 93, p2918.

[2] Kawase, Virol, 221 p189-198

Revised HPV5

```

FEATURES             Location/Qualifiers
    protein_bind     join(7741..7746,1..6)
                    /function="gene regulation"
                    /bound_moiety="E2 protein"
                    /note="putative"
    polyA_signal     146..151
                    /note="putative"
    TATA_signal      152..158
                    /note="putative"
    TATA_signal      154..160
                    /note="putative"
    TATA_signal      156..162
                    /note="putative"
    TATA_signal      158..164
                    /note="putative"
    TATA_signal      160..166
                    /note="putative"
    CDS              200..673
                    /note="ORF E6 from bp 167 to 673"
                    /product="transforming protein"
                    /gene="E6"
                    /note="putative"
                    /codon_start=1
                    /translation="MAEGAEHQKLTTEKDKAELPLSIRDLAELGIPVIDCLIPCNFC
GNFLNYLEACEFDYKRLSLIWKDYCVFACCRVCCGATATYEFNQFYEQTVLGRDIELA
SGLSIFDIDIRCQTCLAFLDIIIEKLDCCGRGLPFHKVRNAWKICRQCKHFYHDW"
    CDS              663..974
                    /note="ORF E7 from bp 618 to 974"
                    /product="transforming protein"
                    /gene="E7"
                    /note="putative"
                    /codon_start=1
                    /translation="MIGKEVTVQDIIILELSEVQPEVLPVDLFCHEELPNEQETEEEPD
NERISYKVIAPCGCRNCEVKLRIFVHATEFGIRAFQQLLTGDLQLLCPDCRGNCKHDS
S"
    CDS              961..2781
                    /note="ORF E1 from bp 913 to 2781"
                    /product="replication protein"
                    /gene="E1"
                    /note="putative"
                    /codon_start=1
                    /translation="MTDPNSKGSTSKEGFGDWCLLEADCSDVENDLGQLFERDSDSI
SDLLDDTELEQNSLELHFHQCEQSEEQQLKLRKYLSPKAVAQLSPRLESISLSPQ
QSKRRLFAEQDSGLELTLNNEAEDVTPVEVEVPAIDSRPDDEGGSGDVDIHYTALLRS
SNKKATLMAKFKESFGVGFNELTRQFKSHKTCCKDWVVSVYAVHDDLFESSKQLLQQH
CDYIWRGIGAMSLYLLCFKAGKNRGTVHKLITSMNLVHEQQILSEPPKLRNTAAALF
WYKGCMSGAFSHGYPDWIAQQTILGHKSAEASTFDFSAMVQWAFHNHLLDEADIAY
QYARLAPEDANAVAWLAHNNQAKFVRECAYMVRFYKKGQMRDMSISEWIYTKINEVEG
EGHWSDIVKFIYQINIFIVFLTALKEFLHSVPKKNCILYGPNSGKSSFAMSLIRV
LKGRVLSFVNSKSQFWLQPLSECKIALLLDDVTDPCWIYMDTYLRNGLDGHYVSLDCKY
RAPTQMKFPPLLLTSNINVHGETNYRYLHTTIKGFEPNPFPMKADNTPQFELTDQSW
KSFFTRLWTQLDLSDQEEEGEDGESQRAFAQCSARSANEHL"
    protein_bind     1658..1669
                    /function="gene transcription"
                    /bound_moiety="E2"
                    /note="putative"
    CDS              2723..4267
                    /note="ORF E2 from bp 2699 to 4267"
                    /product="regulatory protein"

```

```

/gene="E2"
/feature="ORF"
/feature_start=3285
/feature_end=4022
/translation="MENLSERFNALQDQLMNIYEAAEQTLQAQIKHWQTLRKEPVLly
YAREKGVTRLGYQPVPVKAVSETKAKEAIAMVLQLESLQTSDFAHWPWTLVDTSIETF
RSAPEGHFKKGPLVEVIYDNDPDNANLYTMWTVYVYMDADDKWHKARSGVNHIGIYY
LQGTfKNYVLFADDakRYGTTGWEVVKVNETVVFAPVTSSTPPGSPGGQADTNTTPA
TPTTSTAVDSTSRQLTTSKQPQQTETRGRRYGRRPSSKSRSSQTQQRSSRSRHSRS
RSRRSKsQTHtTRSTTRSRSTSLTKTRALTSRSRSRGRSPPTCRRGGGRSPRRRSRS
PSTSSCTTQRSQRARAESSTTRGARGSRGSRGGSRGRRRRGRSSSSSSPAHKRSR
GGSakLRGVSPGEVGGSLRSVSSKHTGRLGRLLLEARDPPVIIVKGAANTLKCFNRa
KIKYmGLFRSFSTTWSWVAGDGTERTLGRPRMLISFSSYTQRDFDEAVRYPKGVdKAY
GNLDSL"
CDS
<3285..4022
/feature="ORF E4 from bp 3285 to 4022"
/gene="E4"
/feature="ORF"
/feature_start=3406
/feature_end=3912
/translation="KLIRKLCLLLSPAPRLQGRQEDKQTQTPPPRPPPPPPQPLTPRP
DSSPHQNSHNKPKPEEEGTGGPPASQGRDRKRSKGDQGPDTGPGLGPRGSPKPTPL
GPPPGPGRRRSPRLGPLQADRDPEEGPQPPAEGEVEGHGPGDQGHPPPPPPAPHNGHS
GHEPKVQQPEGPEGREGHEEGAVGEGGDEEGHPPPPPPPTNGHEGGLSSVASLLVK
WEGHFDQLVQSIQDDLEDYWKKLATPQ"
CDS
<3406..3912
/feature="ORF E5 from bp 3406 to 3912"
/gene="E5"
/feature="ORF"
/feature_start=3537
/feature_end=3548
/function="gene regulation"
/bound_moiety="E2 protein"
/feature="ORF"
CDS
4348..5904
/feature="ORF L2 from bp 4240 to 5904"
/product="minor capsid protein"
/gene="L2"
/feature="ORF"
/feature_start=4438
/feature_end=4443
/polyA_signal
4438..4443
/feature="ORF"
protein_bind
4594..4605
/function="gene transcription"
/bound_moiety="E2"
/feature="ORF"
protein_bind
5199..5210

```

Revised HPV5

```

/function="gene transcription"
/bound_moiety="E2"
/note="putative"
CDS 5917..7467
/note="ORF L1 from bp 5905 to 7467"
/product="major capsid protein"
/gene="L1"
/note="putative"
/codon_start=1
/translation="MAVWHSANGKVYLPPSTPVARVQSTDEYIQRITNIYYHAFSDRLL
TVGHPYFNVYNINGDKLEVPKVSQHRVFRLLKLPDPNRFALADMSVYNPDKERLWVA
CRGLEIGRGQPLGVGSGTGHYPYFNKVKDTENSNAIITFSKDDRQDTSFDPKQIQMFIVG
CTPCIGEHWDKAVPCAENDQQTGLCPPIELKNTYIEDGDMADIGFGNMNFKALQDSRS
DVSLDIVNETCKYPDFLKMQNDIYGDACFFYARREQCYARHFFVRGGKTGDDIPRAQI
DNGTYKNQFYIPGADGQAQKTIGNSMYFPTVSGSLVSSDAQLFNRPFWLQRAQGHNNG
ILWANQMFITVVDNTRNTNFSISVYNQAGALKDVADYNADQFREYQRHVEEYIEISLIL
QLCKVPLKAQVLAQINAMNSSLLEDWQLGFVPTPDNPIQDITYRYIDSLATRCDPKNPP
KEKEDPYKGLHFWDVLDLTERLSLDDLQYSLGRKFLFQAGLQQTTVNGTKAVSYKGSNR
GTRKRKRKN"
polyA_signal 6289..6294
/note="putative"
protein_bind 7396..7407
/function="gene transcription"
/bound_moiety="E2"
/note="putative"
protein_bind 7564..7575
/function="gene regulation"
/bound_moiety="E2 protein"
/note="putative"
repeat_region 7703..7730
/rpt_type=tandem
/rpt_unit=7703..7717, 7716..7730
/note="putative"
source 1..7746
/organism="Human papillomavirus type 5"
/sequenced_mol="DNA"

```

BASE COUNT 2374 a 1545 c 1739 g 2088 t

ORIGIN 354 bp upstream of HindIII site.

```

1 AACGGTaaagt tgcaatttcc ttgtaccagg tgcggtattg ggatttcaca atTATAATgg
5 sj /\
E2-bind <- signal ->
61 ttggtgccaa ctaccatagg catattcaag tttttgcctg tatcgttttc gtatcctgta
121 ataatatcca atatatgtat acataAATAA ATATATATAT ATATAAgtgt ctaagattgg
signal -> E6 orf start ->
signal ->
181 gttcttctgt aatcaggcaA TGgctgaggg agccgaacac caacagaaac tgacagaaaa
E6 cds ->
241 agataaggca gaattacott taagtattag agacttagct gaagccttag gcatccctgt
301 gattgattgt ttaatacott gcaatttctg tggcaacttt ctaaattatt tggagccttg
361 tgaattcgac taaaaaggc ttagtctaatt ttggaaagat tattgtgtgt ttgctgtctg
421 tcgctgatgc tgtggcgcca ctgcaactta tgaatttaac caattttatg agcagacagt
481 gttaggaaga gatattgaat tagcttcagg actttcaata tttgatattg atatcaggtg
541 tcaaacttgc ttagcatttc ttgacattat agaaaagtta gattgctgtg gcagaggcct
601 tccctttcat aaggTGAgga acgcctggaa gggaaatctgt aggcagtgtg agcattttta
E7 orf start ->
661 tcATGattgg TAAagaggtc accgtgcaag atattattct ggagctcagt gaggtgcagc
E7 cds -> <- E6 end
721 ccgaagtgtc accagttgac ctgtttttgtg aagaggaatt accaaacgag caggaaacgg
781 aggaggagcc tgacaacgaa aggatctctt acaaagttat agctccgtgc ggttgcagga

```

```

841 actgtgaggt caagcttgcg atttttgtcc acgccacaga atttggtatt agagctttcc
901 aacagctact GAccggagat ctgcagctcc tgtgccctga ctgtcgcgga aactgcaaac
E1 orf start ->
961 ATGacggatc cTAAttctaa agGTagtaca tctaagaag ggtttgggtga ttggtgttta
      5 sj /\
E1 cds ->          <- E7 end
1021 ttggaagctg actgtagtga tgtagaaaaat gatttggggac aattatttga gagagataca
1081 gactctgata tatcggattt gttagatgat actgaactgg agcagggcaa ttccctggaa
1141 ctatttcac aacaggagtg tgagcagagc gaggagcaat tgcaaaaact aaaacgaaag
1201 tatcttagtc caaaagctgt cgcacagctt agtccgcgac ttgagtcaat ttcattgtca
1261 cccagcaga agtctaagcg aaggctcttt gcagagcagg acagcggact cgagctgact
1321 ttaacaatg aagctgaaga tgttactcct gagggtggagg taccggctat tgactctcgg
1381 ccggatgacg agggaggttc aggggacgta gatatacatt aactgacatt gttgcttct
1441 agcaacaaaa aagctacatt aatggctaag tttaaagagt cgtttggagt aggttttaat
1501 gaattgacac ggcaattcaa aagccacaaa acctgctgta aggactgggt tgtctctgta
1561 tatgacgtgc atgatgatct attgaaagc tcaaagcagc tattgcaaca gcattgtgac
1621 tatatctggg tccgtgggat aggtgcaatg tcattatACC TATTGTGTTt taaggcggga
      -> E2 bind
1681 aaaaatcgcg ggacagttca taagttaatt acctcaatgt taaatgtgca tgaacagcaa
1741 atattgtctg agccgcaaaa attgagaaa acagccgctg cattgttctg gtataaggg
1801 tgtatgggat cggggcggt tagccatgga ccatatcctg attggattgc ccaacaaact
1861 atattaggtc acaaaagtgc tgaggcaagt acttttgatt tttcagcaat ggtccaatg
1921 gcatttcata atcacttatt agacgaagca gatatagcat accagtatgc aaggcttgc
1981 cccgaagacg cgaatgcagt agcttggctt gcacataaca accaggccaa atttgtgaga
2041 gaatgtgcat atatggtacg atttataag aagggacaaa tgagagacat gagtatact
2101 gaatggatat aactaaaaat caatgaagta gaaggggaag ggactggct agatatagta
2161 aagtttatta gatacaaaa tataaacttt atgtattcc taactgcatt aaaagaattc
2221 ctacactcag tgccaaaaaa aaattgcatt ttaatttatg gtcctccaaa ttctggaaag
2281 tcatcatttg caatgtcatt aataagagtg ttgaagggtg gagtggttgc atttgtaaat
2341 tctaaaagtc agttttggct gcaacccctt tcagagtgca agatagctct attggatgat
2401 gtaacagacc ctgttggat atacatggat acatatttaa gaaatggctt ggatggacat
2461 tatgtttcat tagattgtaa atatagagcc ccaacgcaaa tgaaatttcc ccattatta
2521 ttaactccta acattaatgt gcatggggaa actaattata gatatttaca cactacaata
2581 aaaggatttg aatttccaaa tccttttct atgaaagcag ataatacacc tcagttcgaa
2641 ctaactgacc aaagctggaa atcttttttt acaAGgcttt ggacacaatt agaccTGAgt
      /\ 3 sj      E2 orf start ->
2701 gatcaagaag aggagggcga ggATGgagaa tctcagcgag cgtttcaatg ctctgcaaga
      E2 cds ->
2761 tcagctaatg aacatttaTG Aagctgcaga acaaacattg caggcacaaa ttaaacattg
      <- E1 end
2821 gcaaacctta cgaaaagaac ctgtattact ctactatgct agggagaaaag gtgttacaag
2881 gcttggatat caacctgtgc ctgtaaaggc agtatcagaa acaaaggcta aagaagccat
2941 agcaatggtg ctgcagcttg agtcaactaca gacatctgat tttgctcatg agccatggac
3001 tctagttag accagcatag aacattttag aagcgcctca gaaggctcact tcaaaaaag
3061 cccctccct gtagaagtta tttatgacaa tgatccagat aatgccaatt tgtatacaat
3121 gtggacctat gtgtattata tggatgcgga tgataagtgg cataaggcaa gaagtgggg
3181 gaatcacatt ggcatttatt atttacaagg aactttttaa aactattatg tactgtttgc
3241 tgacgatgcg aaaagatatg gtacaactgg agaatgggaa gTAAaagtta ataaggaaac
      E4 orf start ->
      NH2 terminus unknown
3301 tgtgtttgct cctgtcaccA Gctccacgcc tccagggctg ccaggaggac aagcagacac
      /\ 3 sj
3361 aaacaccacc cccgcgaccc ccaccacctc cacaaccgcc gtTGActcca cgtccagaca
      E5 orf start ->
      NH2 terminus unknown
3421 gctcaccaca tcaaacagc cacaacaaac cgaaaccaga ggaagaagGT acggacggag
      5 sj /\
3481 gccctccagc aagtcaagga gatcgcaaac gcagcaaagg cgatcaaggt cccgacACCG
      E2-bind ->

```

Revised HPV5

```

3541 GTCCCGGTct cggccccggg cgcggtccaa gtcccaaacc cacaccactc ggtccaccac
3601 cagggtcccgg tccacGTgcg tcaccaagac tcggggccctt acaagcagat cgcgatccag
      5 sj /\
3661 aggaaggtcc ccaaccacct gcagaagggg aggtggaagg tcaccaggc ggcgatcaag
3721 gtcacctcc acctcctcct cctgcaccac acaacgggtca cagcggggcac gagccgaaag
3781 ttcaacaacc agagggggccc gagggctcgag aggggtcacga ggaggggagc gtgggggggag
3841 agggcgggcga cgaggaaggt catcctcctc ctcctccccc gccACaaac ggtcacgagg
      /\ 3 sj
3901 ggggtctgcT AAgctccgtg gcgtctctcc tggggaagtg ggagggtcac ttcgatcagt
      <- E5 end
3961 tagttcaaag catacaggac gacttggaag attactggaa gaagctcgcg accccccagT
4021 AAtcattgtc aaagggggcg ctaacacact gaaatgcttc cgcaacagag ctaAAattaa
      <- E4 end
      /\ 3 sj
4081 atacatggga ctgtttaggt catttagtac tacctgggtca tgggtggcag gagatggcac
4141 tgagcgtcta ggcaggccca gaatgctcat tagcttttct tcctatactc aaaggagaga
4201 ttttgatgaa gcggtgcgat accccaaagg agttgaTAAg gcctatggca acctggacag
      L2 orf start ->
4261 tcttTAAcat ttactaatgc tgcttttgct actaacatac taacataccc tagcatttta
      <- E2 end
4321 tatttttttt tacattttgt atttgctATG gcgctgcaa aaacgggtcaa gcgagactct
      L2 cds ->
4381 gtaactcata tttaccaaac ctgcaaacag gcaggcactt gccccctga tgttattAAT
      signal ->
4441 AAgtggaac aaacaacagt tgctgacaat attttaaaat atggcagtg cgggtgtattt
4501 tttgggtggc ttgggtattag tacaggccga ggaactgggg tggtacaggt gtacgtgcca
4561 cttgggggaag gtctctggtg ccgtgtcgga ggaACCCCCA CGGTGtaag gccttccttg
      -> E2 bind
4621 gttcctgaaa caatcggggc cgttgatatt ttgcccattg atacagttaa ccccggtgaa
4681 cctacagcat catccgtggg cctcctaact gaggccacag gcgctgattt acttccaggt
4741 gaagttagaaa caattgctga aatccatcct gtacctgagg ggccatcagt ggataccctt
4801 gtagttacca ctagcacagg ttccagtgct gttttagagg ttgccccaga gcctattcct
4861 ccaacacggg tcagggtttc acgcacacag tatcacaatc catcttttca aataataact
4921 gagtctactc cagcacaagg ggaatcgtct cttgcagatc acgttttggg gacatcgggt
4981 tctggggggc aacgaatagg ggtgatata actgacataa ttgagttaga ggaaattcct
5041 agtaggtata catttgaaat tgaagaacca actcctccac gccgcagcag tactccattg
5101 ccacgcaatc aatctgtagg ccgtaggagg ggtttctctt tgactaatag acgttttagta
5161 cagcaggtac aagtggacaa tccattgttt ctaactcaAC CATCTAAGTT agttcgtttt
      -> E2 bind
5221 gcatttgata atcctgtttt tgaggaagaa gtgactaata tatttgaaaa tgatctggat
5281 gtctttgaaag aacctccaga cagagatttt cttgatgtta ggggaattggg acgtccacaa
5341 tattctacaa caccagcggg atagtgttaga gtaagcagggt tggggactcg agccactatt
5401 cgcactcgct ctggtgcaca gatagggtcg caagtccatt tttacagaga tcttagctct
5461 attaatactg aagatcctat tgaattacaa ttattaggcc aacattcagg tgatgctact
5521 atagtccaag gacctgttga aagcacattt atagatatgg atatttctga aaatccatta
5581 tctgaaagca ttgaaagcata ttcacatgat ttattattag atgaaacggg ggaagatttc
5641 agtgggtctc agctggttat aggtaatcga aggagcacia actcttacac tgttcctagg
5701 tttgaaacta caagaaatgg ttcatactat acacaagaca caaagggata ttatgttgca
5761 tatccagagt cacgtaataa tgcagaaatc atttatccta cacctgatat tcctgtagtc
5821 attatacacc ctcatgacag tacagggggac ttttatttac atcccagtct tcacaggcgc
5881 aaacgtaaaa gaaaatattt gTGAtttgca ttcgAGATGg cagtgtggca ctcgggcta
      /\ 3 sj
      L1 orf start -> L1 cds ->
      <- L2 end
5941 ggtaaagtat atcttccacc atcgacaccg gtggccagag tccaaagcac cgatgaatac
6001 attcaaagaa caaatatcta ctatcatgca tttagtgaca gattgttaac tgtaggctcat
6061 ccttatttca atgtatacaa tattaatggg gataagcttg aggttcctaa ggtttcagga
6121 aatcaacaca gagtatttct cctaaaatta ccagatccta acagatttgc attagctgat
6181 atgtctgttt acaaccctga caaagaacgt ttggtttggg cctgtagagg cttagaata

```

```

6241 ggtaggggcc agccattagg tgtagggagt actggtcacc cttatttcAA TAAAgtaaaa
      signal ->
6301 gatacagaaa acagtaatgc atacataaca ttttctaaag atgacagaca ggatacatct
6361 tttgatccta aacagatcca aatgtttatt gtaggatgca caccttgcac aggagagcat
6421 tgggataaag ctgttccatg tgcagaaaat gatcagcaaa ctggcctttg tcctcctatt
6481 gaactaaaaa acacatatat agaagatggg gatatggcag acatagggtt tgggaacatg
6541 aattttaagg cacttcaaga tagtagatca gatgtcagtt tagacatcgt caatgaaact
6601 tgcaagtatc cagatttttt aaagatgcaa aacgatattt atggcgatgc gtgctttttt
6661 tatgctcgtg gggagcaatg ttatgccaga cacttttttg ttagaggggg aaaaactggg
6721 gatgacattc cacgtgcaca aattgacaat ggtacataca aaaatcagtt ttacattcca
6781 ggggctgatg gccaaagctca aaagactata ggaaattcca tgtatttccc aactgttagt
6841 ggctcattag tatccagtga tgctcaattg tttaacaggc ccttctggct ccaaagagcc
6901 caaggtcata ataatggcat cctgtgggct aatcaaagt ttatcacagt ggttgacaac
6961 acaagaata ctaatttcag ttttctgta tataatcagg ctggagcact aaaagatggt
7021 gcagactata atgcagatca atttagagaa tatcaaagac atgtagaaga atatgaaata
7081 tctttaattc tacaactctg taaggttcct ttaaaggcac aggtattggc acagatcaat
7141 gcaatgaact cttcgttatt ggaggattgg cagttaggat ttgttcccac tcctgataat
7201 ccaattcagg acacctacag atatattgac tctttggcta cacgggtgtc agataagaat
7261 cctccgaaag aaaaggaaga cccttataag ggcttacatt tttgggatgt agatttaact
7321 gaaagattgt cattagattt agatcaatat tccttaggca gaaaattttt attccaagct
7381 gggttacaac aaacgACCGT TAACGGTaca aaagcagtgt cttataaagg gtctaataga
      -> E2 bind
7441 ggaacaaaac gcaaacgtaa aaatTGAggt ctgaccgaaa gtggtacatt tttataaact
      <- L1 end
7501 tttacacagt attcaaggaa tgtttgttta ctctgactaa gtataagtct tccaaggata
7561 ccgACCGCAC CCGGTacact cagtcaagtt gttgccaata tagaatcaga tcagtgccaa
      -> E2-bind
7621 acacaccgtc ttggactcag aacagaccgt gttcgttata acatgctcgg attagggacc
7681 tccccaaaga agatttaatc taCAATCGCT TTTGGCAATC GCATTTGGCA ctgctaaaag
      -> overlapping repeat <-
7741 ACCGTT
      -> E2-bind

```


Revised HPV6b

LOCUS HPV6bR 7996 bp ds-DNA circular VRL 11-MAR-1994
DEFINITION Human papillomavirus type 6b (HPV-6b), complete genome.
ACCESSION X00203
KEYWORDS complete genome; overlapping genes.
SOURCE Human papilloma virus type 6b DNA.
ORGANISM Human papillomavirus type 6b
Viridae; ds-DNA nonenveloped viruses; Papovaviridae;
Papillomavirus.

REFERENCE 1 (bases 1 to 7996)
AUTHORS Schwarz,E., Duerst,M., Demankowski,C., Lattermann,O., Zech,R.,
Wolfsperger,E., Suhai,S. and Zur Hausen,H.
TITLE DNA sequence and genome organization of genital human
papillomavirus type 6b
JOURNAL EMBO J. 2, 2341-2348 (1983)
STANDARD full staff_review
COMMENT HPV-11 and HPV-6 are responsible for the large majority of
exophytic condylomas in the genital tract. Even though these
lesions are frequently present in the genital tract, they are
virtually absent in higher grade neoplasias and in cervical
cancers. HPV-6 also infects other mucosal types; the respiratory
tract, oral cavity and conjunctiva. It has been recovered from
approximately 50% of respiratory tract lesions and 50% of all
childhood conjunctival papillomas. Respiratory papillomatosis is
a rare disease that can be life-threatening because of its
recurrent nature and the possibility of obstruction of the airways
and respiratory distress. The most frequent sites of infection are
the vocal cords in the larynx, but papillomas may also be present
in the trachea, lungs, nose and oral cavity. These respiratory
papillomas progress to malignancy rarely, as they account for less
than 0.1% of all respiratory cancers.

The 7996 bp complete genome of HPV-6b has been cloned in pBR322 and
in lambda and was originally recovered from a genital wart. The
sense strand has been numbered by comparative analysis with BPV-1
and HPV-1a. Both the E6 and E7 ORFs contain conserved Cys-X-X-Cys
cysteine doublet motifs. The E6 ORF contains four of these motifs
separated by 29, 36 and 29 intervening amino acids, while the E7
ORF contains just two separated by 29 amino acids. The E6 ORF also
contains a small intron. The E5a ORF codes for a protein of 91
amino acids. It has a stretch of 13 amino acids which is very rich
in leucine. The L2 ORF contains an extremely conserved cluster of
basic residues both at the N terminus and at the C terminus ends.
The authors feel that the conserved region of this part of this
peptide may interact with the conserved L1 structural peptide,
where the variable region may be involved with host or tissue
specific functions.

Between the end of L1 and the beginning of E6 lies a small open
reading frame E8. The first methionine is located in the middle of
the ORF and it has no analog to other papillomaviruses sequenced at
the time of publication. In light of these facts, this ORF is
probably not functional. Thus, the region from the end of L1 to
the beginning of E6 is probably the noncoding region containing the
promoter and origin of replication. Within the first segment of
this region lies a monotonous repetition of thymine-purine which is
just slightly disturbed. Two repeats can be identified within the
LCR; a 24 bp tandem repeat and a 9 bp direct repeat. A TATA box is
located at position 64 and a cap site is located directly in front
of the E6 methionine codon.

SEQUENCE CORRECTION: The original GenBank entry of HPV-6b was missing 94 nucleotides from the LCR following nt 7349. These have been restored in the sequence below (see Fig. 1, Heinz et al., J. Clin. Microbiol., 33:1746-1754).

```

FEATURES             Location/Qualifiers
     5'UTR            join(7292..7996,1..29)
                     /note="putative"
     CDS              join(7840..7996,1..5)
                     /note="probably not functional"
                     /note="E8 from bp 7705 to 5"
                     /gene="E8"
                     /note="putative"
                     /codon_start=1
                     /translation="MHCSQLLKAFACSSSTFFCSYCLVYNNIKMSNLRSHTCDFRFLS
TPYIFPSYS"
     TATA_signal      64..70
     misc_feature     98..104
                     /note="cap site"
                     /note="putative"
     CDS              102..554
                     /note="ORF E6 from bp 30 to 554"
                     /product="transforming protein"
                     /gene="E6"
                     /note="putative"
                     /codon_start=1
                     /translation="MESANASTSATIDQLCKTFNLSMHTLQINCVFCKNALTTAEIY
SYAYKHLKVLFRGGYPYAACACCLEFHGKINQYRHFYDAGYATTVEEETKQDILDVLI
RCYLCHKPLCEVEKVKHILTKARFIKLNCTWKGRCLHCWTTMEDMLP"
     misc_feature     141..142
                     /note="splice acceptor following E7"
                     /note="putative"
     intron           450..503
                     /note="contained in the E6 ORF"
                     /note="putative"
     CDS              530..826
                     /note="ORF E7 from 440 to 826"
                     /product="transforming protein"
                     /gene="E7"
                     /note="putative"
                     /codon_start=1
                     /translation="MHGRHVTLKDIVLDLQPPDPVGLHCYEQLVDSSEDEVDEVDGQD
SQPLKQHFQIVTCCCGCDSNVRLVVQCTETDIREVQQLLLGTLNIVCPICAPKT"
     misc_feature     820..821
                     /note="splice acceptor following E1"
                     /note="putative"
     CDS              832..2781
                     /note="ORF E1 from bp 715 to 2781"
                     /product="replication protein"
                     /gene="E1"
                     /note="putative"
                     /codon_start=1
                     /translation="MADDSGTENEGSGCTGWFVVEAIVQHPTGTQISDDEDEEVEDSG
YDMVDFIDDSNITHNSLEAQALFNRQEADTHYATVQDLKRYLGSYPVSPINTIAEAV
ESEISPRLDIAIKLTRQPKVKRRLFQTRRLTDSGYGYSEVEAGTGTQVEKHGVPENGG
DGQEKDTGRDIEGEEHTEAEAPTNSVREHAGTAGILELLKCKDLRAALLGKFKFCFGL
SFIDLIRPFKSDKTTCLDWVVAGFGIHHSEAFQKLIPLSLYAHIQWLTNAWGMVL
LVLLRFVKNKSRSTVARTLATLLNIPENQMLIEPKIQSGVAALYWFRTGISNASTVI
GEAPEWITRQTVIEHGLADSQFKLTEMVQWAYDNDICEESEIAFEYAQRGDFDSNARA

```

Revised HPV6b

```
FLNSNMQAKYVKDCATMCRHYKHAEMRKMSIKQWIKHRGSKIEGTGNWKPIVQFLRHQ
NIEFIPFLTKFKLWLHGTPKKNCIAIVGPPDTGKSYFCMSLISFLGGTVISHVNSSSH
FWLQPLVDAKVALDDATQPCWIYMDTYMRNLLDGNPMSIDRKHKALTLIKCPPLLV
SNIDITKEDKYKYLHTRVTTFTFPNPFPPDRNGNAVYELSNNTNWKCFERLSSSLDIQ
DSEDEEDGSNSQAFRCVPGTVVRTL"
misc_feature 1279..1280
              /note="splice donor in E1"
              /note="putative"
misc_feature 2678..2679
              /note="splice acceptor following E2"
              /note="putative"
CDS          2723..3829
              /note="ORF E2 from bp 2696 to 3829"
              /product="regulatory protein"
              /gene="E2"
              /note="putative"
              /codon_start=1
              /translation="MEAIARLDACQEQLLELYEENSTDLHKHVLHWKCMRHESVLLY
KAKQMGLSHIGMQVVPPLKVSEAKGHNAIEMQMHLESLLRTEYSMEPWTLOETSSEMWT
QTPPKRCFKKRGKTVEVKFDGCANNMTMDYVWVTDVYVQDNDTWVKVHSMVDAGIYYT
CGQFKTYVNFVKEAEKYGSTKHWEVCYGSTVICSPASVSTTQEVSIPESTTYTPAQ
TSTLVSSSTKEDAVQTPPRKRARGVQQSPCNALCVAHIGPVDSGNHNLIITNNHDQHQ
RNNNSNSATPIVQFQGESNCLKCFRYRLNDRHRHLFDLISSTWHWASSKAPHKHAIVT
VTYDSEEQRQQLDVVKIPPTISHKLGFMSSLHLL"
misc_feature 3242..3243
              /note="splice acceptor following E4"
              /note="putative"
CDS          3255..3584
              /note="ORF E4 from bp 3240 to 3584"
              /gene="E4"
              /note="putative"
              /codon_start=1
              /translation="MGAPNIGKYVMAAQLYVLLHLYLALHKKYPFLNLLHTPPHRPPP
LCPQAPRKTQCKRRLGNEHEESNSPLATPCVWPTLDPWTVEVTTSSLTITTTSTKDGT
VTVQLRL"
misc_feature 3596..3597
              /note="splice donor behind E4"
              /note="putative"
CDS          3887..4162
              /note="ORF E5a from bp 3872 to 4162"
              /gene="E5a"
              /note="putative"
              /codon_start=1
              /translation="MEVVPVQIAAGTTSTFILPVIIAFVVCVFSIILIVWISEFIVYT
SVLVLTLLLYLLLWLLLTTPQLFLLTLLVCYCPALYIHYYIVTTQQ"
CDS          4159..4377
              /note="ORF E5b from bp 4003 to 4377"
              /gene="E5b"
              /note="putative"
              /codon_start=1
              /translation="MMLTCQFNDGDTWLGLWLLCAFIVGMLGLLMHYRAVQGDKHTK
CKKCNKHNCNDDYVTMHYTTDGDYIYMN"
misc_feature 4405..4406
              /note="splice acceptor following L2"
              /note="putative"
CDS          4423..5802
              /note="ORF L2 from bp 4378 to 5802"
              /product="minor capsid protein"
              /gene="L2"
```

```

/note="putative"
/codon_start=1
/translation="MAHSRARRRKRASATQLYQTCKLTGTCPDVIKVEHNTIADQI
LKWGSLGVFFGGGLGIGTGSSTGGRTGYVPLQTSKPSITSGPMARPPVVVEPVAPSDP
SIVSLIEESAIINAGAPEIVPPAHGGFTITTSSETTTPAILDVSVTSHTTTSIFRNPVF
TEPSVTQPQPPVEANGHILISAPTVTSHPIEEIPLDTFVVSSSDSGPTSSTPVPGTAP
RPRVGLYSRALHQVQVTDPAFLSTPQRLITYDNPVYEGEDVSVQFSDHSIHNAPDEAF
MDIIRLHRPAIASRRGLVRYSRIGQRGSMHTRSCHKHIGARIHYFYDISPIAQAAEIEI
MHPLVAAQDDTFDIYAESFEPGINPTQHPVTNISDTYLTSTPNTVTQPWGNTTVPLSL
PNDLFLQSGPDITFPTAPMGTPFSPVTPALPTGPVFITGSGFYLHPAWYFARKRRKRI
PLFFSDVAA"
polyA_signal 4554..4560
/note="putative"
misc_feature 5788..5789
/note="splice acceptor following L1"
/note="putative"
CDS 5789..7291
/note="ORF L1 from bp 5678 to 7291"
/product="major capsid protein"
/gene="L1"
/note="putative"
/codon_start=1
/translation="MWRPSDSTVYVPPNPVSKVVATDAYVTRTNIFYHASSRLLAV
GHPYFSIKRANKTVVPKVSQYQYRVFKVVLDPDNKFPALPDSLSLFDPTTQRLVWACTGL
EVGRGQPLGVGVSQHPFLNKYDDVENSQSGGNPQDNRVNVGMDYKQTLQCMVGCAPP
LGEHWGKQCTNTPVQAGDCPPELITSVIQDGMVDTGFGAMNFADLQTNKSDVPI
DICGTTCQYPDYLQMAADPYGDRLEFFLRKEQMFARHFFNRAGEVGEVPPDTLIIKGS
GNRTSVGSSIYVNTPSGSLVSSEAQLFNKPYWLQKAQGHNNGICWGNQLFVTVVDTR
STNMTLCASVTTSSTYTNSDYKEYMRHVEEYDLQFIFQLCSITLSAEVMAYIHTMNP
VLEDWNFGLSPPPNGTLEDTYRYVQSQAITCQKPTPEKEKPDYKNSLFWEVNLEKEF
SSELDQYPLGRKFLQSGYRGRSSIRTVGKRPVSKASAPKRKRKRAKTKR"
polyA_signal 6422..6427
/note="putative"
repeat_region 7292..7339
/rpt_unit=7292..7315,7316..7339
/standard_name="24 bp tandem repeat"
/note="putative"
misc_feature 7462..7463
/note="splice donor noncoding region"
/note="putative"
polyA_signal 7501..7506
/note="putative"
repeat_region 7544..7568
/rpt_unit=7544..7552, 7560..7568
/standard_name="nonanucleotide direct repeat"
/note="putative"
source 1..7996
/organism="Human papillomavirus type 6"
/sequenced_mol="DNA"

```

BASE COUNT 2458 a 1532 c 1723 g 2283 t

ORIGIN

```

1 GTTAATAACA ATCttggttt aaaaaTAGg agggaccgaa aacggttcaa ccgaaaacgg
  <- E8 end E6 orf start ->
-> CAAT-box <-
(beginns at bp 7898)
61 ttgTATATAA accagcccta aaatntagca aacgaggCAT TATGgaaagt gcaaatgcct
  -> signal E6 cds ->
cap site -> <-

```

Revised HPV6b

```

121 ccacgtctgc aacgaccata Gaccagttgt gcaagacgtt taatctatct atgcatacgt
      /\ 3 sj
181 tgcaaattaa ttgtgtgttt tgcaagaatg cactgaccac agcagagatt tattcatatg
241 catataaaca cctaaaggtc ctgtttcgag gcggctatcc atatgcagcc tgcgctgtgct
301 gcctagaatt tcatggaaaa ataaaccaat atagacactt tgattatgct ggatattgcaa
361 caacagttga agaagaaact aaacaagaca tcttagacgt gctaattcgg tgctacctgt
421 gtcacaaacc gctgtgTGAA gtagaaaagG Taaaacatat actaaccaag gcgcggttca
      E7 orf start -> /\ 5 sj
481 taaagctaaa ttgtacgtgg aAGggctcgt gcctacactg ctggacaacA TGcatggaag
      /\ 3 sj E7 cds ->
541 acatgttacc cTAAaggata ttgtattaga cctgcaacct ccagaccctg tagggttaca
      <- E6 end
601 ttgctatgag caattagtag acagctcaga agatgagggtg gacgaagtgg acggacaaga
661 ttcacaacct ttaaaacaac atttccaaat agtgacctgt tgctgtggat gTGAcagcaa
      E1 orf start ->
721 cgttcgcactg gttgtgcagt gtacagaaac agacatcaga gaagtgcaac agcttctgtt
781 gggaaacta aacatagtggt gtcccatctg cgcaccgaAG accTAAcaac gATGgccggac
      3 sj /\ E1 cds ->
      <- E7 end
841 gattcaghta cagaaaatga ggggtctggg tgtacaggat ggtttatggt agaagctata
901 gtgcaacacc caacaggtac acaaatatca gacgatgagg atgaggagggt ggaggacagt
961 gggatgaca tgggtggactt tattgatgac agcaatatta cacacaattc actggaagca
1021 caggcattgt ttaacaggca ggaggcggac acccattatg cgactgtgca ggacctaaaa
1081 cgaaagtatt taggtagtcc atatgttagt cctataaaca ctatagccga ggcagtgga
1141 agtgaataa gtccacgatt ggacgccatt aaacttacia gacagccaaa aaaggtaaag
1201 cgacggctgt ttcaaaccag ggaactaacg gacagtggat atggctattc tgaagtggaa
1261 gctggaacgg gaacgcagGT agagaaacat ggcgtaccgg aaaatggggg agatgggtcag
      5 sj /\
1321 gaaaaggaca caggaaggga catagagggg gaggaacata cagagggcga agcgcccaca
1381 aacagtgtac gggagcatgc aggcacagca ggaatattgg aattgttaaa atgtaaagat
1441 ttacgggcag cactacttgg taagtttaa gaatgctttg ggctgtcttt tatagattta
1501 attagccat ttaaaagtga taaaacaaca tgtttagatt gggtggtagc aggggttgggt
1561 atacatcata gcatatcaga ggcatttcaa aaattaattg agccattaag tttatagca
1621 catatacaat ggctaacaaa tgcatgggga atggattgt tagtattatt aagatttaaa
1681 gtaataaaaa gtagaagtac cgttgcacgt aacttgcaa cgctattaaa tatacctgaa
1741 aaccaaagt taatagagcc accaaaaata caaagtgggt ttgcagccct gtattggttt
1801 cgtacaggta tatcaaatgc cagtacagtt ataggggaag caccagaatg gataacacgc
1861 caaacagtta ttgaacacgg gttggcagac agtcagttta aattaacaga aatgggtgcag
1921 tgggctgatg ataatgacat atgagaggag agtgaaattg catttgaata tgcacaaagg
1981 ggagattttg attctaagtc acgagcattt ttaaatagca atatgcaggc aaaatatgtg
2041 aaagattgtg caactatgtg tagacattat aaacatgcag aatgaggaa gatgtctata
2101 aaacaatgga taaaacatag gggttctaaa atagaaggca caggaaattg gaaaccaatt
2161 gtacaattcc tacgacatca aaatatagaa ttcattcctt ttttaactaa atttaaatta
2221 tggctgcacg gtacgcaaaa aaaaactgc atagccatag taggcccctc agatactggg
2281 aaatcgtact tttgtatgag ttaataagc tttctaggag gtacagttat tagtcatgta
2341 aattccagca gccatttttg gttgcaaccg ttagtagatg ctaaggtagc attgttagat
2401 gatgcaacac agccatgttg gatataatg gatacatata tgagaaattt gttagatggt
2461 aatcctatga gtattgacag aaagcataaa gcattgacat taattaaatg tccacctctg
2521 ctagtaacgt ccaacataga tattactaaa gaagataaat ataagtattt acatactaga
2581 gtaacaacat ttacatttcc aaatccattc ctttttgaca gaaatgggaa tgcagtgat
2641 gaactgtcaa atacaaactg gaaatgtttt tttgaaAGac tgtcgtcaag ccTAGacatt
      E2 orf start ->
      /\ 3 sj
2701 caggattctg aggacgagga agATGgaagc aatagccaag cgtttagatg cgtgccagga
      E2 cds ->
2761 acagttgtta gaactttaTG Aagaaaacag tactgaccta cacaaacatg tattgcattg
      <- E1 end
2821 gaaatgcatg agacatgaaa gtgtattatt atataaagca aaacaaatgg gcctaagcca
2881 cataggaatg caagtagtgc caccattaaa ggtgtccgaa gcaaaaggac ataatgccat

```

```

2941 tgaaatgcaa atgcatttag aatcattatt aaggactgag tatagtatgg aaccgtggac
3001 attacaagaa acaagttagt aaatgtggca aacaccacct aaacgctggt ttaaaaaacg
3061 gggcaaaact gtagaagtta aatttgatgg ctgtgcaaac aatacaatgg attatgtggt
3121 atggacagat gtgtatgtgc aggacaatga cacctgggta aagggtgcata gtatggtaga
3181 tgctaagggt atatattaca catgtggaca atttaaaaca tattatgtaa actttgTAAa
                                     E4 orf start ->
3241 AGaggcagaa aagtATGgga gcaccaaaca ttgggaagta tgttatggca gcacagttat
      E4 cds ->
      /\ 3 sj
3301 atgttctcct gcatctgtat ctagcactac acaagaagta tccattcctg aatctactac
3361 atacaccccc gcacagacct ccacccttgt gtctcaagc accaaggaag acgcagtgca
3421 aacgccgctt aggaaacgag cacgaggagt ccaacagtc ccttgcaacg ccttgtgtgt
3481 ggccccacatt ggaccctggt acagtggaaa ccacaacctc atcactaaca atcacgacca
3541 gcaccaaaga cggaacaaca gtaacagttc agctacgcct aTAGtgcaat ttcaagGTga
                                     <- E4 end
                                     5 sj /\
3601 atccaattgt ttaaagtgtt ttagatatag gctaaatgac agacacagac atttatttga
3661 tttaatatca tcaacgtggc actgggcctc ctcaaaggca ccacataaac atgccattgt
3721 aactgtaaca tatgatagtg aggaacaaag gcaacagttt ttagatgttg taaaaatacc
3781 ccctaccatt agccacaaac tgggatttat gtcactgcac ctattgTAAt ttgtatatat
                                     <- E2 end
3841 gtaaatgtgt aaatatatgg tattggtgTA Atacaactgt acatgtATGg aagtggtgcc
      E5a orf start ->      E5a cds ->
3901 tgtacaaata gctgcaggaa caaccagcac attcatactg cctgttataa ttgcatttgt
3961 tgtatgtttt gttagcatca tacttattgt atggatattcT GAgtttattg tgtacacatc
      E5b orf start ->
4021 tgtgctagta ctaacactgc ttttatattt actattgtgg ctgctattaa caacccccctt
4081 gcaatttttc ctactaactc tacttgtgtg ttactgtccc gcattgtata tacactacta
4141 tattgttacc acacagcaAT GATgctaaca tgtcaattta atgatggaga tacctggctg
      E5b cds -> <- E5a end
4201 ggtttgtggt tgttatgtgc ctttattgta gggatgttgg ggttattatt gatgcactat
4261 agagctgtac aaggggataa acacacaaa tgtaagaagt gtaacaaaca caactgtaat
4321 gatgattatg taactatgca ttatactact gatggtgatt atatatatat gaatTAGagt
                                     <- E5b end
                                     L2 orf start ->
4381 aaaccgtttt ttatatattgt aacAGtgtat gctttgtata ccATGgcaca tagtagggcc
      /\ 3 sj      L2 cds ->
4441 cgacgacgca agcgtgcgtc agctacacag ctatatcaaa catgtaaact cactggaaca
4501 tgccccccag atgtaattcc taagggtggag cacaacacca ttgcagatca aatATTAATAA
      signal ->
4561 tggggaagtt tgggggtggt ttttggaggg ttgggtatag gcacggggtc cggcactggg
4621 ggtcgtactg gctatgttcc cttacaaact tctgcaaaac cttctattac tagtgggcct
4681 atggctcgtc ctctgtggtt ggtggagcct gtggcccctt cggatccatc tattgtgtct
4741 ttaattgaag aatcggcaat cattaacgca ggggcgcctg aaattgtgcc cctgacacac
4801 ggtgggttta caattacatc ctctgaaaca actaccctg caatattgga tgtatcagtt
4861 actagtcaca ctactactag tatatttaga aatcctgtct ttacagaacc ttctgtaaca
4921 caacccccaa caccctgga ggctaattgga catatattaa tttctgcacc cactgtaacg
4981 tcacacccta tagaggaaat tcctttagat acttttgtgg tatcatctag tgatagcggg
5041 cctacatcca gtaccctgt tcctggtact gcacctcggc ctcgtgtggg cctatatagt
5101 cgtgcattgc accagtgca ggttacagac cctgcatttc tttocactcc tcaacgctta
5161 attacatatg ataaccctgt atatgaaggg gaggatgta gtgtacaatt tagtcatgat
5221 tctatacaca atgcacctga tgaggctttt atggacataa ttcgtttgca cagacctgcc
5281 attgctccc gacgtggcct tgtgcggtac agtcgcattg gacaacgggg gtctatgcac
5341 actcgcagcg gaaagcacat aggggcccgc attcattatt tttatgatat ttcacctatt
5401 gcacaggctg cagaagaaat agaaatgac cctcttgtgg ctgcacagga tgatacaatt
5461 gatatttat ctgaatcttt tgaacctggc ataacccta ccaacacccc tgttacaat
5521 atatcagata catatttaac ttccacacct aatacagtta cacaaccgtg gggtaacacc
5581 acagttccat tgtcacttcc taatgacctg tttttacaat ctggccctga tataactttt

```

Revised HPV6b

```

5641 cctactgcac ctatgggaac accctttagt cctgTAAActc ctgctttacc tacaggccct
      L1 orf start ->
5701 gttttcatta caggttctgg attttatttg catcctgcat ggtattttgc acgtaaacgc
5761 cgtaaacgta ttccttatt tttttcAGAT GtggcggccT AGcgacagca cagtatatgt
      L1 cds -> <- L2 end
      /\ 3 sj
5821 gcctcctcct aaccctgtat ccaaagttgt tgccacggat gcttatgtta ctcgcaccaa
5881 catattttat catgccagca gttctagact tcttgcagtg ggacatcctt atttttccat
5941 aaaacgggct aacaaaactg ttgtgccaaa ggtgtcagga tatcaataca gggatatttaa
6001 ggtgggtgta ccagatccta acaaatttgc attgcctgac tegtctcttt tcgatcccac
6061 aacacaacgt ttagtatggg catgcacagg cctagagggtg ggcagggggac agccattagg
6121 tgtgggtgta agtggacatc ctttctataa taaatatgat gatggtgaaa attcagggag
6181 tgggtgtaac cctggacagg ataacagggt taatgtaggt atggattata aacaaacaca
6241 attatgcatg gttggatgtg ccccccttt gggcgagcat tggggtaaag gtaaacagtg
6301 tactaataca cctgtacagg ctggtgactg cccgccctta gaacttatta ccagtgttat
6361 acaggatggc gatatggttg acacaggcct tgggtgctatg aattttgctg atttgcagac
6421 cAATAAAAtca gatgttctta ttgacatatg tggcactaca tgtaaatac cagattattt
signal ->
6481 acaaatggct gcagacccat atggtgatag attatTTTTT tttctacgga aggaacaaat
6541 gtttgcaga cattttttta acagggctgg cgaggtgggg gaacctgtgc ctgatacact
6601 tataattaag ggtagtggaa atcgcacgta tgtagggagt agtatatatg ttaacacccc
6661 gagcggctct ttggtgtcct ctgaggcaca attgtttaat aagccatatt ggctacaaaa
6721 agcccaggga cataacaatg gtatttggtt gggtaatcaa ctgtttgta ctgtggtaga
6781 taccacacgc agtaccaca tgacattatg tgcacccgta actacatcct ccacatacac
6841 caattctgat tataaagagt acatgcgtca tgtggaagag tatgatttac aatttatttt
6901 tcaattatgt agcattacat tgtctgctga agtaatggcc tatattcaca caatgaatcc
6961 ctctgttttg gaagactgga actttggggt atcgcctccc ccaaatggta cattagaaga
7021 tacctatagg tatgtgcagt cacaggccat tacctgtcaa aagcccactc ctgaaaagga
7081 aaagccagat ccctataaga accttagttt ttgggagggt aattttaaag aaaagttttc
7141 tagtgaattg gatcagtatc ctttgggacg caagtTTTTT ttacaaagtg gatatagggg
7201 acggtcctct attcgtacag gtgttaagcg ccctgctgtt tccaaagcct ctgctgcccc
7261 taaacgtaag cgcgcacaaa ctaaaaggTA ATATATGTGT ATATGTACTG TTATATATAT
      <- L1 end
      -> 24 bp tandem repeat <-->
7321 GTGTGTATGT ACTGTTATGT atatgtgtgt atgtactgtt atatgtatgt gtggtgtata
      24 bp tandem repeat<-
7381 tatgtgtgta tatatgtgtc tgtgtgtata tgtatatgta tgtgttgtgt atatatatgt
7441 gtgtgtgtgt tctgtgtgta atGTAagtta tttgtgtaat gtgtatgtgt gtttatgtgc
      5 sj /\
7501 AATAAAcaat tacctcttgt tacaccctgt gactcagtggt ctgTTGCACG CGttttggT
signal -> repeat -> repeat ->
7561 TGCACGCGcc ttacacacat aagtaataa catgcacaa atatatattt ttgtttaaaa
7621 tactatactt ttatatattg aACCGTTTTT GGTtgcctt agcatacact ttccaccaat
      E2 bind ->
7681 ttgttacaac gtgtttcctc tTAAAtcctat atattttgtg ccagggtacac attgccctgc
      E8 orf start ->
7741 caagttgctt gccaaagtga tcatatcctg ccaaccacac acctggcgcc aggggtgcggt
7801 attgccttac tcataaacct gtctttgtgt tatacttttA TGcactgtag ccaactctta
      E8 cds ->
7861 aaagcatttt tggctttag cagcacattt ttttgcctt actgtttggg atacaataac
7921 ataaaaatga gtaacctaa gtcacacacc tgcgACCGGT TFCGGTtacc cacaccctac
      E2 bind ->
7981 atatttcctt cTTATA
      -> CAAT-box start

```

//

LOCUS HPV18R 7857 bp ds-DNA VRL 11-DEC-1992

DEFINITION Human papillomavirus type 18 (HPV-18), complete genome.

ACCESSION X05015

SOURCE Human papillomavirus type 18 DNA recovered from a cervical carcinoma of a Brazilian patient.

ORGANISM Human papillomavirus type 18
Viridae; ds-DNA nonenveloped viruses; Papovaviridae;
Papillomavirus.

REFERENCE 1 (bases 1 to 7857)

AUTHORS Cole,S.T. and Danos,O.

TITLE Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome

JOURNAL J. Mol. Biol. 193, 599-608 (1987)

REFERENCE 2 (bases 2855-2860; revision)

AUTHORS Baker,C.C.

TITLE The Genomes of the Papillomaviruses

JOURNAL (in) O'Brien,S.J. (Ed.);
Genetic Maps; Locus Maps of Complex Genomes: 1-1,
Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1993)

STANDARD full staff_review

COMMENT Data kindly reviewed (14-AUG-1987) by O. Danos.

HPV-18 is most often found in lesions of the genital mucosa with considerable risk for malignant progression. Estimates indicate that HPV-18 has been recovered from about 10-20% of all invasive cervical cancers. Studies show that the predominance of HPV-18 in high-grade anogenital lesions and invasive cancers is also observed in tissues of the vulva, the penis, and the anus.

HPV-18 has been found in high prevalence in adenocarcinomas and in moderate prevalence in squamous cell carcinomas. Consistently, HPV-18 and HPV-16 are the most prevalent HPV types in adenocarcinomas and adenosquamous carcinomas. Relative frequencies of the two types vary among studies. However, in most studies HPV-18 is detected at least as frequently as HPV-16. In contrast, HPV-18 is found much less frequently than HPV-16 in squamous cell carcinomas of the genital tract, in some cases up to five times less.

The 7857 bp genome of HPV-18 was originally recovered and cloned from a cervical carcinoma of a Brazilian patient. This sequence has been corrected as stated in [2]; at nt 2855-2860 change from "TTGCGT" to "TGCGTT". The E7 ORF is situated immediately in front of E1, a characteristic common to all genital papillomaviruses sequenced at the time of publication. Whereas in the other subgroups, E7 is located in one of the other reading frames. HPV-18 and other genital papillomaviruses and fibropapillomaviruses encode a hydrophobic E5 gene product. The cutaneous papillomaviruses do not possess a homologous E5 ORF.

The long control region (LCR) of HPV-18 can be analyzed in three sections. Segment 1 is a purine + thymidine rich area, which contains the polyadenylation signal for the late genes. Segment 2 is about 200 bp long and only appears in genital papillomaviruses. The third segment is the best conserved among all HPVs. It contains three PV-specific palindromes, and TATA and CAAT boxes; the genital HPVs have one TATA box.

The E6 and E7 ORFs contain regularly spaced cysteine doublet motifs with the form (Cys-X-X-Cys). Also found in E6 of HPV-18, the sequence (XXXLXXXE) is found immediately after the first and third doublet. Cole et al. believe these regions were derived from a duplication of a 33 amino acid peptide including the

Revised HPV18

cysteine doublet. E6 has four of these units, while E7 has three units, the first unit is degenerate.

ERRORS: Several errors in the original GenBank sequence of this file resulted from transcribing the sequence from a figure published in [1]. Cole, S.T. and Danos, O., J. Mol. Biol. 193, 599-608 (1987). In the transcription process G was mistakenly recorded as C.

SITE	CHANGE	EFFECTS	REF
287	C->G	E6aa:(silent)	[2]
2856-60	TGCGT->GCGTT	E1aa:LR->CV, E2aa:CV->AL	[1,3]
3084-3085	CG->GC	E2aa:R->A	[2,4]
3275	C->G	E2aa:(silent)	[2]
5701	C->G	L1aa:P->R	[2]
6460	C->G	L1aa:P->R	[2]
6625	C->G	L1aa:P->R	[2,5]
6842	C->G	L1aa:(silent)	[2,5]

Nt 7592 is indeed "T" in the reference clone even though most variants sequenced so far have a "C" here (Meissner personal communication).

REFERENCES:

1. Nts 2856-2860 corrected in C.C. Baker, "The Genomes of the Papillomaviruses", in O'Brien, S.J. (ed.), Genetic Maps: Locus Maps of Complex Genomes 1-1. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1993.
2. Meissner, John, personal communication: error in creation of GenBank entry from fuzzy figure in publication. Based on resequencing of complete genome of reference clone.
3. Meissner, John, personal communication: true error in original publication (Cole and Danos). Based on resequencing of complete genome of reference clone.
4. Meissner, J., Nuc. Acids Res. 21(4):1041.
5. Stewart, A.-C. et al, J Virol 70(5):3127-3136, 1996.

```
FEATURES             Location/Qualifiers
     5'UTR            join(7137..7857,1..104)
                    /standard_name="LCR"
                    /note="putative"
     repeat_region    41..53
                    /standard_name="PV-specific palindrome"
                    /note="putative"
     repeat_region    58..69
                    /standard_name="PV-specific palindrome"
     CDS              105..581
                    /note="E6 ORF from bp 87 to 581"
                    /product="transforming protein"
                    /gene="E6"
                    /note="putative"
                    /codon_start=1
                    /translation="MARFEDPTRRPYKLPDLCTELNTSLQDIEITCVYCKTVLELTEV
                    FEFAFKDLFVVYRDSIPHAACHKCIDFYSRIRELRHYSDSVYGDITLEKLTNTGLYNLL
                    IRCLRCQKPLNPAEKLRLHLEKRRFHNIAAGHYRGQCHSCCNRRARQERLQRRRETQV"
     CDS              590..907
                    /note="E7 ORF from bp 509 to 907"
                    /product="transforming protein"
                    /gene="E7"
                    /note="putative"
                    /codon_start=1
                    /translation="MHGPKATLQDIVLHLEPQNEIPVDLLCHEQLSDSEEENDEIDGV"
```

NHQHLPARRAEPQRHTMLCMCKCEARIELVVESSADDLRAFQQQLFNLNTLSFVCPWCA
 SQQ"
 CDS 914..2887
 /note="E1 ORF from 908 to 2887"
 /product="replication protein"
 /gene="E1"
 /note="putative"
 /codon_start=1
 /translation="MADPEGTDGEGTGCNGWFYVQAIVDKKTGDVISEDDEDENATDTG
 SDMVDFIDTQGTFCQAELETAQALFHAQEVHNDQVLHVLKRKFAGGSTENSPLGER
 LEVDTELSRPLQEISLNSGQKKAKRRLFTISDSGYGCSEVEATQIQVTTNGEHGNCV
 SGGSTEADIDNGGTEGNSSVDGTSDNSNIENVNPQCTIAQLKDLLKVNKQGAMLAVF
 KDITYGLSFTDLVRNFKSDKTTCTDWVTAIFGVNPTIAEGFKTLIQPFILYAHIQCLDC
 KWGVLILALLRYKCGKSRLTVAKGLSTLLHVPETCMLIQPPKLRSSVAALYWYRTGIS
 NISEVMGDTPEWIIQRLTIIQHGIIDSNFDLSEMVQWAFDNELTDESMAFEYALLADS
 NSNAAFLKSNCAKYLKDCATMCKHYRRAQKRQMNMSQWIRFRCISKIDEGGDWRPIV
 QFLRYQQIEFITFLGALKSFLKGTPKKNCLVFCGPANTGKSYFGMSFIHFIQGAVISF
 VNSTSHFWLEPLTDTKVAMLDDATTCWTFYFDTYMRNALDGNPISIDRKHKPLIQLKC
 PPILLTNIHPAKDNRWPYLESRITVFEFPNAFPFDKNGNPVYEINDKNWKCFERTW
 SRLDLHEEEEDADTEGNPFGTFCVAGQNRPL"
 CDS 2817..3914
 /note="E2 ORF from bp 2796 to 3914"
 /product="regulatory protein"
 /gene="E2"
 /note="putative"
 /codon_start=1
 /translation="MQTPKETLSERLSALQDKIIDHYENDSKDIDSQIQYWQLIRWEN
 AIFFAAREHGIQTLNHQVVPAYNISKSKAHKAIELQMALQGLAQSAKTEDWTLQDTC
 EELWNTEPTHCFKKGQTVQVYFDGNKDNMCTYVAWDSVYYMTDAGTWDKTATCVSHR
 GLYYVKEGYNTFYIEFKSECEKYGNTGTWEVHFGNVVIDCNSMCSSTSDTIVSATQLV
 KQLQHTPSPYSSTVSVGTAKTYGQTSAAATRPGHCGLAEKQHCQVNVNPLGAAATPTGNN
 KRRKLCSGNTTPIIHLKGDNRSLKCLRYRLRKHSDDHYRDISSTWHWTGAGNEKTGILT
 VTYHSETQRTKFLNTVAIPDSVQILVGYMTM"
 CDS 3418..3684
 /note="E4 ORF from bp 3409 to 3684"
 /gene="E4"
 /note="putative"
 /codon_start=1
 /translation="MTLCAVPVTRYPLLSLLNSYSTPPHRI PAPCPWAPQRP TARRR
 LLHDLDTVDSRRSSIVDLSTHFSVQLHLQATTKDGNSSVVTLRL"
 CDS 3936..4157
 /note="E5 ORF from bp 3915 to 4157"
 /gene="E5"
 /note="putative"
 /codon_start=1
 /translation="MLSLIFLFCFCVCMYVCCHVPLLPVCMCAYAWLVFVYIVVIT
 SPATAFTVYVFCFLPMLLLHIHAILSLQ"
 CDS 4244..5632
 /note="L2 ORF from bp 4166 to 5632"
 /product="minor capsid protein"
 /gene="L2"
 /note="putative"
 /codon_start=1
 /translation="MVSHRAARRKRASVTDLYKTCKQSGTCCPPDVVPKVEGTTLADKI
 LQWSSLGIFLGLGIGTGSSTGGRTGYIPLGGRSNTVVVDVGPTRPPVIEPVGPTDPS
 IVTLIEDSSVVTSGAPRPTFTGTSGFDITSAGTTT PAVLDITPSSTSVSISTNFTNP
 AFSDPSIIIEVPQTGEVAGNVFVGTPTSSTHGYEIEPLQTFASSGTGEEPISSTPLPTV
 RRVAGPRLYSRAYQQVSVANPEFLTRPSSLITYDNPAFEPVDTTLTFDPRSDVPDSDF
 MDIIRLHRPALTSRRGTVRF SRLGQRATMFTSRGTQIGARVHFYHDISPIAPSPEYIE

Revised HPV18

```

LQPLVSATEDNDLFDIYADDMDPAVFPVPSRSTTSFAFFKYSPTISSASSYSNVTVPLT
SSWDVPVYTGPDITLPSTTSVWPIVSPTAPASTQYIGIHGTHYLLWPLYFYFIPKKRKR
VPYFFADGFVAA"
CDS      5430..7136
        /note="L1 ORF from bp 5418 to 7136"
        /product="major capsid protein"
        /gene="L1"
        /note="putative"
        /codon_start=1
        /translation="MCLYTRVLILHYHLLPLYGPLYHPRPLPLHSILVYMVHIIICGH
YIILFLRNVNVPFIPLQMALWRPSDNTVYLPSPVARVVNTDDYVTRTSIFYHAGSSR
LLTVGNPFYFRVPAGGGNKQDIPKVSAYQYRVFRVQLPDPNKFGLPDTSIYNPETQRLV
WACAGVEIGRGQPLGVLSGHPFYNKLDDESSHAATSNVSEDVRDNVSVDYKQTQLC
ILGCAPAIGEHWAKGTACKSRPLSQGDCPPLELKNVLEEDGDMVDTGYGAMDFSTLQD
TKCEVPLDICQSICKYPDYLQMSADPYGDSMFCLRREQLFARHFVNRAGTMGDVTPQ
SLYIKGTGMRASPGSCVYSPSPSGSIVTSDSQLFNKPYWLHKAQGHNNNGVCWHNQLFV
TVVDTRSTNLTICASTQSPVPGQYDATKFKQYSRHVEEYDLQFIFQLCTITLTADV
SYIHSMNSSILEDWNFGVPPPTTSLVDTYRFVQSVAITCQKDAAPAENKDPYDKLKF
WNVDLKEKFSLDLQYPLGRKFLVQAGLRKPTIGPRKRSAPSATTSSKPAKRVRVRA
RK"
repeat_region 7458..7469
        /standard_name="PV-specific palindrome"
repeat_region 7821..7833
        /standard_name="PV-specific palindrome"
source      1..7857
        /organism="Human papillomavirus type 18"
BASE COUNT  2365 a  1491 c  1686 g  2315 t
ORIGIN
      1  attaatactt  ttaacaattg  tagtatataa  aaaagggagt  aACCGAAAAC  GGTcgggACC
                               E2 bind ->      E2 bind ->
     61  GAAACCGGTg  tatataaaag  atgTGAgaaa  cacaccacia  tactATGgcg  cgctttgagg
                               E6 orf start ->      E6 cds ->
                                               |-> mRNA start site from
                                               P(105) promoter
    121  atccaacacg  ggcaccctac  aagctacctg  atctgtgcac  ggaactgaac  acttcactgc
    181  aagacataga  aataacctgt  gtatattgca  agacagtatt  ggaacttaca  gaggtatttg
    241  aatttgcatt  taaagattta  tttgtggtgt  atagagacag  tataccgcat  gctgcatgcc
    301  ataaatgtat  agatttttat  tctagaatta  gagaattaag  acattattca  gactctgtgt
    361  atggagacac  attggaaaaa  ctaactaaca  ctgggttata  caatttatta  ataagggtgc
    421  tgcggtgcca  gaaaccgttg  aatccagcag  aaaaacttag  acaccttaat  gaaaaacgac
    481  gatttcacaa  catagctggg  cactaTAGag  gccagtgcca  ttcgtgctgc  aaccgagcac
                               E7 orf start ->
    541  gacaggaacg  actccaacga  cgagagaaa  cacaagtaTA  AtattaagtA  TGcatggacc
                               <- E6 end -> E7 cds
    601  taaggcaaca  ttgcaagaca  ttgtattgca  ttagagccc  caaaatgaaa  ttccggttga
    661  ctttctatgt  cagcagcaat  taagcgactc  agaggaagaa  aacgatgaaa  tagatggagt
    721  taatcatcaa  catttaccag  cccgacgagc  cgaaccacia  cgtcacacia  tgttgtgtat
    781  gtgttgtaag  tgtgaagcca  gaattgagct  agtagtagaa  agctcagcag  acgaccttcg
    841  agcattccag  cagctgtttc  tgaacacct  gtcctttgtg  tgtccgtggt  gtgcatcca
    901  gcagTAAGca  acaATGgctg  atccagaagg  tacagacggg  gagggcacgg  gttgtaacgg
                               <- E7 end
E1 orf start ->      -> E1 cds
    961  ctggttttat  gtacaagcta  ttgtagacaa  aaaaacagga  gatgtaatat  cagatgacga
   1021  ggacgaaaat  gcaacagaca  cagggtcgga  tatggtagat  tttattgata  cacaaggaac
   1081  attttgtgaa  caggcagagc  tagagacagc  acaggcattg  ttccatgcgc  aggaggtcca
   1141  caatgatgca  caagtgttgc  atgttttaaa  acgaaagtgt  gcaggaggca  gcacagaaaa
   1201  cagtcacata  ggggagcggc  tggaggtgga  tacagagtta  agtcacaggt  tacaagaaat
   1261  atctttaaag  agtgggcaga  aaaaggcaaa  aaggcggctg  tttacaatat  cagatagtgg
   1321  ctatggctgt  tctgaagtgg  aagcaacaca  gattcaggta  actacaaatg  gcgaacatgg

```

```

1381 cggcaatgta tgtagtggcg gcagtacgga ggctatagac aacgggggca cagaggggcaa
1441 caacagcagt gtagacggta caagtacaa tagcaatata gaaaatgtaa atccacaatg
1501 taccatagca caattaaag acttgttaaa agtaacaat aaacaaggag ctatgttagc
1561 agtattttaa gacacatat ggctatcatt tacagattta gttagaaatt ttaaaagtga
1621 taaaaccacg tgtacagatt gggttacagc tatatttgga gtaaacccaa caatagcaga
1681 aggattttaa aactaatac agccatttat attatatgcc catattcaat gtctagactg
1741 taaatgggga gtattaatat tagccctggt gcgttacaaa tgtggtaaga gtagactaac
1801 agttgctaaa ggtttaagta cgttgttaca cgtacctgaa acttgatgt taattcaacc
1861 accaaaattg cgaagtagtg ttgcagcact atattggtat agaacaggaa tatcaaata
1921 tagtgaagta atgggagaca cacctgagtg gatacaaaga cttactatta tacaacatgg
1981 aatagatgat agcaattttg atttgtcaga aatggtaaaa tgggcatttg ataatgagct
2041 gacagatgaa agcgatatgg catttgaata tgccttatta gcagacagca acagcaatgc
2101 agctgccttt ttaaaaagca attgccaaagc taaatattta aaagattgtg ccacaatgtg
2161 caaacattat aggcgagccc aaaaacgaca aatgaatatg tcacagtggg tacgatttag
2221 atgttcaaaa atagatgaag ggggagattg gagaccaata gtgcaattcc tgcgatacca
2281 acaaatagag tttataacat ttttaggagc cttaaaatca tttttaaag gaaccccaa
2341 aaaaaattgt ttagtatttt gtggaccagc aaatacagga aaatcatatt ttggaatgag
2401 tttatacac tttatacaag gacagtaat atcatttgtg aattccacta gtcatttttg
2461 gttggaaccg ttaacagata ctaaggtggc catgttagat gatgcaacga ccacgtgttg
2521 gacatacttt gatacctata tgagaaatgc gttagatggc aatccaataa gtattgatag
2581 aaagcacaaa ccattaatac aactaaaatg tcctccaata ctactaacca caaatataca
      |-> mRNA start site from P(2598) promoter
2641 tccagcaaag gataatagat ggccatattt agaaagtaga ataacagtat ttgaatttcc
2701 aaatgcattt ccatttgata aaaatggcaa tccagtatat gaaataaatg acaaaaattg
2761 gaaatgtttt tttgaaagga catgggtccag atTAGatttg cacgaggaag aggaagATGc
      E2 orf start -> E2 cds ->
2821 agacaccgaa ggaaaccctt tcggaacggt taagtgcggt gcaggacaaa atcatagacc
2881 actaTGAaaa tgacagtaaa gacatagaca gccaaataca gtattggcaa ctaatacgtt
      <- E1 end
2941 gggaaaatgc aatattcttt gcagcaaggg aacatggcat acagacatta aaccaccagg
3001 tgggtgccagc ctataacatt tcaaaaagta aagcacataa agctattgaa ctgcaaatgg
      |-> mRNA start site from
      P(3036) promoter
3061 ccctacaagg ccttgacaaa agtgcataca aaaccgagga ttggacactg caagacacat
3121 gcgaggaact atggaataca gaacctactc actgctttaa aaaagggtggc caaacagtac
3181 aagtatattt tgatggcaac aaagacaatt gtatgacctg ttagcatgag gacagtggtg
3241 attatatgac tgatgcagga acatgggaca aaacggctac ctgtgtaagt cacaggggat
3301 tgtattatgt aaaggaaggg tacaacacgt tttatataga atttaaaagt gaatgtgaaa
3361 aatatgggaa cacaggtacg tgggaagtac attttgggaa taatgTAAtt gattgtaATG
      E4 orf start -> E4 cds ->
3421 actctatgtg cagtaccagt gacgacacgg tatccgctac tcagcttgtt aaacagctac
3481 agcacacccc ctaccgctat tccagcaccg tgtccgtggg caccgcaaag acctacggcc
3541 agacgtcggc tgctacacga cctggacact gtggactcgc ggagaagcag cattgtggac
3601 ctgtcaaccc acttctcggg gcagctacac ctacaggcaa caacaaaaga cggaaactct
3661 gtagtggtaa cactacgcct aTAAtacatt taaaagggtg cagaacagat ttaaaatggt
      <- E4 end
3721 tacggtacag attgcgaaaa catagcgacc actatagaga tatatcatcc acctggcatt
3781 ggacaggtgc aggcaatgaa aaaacaggaa tactgactgt aacataccat agtgaaacac
3841 aaagaacaaa atttttaaact actgttgcaa ttcagatag tgtacaaata ttgggtgggat
3901 acatgacaat gTAAtacata tgctgtagta ccaatATGtt atcaattatt tttttatttt
      <- E2 end E5 cds ->
      E5 orf start ->
3961 gcttttgtgt atgcatgat gtgtgctgcc atgtcccgt tttgccatct gtctgtatgt
4021 gtgcgtatgc atgggtattg gtatttgtgt atattgtggg aataacgtcc cctgccacag
4081 cattcacagt atatgtattt tgttttttat tgcccattgt actattgcat atacatgcta
4141 tattgtcttt acagTAAttg taTAGgttgt tttatacagt gtattgtaca ttgtatattt
      <- E5 end
      L2 orf start ->

```

Revised HPV18

```

4201 tgttttatac cttttatgct ttttgtattt ttgtaataaa agtATGgtat cccaccgtgc
                                L2 cds ->
4261 cgcacgacgc aaacgggctt cggtaactga cttatataaa acatgtaaac aatctggtag
4321 atgtccacct gatgttgttc ctaagggtgga gggcaccacg ttagcagata aatatattgca
4381 atggccaagc cttggtatat ttttgggtgg acttggcata ggtactggca gtggtacagg
4441 gggctgtaca gggtagattc cattgggtgg gcgttccaat acagtgggtg atggttggtcc
4501 tacacgtccc ccagtgggta ttgaacctgt gggccccaca gacccatcta ttgttacatt
4561 aatagaggac tccagtgtgg ttacatcagg tgcacctagg cctacgttta ctggcacgtc
4621 tgggtttgat ataacatctg cgggtacaac tacacctgcg gttttggata tcacacctc
4681 gtctacctct gtgtctatth ccacaaccaa tttaccaat cctgcatttt ctgatccgtc
4741 cattattgaa gttccacaaa ctggggaggt ggaggtaat gtatttgttg gtaccctac
4801 atctggaaca catgggtatg aggaaatacc tttacaaaca tttgcttctt ctggtagcgg
4861 ggaggaaccc attagtagta ccccatggcc tactgtgagg cgtgtagcag gtccccgcct
4921 ttacagtagg gcctaccaac aagtgtcagt ggctaaccct gagtttctta cacgtccatc
4981 ctctttaatt acatatgaca acccggcctt tgagcctgtg gacactacat taacatttga
5041 tcctcgtagt gatgttctct attcagattt tatggatatt atccgtctac ataggcctgc
5101 tttaacatoc aggcgtggga ctgttcgctt tagtagatta ggtcaacggg caactatggt
5161 taccgcagc ggtacacaaa taggtgctag ggttcacttt tatcatgata taagtcctat
5221 tgcaccttcc ccagaatata ttgaactgca gcctttagta tctgccacgg aggacaatga
5281 cttgtttgat atatatgcag atgacatgga ccctgcagtg cctgtacct cgcgttctac
5341 tacctccttt gcatttttta aatattcgcc cactatatct tctgcctctt cctatagtaa
5401 tgtaacggtc ccttTAAcct cctctgggA TGtgcctgta tacacgggtc ctgatattac
                                L1 orf start ->      L1 cds ->
5461 attaccatct actacctctg tatggcccat tgtatcacc acggccctct cctctacaca
5521 gtatattggt atacatggta cacattatta tttgtggcca ttatattatt ttattcetaa
5581 gaaacgtaaa cgtgttccct attttttgc agatggcttt gtggcggccT AGtgacaata
                                |-> mRNA start site from P(5600) promoter
                                <- L2 end
5641 ccgtatatct tccacctcct tctgtggcaa gagttgtaaa taccgatgat tatgtgactc
5701 gcacaagcat attttatcat gctggcagct ctagattatt aactgttggt aatccatatt
5761 ttagggttcc tgcaggtggt ggcaataagc aggatattcc taaggtttct gcataccaat
5821 atagagtatt tagggtgcag ttacctgacc caaataaatt tggtttacct gatactagta
5881 tttataatoc tgaacacaaa cgtttagtggt gggcctgtgc tggagtgga atggccctg
5941 gtcagccttt aggtgtggc cttagtgggc atccatttta taataaatta gatgacactg
6001 aaagttccca tgcgccacg tctaattggt ctgaggacgt tagggacaat gtgtctgtag
6061 attataagca gacacagtta tgtattttgg gctgtgcccc tgctattggg gaacactggg
6121 ctaaaggcac tgettgtaaa tgcgctcctt tatcacaggg cgattgcccc cctttagaac
6181 ttaaaaacac agttttgga gatggtgata tggtagatac tggatatggt gccatggact
6241 ttagtacatt gcaagatact aaatgtgagg taccattgga tatttgtcag tctatattga
6301 aatatcctga ttatttacia atgtctgcag atccttatgg ggattccatg tttttttgct
6361 tacggcgtga gcagcttttt gctaggcatt tttggaatag agcaggtagt atgggtgaca
6421 ctgtgcctca atccttata attaaaggca caggtagtgc tgcttcacct ggcagctgtg
6481 tgtattctcc ctctccaagt ggtctattg ttacctctga ctcccagttg ttttaataaac
6541 catattggtt acataaggca cagggtcata acaatgggtg ttgctggcat aatcaattat
6601 ttgttactgt ggtagatacc actcgcagta ccaatttaac aatatgtgct tctacacagt
6661 ctctgttacc tgggcaatat gatgctacca aatttaagca gtatagcaga catgttgagg
6721 aatatgattt gcagtttatt tttcagttgt gtactattac tttactgca gatgttatgt
6781 cctatatcoa tagtatgaat agcagtattt tagaggattg gaactttggt gttccccccc
6841 cgccaactac tagtttgggt gatacatatc gttttgtaca atctgttgct attacctgtc
6901 aaaaggatgc tgcaccggt gaaaataagg atccctatga taagttaaag ttttggaaatg
6961 tggatttaaa ggaaaagttt tctttagact tagatcaata tccccttggc cgtaaatatt
7021 tggttcaggc tggattgctg cgcaagccca ccataggccc tcgcaaacgt tctgctccat
7081 ctgccactac gtcttctaaa cctgccaaagc gtgtgctgtg acgtgcccagg aagTAAatg
                                <- L1 end
7141 tgtgtgtgta tatatatata catctattgt tgtgtttgta tgtcctgtgt ttgtgttttt
7201 tgtatgattg cattgtatgg tatgtatggt tgtgtttgta tgtgtatgt tactatattt
7261 gttggtatgt ggcattaaat aaaaatagtt ttgtggttct gtgtgttatg tggttgcgcc
7321 ctagtgagta acaactgtat ttgtgtttgt ggtatgggtg ttgcttgttg ggctatatat
7381 tgtcctgtat ttcaagttat aaaactgcac accttacagc atccatttta tccataatc

```

```
7441 ctccatTTTg ctgtgcaACC GATTTCGGTt gcctttggct tatgtctgtg gttttctgca
      E2 bind ->
7501 caatacagta cgctggcact attgcaaact ttaatctttt gggcactgct cctacatatt
7561 ttgaacaatt ggcgcgcctc tttggcgcat ataaggcgca cctgggtatta gtcattttcc
7621 tgtccaggTg cgctacaaca attgcttgca taactatata cactccctaa gtaataaaaac
7681 tgcttttagg cacatatttt agtttgTTTT tacttaagct aattgcatac ttggcttgta
7741 caactacttt catgtccaac attctgtcta cccttaacat gaactataat atgactaagc
7801 tgtgcataca tagtttatgc aACCGAAATA GGTtgggcag cacatactat acttttc
      E2 bind ->
```