
HPV and Animal PV Nucleic Acid Sequences

Papillomavirus sequences in Part I have been grouped for purposes of presentation. No taxonomic argument is intended by the groupings, although the principal basis for them was phylogenetic tree analysis of L1 coding sequences. One group, F, simply comprises all the diverse HPVs that cannot be satisfactorily assigned to any well-defined phylogenetic clade at this time. The cutaneous and EV sequences were each placed into groups by virtue of their phenotypic properties. The analyses supporting the groups—typically “weighted parsimony” and “linear correlation”—are presented in detail in Part III. With few exceptions, these analyses are in accord with the published findings of H.-U. Bernard and coworkers [1,2] and of van Ranst and coworkers [3]. The recent review by De Villiers [4] discusses the extent to which the phylogenetic relationships agree with, in some instances, and deviate from, in other instances, phenotypic and clinical observations.

-
- [1] Bernard H.-U., Chan S.-Y., and Delius H: Evolution of Papillomaviruses. In: *Human Pathogenic Papillomaviruses* H. zur Hausen (Ed.). Springer-Verlag, Berlin, 1994; pp. 33–54.
 - [2] Bernard H.-U., Chan S.-Y., Ong C.-K., Villa L.L., Delius H., Peyton C.L., Bauer H.M., Manos M.M., and Wheeler C.M: Identification and Assessment of Known and Novel Papillomaviruses by Polymerase Chain Reaction, Restriction Digestion Fingerprinting, Nucleotide Sequence, and Phylogenetic Algorithms. *Journal of Infectious Diseases* 1994 (in press, Nov 1994).
 - [3] van Ranst M., Kaplan J.B., and Burk R.D: Phylogenetic Classification of Human Papillomaviruses: Correlation with Clinical Manifestations. *Journal of General Virology* 1992;**73**:2653–2660.
 - [4] De Villiers E.-M: Human Pathogenic Papillomavirus Types: An Update. In: *Human Pathogenic Papillomaviruses* H. zur Hausen (Ed.). Springer-Verlag, Berlin, 1994; pp. 1–12.

PART I HPV and Animal PV Nucleotide Sequences

Introduction	1
Contents	2
Group A Sequences	
Introduction	I-A-1
HPV16	I-A-3
HPV31	I-A-9
HPV33	I-A-14
HPV35	I-A-19
HPV35h	I-A-24
HPV52	I-A-29
HPV58	I-A-33
HPV67MY911	I-A-37
Group B Sequences	
Introduction	I-B-1
HPV6b	I-B-5
HPV11	I-B-10
HPV13	I-B-14
HPV34	I-B-19
HPV44E6	I-B-23
HPV44MY911	I-B-24
HPV55MY911	I-B-25
HPV64MY911	I-B-26
HPVMM9	I-B-27
Group C Sequences	
Introduction	I-C-1
HPV18	I-C-4
HPV39	I-C-8
HPV45	I-C-12
HPV59MY911	I-C-16
HPV68ME180	I-C-17
HPVCP141	I-C-20
HPVL1AE1	I-C-22
HPVLVX160	I-C-23
Group D Sequences	
Introduction	I-D-1
HPV26	I-D-3
HPV30	I-D-7
HPV51	I-D-11
HPV53	I-D-15
HPV56	I-D-19
HPV66L1AE3	I-D-23

HPV66MY911	I-D-24
HPV69MY911	I-D-25
HPVIS039	I-D-26
HPVL1AE2	I-D-27
HPVMM4	I-D-28

Group E Sequences

Introduction	I-E-1
HPV61L1AE4	I-E-2
HPV61MY911	I-E-3
HPV62MY911	I-E-4
HPVCP4173	I-E-5
HPVCP6108	I-E-7
HPVCP8304	I-E-9
HPVLVX100	I-E-11
HPVLVX82	I-E-12
HPVMM7	I-E-13
HPVMM8	I-E-14

Group F Sequences

Introduction	I-F-1
HPV2a	I-F-4
HPV3	I-F-8
HPV7	I-F-12
HPV10	I-F-16
HPV27	I-F-20
HPV28MY911	I-F-24
HPV29MY911	I-F-25
HPV32	I-F-26
HPV40	I-F-30
HPV42	I-F-34
HPV43E6	I-F-38
HPV43MY911	I-F-39
HPV54MY911	I-F-40
HPV57	I-F-41
HPVCP8061	I-F-45

Group G Sequences

Introduction	I-G-1
HPV1a	I-G-3
HPV4	I-G-8
HPV41	I-G-12
HPV63	I-G-16
HPV65	I-G-20

Group H Sequences

Introduction	I-H-1
------------------------	-------

HPV5	I-H-4
HPV5b	I-H-8
HPV5d	I-H-12
HPV8	I-H-16
HPV9	I-H-20
HPV12	I-H-24
HPV14d	I-H-28
HPV15	I-H-32
HPV17	I-H-36
HPV19	I-H-40
HPV20E6	I-H-44
HPV21E6	I-H-46
HPV25	I-H-48
HPV47	I-H-52
HPV49	I-H-56

Group I Sequences

Introduction	I-I-1
COPV	I-I-5
CRPV	I-I-9
EEPV	I-I-13
DPV	I-I-18
BPV1	I-I-23
BPV2	I-I-28
BPV4	I-I-32
PCPV1	I-I-37
MnPV	I-I-42
RhPV1	I-I-46

Group A Sequences

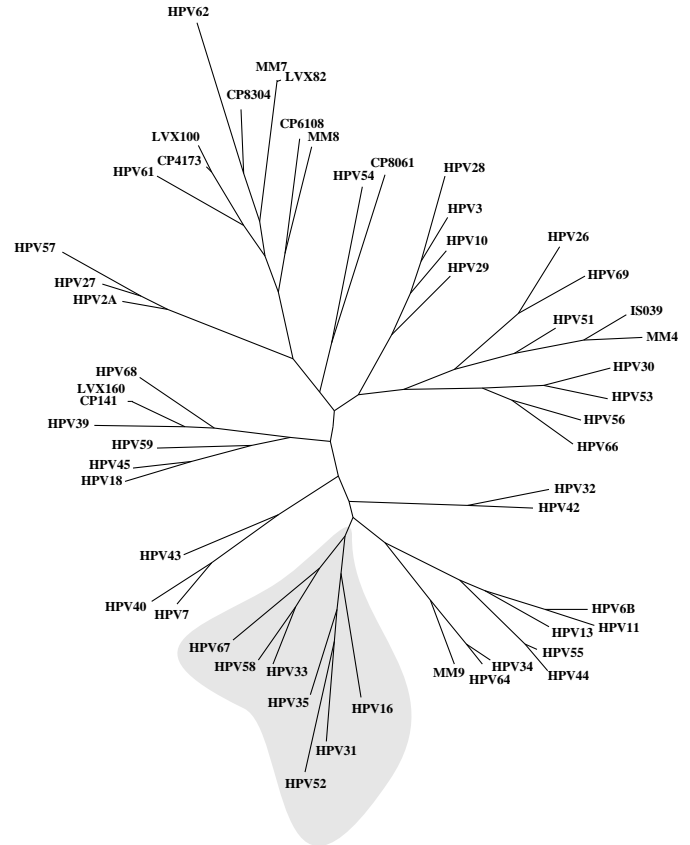
HPV16	HPV31
HPV33	HPV35
HPV35h	HPV52
HPV58	HPV67

INTRODUCTION

Group A consists of human papillomavirus types 16, 31, 33, 35, 35h, 52, 58, and 67, a subset of the “high-risk and intermediate-risk” anogenital viruses as classified by Lorincz et al. The Lorincz study used probes from 15 common anogenital HPV types to screen 2627 U.S. subjects. These data were analyzed to determine the carcinogenic potential of each of the 15 HPV types [1]. HPV-16, the only “high-risk” type in this group, is predominantly associated with invasive cancer. Hybridization positive for HPV-16 DNA occurred in 47% of all invasive cancers, 47% of all high-grade lesions (CIN II or III), and 16% of all low-grade lesions (CIN I or condyloma) [1]. “Intermediate risk” HPV species: 31, 33, 35, 35h, 52, and 58 are most prevalent in high-grade intraepithelial lesions and, unlike HPV-16, are less prevalent in invasive cancers. Several Japanese studies indicate the Asian prevalence of HPV-52 and HPV-58 in invasive cervical carcinomas is higher than that reported in U.S. studies [2]. Due to the recent identification of HPV-67, an assessment of its risk does not exist.

Group A viruses predominantly infect tissues of the cervix and other anogenital tract sites: the vulva, the vagina, the penis, the perineum and the anus. Within the anogenital tract, HPV-16 is the single most prevalent type. In addition to these sites, HPV-16 has been frequently detected in carcinomas of the oral cavity, larynx and lung, and less often, in lesions of the skin [2,3]. Carcinomas of the larynx and lung have also been linked to HPV-33 infection [3]. Bowenoid lesions of the finger are associated with HPV-35 and HPV-35h [2]. Conversely, types 31, 52, 58 and 67 have been detected in anogenital tissue only (HPV-67 tissue restriction is based on limited data). Multiple site infection is common with all Group A viruses, but is most prevalent for HPV-16 [4]. Carcinomas of the cervix occur more frequently than those at other anogenital tract sites. These latter sites lack an area that is comparable in vulnerability to the transformation zone of the cervix, which is often the site of origination or transgression of cervical neoplasia [4].

Because of the acute and long-term pathogenic potential of HPV-16, it is one of the best studied of the HPV types. A considerable body of literature exists which elucidates the molecular mechanisms of HPV-16 transcription, regulation and interaction with host cellular factors, see Chapter V in this compendium and [5]. Transcription of the major HPV-16 RNA species begins at nucleotide 97, from the only known HPV-16 promoter, P₉₇. Regulation of the promoter is dependent on the location of the occupied E2-binding site and the nature of the E2 proteins. HPV-16 codes an E2 protein homologous to the BPV-1 E2 transactivator. Negative regulation is mediated by the binding of the E2 protein to one or both of the proximal E2-binding sites. These sites lie adjacent to the TATA box of the promoter and downstream of an essential Sp-1 binding site. Sp-1 can enhance E2 binding to sites possessing only a weak affinity (determined by the specific sequence of the binding site). Binding of E2 to the proximal sites may sterically hinder cellular factor TFIID binding or interfere with the assembly of the transcription preinitiation complex. If the two more distal upstream E2-binding sites are occupied, transcriptional activation may be observed [6]. Disruption of the E1 and



E2 genes occurs in cervical cancers as the result of genomic linearization and integration into the host chromosome. In the absence of E2 repression, the E6 and E7 genes are actively transcribed [4].

Post-transcriptional processing of the transcribed DNA results in several mRNA species which have been catalogued by extensive R-loop mapping studies. These studies have characterized the splicing patterns required to produce these species (Figure I-1) [7]. Splice sites within the E6 gene potentially produce a truncated E6 product, E6*. After translation, the proteins are directed from the cytoplasm to the nucleus by the recognition of a basic cluster of amino acids which form the nuclear localization signal. Two such signals are present in the L1 coding region [8]. In the nucleus, the viral proteins interact with various cellular proteins to redirect the cell replication machinery and, in many cases, the cell cycle.

Complete genomic sequences are available for all the current members of Group A except HPV-67, which has been sequenced only over the My09-My11 region of L1. Of the sequences in this group, HPV-35 and 35h are variants and HPV-33 and HPV-58 are “close types”—sequences which qualify to be distinct types under the criterion of ten percent dissimilarity at the nucleotide level, but between which most of the changes are “silent,” causing no difference at the amino acid level (Part III-C). The Rhesus monkey papillomavirus (RhPV-1) often clusters with this group during phylogenetic analysis, although we have put it with the other animal papillomaviruses in Group I for the purposes of presentation.

-
- [1] Lorincz,A.T., Reid,R., Jenson,A.B., Greenberg,M.D., Lancaster,WD, Kurman,R.J. Human papillomavirus infection of the cervix: relative risk associations of 15 common anogenital types. *Obstet Gynecol* **79**:328–337
 - [2] de Villiers,E.M. Human pathogenic papillomavirus types: an update. in *Human Papillomaviruses*, edited by Harald zur Hausen, Springer-Verlag, Heidelberg, 1994, pp 1–12
 - [3] Snijders,P.J.F., van den Brule, A.J.C., Meijer,C.J.L.M., and Walboomers, J.M.M. Papillomaviruses and cancer of the upper digestive and respiratory tracts. in *Human Papillomaviruses*, edited by Harald zur Hausen, Springer-Verlag, Heidelberg, 1994, pp 177–197
 - [4] Shah,K. and Howley,P. Papillomaviruses. in *Virology*, second edition, edited by B.N. Fields, Raven Press, N.Y., pp 1651–1676
 - [5] Scheffner,M., Romanczuk, H., Munger, K., Huibregtse,J.M., Mietz,J.A., and Howley,P.M. Functions of human papillomavirus proteins. in *Human pathogenic papillomaviruses*, edited by Harald zur Hausen, Springer-Verlag, Heidelberg, 1994, pp 55–81
 - [6] McBride, A., Romanczuk,H., and Howley,P. The papillomavirus E2 regulatory proteins. *J. Biol. Chem.* **266**:18411–18414
 - [7] Baker,C.C. The genomes of the papillomaviruses. in *Genetic Maps; Locus Maps of Complex Genomes: 1.134*, edited by S.J O’Brien, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1993, 1.134–1.146
 - [8] Zhou,J., Doorbar,J., Yi Sun,X., Crawford,L., McLean,C., and Frazer,I. Identification of the nuclear localization signal of human papillomavirus type 16 L1 protein. *Virology* **185**:625–632

LOCUS HPV16 7905 bp ds-DNA circular VRL 15-APR-1991
DEFINITION Human papillomavirus type 16 (HPV16), complete genome.
ACCESSION K02718
KEYWORDS circular; complete genome.
SOURCE Papilloma virus type 16 DNA recovered from a human invasive cervical carcinoma.

REFERENCE 1 (bases 1 to 7904)
AUTHORS Seedorf,K., Kraemmer,G., Duerst,M., Suhai,S. and Roewekamp,W.G.
TITLE Human papillomavirus type 16 DNA sequence
JOURNAL Virology 145, 181-185 (1985)

REFERENCE 2 (sites)
AUTHORS Kennedy,I.M., Haddow,J.K. and Clements,J.B.
TITLE A negative element in the human papillomavirus type 16 genome acts at the level of late mRNA stability
JOURNAL J. Virol. 65, 2093-2097 (1991)

REFERENCE 3 (base 4363 and sites; 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: 1.134, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1993)

REFERENCE 4 (base 3906; revision)
AUTHORS Bubb,V., McCance,D.J., and Schlegel,R.
TITLE DNA sequence of the HPV-16 E5 ORF and the structural conservation of its encoded protein
JOURNAL Virology 163, 243-246 (1988)

REFERENCE 5 (sites)
AUTHORS Zhou,J., Doorbar,J., Yi Sun,X., Crawford,L., McLean,C., and Frazer,I.
TITLE Identification of the nuclear localization signal of human papillomavirus type 16 L1 protein
JOURNAL Virology 185, 625-632 (1991)

COMMENT HPV-16 is considered to be a member of the "high risk" mucosal group. All of the viruses in this group are predominantly associated with invasive cancer. In a study which used 15 common genital HPV types as probes, a positive HPV-16 hybridization occurred for 47% of all invasive cancers, 47% of high-grade lesions (CIN II or III), and 16% of low-grade lesions (CIN I or condyloma). In addition to being a "high risk" type, HPV-16 has a high prevalence of infection. Of all invasive cancer biopsies tested, 47% were HPV-16 positive. Out of the transformation zone, HPV-16 was found in 85% of all vulvar, vaginal, and anal neoplasias. HPV-16 is usually found in lesions of the genital mucosa; most often of the cervix, and more rarely of the vulva, penis and anus. HPV16 has also been demonstrated in cutaneous lesions including Bowen's disease of the foot, an epidermal naevus of the foot, a case of arsenic keratosis, bowenoid lesions of the hand, bowenoid papulosis of the face, squamous cell carcinomas of the finger and keratoacanthoma of the finger.

In cervical cancers, the HPV DNA is usually found integrated into the host chromosome. This is in contrast to the extrachromosomal form of the viral genome found in premalignant CIN lesions. The site of integration in the host appears to be random. As an interesting note, the integration of HPV-18 (observed in the HeLa HPV-18 positive cervical carcinoma cell line) has occurred near the c-myc locus. The integration of the HPV genome requires linearization, which most often occurs in the E1/E2 region. In the case of HPV16 and HPV18 this may result in the inactivation of the E2 repressor of the E6 and E7 genes. The E6 and E7 genes of the "high risk" group are transforming genes. With repression removed, immortalization may result.

The DNA of HPV-16 has been molecularly cloned in bacteriophage lambda from a genomic library of an invasive cervical carcinoma. This sequence has been corrected as stated in [3] and [4]; at nt 4363 a substitution of a T for an A, and at nt 3906 an additional T,

HPV16

respectively. The sense strand is shown, with a numbering system matching the first 60 bp of HPV1a, HPV6b and BPV1. The designation of the early and late reading frames of HPV-16 is based solely upon homology comparison with HPV-6b and BPV-1. Two TATA boxes have been identified upstream from the E6 open reading frame, at bp 17 and 65, and another upstream from the L1 coding region, at bp 4289. A potential CAT box precedes the TATA boxes near the E6 region, at bp 7895. Polyadenylation signals can be found at the end of the early and late gene regions, at bp 4213 and 7260. Both open reading frames E4 and E5 do not have a translation initiation codon. This is true for most other sequenced papilloma viruses. Also, the open reading frame for a possible E1 gene product, thought to be involved in DNA replication, seems to be interrupted. Seedorf et al. contend that an insertion of one nucleotide somewhere in the region between positions 1101 and 1168 would generate an E1 open reading frame of normal size. Based on our alignment of the E1 gene for several HPV genomes the location of the insertion point has been identified and annotated in the sequence.

Zhou et al. identified two nuclear localization signals (NLS) in the HPV-16 genome, each consisting of two basic aa clusters. One is located at the carboxy terminal of L1 with the form of KRKKRK at nt 7132-7149 (within L1) and the other is a bipartite group also in the L1 gene with KRK at nt 7087-7095 and KR at nt 7132-7137 [5].

The deduced transcription structure of HPV-16 has been reviewed in [3]. Transcription of the major HPV-16 RNA species begins at nucleotide 97 from the only known HPV-16 promoter, P₉₇. A cap site is located at nucleotide 97, 30 nt downstream of the promoter TATA box. Extensive R-loop mapping studies have made it possible to catalog the mRNA species produced by various splicing patterns. Splice junction pairs and their coding potentials are shown in the table below. All splice sites annotated in the sequence have been experimentally determined [3].

mRNA species	Splice donor/acceptor pair	Coding potential
a	880/3357	E6, E7, E1 ^{E4} , E5
b	226/409, 880/3357	E6* ^I , E7, E1 ^{E4} , E5
c	226/526, 880/3357	E6* ^{II} , E7, E1 ^{E4} , E5
d	226/3357	E6* ^{III} , E5
e	880/3357	E1 ^{E4} , E5
f	880/2709	E2, E5
g	1302/3357	E2C, E5
h	1302/3357, 3632/5638	E2M, L1
i	1302/5638	L1

BASE COUNT 2601 a 1377 c 1509 g 2417 t

ORIGIN Unreported.

```

1 actacaataa ttcattgTATA AAAActaaggg cgtaACCGAA ATCGGTtgaA CCGAAACCGG
      signal ->                -> E2 bind      -> E2 bind
61 TTAGTATAAA Agcagacatt ttATGcacca aaagagaact gcaatgtttc aggaccaca
signal ->                |-> mRNA start site from
      <- LCR                    P97 promoter
E6 orf start->                E6 cds ->
121 ggagcgACCC AGAAAGTTac cacagttatg cacagagctg caaacaacta tacatgatat
      -> E2 bind
181 aatattagaa tgtgtgtact gcaagcaaca gttactgcca cgtgagGTat atgactttgc
      5' sj /\
241 ttttcgggat ttatgcatag tatatagaga tggaatcca tatgctgtat gtgataaatg
301 tttaaagttt tattctaaaa ttatgtgagta tagacattat tgttatagtt tgtatggaac
361 aacattagaa cagcaataca acaaacggtt gtgtgatttg ttaattAGgt gtattaactg
      /\ 3' sj
421 tcaaaagcca ctgtgtcctg aagaaaagca aagacatctg gacaaaaagc aaagattcca
481 taatataagg ggtcgggtgga ccggtcagatg tatgtcttgt tgcAGatcat caagaacacg
      /\ 3' sj
541 TAGagaaacc cagctgTAAt cATGcatgga gatacaccta cattgcatga atatatgtta
E7 orf start ->                E6 end <-      -> E7 cds

```



```

601 gatttgcaac cagagacaac tgatctctac tgttatgagc aattaaatga cagctcagag
661 gaggaggatg aatagatgg tccagctgga caagcagaac cggacagagc ccattacaat
721 attgtaacct tttgttgcaa gtgtgactct acgcttcggg tgtgctgaca aagcacacac
781 gtagacattc gtactttgga agacctgtta atggggcacac taggaattgt gtgccccatc
841 tgttctcaga aaccaTAAtc taccATGgct gatcctgcag GTaccAatgg ggaagagggg
      El orf start ->          -> El cds 5' sj /\
      <- E7 end
901 acgggatgta atggatgggt ttatgtagag gctgtagtgg aaaaaaaaa acgggatgct
961 atatcagatg acgagaacga aatgacagt gatacagggt aagatttggg agattttata
1021 gtaaatagata atgattatct aacacaggca gaaacagaga cagcacatgc gttgtttact
1081 gcacaggaag caaaaacaaca tagagatgca gtacagggtc taaaacgaaa gtatttggta
      /\
      deletion
      causing premature
      termination of El
1141 gtccacttag tgatattagt ggatgtgTAG acaataatat tagtcctaga ttaaaagcta
      <- premature termination
      of El cds
1201 tatgtataga aaaacaaagt agagctgcaa aaaggagatt atttgaaagc gaagacagcg
1261 ggtatggcaa tactgaagtg gaaactcagc agatgttaca gGTagaaggg cgccatgaga
      5' sj /\
1321 ctgaaacacc atgtagtcag tatagtgggtg gaagtggggg tggttgcagt cagtacagta
1381 gtggaaagtgg gggagagggt gttagtgaaa gacacactat atgccaaaca ccacttacia
1441 atattttaaa tgtactaaaa actagtaatg caaaggcagc aatgttagca aaattttaaag
1501 agttatacgg ggtgagtttt tcagaattag taagaccatt taaaagtaat aatcaacgt
1561 gttgcatggt gtgtattgct gcatttggac ttacaccagc tatagctgac agtataaaaa
1621 cactattaca acaatattgt ttatatttac acattcaaag tttagcatgt tcatggggaa
1681 tggttgtggt actattagta agatataaat gtggaaaaaa tagagaaaca attgaaaaat
1741 tgcctgtctaa actattatgt gtgtctccaa tgtgtatgat gatagagcct caaaaattgc
1801 gtagtacagc agcagcatta tattgggata aaacaggtat atcaaataat agtgaagtgt
1861 atggagacac gccagaatgg atacaaagac aaacagcatt acaacatagt ttaatgatt
1921 gtacatttga attatcacag atggtacaat gggcctacga taatgacata gtacagata
1981 gtgaaattgc atataaatat gcacaattgg cagacactaa tagtaatgca agtgcctttc
2041 taaaagtaa ttcacaggca aaaattgtaa aggattgtgc aacaatgtgt agacattata
2101 aacgagcaga aaaaaaaca atgagtatga gtcaatggat aaaaatagca tgtgataggg
2161 tagatgatgg aggtgatggg aagcaaattg ttatgttttt aaggatcaa ggtgtagagt
2221 ttatgtcatt tttactgca ttaaaaagat tttgcaagg catacctaaa aaaaattgca
2281 tattactata tgggtgcagct aacacaggta aatcattatt tggtagagt ttaatgaaat
2341 tctgcaagg gtctgtaata tgttttgtaa attctaaaag ccatttttgg ttacaacct
2401 tagcagatgc caaaataggt atgttagatg atgctacagt gccctgttgg aactacatag
2461 atgacaattt aagaaatgca ttggatggaa atttagtttc tatggatgta aagcatagac
2521 cattggtaca actaaaatgc cctccattat taattacatc taacattaat gctggtacag
2581 attctaggtg gccttattta cataatagat tgggtggtgt tacatttctt aatgagtttc
2641 gacttgacga aaacggaaat ccagtgtatg agcttaatga taagaactgg aatcctttt
2701 tctcaagGAC gtggtccaga tTAAgtttgc acgaggacga ggacaaggaa aacATGgag
      5' sj /\      E2 orf start ->          E2 cds ->
2761 actctttgcc aacgtttaa tgtgtgtcag gacaaaatac taacacatta TGAaaatgat
      <- El end
2821 agtacagacc tacgtgacca tatagactat tggaaacaca tgcgcctaga atgtgctatt
2881 tattacaagg ccagagaaat gggattttaa catattaacc accaagtggg gccaacactg
2941 gctgtatcaa agaataaagc attacaagca attgaactgc aactaacgtt agaaacaata
3001 tataactcac aatatagtaa tgaaaagtgg acattacaag acgttagcct tgaagtgtat
3061 ttaactgcac caacaggatg tataaaaaaa catggatata cagtggaagt gcagtttgat
3121 ggagacatat gcaatacaat gcattataca aactggacac atatatatat ttgtgaagaa
3181 gcatcagtaa ctgtggtaga ggggtcaagt gactattatg gtttatatta tgttcatgaa
3241 ggaatacga catattttgt gcagttaaa gatgatgcag aaaaatag tagaaataaa
3301 gtatgggaag ttcattcggg tggcaggTA Atattatgtc ctacatctgt gttTAGcagc
      E4 orf start ->          /\ 3' sj
      NH2 terminus unknown
3361 aacgaagtat cctctcctga aattattagg cagcacttgg ccaaccaccc cgccgcgacc
3421 cataccaaag ccgtcgcctt gggcaccgaa gaaacacaga cgactatcca gcgaccaaga
3481 tcagagccag acaccggaaa cccctgccac accactaagt tgttgacag agactcagtg
3541 gacagtgcct caatcctcac tgcatttaac agctcacaca aaggacggat taactgtaat
3601 agtaaacacta caccctaTAGt acatttaaaa gGTgatgcta atactttaa atgtttaaga

```

HPV16

```

<- E4    5' sj /\
end
3661 tatagattta aaaagcattg tacattgtat actgcagtgt cgtctacatg gcattggaca
3721 ggacataatg taaaacataa aagtgcaatt gttacactta catatgatag tgaatggcaa
3781 cgtgaccaat ttttgtctca agtTAAaata ccaaaaaacta ttacagtgtc tactggattt
      E5 orf start ->
      NH2 terminus unknown
3841 atgtctatAT GACaaatctt gatactgcat ccacaacatt actggcgtgc tttttgcttt
      -> E5 cds
      <- E2 end
3901 gcttttgtgt gcttttgtgt gtctgcctat taatacgtcc gctgcttttg tctgtgtcta
3961 catacacatc attaataata ttggatttac tattgtggat aacagcagcc tctgcgttta
4021 ggtgttttat tgtatataat atatttgttt atataccatt atttttaata catacacatg
4081 cacgcttttt aattacaTAA tgtatatgta cataatgtaa ttgttacata TAAttgttgt
      <-E5 end          L2 orf start ->
4141 ataccataac ttactatttt ttctttttta ttttcatata taattttttt ttttgtttgt
4201 ttgtttgttt tttAATAAAc tgttattact taacaATGcg acacaaaacgt tctgcaaaaac
early poly-A signal ->          L2 cds ->
4261 gcacaaaacg tgcacggtct acccaacttt ATAAAcatg caaacaggca ggtacatgtc
      signal ->
4321 cacctgacat tatacctaag gttgaaggca aaactattgc tgaacaaata ttacaatag
4381 gaagtatggg tgtatttttt ggtgggtag gaattggaac agggcgggt acaggcggac
4441 gcaactggga tattccattg ggaacaaggc cccccacagc tacagataca cttgctcctg
4501 taagaccccc ttaacagta gatcctgtgg gcccttctga tccttctata gtttctttag
4561 tggaagaaac tagttttatt gatgctggtg caccaacatc tgtaccttcc attccccag
4621 atgtatcagg atttagtatt actacttcaa ctgataccac acctgctata ttagatatta
4681 ataatactgt tactactgtt actacacata ataatcccac tttcaactgac ccatctgtat
4741 tgcagcctcc aacacctgca gaaactggag ggcattttac acttcatca tccactatta
4801 gtacacataa ttatgaagaa attcctatgg atacatttat tggtagcaca aaccttaaca
4861 cagtaactag tagcacaccc ataccagggt ctgcccagc ggcacgccta ggattatata
4921 ctgcacacac acaacaggtt aaagtgtgag accctgcttt tgaaccact cccactaaac
4981 ttattacata tgataatcct gcatatgaag gtatagatgt ggataataca ttatattttt
5041 ctagtaaatga taatagtatt aatatagctc cagatcctga ctttttggat atagttgctt
5101 tacataggcc agcattaacc tctaggcgtg ctggcattag gtacagtaga attggtaata
5161 aacaaacact acgtactcgt agtggaaaat ctatagggtgc taaggtagcat tattattatg
5221 atttaagtac tattgatcct gcagaagaaa tagaattaca aactataaca ccttctacat
5281 atactaccac ttcacatgca gcctcaccta cttctattaa taatggatta tatgatattt
5341 atgcagatga ctttattaca gatacttcta caaccocggg accatctgta cctctacat
5401 ctttatcagg ttatattcct gcaaatataca caattccttt tgggtgggca tacaatatc
5461 ctttagtata aggtcctgat ataccatta atataactga ccaagctcct tcattaattc
5521 ctaTAGttcc aggtctcca caatatacaa ttattgctgA TGcagggtgac ttttatttac
L1 orf start ->          L1 cds ->
5581 atcctagtta ttacatgta cgaaaacgac gtaaacgttt accatatttt ttttcAGatg
      /\ 3' sj
5641 tctctttggc tgcctAGtga ggccactgtc tacttgcttc ctgtccagc atctaaggtt
      <- L2 end
5701 gtaagcacgg atgaatatgt tgcacgcaca aacatatatt atcatgcagg aacatccaga
5761 ctacttgtag ttggacatcc ctattttctt attaaaaaac ctaacaataa caaaatatta
5821 gttcctaaag tadcaggatt acaatacagg gtatttagaa tacatttacc tgaccccaat
5881 aagtttggtt ttcctgacac ctcattttat aatccagata cacagcggct ggtttgggcc
5941 tgtgtagggt ttgaggtagg tctgtgctcag ccattagggtg tgggcattag tggccatcct
6001 ttattaaata aattggatga cacagaaaaa gctagtgtct atgcagcaaa tgcagggtgtg
6061 gataatagag aatgtatata tatggattac aaacaacac aattgtgttt aattgggtgc
6121 aaaccaccta taggggaaca ctggggcaaa ggatccccc gtaccaatgt tgcagtaaat
6181 ccagggtgatt gtccACCATT AGAGTTaata aacacagtta ttcaggatgg tgatattggt
      -> E2 bind
6241 catactggct ttgggtgctat ggactttact acattacagg ctaacaaaaag tgaagttcca
6301 ctggatattt gtacatctat ttgcaaatat ccagattata ttaaaatggt gtcagaacca
6361 tatggcgaca gcttattttt ttatttaoga agggaacaaa tggtttggtag acatttattt
6421 aatagggtctg gtactgttgg tgaaaatgta ccagacgatt tatacattaa aggctctggg
6481 tctactgcaa atttagccag ttcaaattat tttctacac ctagtgggtc tatgggttacc
6541 tctgatgccc aaatattcaa taaACCTTAT TGGTTacaac gagcacaggg ccacaataat
      -> E2 bind
6601 ggcatttgtt ggggtaacca actatttgtt actgttgttg atactacagc cagtacaaat

```

```
6661 atgtcattat gtgctgccat atctacttca gaaactacaT ATAAAAatac taactttaag
      signal ->
6721 gagtacctac gacatgggga ggaatatgat ttacagttta tttttcaact gtgcaaaata
6781 accttaactg cagacgttat gacatacata cattctatga attccactat tttggaggac
6841 tggattttg gtctacaacc tccccagga ggcacactag aagatactta taggtttgta
6901 acccaggcaa ttgcttgtca aaaacataca cctccagcac ctaaagaaga tgatcccctt
6961 aaaaaataca ctttttggga agtaaattha aaggaaaagt tttctgcaga cctagatcag
7021 tttcctttag gacgcaaatt tttactacaa gcaggattga aggocaaacc aaaatttaca
```

HPV16

```
7081 ttaggaAAAC GAAAagctac acccaccacc tcatctacct ctacaactgc TAAACGCAAA
    basic aa ->                nuclear localization signal (NLS) ->
    cluster 1                    basic aa cluster 2 of -> <-
    of bipartite NLS              bipartite NLS
7141 AAACGTAAGc tgTAAgtatt gtagtatgt tgaattagtg ttgtttggtg tgtatatggt
    <- L1 end
    -> LCR
7201 tgtatgtgct tgtatgtgct tgtaaatatt aagttgtatg tgtgtttgta tgtatggtat
7261 AATAAAcacg tgtgtatgtg tttttaaatg cttgtgtaac tattgtgtca tgcaacataa
poly-A ->
7321 ataaacttat tgtttcaaca cctactaatt gtggttggtg tattcattgt atataaacta
7381 tatttgctac atcctgtttt tgttttatat atactatatt ttgtagcgcc aggccattt
7441 tgtagcttca ACCGAATTCG GTtgcattgt ttttggcaca aaatgtggtt ttttaaatag
    -> E2 bind
7501 ttctatgtca gcaactatgg tttaaacttg tacgtttcct gcttgccatg cgtgccaat
7561 ccctgttttc ctgacctgca ctgcttgcca accattocat tgttttttac actgcactat
7621 gtgcaactac tgaatcacta tgtacattgt gtcaTATAAA Ataaatcact atgcgccaac
    signal ->
7681 gccttacata ccgctgtag gcacatattt ttggcttggt ttaactaacc taattgcata
7741 tttggcataa ggtttaaact tctaaggcca actaaatgac accctagttc atacatgaac
7801 tgtgtaaagg ttagtcatac attgttcatt tgtaaaactg cacatgggtg tgtgcaaacc
7861 gattttgggt tacacattta caagcaactt aTATAATAat actaa
    signal ->
```

LOCUS HPV31 7912 bp ds-DNA circular VRL 15-DEC-1989
 DEFINITION Human papillomavirus type 31 (HPV-31), complete genome.
 ACCESSION J04353
 KEYWORDS circular; complete genome.
 SOURCE Human papillomavirus type 31 DNA recovered from a cervical biopsy
 from a woman with CIN.
 REFERENCE 1 (bases 1 to 7912)
 AUTHORS Goldsborough,M.D., DiSelvestre,D., Temple,G.F. and Lorincz,A.T.
 TITLE Nucleotide sequence of human papillomavirus type 31: A cervical
 neoplasia associated virus
 JOURNAL Virology 171, 306-311 (1989)
 COMMENT Draft entry and computer-readable copy of sequence [1] kindly
 submitted by M.D.Goldsborough, 05-JUL-1989.

HPV-31 is most often found in lesions of the genital mucosa which may have a risk for malignant progression. Estimates indicate that HPV-31 and other less thoroughly studied types (33, -45, -51, and -56) have been recovered from about 15% of all invasive cervical cancers.

HPV31 DNA was initially identified in biopsy tissue recovered from a woman with cervical intraepithelial neoplasia (CIN) and was cloned in vector lambda. The coding strand of the 7912 bp genome of HPV31 was aligned with the nucleotide sequences of HPV types 6, 11, and 16 and was numbered according to the system devised for BPV1. As is common with all other sequenced papilloma viruses, all ORFs are located on one strand of DNA. The E4 ORF lacks an initiating methionine, which is also the case for many of the other sequenced papillomaviruses.

The upstream regulatory region (URR) of HPV-31 exhibits many features conserved among either all papillomaviruses or among those which are associated with anogenital lesions. The first 200 nucleotides of the upstream regulatory region (URR) - bp 7076 to 7276 has a constitution of 49% T and 5% C. This feature is common among the HPVs associated with anogenital lesions (6, 11, 16, 18, 31, and 33). The URR contains two sets of direct repeats: two copies of a 6 nucleotide sequence and two copies of a ten nucleotide sequence. An octamer which is a putative keratinocyte enhancer, nt 7542-7549, also exists in HPV-16. This enhancer is thought to play a role in the epithelial cell tropism seen in genital HPVs. Also, a glucocorticoid receptor sequence, nt 7406-7420, exhibits a high degree of similarity to a sequence identified in HPV-16. Features which have been found in the URRs of all sequenced HPVs and which have been identified in HPV-31 include: 4 copies of the recognition sequence for the E2 transactivator and repressor gene products, and two TATAAAA sequences.

The early genes, especially E6 and E7, may be involved with the oncogenic properties of the viruses which cause malignancy. Both E6 and E7 of HPV-31 exhibit repeats of the 4-amino acid motif, Cys-X-X-Cys, which may be representative of a metal binding domain. The E6 gene contains three repeats of the motif separated by 29 and 36 nucleotides, whereas the E7 gene contains two motifs separated by 29 nucleotides. The same motif and spacing between motifs is found in all sequenced HPVs. It was noticed that both HPV-16 and HPV-18 cell lines had splice acceptor and donor sites contained within the E6 gene, and that they were not found in types 6 and 11. It was first thought that the presence of these sites may be required for oncogenic progression. These splice sites have been found in the following types associated with anogenital cancer: 16, 18 31, 33, 52 and 56 and with the exception of HPV43, have not been found in the following cutaneous types: 1, 5, 8, and 43. Goldsborough et al. notes that the presence of a possible functional splice in type 43 makes it unclear as to whether this feature distinguishes HPVs with respect to oncogenic potential.

Goldsborough et al. believe that the absence or presence of a cell

HPV31

division motif (CD) in the E7 gene may alone or in part determine viral oncogenic potential. The HPV types which are associated with anogenital cancers possess this motif, and those which rarely cause mucosal malignancy do not. This distinction, however, can only be made in mucosal types as this rule breaks down for the cutaneous types. For example, types 1, 5 and 8 all exhibit this motif, but only 5 and 8 are associated with malignancy [1].

```

BASE COUNT      2528 a   1364 c   1572 g   2448 t
ORIGIN
  1 taataataat aatccttagTA TAAAAaaagta gggagTGACC GAAAGTGGTg aACCGAAAAAC
      signal ->                                <- LCR                                -> E2 bind
      E6 orf start->
      -> E2 bind
 61 GGTtggtata TAAAGcacat agtattttgt gcaaacctac agacgccATG ttcaaaaatc
      signal ->                                E6 cds ->
121 ctgcagaaag acctcggaaa ttgcatgaac taagctcggc attggaaata cctcagatg
181 aactaagatt gaattgtgtc tactgcaaag gtcagttaac agaaacagag GTattagatt
      5' sj /\
241 ttgcatttac agatttaaca atagtatata gggacgacac accacacgga gtgtgtacaa
301 aatgtttaag attttattca aaagtaagtg aatttagatg gtatagatat agtgtgtagt
361 gaacaacatt agaaaaattg acaaacaaag gtatatgtga tttcttaatt AGgtgtataa
      /\ 3' sj
421 cgtgtcaaaag accgtttgtg ccagaagaaa aacaaagaca tttggataaa aagaaacgat
481 tccacaacat aggaggaaagg tggacaggac gttgcatagc atgttggaga agacctcgta
541 cTGAaaccca agtgTAAacA TGcgtggaga aacacctacg ttgcaagact atgtgttaga
E7 orf start ->                                <- E6 end
      E7 cds ->
601 tttgcaacct gaggcaactg ACCTCCACTG TTatgagcaa ttaccogaca gctcagatga
      ->                                <- E2 bind site
661 ggaggatgtc atagacagtc cagctggaca agcagaaccg gacacatcca attacaatat
721 cgttaccttt tgttgtcagt gtaagtctac acttcgtttg tgtgtacaga gcacacaagt
781 agatattcgc atattgcaag agctgttaat gggctcattt ggaatcgtgt gccccaactg
841 ttctacTAGA ctgTAAactac aATGgctgat ccagcaggta cagatgggga ggggacggga
E1 orf start ->                                <- E7 end
      E1 cds ->
901 tgcaatgggt ggttttatgt agaagcagta attgacagac agacagggga caacatttca
961 gaggacgaaa atgaagacag tagtgatact ggggaggata tggttgactt tattgacaat
1021 tgtaatgtat acaacaatca ggcagaagca gagacagcac aggcattggt tcatgacacag
1081 gaagcggagg aacatgcaga ggctgtgcag gttctaaaac gaaagtatgt aggtagtctt
1141 ttaagtgata ttagtagttg tgtggattat aatattagtc cacggttaaa agctatatgc
1201 atagaaaaata acagtaaaac agcaaaaacga agactcttgg aacttccaga cagcgggtat
1261 ggcaataactg aagtggaaac gcagcagatg gtacaggtag aggagcaaca aacaacatta
1321 agttgtaatg gtagtgcagg gacacatagt gaacgagaga atgaaactcc aacacgtaat
1381 agtgacgaag tgttaaaaac tagcaatggt aaagctgcta tgttaggtaa atttaaaaga
1441 ttatatgggt taagttttat ggaactaatt aggccatttc aaagcaataa aagcacatgt
1501 actgattgggt gtgtagctgc gtttggagtt acaggtacag ttgcagaagg atttaaaacc
1561 ctattgcaac catattgttt gtattgccat ttacaaagtt tagcatggtc ctggggcatg
1621 gttatgtaa tgcttgtgag attttaaagt gcaaaaaata gaataacaat tgaaaaatta
1681 ttagaaaaat tattgtgtat atctacaaat tgtatgtaa ttcagccacc caaattacgt
1741 agcacagctg cagcattata ttggtagaca acaggaatgt caaacattag cgatgtatat
1801 ggtgaaacac cagaatggat agaaagacaa acagtattac agcatagttt taatgacaca
1861 acatttgatt tgtoccaaat ggtacaatgg gcatatgaca atgatgttat ggatgatagt
1921 gaaattgcct ataaatatgc acaattagct gacagtgata gtaatgcatg tcatttttta
1981 aaaagtaatt cgcaggcaaa aatagttaaa gattgtggaa caatgtgtag acattataaa
2041 cgagcagaaa aacgacaaat gtccatggga cagtggatta aaagtagatg tgacaaagtt
2101 agtgacgaag gtgactggag ggacatagta aagtttttaa gatatcaaca aatagaattt
2161 gtgtcatttt tatctgcatt aaagctgttt ttaaaaggag tgccaaagaa aaactgtatt
2221 ttaatacatg gtgcacctaa tacaggtaaa tcatattttg gaatgagcct tattagcttt
2281 ttacaaggat gtataatatc atatgcaaat tcaaaaagtc atttttgggt acaaccactg
2341 gctgatgcta aaataggcat gttagatgat gctacaacgc catgttggca ttatatagac
2401 aattacctac gaaatgcact agatggcaac cctgtatcta tagatgtaa gcataaagct
2461 ttaatgcagt taaaatgtcc tcccttattg attacatcta atataaatgc aggtaaggat
2521 gacagatggc catacctaca tagcagactg gtggttttta catttccaaa tccatttcca

```

```

2581 tttgacaaaa acggaaatcc agtatatgaa ttaagtgata aaaactggaa atcctttttc
2641 tcaaggacgt ggtgcagatT AAatttgcac gaggaagagg acaaagaaaa cgATGgagac
      E2 orf start ->                E2 cds ->
2701 tcttttctcaa cgttttaaattg tgtgtcagga caaaatatta gaacattaTG Aaaatgatag
      <- E1 end
2761 taaacgactt tgtgatcata tagactattg gaaacatatt cgacttgaat gtgtattaat
2821 gTATAAAgca agagaaatgg gaatacacag tattaaccac cagggtggtgc cagcgttgtc
signal ->
2881 agtatcaaag gccaaagcct tacaagctat tgaactacaa atgatgttgg aaacattaaa
2941 taacactgaa tacaaaaatg aggactggac aatgcagcaa acaagtcttg aactgtattt
3001 aactgcacct acagggtggt taaaaaaaca tggatatact gtagaggtgc aatttgatgg
3061 tgatgtacac aacaccatgc attatactaa ctggaaaattt atatacctat gtatagatgg
3121 ccaatgtact gttgtggaag ggcaagttaa ttgtaagggc atttattatg tacatgaagg
3181 acatataaca tattttgtaa attttacaga agaggcaaaa aaatatggga ctggtaaaaa
3241 atggggaagtg catgcgggtg gtcaggTAAT tgtttttcct gaactctgat ttagcagtga
      E4 orf start ->
      NH2 terminus unknown
3301 cgaaatatcc tttgctggga ttggtacaaa gttaccaaca gccacaaca ccaccacatc
3361 gaattccaaa acctgcgcct tgggcaccag tgaagggtgtg cggcgggcca cgacgtctac
3421 taagcgacca agaacagagc cagagcacag aacacccac cacccaaca agttgttgcg
3481 aggcgactcc gtggacagtg tcaactgtgg ggttatcagt gcagctgcat gcacaaacca
3541 aacaagggct gtcagttgtc ctgcaactac acctATAAta cacttaaaag gtgatgcaaa
      <- E4 end
3601 tatattaataa tgtttaagat ataggctgtc aaaatataaa caattgtatg aacaagtgtc
3661 atctacatgg cattggacat gtacagatgg aaaacataaa aatgctattg taaccttaac
3721 atatataagt acatcaciaa gagacgattt tttaaatact gtaaaaatac ctaacacagt
3781 atcagtgatca acaggatata TGAactattTA GcctaATGat tgaactaaat atttctacag
      E5 orf start ->                <- E2 end
      E5 cds ->
3841 taagcattgt gctatgcttt ttgctttgct tttgtgtgct actatttgtg tgtccttgca
3901 tacgtccact tgtgctgtct gtgtcgggat atgcaacact actattatta attgtgattt
3961 tatgggttat tgcaacctct ccattacggt gtttttgtat atatggtgtg tttatatata

```

HPV31

```

4021 ttccattatt tgtaattcat acacatgcat ctttttTAAG tcaacagTAA CTTTTTact
                                L2 orf start ->                <- E5 end
                                                                <- repeat region
4081 tgtgtatact gttgtttTA TTGGTATTGG TATTGGTATT GGTATTGGTA TTGGtataAT
                                                                poly-A signal ->
4141 AAACTTTTTT ACTTTTTTtt tattattacc ATGcgggtcca aacgctctac aaaacgcact
                                L2 cds ->
    end repeat region <-
4201 aaacgtgcgt ctgctacaca attatatcaa acatgtaaag cagcagggtac ttgtccatca
4261 gacgttatac ctaaaataga acatactacc attgcagacc aaatattaag gtatggtagt
4321 atgggtgttt tttttgggtg gttgggtatt ggggtccggct ctgggtactgg ggggtcgcact
4381 ggatatgtcc ctcttagtac acgtccttct acagtatctg aggcaagtat acctattaga
4441 ccaccagtta gcattgaccg tgtaggtccc ttggaccocct ctatagtaag tcttgttgaa
4501 gaatctggaa ttgttgatgt tgggtgcccct gctcctatac cacaccctcc tacaacatct
4561 ggggttgaca ttgctacaac tgcagacaca acacctgcaa ttttagatgt acaagtggtt
4621 agcacacatg aaaatcctac ttttactgat ccactctgat tgcagcctcc tacacctgca
4681 gaaacatcag gtcatttact actttcatca tcatctatta gcacacataa ttatgaggaa
4741 atacctatgg atacatttat tgtttctact aataatgaaa acataacaag tagcacaccc
4801 attccagggg tgcgccctcc tgcacgttta ggggtatata gtaaggctac acaacaagta
4861 aaagtatttg atccaacggt tcttagtgct ccaaaacagc taattacata tgaaaacct
4921 gcctatgaaa ctgtaaagtc tgaagaatct ttatactttt ccaatacatc gcataatata
4981 gccctgatc ccgactttct agatattata gcattacata ggctgcccct tacctcagct
5041 aggaacactg ttagatatag tagactaggt aataaacaac ctttgccgac tctgtatggt
5101 gctactattg gtgcaagggt gcattattat tatgatatta gtagtattaa tctgcaggt
5161 gaaagtattg aaatgcaacc tttagggcgc tctgcaacta ctacttctac tttaaattag
5221 ggcttatatg acatttatgc agacactgat tttactgtgg atacacctgc cacacataat
5281 gtttcccctt ctactgctgt acagtccaca tctgctgtgt ctgcctatgt acctacaaat
5341 accactgtgc cactaagtac aggttttgac attcccatat tttctgggcc tgatgtacct
5401 atagagcatg cacctacaca ggttttccca tttcctttgg cccctacaac gccacaagtg
5461 tctatttttg ttgatggggg tgatttttat ttgcacccta gttattatat gtTAAacgt
                                L1 orf start ->
5521 cgacgtaaac gtgtatcata tttttttaca gATGtctctg tggcggccTA Gcgaggctac
                                L1 cds ->                <- L2 end
5581 tgtctactta ccactgtcc cagtgctctaa agttgtaagc acggatgaat atgtaacacg
5641 aaccaacata tattatcacg caggcagtgcc taggctgctt acagtaggcc atccatatta
5701 ttccatacct aaatctgaca atcctaataaa aatagttgta ccaaagggtgt caggattaca
5761 atatagggta tttagggttc gtttaccaga tccaaacaaa tttggatttc ctgatacatc
5821 tttttataat cctgaaactc aacgccttagt ttgggcctgt gttggtttag aggtaggtcg
5881 cgggcagcca ttaggtgtag gtattagtgg tcatccatta ttaaataaat ttgatgacac
5941 tgaaaactct aatagatatg ccggtggtcc tggcactgat aatagggat gtatatcaat
6001 ggattataaa caaacacaac tgtgtttact tgggtgcaaa ccacctattg gagagcattg
6061 gggtaaagggt agtccttgta gtaacaatgc tattaccctt ggtgattgtc ctccattaga
6121 attaaaaaat tcagttatac aagatgggga tatggttgat acaggctttg gagctatgga
6181 ttttactgct ttacaagaca ctaaaagtaa tgttcctttg gacatttgta attctatttg
6241 taaatatcca gattatctta aaatggttgc tgagccatat ggcgatacat tttttttta
6301 tttacgtagg gaacaaatgt ttgtaaggca tttttttaat agatcaggca cggttggtga
6361 atcgggtccc actgacttat atattaaagg ctccggttca acagctactt tagctaacag
6421 tacatacttt cctacaccta gcggctccat ggttacttca gatgcacaaa ttttaataa
6481 accatattgg atgcaacgtg ctcagggaca caataatggt atttgttggg gcaatcagtt
6541 atttgttact gtggtagata ccacacgtag taccaatag tctgttttg tgcgaattgc
6601 aaacagtgat actacattta aaagtagtaa ttttaagag tatttaagac atggtgagga
6661 atttgattta caatttata ttcagttatg caaaataaca ttatctgcag acataatgac
6721 atatatccac agtatgaatc ctgctatttt ggaagattgg aattttggat tgaccacacc
6781 tccctcaggt tctttggagg atacctatag gtttgcacc tcacaggcca ttacatgtca
6841 aaaaactgcc ccccaaaagc ccaaggaaga tccatttaaa gattatgtat tttgggaggt
6901 taatttaaaa gaaaagtttt ctgcagattt agatcagttt ccaactgggtc gcaaattttt
6961 attacagga ggatataggg cacgtcctaa atttaagca ggtaaacgta gtgcacctc
7021 agcatctacc actacaccag caaaacgtaa aaaaactaaa aagTAAtgga tgtgtatgta
                                <- L1 end
                                LCR ->
7081 atacatgtgt ctgtatgtgt atgtgcttgt gctgtattgt atatgtgtgt gtttgtgtgt
7141 tataatgggt atatgtatgt ttatgtatgc gtgtgtactt gtatatatgt atagtatggt
7201 atgtgtgtat gtatgctatg tatggtAATA Aatatgtgta tacctgtgtg tgttgtgtat
    signal ->

```



```

7261 gttgtcctta tatacacct attagtaaca TACTATTACT ATtttataaa ctaTTGTTCC
      -> repeat region
7321 TACTTGTTCC TACTtgttcc tgctcctccc aatagtcacg tacttatttc tgcctataat
end repeat region <-
7381 ttaggtgtca cgccatagta aaagtTGTAC ACCCGGTCCG ttttttgcaa ctaaagctac
      glucocorticoid receptor ->
7441 tccattttga ttttatgcag ccattttaa tcctaACCG TTTTCGGTg cattgtttaa
      E2 bind ->
7501 acatgctagt acaactatgc tgatgcagta gttctgcggt tTTTGTTTc ctgaatacta
      keratinocyte-dependent enhancer ->
7561 gtttttgcca acattctggc ttgtagtttc ctgcctaaca cacottgcca acatataatc
7621 cagtccaact ttgcaattat actatgaatc atgtttgttt aaatacaact gtagttcaac
7681 tatgtgtcat gcacatatat tatattatcc tacacacctt aaactgcttt taggcacata
7741 tttttagat tatctatatc cttgattgca gtgctggctt ttgcacatgt ttaaactgcc
7801 aaggttgtgt catgcattat aaataagttg tatgttactc atataattaa ttgcatatag
7861 gtattacACC GTTTTCGGTt acagttttac aagcaattgt tctttttata ct
      -> E2 bind

```

//

HPV33

LOCUS HPV33 7909 bp ds-DNA circular VRL 16-FEB-1987
DEFINITION Human papillomavirus type 33 (HPV-33), complete genome.
ACCESSION M12732
SOURCE Human papillomavirus type 33 DNA recovered from a human invasive cervical carcinoma, clone p15-5.
REFERENCE 1 (bases 1 to 7909)
AUTHORS Cole,S.T. and Streeck,R.E.
TITLE Genome organization and nucleotide sequence of human papillomavirus type 33, which is associated with cervical cancer
JOURNAL J. Virol. 58, 991-995 (1986)
COMMENT Draft entry and computer-readable copy of the sequence were kindly provided by S.T.Cole, 04-AUG-1986.

HPV-33 is considered to be a member of the "intermediate risk" anogenital group. All of the viruses in this group are most prevalent in high-grade intraepithelial lesions and less prevalent in invasive cancers. In a study which used 15 common anogenital probes to screen 2627 subjects, hybridization to the "intermediate risk" probes occurred in 26.4% of all high-grade intraepithelial neoplasias, 23.3% of all low-grade intraepithelial neoplasias, 13.7% of all invasive cancers, and in 3.3% of all normal samples. The relatively high percentage of high-grade lesions with respect to invasive cancers justifies the placement of type 33 in the "intermediate risk category. HPV33 is considered to be of moderate prevalence. Of all invasive cancer biopsies tested, 11% were positive for DNA in the "intermediate risk" group. HPV33 predominantly infects tissues of the cervix and other lower-anogenital tract sites: the vulva, the vagina, the penis, the perineum and the anus.

HPV-33 DNA was initially cloned from an invasive cervical carcinoma, using HPV-16 as a probe with reduced stringency. The complete 7909 base pair sequence begins at a site resembling the recognition sequence for HpaI, which is consistently used as the origin for other papillomavirus sequences. The similarity between HPV-16 and HPV-33 L1 peptides is 90% if the conservative substitutions are included. Cole et al. suggest that the two viruses' capsids are antigenically related.

The noncoding region of HPV-33 contains many features either common to all of the papilloma viruses or are common to the anogenital viruses. The first 223 base pairs of the noncoding region are unusually high in thymine plus guanine composition (79%). This feature is common among those types associated with anogenital lesions (6, 11, 16, 18, 31, and 33). Also found in this region are two copies of a 19 base pair repeat and seven copies of the motif TTGTRTR (where R is A or G). Three copies of the binding site for E2 activation or repression have been identified in HPV-33.

The most unique feature found in HPV-33 is a perfect 78- bp tandem repeat located 200 bp from the origin. No other repeats of this size or sequence had been found at the time of publication in any other papillomavirus sequences. In other viruses such as SV 40, Moloney murine leukemia virus and BK virus tandem repeats of a similar size have been shown to enhance transcription from Pol-II dependant promoters in a cis-active manner. Within the repeat itself is a section which may adopt a Z DNA structure. Z DNA regions are thought to attract regulatory molecules to eukaryotic promoters.

BASE COUNT 2544 a 1354 c 1537 g 2474 t
ORIGIN Site resembling recognition sequence for HpaI.
1 gtaaacTATA atgccaagtt ttaaaaaagt aggggtgtaAC CGAAAGCGGT tcaACCGAAA
signal -> -> E2 bind -> E2 bind
61 ACGGTgcata TATAAAgcaa acattttgca gtaaggctac gcacgactAT Gtttcaagac
-> signal

```

                <- LCR
    E6 orf start ->
121 actgaggaaa aaccacgaac attgcatgat ttgtgccaaag cattggagac aactatacac
181 aacattgaaac tacagtgcgt ggaatgcaaaa aaacctttgc aacgatctga ggtatatgat
241 tttgcatttg cagatttaac agttgtatat agagagggaa atccatttgg aatatgtaaa
301 ctgtgtttgc ggttcttatac taaaattagt gaatatagac attataatta ttctgtatat
361 ggaaatacat tagaacaac agttaaaaaa cctttaaagc aaatattaat taggtgtatt
421 atatgtcaaa gacctttgtg tcctcaagaa aaaaaacgac atgtggattt aaacaaacga
481 tttcataata tttcgggtcg ttgggcaggg cgctgtgcgg cgtgttgagg gtcccagcGT
541 AGagaaactg cactgTGAcg tgtaaaaacg ccATGagagg acacaagcca acgttaaagg
E7 orf start->                <- E6 end E7 cds ->
601 aatattgttt agatttatat cctgaaccaa ctgacctata ctgctatgag caattaagtg
661 acagctcaga tgaggatgaa ggcttgacc ggccagatgg acaagcacia ccagccacag
721 ctgattacta cattgtaacc tgttgtcaca cttgtaacac cacagttcgt ttatgtgtca
781 acagtacagc aagtgcacta cgaaccatac agcaactact tatgggcaca gtgaatattg
841 tgtgccctac ctgtgcacaa caaTAAacat catctacaAT Ggccgatcct gaaggtacaa
                <- E7 end
    E1 orf start ->          E1 cds ->
901 atggggctgg gatggggtgt actggttggg ttgaggtaga agcagtcata gagagaagaa
961 caggagataa tatttcagaa gatgaggatg aaacagcaga tgacagtggc acggatttac
1021 tagagtttat agatgattct atggaaaata gtatacaggc agacacagag gcagcccggg
1081 cattgtttaa tatacaggaa ggggaggatg atttaaattgc tgtgtgtgca ctaaaaacgaa
1141 agtttgccgc atgttcacaa agtgcctgcgg aggacgttgt tgatcgtgct gcaaaccctg
1201 gtagaacgct tattaataaa aataaagaat gcacatacag aaaacgaaa atagatgagc
1261 tagaagacag cggatatggc aatactgaag tggaaactca gcagatggta caacaggtag
1321 aaagtcaaaa tggcgacaca aacttaaagc acttagaatc tagtggggtg ggggatgatt
1381 cagaagtaag ctgtgagaca aatgtagata gctgtgaaaa tgttacgttg caggaaatta
1441 gtaatgttct acatagtagt aatacaaaaag caaatatatt atataaattt aaagaggcct
1501 atggataaag ttttatggaa ttagtaagac catttaaaag tgataaaaca agctgtacag
1561 attggtgtat aacaggatat ggaattagtc catcagtagc agaaagtta aaagtattaa
1621 taaacagca tagtttgtat actcatttac aatgtttaac ttgogataga ggaataataa
1681 tattattgtt aattagattt aggtgtagca aaaacaggtt aacagtagca aaactaatga
1741 gtaatttatt atcaatacct gaaacatgta tggttataga gccacaaaa ttacggagcc
1801 aaacatgtgc attgtattgg tttagaacag caatgtcaaa cattagtgat gtacaaggta
1861 caacacctga atggatagat agactaactg tttacaaca tagctttaat gataatatat
1921 ttgatttaag tgaatgggta cagtgggcat atgataacga gttaacggac gatagtgaca
1981 ttgcatatta ttatgcacaa cttgcagatt caaatagtaa tgctgctgca tttttaaaaa
2041 gtaactcaca agcaaaaata gtaaaggact gtggaataat gtgtagacat tataaaaaag
2101 cagaaaaacg taaaatgtca ataggacaat ggatacaaa tagatgtgaa aaaacaaatg
2161 atggaggaaa ttggagacca atagtacagt tgtaagata tcaaacatt gaatttacag
2221 catttttagg tgcatttaaa aagtttttaa aaggtatacc aaaaaaacg tgatgtctaa
2281 tttgtggacc agcaaataca ggaaagtc attttggaaat gaggtttaata cagtttttaa
2341 aagggtgtgt tatatcatgt gtaaattcta aaagtcactt ttggttgac ccattatcag
2401 atgcaaaaat aggaatgata gatgatgtaa cgccaataag ttggacatat atagatgatt
2461 acatgagaaa tgcgttagat ggaaatgaaa tttcaataga tgtgaaacat agggcattag
2521 tgcaattaaa atgtccacca ctgcttctta cctcaaac aaatgcaggc acagactcta
2581 gatggccata tttacatagt agattaacag tatttgaatt taaaaatcca ttoccatttg
2641 atgaaaatgg taaccagtg tatgcaataa atgatgaaaa ttggaaatcc tttttctcaa
2701 ggacgtggtg caaattagat ttaaTAGagg aagaggacaa ggaaaaccAT Ggaggaata
                E2 orf start ->          E2 cds ->
2761 tcagcacgtt taaatgcagt gcaggagaaa atactagatc ttacgaagc TGAtaaaaact
                <- E1 end
2821 gatttaccat cacaaattga acattggaaa ctgatacgc tggagtgtgc tttattgtat
2881 acagccaaac aaatgggatt ttcacattta tgccaccagg tgggtgcctc tttgttagca
2941 tcaaagacca aagcatttca agtaattgaa ctacaaatgg cattagagac attaagtaaa
3001 tcacagtata gtacaagcca atggacattg caacaacaa gcttagaggt gtggctttgt
3061 gaaccaccaa aatgttttaa aaaacaagga gaaacagtaa ctgtgcaata tgacaatgac
3121 aaaaaaata caatggatta tacaactgg ggtgaaatat atattataga ggaagataca
3181 tgtactatgg ttacagggaa agtagattat ataggtatgt attatataca taactgtgaa
3241 aaggatattt taaatattt taaagaggat gctgcaaaat attctaaaac acaaatgtgg
3301 gaagtacatg tgggtgggtca ggTAAAttgt tgtcctacgt ctatatctag caaccaataa
    E4 orf start ->
    NH2 terminus unknown
3361 tccactactg aaactgctga catacagaca gacaacgata accgaccacc acaagcagcg

```

HPV33

```

3421 gccaaacgac gacgacctgc agacaccaca gacaccgccc agccccttac aaagctgttc
3481 tgtgcagacc ccgccttggc caatagaaca gcacgtactg caactaactg cacaaacaag
3541 cagcggactg tgtgtagttc taacgttgca cctaTAGtgc atttaaaagg tgaatcaaat
                                     <- E4 end
3601 agtntaaat  gtttaagata  cagattaaaa  ccttataaag  agttgtatag  ttctatgtca
3661 tccacctggc  attggaccag  tgacaacaaa  aatagtaaaa  atggaattgt  aactgtaaca
3721 tttgtaactg  aacagcaaca  acaaatgttt  ttaggtaccg  taaaaatacc  acctactgtg
3781 caaataagta  ctggatttat  gacattaTAA  gtgtacatca  caagccaata  tgtgtgtcTA
                                     <- E2 end
3841 Attgtatata  accATGatat  ttgtttttgt  attatgtttt  atattgtttt  tatgcttatac
      -> E5 orf start
      E5 cds ->
3901 cttattatta  cgtcctttaa  tactttccat  ttctacctat  gcttggttgc  tgggtgttgg
3961 attgctgctt  tgggtgtttg  tgggatctcc  tttaaaaaatt  tttttttgct  atttgttgtt
4021 tttatattta  ccaatgatgt  gtattaattt  tcatgcacag  catatgacac  aacaagagTA
4081 Atgtatatac  atgtatata  tgtttgtata  tatgtgcaca  tgggtgtgtt  ttaacattgt
      <- E5 end
4141 tgtgttatt  ttagtttttt  tttttttgta  ttactaataa  atacctttat  atttTAGcag
                                     L2 orf start ->
4201 tgtattattA  TGagacacaa  acgatctaca  aggcgcaagc  gtgcatctgc  aacacaacta
      L2 cds ->
4261 taccaaacat  gcaaggccac  aggcacctgc  ccaccgatg  ttattcctaa  agtgggaagga
4321 agtaccatag  cagatcaaat  tcttaaatat  ggcagttag  gggttttttt  tgggtgttta
4381 ggtattggca  caggctctgg  ttcagggtga  aggactggct  atgtacctat  tggtagctac
4441 ccacctacag  ctgcaatccc  cttgcagcct  atacgtctc  cggttactgt  agacactgtt
4501 ggacctttag  actcgtctat  agtgtcatta  atagaagaaa  caagttttat  agaggcaggt
4561 gcaccagccc  catctattcc  tacaccatca  ggttttgatg  ttactacatc  tgcagatact
4621 acacctgcaa  ttattaatgt  ttcactctgt  ggggagtcac  ctatcctaac  tatttctaca
4681 catttaaate  ccacatttac  tgaaccatct  gtactacacc  ctccagcgcc  tgcagaagcc
4741 tctggacatt  ttatattttc  ttcccctact  gttagcacac  aaagttatga  aaacatacca
4801 atggatacct  ttgttgtttc  cacagacagt  agtaatgtaa  catcaagcac  gccatttcca
4861 gggctctgcc  ctgtggcagc  ccttggttta  tatagtcgca  atACCCAACA  GGTaaaggt
                                     E2 bind site ->      <-
4921 gttgacctg  cttttttaac  atgcctcat  aaactataa  catatgataa  tcctgcattt
4981 gaaagctttg  acctgaaga  cacattacaa  tttcaacata  gtgatatac  acctgctcct
5041 gatcctgact  ttctagata  tattgcatta  cataggcctg  ctattacatc  tcttagacat
5101 actgtgcgtt  ttagtagagt  aggtcaaaaa  gccacactta  aaactcgcag  tggtaaacaa
5161 attggagcta  gaatacatta  ttatcaggat  ttaagtctca  ttgtgccttt  agaccacacc
5221 gtgcaaatg  aacaatatga  attacagcct  ttacatgata  cttctacatc  gtcttatagt
5281 attaatgatg  gtttgtatga  tgtttatgct  gacgatgtgg  ataattgata  caccctaatg
5341 caacactcat  acagtacgtt  tgcaacaaca  cgtaccagca  atgtgtctat  acctttaaate
5401 acaggatttg  atactcctgt  tatgtctggc  cctgatatac  cttccccttt  atttcccaca
5461 tctagcccat  ttgttctat  ttgcctttt  tttcctttt  acaccattgt  tgTAGacggt
                                     L1 orf start ->
5521 gctgactttg  ttttacatcc  tagttatttt  attttacgtc  gcaggcgtaa  acgttttcca
5581 tattttttta  cagATGtccg  tgtggcggcc  TAGtgaggcc  acagtgtacc  tgctcctgt
      L1 cds ->      <- L2 end
5641 acctgtatct  aaagtgtgca  gcaactgatga  atatgtgtct  cgcacaagca  tttattatta
5701 tgctggtagt  tccagacttc  ttgctgttgg  ccatccatat  ttttctatta  aaaatcctac
5761 taacgctaaa  aaattattgg  taccctaaagt  atcaggcttg  caatataggg  tttttagggt
5821 ccgtttacca  gatcctaata  aatttggatt  tcctgacacc  tccttttata  acctgatac
5881 acaacgatta  gtatgggcat  gtgtaggcct  tgaaataggt  agagggcagc  cattaggcgt
5941 tggcataagt  ggtcatcctt  tattaacaaa  atttgatgac  actgaaaccg  gtaacaagta
6001 tcctggacaa  ccgggtgctg  ataataggga  atgtttatcc  atggattata  aacaaacaca
6061 gttatgttta  cttggatgta  agcctccaac  aggggaacat  tggggtaaaag  gtgttgcctg
6121 tactaatgca  gcacctgcca  atgattgtcc  accttagaa  cttataaata  ctattatgta
6181 ggtatggatg  atgggtgaca  caggatttgg  ttgcatggat  tttaaaacat  tgcaggctaa
6241 taaaagtgat  gttcctattg  atatttggg  cagtacatgc  aaatatccag  attattttaa
6301 aatgactagt  gagccttatg  gtgatagttt  atttttctt  cttcgacgtg  acaaatgtt
6361 tgtaagacac  ttttttaata  gggctgttac  attaggagag  gctgttccc  atgacctgta
6421 cattaaaggt  tcaggaacta  ctgcctctat  tcaaagcagt  gctttttttc  ccactcctag
6481 tggatcaatg  gttacttccg  aatctcagtt  atttaataag  ccatattggc  tacaacgtgc
6541 acaaggtcat  aataatggta  tttgttggg  caatcaggta  tttgttactg  tggtagatac
6601 cactcgcagt  actaatatga  ctttatgcac  acaagtaact  agtgacagta  caTATAAAAA

```

```

                                signal ->
6661 tgaaaat ttt aaagaatata taagacatgt tgaagaatat gatctacagt ttgtttttca
6721 actatgcaaa gttaccttaa ctgcagaagt tatgacatat attcatgcta tgaatccaga
6781 tatttttagaa gattggcaat ttggtttaac acctcctcca tctgctagtt tacaggatac
6841 ctataggttt gttacctctc aggctattac gtgtcaaaaa acagtacctc caaaggaaaa
6901 ggaagacccc ttaggtaaaat atacat tttg ggaagtggat ttaaaggaaa aat tttcagc
6961 agatttagat cagtttcctt tgggacgcaa gtttttatta caggcaggtc ttaaagcaaa
7021 acctaaactt aaacgtgcag cccccacatc caccgcaca tctgtctgcaa aacgcaaaaa
7081 ggttaaaaaa TAAcactTTG TGTAattgtg ttagt tttgtt gttttgttct gtctatgtac
                                LCR -> -> repeat region start
                                <- L1 end
7141 tttgtgttgt tgtg tttgtgt tgttgtttgt tttTTGTGTA tgtgttacia tgtatgttat
7201 gTTGTATGtt actgtgtttg ttttatgtgt acttgtTTGT GTGcatgttc tatgtacttg
7261 tcagtttctt gtTTGTGTA atgttAATAA AacaTTGTGT Gtatttg tta aactatTTGT
                                signal ->
7321 ATGtatgtta tgtatatggg tgtacctata tgagtaagga gttgtattgc ttgccctacc
                                <- repeat region end
7381 ctgcattgca atgtacctac ctttatttcc ctatatttgt agtacctaca tgtttagtat
7441 tgctttacct tttgacatac tagtgtccat attgtacaat ttcctccatt ttgtatgcct
7501 aACCGTTTTTC GGTtacttgg catacatacc ctatgacatt ggcagaacag ttaatccttt
                                -> E2 bind                                -> 78 base pair repeat
7561 tctttcctgc actgtgtttg tctgtacttg ctgcattggc atacatacco tatgacattg
                                78 base pair repeat                                <--> 78 base pair repeat

```

HPV33

```
7621 gcagaacagt taatcctttt ctttcctgca ctgtgtttgt ctgtacttgc tgcattgact
      78 base pair repeat <-
7681 catatataca tgcagtgcaa ttgcaaaata ctttaattgta ctaatagttt acacatgctt
7741 ttaggcacat atttttactt tactttcaaa ccttaagtgc agttttggct tacacaattg
7801 ctttgtatgc caaactatgc cttgtaaaag tgagtcacta cctgtttatt accagggtgtg
7861 gactaacctg tttaggcat attggtcatt TATAatcttt TATAaata
      signal -> -> signal
```

LOCUS HPV35 7851 bp ds-DNA VRL 30-JAN-1992
DEFINITION Human papillomavirus type 35 (HPV35), complete genome.
ACCESSION M74117
KEYWORDS complete genome; major capsid protein; minor capsid protein;
regulatory protein; replication protein; transformer protein.
SOURCE Human papillomavirus type 35 DNA recovered from a cervical
adenocarcinoma.
REFERENCE 1 (sites)
AUTHORS Lorincz,A.T., Quinn,A.P., Lancaster,W.D. and Temple,G.F.
TITLE A New Type of Papillomavirus Associated with Cancer of the Uterine
Cervix
JOURNAL Virology 159, 187-190 (1991)
REFERENCE 2 (bases 1 to 7851)
AUTHORS Marich,J.E., Pontsler,A.V., Rice,S.M., McGraw,K.A. and
Dubensky,T.W.
TITLE The phylogenetic relationship and complete nucleotide sequence of
human papillomavirus type 35
JOURNAL Virology 186, 770-776 (1992)
COMMENT HPV-35 is considered to be a member of the "intermediate risk"
anogenital group. All of the viruses in this group are most
prevalent in high-grade intraepithelial lesions and less prevalent
in invasive cancers. In a study which used 15 common anogenital
probes to screen 2627 subjects, hybridization to the "intermediate
risk" probes occurred in 26.4% of all high-grade intraepithelial
neoplasias, 23.3% of all low-grade intrepithelial neoplasias, 13.7%
of all invasive cancers, and in 3.3% of all normal samples. The
relatively high percentage of high-grade lesions with respect to
invasive cancers justifies the placement of type 35 in the
"intermediate risk category. HPV35 is considered to be of moderate
prevalence. Of all invasive cancer biopsies tested, 11% were
positive for DNA in the "intermediate risk" group. HPV35
predominantly infects tissues of the cervix and other
lower-anogenital tract sites: the vulva, the vagina, the penis, the
perineum and the anus. It has also been detected in a bowenoid
lesion of the finger.

The 7851 base pair genome of HPV-35 was first recovered from a
cervical adenocarcinoma. The numbering of the sequence was
determined by similarity to HPV-31. In the E6 orf, Marich et al.
identified a putative splice/donor acceptor pair which corresponds
to those seen in other oncogenic HPV types (16, 18, 31, and 33).
They also noted sequence similarity between E7 and the adenovirus 5
E1A retinoblastoma binding site, as seen in HPV-16.

The long control region (LCR) of HPV-35 contains features conserved
for all or many of the papillomaviruses or for just those associated
with anogenital lesions. A feature which appears to be common among
the mucosal types is the glucocorticoid response element. These
elements have been shown to mediate hormonal response in the
presence of glucocorticoids. Besides the presence in HPV-35,
potential GREs have been identified in types 6, 11, 16, 18, 31, 33,
and 39. However, Marich et al. believe that their irregular
distribution within the LCR brings up questions about their role in
the life cycle of the mucosal HPVs. In HPV-35, two putative AP-1
regions, and five putative NF-1 regions are present. Also found in
the LCR of many HPV types are tandem direct repeats, direct repeats,
and inverted repeats. HPV-35 contains tandem direct repeats of 8
bp, direct repeats of 11 bp, and inverted repeats of 7 and 8 bp.
The CK-octamer motif has been identified in HPV type 6, 11, 16, 18,
and 33. This element is present in HPV-35 and may explain the
tissue specificity of HPV infection. Three putative E2 binding
sites with the consensus ACCN₆GGT have been identified in HPV-35.

Also of particular interest, Marich et al. identified a 20 bp
sequence which is conserved between the oncogenic mucosal types
(16, 18, 31, 33, 35, 39, and 51) and has not been found in other

HPV35

nononcogenic types (1a, 2a, 5, 6, 8, 9, 11, 17, 19, 20, 25, 36, 47 or 57). This 20 bp region is located approximately 30 bp 5' to the keratinocyte-specific octamer.

```

BASE COUNT      2553 a   1343 c   1568 g   2387 t
ORIGIN
  1 ccctataaaa aaaacaGGGA GTgACCGAAA ACGGTcgtAC CGAAAACGGT tgcCATAAAA
      Sp-1 bind ->      -> E2 bind      -> E2 bind      -> signal
                                E6 orf start ->
 61 gcagaagtgc acaaaaaagc agaagtggac agacattgta aggtgcggTA TGtttcagga
                                E6 cds ->
                                <- LCR
121 cccagctgaa cgaccttaca aactgcatga tttgtgcaac gaggtagaag aaagcatcca
181 tgaatttgtg ttgaattgtg tatactgcaa acaagaatta cagcggagtg agGTatatga
                                5' sj /\
241 ctttgcattg tatgatttgt gtatagtata tagagaaggc cagccatagt gagtattgcat
301 gaaatgttta aaattttatt caaaaataag tgaatataga tggatatagat atagtgtgta
361 tggagaaacg ttagaaaaac aatgcaacaa acagttatgt catttattaa ttAGgtgtat
                                /\ 3' sj
421 tacatgtcaa aaaccgctgt gtccagttga aaagcaaaga catttagaag aaaaaaacg
481 attccataac atcgggtggc ggtggacagg tcggtgtatg tcctgttggg aaccaacacg
541 tagagaaacc gaggtgTAAt cATGcatgga gaaataacta cattgcaaga ctatgtttaa
      <- E6 end
      E7 cds ->
601 gatttggAAC cggaggcaac tgACCTATAC TGTtatgagc aattgtgtga cagctcagag
      -> E2 bind
661 gaggaggaag atactattga cggctccagct ggacaagcaa aaccagacac ctccaattat
721 aatattgtaa cgtcctgttg taaatgtgag gcgacactac gtctgtgtgt acagagcaca
781 cacattgaca tacgtaaatt ggaagattta ttaatgggca catttggaat agtgtgcccc
841 ggctgttcac agagagcaTA AtctacaATG gctgatcctg caggtacaga tgaaggggag
      <-E7 end
      E1 cds ->
901 gggacgggat gtaatggatg gttttttgtg gaagcagtag ttagtagacg tacgggatcc
961 agttagagag acgaaaaatga agatgactgt gacagggggg aggatatggt ggactttata
1021 aatgatacag atatatataaa catacaggca gaaacagaga cagcacaagc attatttcat
1081 gcacaggagg agcaaacaca caaagaggct gtacaggtcc taaaacgaaa gtatgctagt
1141 agtccactta gcagcgtgag cttatgtggt aataataaca taagtccacg tttaaaagct
1201 atttgcattg aaaataaaaa tacagcagca aagcgacgat tatttgaact accagacagc
1261 ggttatggca attctgaagt ggaaatacac gagatacaac aggtagaggg gcatgataca
1321 gttgaacaat gtagtatggg cagtggggat agtataacct ctagttagcga tgaagacat
1381 gatgagactc caacgcgaga cataatacaa atactaaaat gtagttaatgc aaacgcagct
1441 atgttggcta aatttaaaga actatttggg attagtttta cagaacttat tagaccattt
1501 aagagtgata aatccacatg tacagattgg tgtgtggccg catttggaat agccccaggt
1561 gtggcgaact ttaaacatat aacatatgta tacatataca atgtttatcg tgttcattggg
1621 gctatggtaa ttctagcatt attacgattt aaagtcgaaa aacgagaaca acaattgaaa
1681 actattgatg ctaaattgct atgtatttca gctgcaagta tgctaataca accaccaaaa
1741 ttacgtagta cccagctgac gttatattgg tttaaaacag caatgtcaaa tattagttag
1801 gttgatggag aaacaccaga atggattcaa agacaaacag tattacagca tagttttaa
1861 gatgcaatat ttgacctatc tgaaatggta caatgggcat atgacaatga ttttatagat
1921 gatagtgata tagcatataa atatgcacaa ttggcagaaa ctaaatagtaa tgcattgtgct
1981 tttttaaaaa gtaattcgca agctaaaatt gtaaaagatt gtgcaacaat gtgtagacat
2041 tataaacgag ctgaaaaaag agaaatgaca atgtcacagt ggattaaaag gcgatgtgca
2101 caggtggagc atgacgggta ctggagggac atagtacgat ttttaagata tcaacaagta
2161 gattttgtgg catttttatc tgcactaaaa aattttttac atgggtgtgcc taaaaaaaat
2221 tgcatactaa tatatggagc accaaacaca ggtaaatcat tatttggaat gagtctaattg
2281 catttcttac aaggagctat tatatcctat gtaaatctta aaagccattt ttgggtgcag
2341 ccattatag atgccccaaat agctatgta gatgatgcta catcgccatg tggcatatat
2401 agaccaatat ttaagaaatg cactagatgg aaatcctata tttcatttag atgtaaagca
2461 ttaagcatag tgcataataa gccaccttt acttattaca tcaatataaa tgcaggcaaa
2521 gatgacaggt ggccatactt acatagcagg gtatgtgtct ttacatttca caatgaattc
2581 ccatttgata aaaatggaaa cccagagtat gggcttaatg ataaaaactg gaaatccttt
2641 ttctcaagga cgtgggtgac attaaatttg cacgaggaag aggtcaaaaga aaATGatgga
                                E2 cds ->
2701 gacgctttcc cagcgtttta gtgtgtgtca ggacaaaata ctagaacatt acgagacTGA
                                E1 end <-

```



```

2761 tagcacatgt ttgtctgac acatacagta ttggaaactg attcgtcttg aatgtgcagt
2821 atttTATAAA gcaagagaaa tgggaattaa aactcttaac caccaagtgg ttccaacgca
signal ->
2881 ggccatttca aaagccaaag caatgcaagc aattgaaactg caattaatgt tagagacatt
2941 aaatacaact gagtatagca cagaggactg gacactgcaa gaacaagta ttgaactata
3001 tacaacagtt cctacaagat gtttaaaaaa agatgtttat actgtggaag cacaatttga
3061 tggtgataaa caaaatacta tgcattatac taattggaca catatatata tattagagga
3121 cagtatatgt actgttgtaa agggactggg aaattataaa ggtatttatt atgtgcatca
3181 ggggttagaa acatattatg ttacttttag ggaagaggct aaaaagtatg gaaaaaaaaa
3241 tatatgggaa gtgcatgtgg gtggtcaggT AAttgtttgt cctgaatctg tatttagcag
E4 orf start ->
NH2 terminus unknown
3301 cacagaacta tccactgctg aaattgctac acagctacac gcctacaaca ccaccgagac
3361 ccataccaaa gcctgctccg tgggcaccac agaaaccag aagacaaatc acaaacgact
3421 tcgagggggg accgagctcc cctacaaccc caccaagcga gtgogactca gtgccgtgga
3481 cagtgttgac agaggggtct actctacatc tgactgcaca aacaaagacc ggtgtggtag
3541 ttgtagtaca actacaccta TAGTtacattt aaaagggtgat gcaaatacat taaagtgttc
<- E4 end
3601 aagatataga ttgggtaaat ataaagcatt gtatcaagat gcttcatcta catggagatg
3661 gacatgtaca aacgataaaa aacaaatagc aattgtaaca ttaacttaca caacagaata
3721 tcaaagggat aaatttttaa ctacagtaaa aataoctaac acagttacag tgtctaagg
3781 atatatgtct atATGAtaga ccttacagct tccagtactg tgttgctgtg ctttttgttg
<- E2 end
E5 cds ->
3841 tgcttttgtg tgcttttgtg cttgtgtctg cttgtacgtt cgctattgct atctgtgtca
3901 ttatactcag cattaatatt actggtttta atactgtggg ttactgtagc aacaccacta
3961 cttgcttttg ttgtttcttg cttttgtata tacctatgga tgattaacgc tcatgcacaa
4021 tatttggcag tacagTAAtt gtatacaaac attgtgtttg gtactgtgta acatgtgtgt
<- E5 end
4081 atgggtggtt tatttttgtg tgttcattgt atattttgtt tttttactgt ttttaaacat
4141 ttttatttct gtgttttAA TAAAtgac acatggtata accATGcgac acaaaaagtc
signal -> <- L2 cds ->
4201 taaaaaacgt gttaaactg catctgcaac acaactatat cgtacttga aagctgcagg
4261 aacttgcca ccagatgta tacctaagg tggggaat actgttgctg atcaaatttt
4321 aaaatatggc agcatggctg tgtttttgg ggggttagga attggttctg gatctggcac
4381 aggtggaaga tctggatag ttccactggg tacaacacct ccaacggctg ccacaaacat
4441 tcctatacga cccctgtaa ctgtgaaaag tataccatta gacacaattg gccctttaga
4501 ttcttctata gtgtcattag tagaggaaac tagttttatt gagtctggtg cccctgttgt
4561 tacaccaagg gtcccaccta caacaggttt tacaataacc acatctacag ataccacacc
4621 tgctatttta gatgtgacat ccataagtac acatgataat cctacttca ctgacccctc
4681 tgttttacac ccaccacgc ctgcagaaac ttcaggatcat ttgtacttt catcatctc
4741 tattagtaca cataattatg aagaaatccc tatggatact tttattgtt ccacagacag
4801 caataatata actaatagca cgcctattcc agggctctgc cctacgacac gcctaggatt
4861 atatatgaaa ggtACCCAGC AGGTTaagg tgttgaccct gcctttatga cttctcctgc
-> E2 bind
4921 aaaacttatt acatatgata atcctgcata tgaaggcctt aacctgata caaccttaca
4981 atttgagcat gaggatatta gcttagctcc ggatocctgac ttatggaca ttatagcttt
5041 acataggcct gcactaacat ctaggaaaag cactattaga tatagttag taggtaataa
5101 acgtactatg catacacgaa gtggaaaagc tataggggca cgggtacatt attatcagga
5161 tttaagtagt attactgaag atatagaatt acaacccta caacatgtac catcctcttt
5221 accacatacc actgtttcaa catcattaaa tgatggtag tttgatatt atgctcctat
5281 agatactgag gaagatatta tttttcagc atcttctaac aatactttat atactacatc
5341 taacactgca tatgttctca gcaactactac tataccatta agtagtggct atgatattcc
5401 tataacagca gggccagaca ttgtatttaa ctctaatact attactaact ctgtactacc
5461 ggtaccacaca ggtcctatat attctattat tgcagatggg ggtgactttt atttacacc
5521 tagttattat ttattaaaac gacgtcgtaa agctatccca ttttttttg cagATGctc
L1 cds ->
5581 tgtggcggtc TAAcgaagcc actgtctacc tgcctccagt gtcagtgtct aaggttgtta
<- L2 end
5641 gcaactgatga atatgtaaca cgcacaaaca tctactatca tgcaggcagt tctaggctat
5701 tagctgtggg tcaccatac tatgtctatta aaaaacaaga ttctaataaa atagcagtac
5761 ccaaggtatc tggtttgcga tacagagtat ttagagtaaa attaccagat cctaataagt
5821 ttggatttcc agacacatca ttttatgac cctgcctcca cgttttgggt tgggctgta
5881 caggagtga agtaggtcgt ggtcagccat taggagttag tattagtgg catcctttat

```

HPV35

```
5941 taaataaatt ggatgatact gaaaatctta ataaatatgt tggtaactct ggtaactctg
6001 gtacagataa cagggaatgc atttctatgg attataaaca aacacaattg tgtttaatag
6061 gttgtaggcc tcctataggt gaacattggg gaaaaggcac accttgtaat gctaaccagg
6121 taaaagcagg agaatgtcct cctttggagt tactaaacac tgtactacaa gacggggaca
6181 tggtagacac aggatttggt gcaatggatt ttactacatt acaagctaat aaaagtgatg
6241 ttcccctaga tatatgcagt tccatttgca aatatcctga ttatctaaaa atggtttctg
6301 agccatatgg agatatgtta tttttttatt tacgtaggga gcaaatgttt gttagacatt
6361 tatttaatag ggctggaact gtaggtgaaa cagtacctgc agacctatat attaagggta
6421 ccactggcac attgcctagt actagttatt ttcctactcc tagtggctct atggtaacct
6481 ccgatgcaca aatatttaat aaACCATATT GGTGcaacg tgcacaaggc cataataatg
      -> E2 bind
6541 gtatttggtg gagtaaccaa ttgtttgtaa ctgtagttga tacaaccctg agtacaaata
6601 tgtctgtgtg ttctgctgtg tcttctagtg acagtacaTA TAAAAtgac aattttaagg
      -> signal
6661 aatatttaag gcatggtgaa gaatatgatt tacagtttat ttttcagtta tgtaaaataa
6721 cactaacagc agatggtatg acatatattc atagtatgaa cccgtccatt ttagaggatt
6781 ggaattttgg cttacacca ccgccttctg gtaccttaga ggacacatat cgctatgtaa
6841 catcacaggc tgtaacttgt caaaaaccca gtgcacaaa acctaaagat gatccattaa
6901 aaaattatac tttttgggag gttgatttaa aggaaaagt tctcgcagac ttagatcagt
6961 ttccggttgg ccgtaaattt ttgttacaag caggactaaa ggccaggcct aattttagat
7021 taggcaggcg tgcagctcca gcatctacat ctaaaaaatc ttctactaaa cgtagaaaag
```

```

7081 taaaaagtTA ATGTGTAAAT GTGTAtgcat gtatactgtg tgTTATGTGT TGTagtgctt
      -> repeat region start
      <- L1 end
      LCR ->
7141 gtatatataT TATGTGTTGT ggtgcctggt tgtgttgtag atggcgtgta aatgtgtgta
      repeat region end <-
7201 taatattgtg caatgtggtg tacgtgggtg tttttgtact tagtgtgtag tagttcagta
7261 gccataaagt gatgtgtgtg tttataatta aactgtatt gttgtatgac tatgggtgcac
7321 cgatatgagc ttacataatt acatgacagc tatattgtgt atataaataa tctacctcca
7381 ttttgtgtgt tagtgtcctt tacattacct ttcaACCGAT TTCGGTtgc tttggtaagc
      -> E2 bind
7441 tttatatggt ttttcaaaa acattcctac ctcagcAGAA CACTTAATCC Ttgtgttctt
      glucocorticoid response element -> <-
7501 gatatatatt gttTGCCAAc tttataTTGG CTtTGCCAA tctttaaact tgattcatct
      NF-1 bind -> NF-1 bind -> -> NF-1 bind
7561 tgcagtatta gtcatttttc atacttgggg tccaccaca cttgtaacac ttgtaacagt
7621 gcttttaggc acatattttt tgcatttcta aagggttta attgcacacc TTGGCTttac
      NF-1 bind ->
7681 atattatggt tgttTGCCAA caccacccta cacatcctgc caactttaag ttaaaaacatg
      NF-1 bind ->
7741 catgtaaac attactcact gtattacaca ttgttatatg cacacaggtg tgtccaaccg
7801 atttgatta cagttttata agcatttctt tttattatag ttagtaacaa t

```

HPV35h

LOCUS HPV35h 7879 bp ds-DNA VRL 04-OCT-1993
DEFINITION Human papillomavirus type 35h (HPV-35h), complete genome.
ACCESSION X74477
SOURCE Human papillomavirus type 35h DNA.
REFERENCE 1 (bases 1 to 7879)
AUTHORS Delius,H. and Hofmann,B.
TITLE Primer-directed sequencing of human papillomavirus types
JOURNAL Curr. Top. Microbiol. Immunol. 186, 13-31 (1994)
REFERENCE 2 (bases 1 to 7879)
AUTHORS Delius,H.
TITLE Direct Submission
JOURNAL Submitted (06-AUG-1993) to the EMBL/GenBank/DDBJ databases. H. Delius, Deutsches Krebsforschungszentrum, Abteilung ATV, Im Neuenheimer Feld 506, W 6900 Heidelberg, FRG

COMMENT HPV-35h is considered to be a member of the "intermediate risk" anogenital group. All of the viruses in this group are most prevalent in high-grade intraepithelial lesions and less prevalent in invasive cancers. In a study which used 15 common anogenital probes to screen 2627 subjects, hybridization to the "intermediate risk" probes occurred in 26.4% of all high-grade intraepithelial neoplasias, 23.3% of all low-grade intraepithelial neoplasias, 13.7% of all invasive cancers, and in 3.3% of all normal samples. The relatively high percentage of high-grade lesions with respect to invasive cancers justifies the placement of type 35h in the "intermediate risk category. HPV35h is considered to be of moderate prevalence. Of all invasive cancer biopsies tested, 11% were positive for DNA in the "intermediate risk" group. HPV35h predominantly infects tissues of the cervix and other lower-anogenital tract sites: the vulva, the vagina, the penis, the perineum and the anus. It has also been detected in a Bowenoid lesion of the finger.

The 7879 bp genome of HPV-35h was cloned in vector pBR322. As is common in other papillomaviruses, the E4 ORF lacks an initiating methionine codon. Of note, the E6 ORF exhibits putative splice donor and acceptor sites similar to those seen in other oncogenic types (16, 18, 31, and 33).

The long control region (LCR) of HPV-35h contains features which are either conserved among all or many of the papillomaviruses or are conserved among just those associated with anogenital lesions. A feature which appears to be common among the mucosal types is the glucocorticoid response element. These elements have been shown to mediate the hormonal response in the presence of glucocorticoids. Besides the presence in HPV-35h, potential GREs have been identified in types 6, 11, 16, 18, 31, 33, and 39. In HPV-35h, five putative NF-1 regions are present. Also found in the LCR of many HPV types are tandem direct repeats, direct repeats, and inverted repeats. Like those other HPVs, HPV-35h contains tandem direct repeats of 8 bp, and direct repeats of 11 bp.

Of particular interest is the presence of a 20 bp sequence which is conserved between the oncogenic mucosal types (16, 18, 31, 33, 35h, 35, 39, and 51) and has not been found in other non-oncogenic types (1a, 2a, 5, 6, 8, 9, 11, 17, 19, 20, 25, 36, 47 or 57). This 20 bp region is located approximately 30 bp 5' to the keratinocyte-specific octamer.

BASE COUNT 2570 a 1339 c 1570 g 2400 t
ORIGIN 109 bp upstream from beginning of E6 cds
1 ccctataaaa aaaacaGGGA GTgACCGAAA ACGGTcgtAC CGAAAACGGT tgccaTAAaa
Sp-1 bind -> E6 orf start ->
-> E2 bind -> E2 bind
61 gcagaagtgc acaaaaaagc agaagtggac agacattgta aggtgCGGT TgTttcagga
E6 cds ->
-< LCR

```

121 cccagctgaa cgaccttaca aactgcatga tttgtgcaac gaggtagaag aaagcatcca
181 tgaaatattgt ttgaattgtg tatactgcaa acaagaatta cagcggagtg agGTatatga
                    5' sj /\

241 ctttgcattgc tatgatttgt gtatagtata tagagaaggc cagccatatt gagtatgcat
301 gaaatgttta aaattttatt caaaaataag tgaatataga tggatatagat atagtgtgta
361 tggagaaaacg ttagaaaaac aatgcaaaa acagttatgt catttattaa ttAGgtgtat
                    /\ 3' sj

421 tacatgtcaa aaaccgctgt gtccagttga aaagcaaaga catttagaag aaaaaaacg
481 atccataac atcgggtggac ggtggacagg tccgtgtatg tcctgttgga aaccaaacg
541 TAGagaaaacc gaggtgTAA cATGcatgga gaaataacta cattgcaaga ctatgtttta
E7 orf start ->                E7 cds ->
                                <- E6 end
601 gatttgaac ccgaggcaac tgACCTATAC TGTtatgagc aattgtgtga cagctcagag
                                -> E2 bind
661 gaggaggaag atactattga cgggtccagct ggacaagcaa aaccagacac ctccaattat
721 aatattgtaa cgtcctgttg taaatgtgag gcgacactac gtctgtgtgt acagagcaca
781 cacattgaca tacgtaaatt ggaagattta ttaatgggca catttggaa agtgtgcccc
841 ggctgttcac agagagcaTA AtctacaATG gctgatcctg caggtacaga tgaaggggag
                                E1 cds ->
E1 orf start ->
                                <- E7 end
901 gggacgggat gtaatggatg gttttttgta gaagcagtag ttagtagacg tacgggggat
961 ccagtgctcag aggacgaaaa tgaagatgac tgtgacaggg gggaggatag ggtggacttt
1021 ataaatgata cagatatatt aaacatacag gcagaaacag agacagcaca agcattattt
1081 catgcacagg aggagcaaac acacaaagag gctgtacagc tcctaaaaacg aaagtatgct
1141 agtagtccac ttagcagcgt gagcttatgt gttaataata acataagtc acgtttaaaa
1201 gctatttga ttgaaaaata aaatacagca gcaaagcgac gattatttga actaccagac
1261 agcggttatg gcaattctga agtgaaaata cagcagatag aacaggtaga ggggcatgat
1321 acagttgaac aatgtagtat gggcagtggt gatagtataa cctctagtag cgatgaaaga
1381 catgatgaga ctccaacgag agacataata caaataacta aatgtagtaa tgcaaacgca
1441 gctatgttg ctaaaattaa agaactattt ggtattagtt ttacagaact tattagacca
1501 ttaagagtg ataaatccac atgtacagat tgggtgtgtg cgcatttgg aatagcccca
1561 agtgtggcgg aaagttaaa aacattaatt aaaccatatt gtttatatat acatatacaa
1621 tgtttatcgt gttcatgggg tatggttaatt ctagcattat tacgatttaa atgtgcaaaa
1681 aacagaacaa caattgaaaa actattatca aaattgctat gtatttcagc tgcaagtatg
1741 ctaatacaac caccaaaatt acgtagtacc ccagctgcgt tatattgggt taaaacagca
1801 atgtcaaaa ttagtgaggt tgatggagaa acaccagaat ggattcaag acaaacagta
1861 ttacagcata gttttaatga tgcaaatatt gacctatctg aaatggtaca atgggcatat
1921 gacaatgatt ttatagatga tagtgatata gcatataaat atgcacaatt ggcagaaact
1981 aatagtaatg catgtgcttt tttaaaaagt aattcgcaag ctaaaattgt aaaagattgt
2041 gcaacaatgt gtagacatta taaacgagct gaaaaaagag aaatgacaat gtcacagtg
2101 attaaaaggc gatgtgaaaa ggtggacgat gacggtgact ggagggacat agtacgattt
2161 ttaagatata aacaagtaga ttttgtggca tttttatctg cactaaaaaa ttttttacat
2221 ggtgtgctca aaaaaaattg catacttata tatggagcac caaacacagg taaatcatta
2281 tttggaatga gtctaagcga tttcttacia ggagctatta tctctatgt aaattctaaa
2341 agccattttt ggttgcagcc attatatgat gccaaaatag ctatgtaga tgatgctaca
2401 tcgccatggt gggcatatat agaccaatat ttaagaaatg cactagatgg aaatcctatt
2461 tcattagatg taaagcataa agcattatgt caattaaaat gccacactt acttattaca
2521 tcaaatataa atgcaggcaa agatgacagg tggccatact tacatagcag ggtagtggtc
2581 tttacatttc acaatgaatt cccatttgat aaaaatggaa acccagtgta tgggctaat
2641 gataaaaact ggaaatcctt tttctcaagg acgtggtgca gatTAAattt gcacgaggaa
                                E2 orf start ->
2701 gaggacaaag aaaATGatgg agacgctttc ccagcgttta agtgtgtgtc aggacaaaat
                                E2 cds ->
2761 actagaacat tacgagactG Atagcacatg tttgtctgat cacatacagt attggaaact
                                <- E1 end
2821 gattcgtctt gaatgtgcag tatttTATAA Agcaagagaa atgggaatta aaactcttaa
                                -> signal
2881 ccaccaagtg gttccaacgc aggccatttc aaaagccaaa gcaatgcaag caattgaact
2941 gcaattaatg ttagagacat taaatacaac tgagtatagc acagaaacat ggacactgca
3001 agaacaagat attgaattat atacaacagt tccacaagga tgttttaaaa aacatgggggt
3061 tacagtggaa gtacaatttg atggtgataa acaaaatact atgcattata ctaattggac
3121 acatataat atattagagg acagtatag tactgttga aagggactgg taaattataa
3181 aggtatttat tatgtgcatc aggggtgtag aacatattat gttactttta gggagaggc

```

HPV35h

```

3241 taaaaagtat ggaaaaaaa atatatggga agtgcgatgt ggtggtcagg TAAttgtttg
      E4 orf start ->
      NH2 terminus unknown
3301 tcctgaatct gtatttagca gcacagaact atccactgct gaaattgcta cacagctaca
3361 cgctacaac accaccgaga cccataccaa agcctgctcc gtgggcacca cagaaacca
3421 gaagacaaat cacaaacgac ttcgaggggg taccgagctc ccctacaacc ccaccaagcg
3481 agtgcgactc agtgccgtgg acagtgttga cagaggggtc tactctacat ctgactgcac
3541 aaacaaagac cgtgtgtgta gttgtagtac aactacacct aTAGtacatt taaaaggatga
      <- E4 end
3601 tgcaaatata ttaaagtgtt taagatatag attgggtaaa tataaagcat tgtatcaaga
3661 tgcttcatct acatggagat ggacatgtac aaacgataaa aaacaaatag caattgtaac
3721 attaacttac acaacagaat atcaaaagga taaattttta actacagtaa aaatacctaa
3781 cacagttaca gtgtcTAAag gatatagttc tatATGATag accttacagc ttcagact
      E5 orf start ->
      <- E2 end
      E5 cds ->
3841 gtgttgctgt gctttttgtt gtgcttttgt gtgcttttgt gcttgtgtct gcttgtacgt
3901 tcgctattgc tatctgtgtc attatactca gcattaatat tactggtttt aatactgtgg
3961 gttactgtag caacaccact acgttgcttt tgttgttttc tttgcttttt gtatatacct
4021 atgggaatga ttaacgctca tgcacaatat ttggcagtac agTAAttgta tacaacatt
      <- E5 end
4081 gtgtttggtg ctgtgtaaca tgtgtgtatg gtggttttat tttttgttgt tcattgtata
4141 ttttgttttt ttactgtttt taacattttt tatttctgtg tttttaataa atTGATcaca
      L2 orf start ->
4201 tggataaacc ATGcgacaca aaaggctctac aaaacgtggt aaacgtgcat ctgcaacaca
      L2 cds ->
4261 actatatact acttgcaaag ctgcaggaac ttgtccacca gatgttatac ctaaggttga
4321 gggtaatact gttgctgatc aaattttaaa atatggcagc atggctgtgt tttttggggg
4381 gttaggaatt gttctctggat ctggcacagg tggaaagatct ggatatgttc cactgggtac
4441 aacacctcca acggctgcca caaacattcc tatacgaccc cctgtaactg tggaaagtat
4501 accattagac acaattggcc cttagattc ttctatagtg tcattagtag aggaaactag
4561 tttattagag tctggtgccc ctgttgttac accaagggtc ccactaca caggttttac
4621 aataaccaca tctacagata ccacacctgc tattttagat gtgacatcca taagtacaca
4681 tgataatcct actttcactg atccttctgt tttacacca cccacgctg cagaaacttc
4741 aggtcatttt gtactttcat catcttctat tagtacacat aattatgaag aaatccctat
4801 ggatactttt attgtttcca cagacagcaa taatataact aatagcacgc ctattccagg
4861 gtctgcacct acgacacgcc taggattata tagtaaaggt ACCCAGCAGG TTAaggttgt
      -> E2 bind
4921 tgaccctgcc tttatgactt ctctgcaaa acttattaca tatgataatc ctgcatatga
4981 aggccttaac cctgatacaa ccttacaatt tgagcatgag gatattagct tagctccgga
5041 tcctgacttt atggacatta tagctttaca taggcctgca ctaacatcta ggaaggcac
5101 tattagatat agtagagtag gtaataaacg tactatgcat acacgaagtg gaaagctat
5161 aggggacagc gtacattatt atcaggattt aagtagtatt actgaagata tagaattaca
5221 acccttaaca catgtaccat cctctttacc acataccact gtttcaacat cattaaatga
5281 ttttaggttt gatattttat ctctataga tactgaggaa gatattatat tttcagcatc
5341 ttctaacaat actttatata ctacatctaa cactgcatat gttcctagca atactactat
5401 accattaagt agtggctatg atattcctat aacagcaggg ccagacattg tatttaactc
5461 taatactatt actaacctg tactaccggt acccagagg cctatatatt ctattattgc
5521 agatgggggt gacttttatt tacaccctag ttattattta tTAAacgac gtcgtaaagc
      L1 orf start ->
5581 tatcccatat ttttttgag ATGtctctgt ggcggtcTAA cgaagccact gtctacctgc
      L1 cds ->
      <- L2 end
5641 ctccagtgtc agtgtetaag gttgttagca ctgatgaata tgtaaacacgc acaaacatct
5701 actatcatgc aggcagttct aggcatttag ctgtgggtca ccatactat gctattaaaa
5761 aacaagattc taataaaata gcagtacca aggtatctgg tttgcaatac agagtattta
5821 gagtaaaatt accagatcct aataagtttg gatttccaga cacatcattt tatgatctcg
5881 cctcccagcg tttggtttgg gcctgtacag gagttgaagt aggtcgtggt cagccattgg
5941 gtgtaggat tagtggcat cctttattaa ataaattgga tgatactgaa aattctaata
6001 aatatgttgg taactctggt acagataaca gggaaatgcat ttctatggat tataaacaaa
6061 cacaattgtg ttaaatagg ttaggcctc ctataggtga acattgggga aaaggcacac
6121 cttgtaatgc taaccaggta aaagcaggag aatgtcctcc tttggagtta ctaaacactg
6181 tactacaaga cggggacatg gtgacacag gatttgggtgc aatggatttt actacattac
6241 aagctaataa aagtgatgtt cccctagata tatgcagttc catttgcaaa tatcctgatt
6301 atctaataat ggtttctgag ccatatggag atatgttatt ttttattta cgtagggagc
6361 aaatgtttgt tagacattta ttaaataggg ctggaactgt aggtgaaaca gtacctgcag

```

```

6421 acctatatat taagggtacc actggcacat tgcttagtac tagttatfff cctactccta
6481 gtggctctat ggtaacctcc gatgcacaaa tatttaataa ACCATATTGG TTgcaacgtg
        -> E2 bind
6541 cacaaggcca taataatggt atttgttga gtaaccaatt gtttggtact gtagttgata
6601 caaccctag tagaaaatag tctgtgtgtt ctgctgtgtc ttctagtac agtacaTATA
        signal ->
6661 AAAatgacaa ttttaaggaa tatttaaggc atggtgaaga atatgattta cagtttattt
6721 ttcagttatg taaaataaca ctaacagcag atgttatgac atatattcat agtatgaacc
6781 cgtccatttt agaggattgg aatfffggcc ttacaccacc gccttctggt accttagagg
6841 acacatatcg ctatgtaaca tcacaggctg taacttgtca aaaaccagc gcacaaaaac
6901 ctaaagatga tccattaaaa aattatactt tttgggaggt tgatttaaag gaaaagtttt
6961 ctgcagactt agatcaattt ccgttgggcc gtaaattfff gttacaagca ggactaaagg
7021 ccaggcctaa ttttagatta ggcaagcgtg cagctccagc atctacatct aaaaaatctt
7081 ctactaaacg tagaaaagta aaaagtTAAT GTGTAATGT GTAtgcatgt atactgtgtg
        -> repeat region start
        <- L1 end
7141 TTATGTGTTG Tagtgcttgt atatataTTA TGTGTTGTgg tgctgtttg tgttgtacat
        repeat region end <-
7201 ggcgtgtaaa tgtgtgtata atattgtgca atgtgttgta cgtgggtggt ttttgtatgt
7261 atgttgttgt atgtatgtca gtacgcaata aaagtgatgt gtgtgtttat aattaacact
7321 gtattgttgt atgactatgg gtgcacccat atgacttaca taattacagt acacgctata
7381 tgttgtatat aacaattcta cctccatttt gtgtgttagt gtcctttaca ttaccttca

```

HPV35h

```
7441 ACCGATTTCG GTtgctgttg gcaagcttta tatgtttttt acaaaaacat tcctacctca
    -> E2 bind
7501 gcAGAACACT TAATCCTtgt gttcctgata tatattggtT GCCAActtta taTGGCTtT
          NF1 ->          NF1 ->NF-1 ->
    -> glucocorticoid response element
7561 TGCCAatctt taaacttgat tcatcttgca gtattagtca ttttccatac ttgtggtcca
7621 cccacacttg taacacttgt aacagtgtt ttaggcacat attttttgca tttctaaagg
7681 gctttaattg cacacTTGGC Tttacatatt atgtgtgttT GCCAAcacca ccctacacat
          ->NF1          ->NF1
7741 cctgccaaact ttaagttaaa acatgcatgt aaaacattac tcactgtatt acacattggt
7801 atatgcacac aggtgtgtcc aaccgatttg gattacagtt ttataagcat ttctttttat
7861 tatagttagt aacaattat
```


LOCUS HPV52 7942 bp ds-DNA VRL 04-OCT-1993
 DEFINITION Human papillomavirus type 52 (HPV-52), complete genome.
 ACCESSION X74481
 SOURCE Human papillomavirus type 52 DNA.
 REFERENCE 1 (bases 1 to 7942)
 AUTHORS Delius,H. and Hofmann,B.
 TITLE Primer-directed sequencing of human papillomavirus types
 JOURNAL Curr. Top. Microbiol. Immunol. 186, 13-31 (1994)
 REFERENCE 2 (bases 1 to 7942)
 AUTHORS Delius,H.
 TITLE Direct Submission
 JOURNAL Submitted (06-AUG-1993) to the EMBL/GenBank/DDBJ databases. H. Delius, Deutsches Krebsforschungszentrum, Abteilung ATV, Im Neuenheimer Feld 506, W 6900 Heidelberg, FRG
 COMMENT HPV-52 is considered to be a member of the "intermediate risk" anogenital group. All of the viruses in this group are most prevalent in high-grade intraepithelial lesions and less prevalent in invasive cancers. In a study which used 15 common anogenital probes to screen 2627 subjects, hybridization to the "intermediate risk" probes occurred in 26.4% of all high-grade intraepithelial neoplasias, 23.3% of all low-grade intraepithelial neoplasias, 13.7% of all invasive cancers, and in 3.3% of all normal samples. The relatively high percentage of high-grade lesions with respect to invasive cancers justifies the placement of type 52 in the "intermediate risk category. HPV52 is considered to be of moderate prevalence. Of all invasive cancer biopsies tested, 11% were positive for DNA in the "intermediate risk" group. Several Japanese studies indicate the Asian prevalence of HPV-52 in invasive cervical carcinomas is higher than that reported in U.S. studies. HPV52 predominantly infects tissues of the cervix and other lower-anogenital tract sites: the vulva, the vagina, the penis, the perineum and the anus.

 The HPV-52 genome was first identified by Shimoda et. al. in 1988 (Shimoda et al. J Gen Virol 69: 2925-2928). It was isolated from an early cervical intraepithelial neoplasia lesion obtained from a patient in the Washington D.C. area. Their original clone yielded a 8 kb fragment after EcoRI digestion which hybridized with greatest similarity to a HPV-33 probe. Since it hybridized with less than 50% similarity (28%), it was considered to be a novel type.

 BASE COUNT 2540 a 1449 c 1621 g 2332 t
 ORIGIN 101 bp upstream from beginning of E6 cds
 1 taaattataa tcttatacta gtaaaaaata gggtgtaACC GAAAACGGTc agACCGAAAC
 -> E2 bind -> E2 bind
 61 CGGTgtatat atatagaaca cagtgtagcT AACgcacggc cATGtttgag gatccagcaa
 E6 orf start -> E6 cds ->
 121 caccgaccccg gaccctgcac gaattgtgtg aggtgctgga agaatcggtg catgaaataa
 181 ggctgcagtg tgtgcagtg aaaaaagagc tacaacgaag agaggtatac aagtttctat
 241 ttacagattt acgaatagta tatagagaca ataatccata tggcgtgtgt attatgtgcc
 301 tacgcttttt atctaagata agtgaatata ggcattatca atattcactg tatgggaaaa
 361 cattagaaga gagggtaaaa aaaccattaa gtgaaataac tattagatgt ataattgtc
 421 aaacgccatt atgtcctgaa gaaaaagaaa gacatgtaa tgcaacaag cgatttcata

HPV52

```

481 atattatggg tcgttgaca gggcgtggt cagagtgtt gagacccga cctgTGAccc
                                     E7 orf start ->
541 aagtgtAAcG tcATGcgtgg agacaaagca actataaaag attatatatt agatctgcaa
    <- E6 end
    E7 cds ->
601 cctgaaacaa ctgacctaca ctgctatgag caattaggtg acagctcaga tgaaggaggat
661 acagatgggtg tggaccggcc agatggacaa gcagaacaag ccacaagcaa ttactacatt
721 gtgacatatt gtcacagttg tgatagcaca ctacggctat gcattcaTAG cactgcgacg
                                     E1 orf start ->
781 gaccttcgta ctctacagca aatgctgttg ggcacattac aagttgtgtg ccccggtgtg
841 gcacggctaT AAacaaccct gcaATGgagg accctgaagg tacagagggc gaaagggagg
    E1 cds ->
    <- E7 end
901 gatgtacagg ctggtttgaa gttagaggcaa taatagaaaa acaaacagga gataacattt
961 cagaggacga ggatgaaaat gcatatgata gtggaacaga tctaatagat tttatagatg
1021 attcaaatat aaataatgaa caggcagaac atgaggcagc ccgggcattg tttaatgcac
1081 aggaagggga ggatgattta catgctgtgt ctgacgtaaa acgaaagttt acaagcagtc
1141 cggaaagtgc tgggcaagat ggtgtagaaa aacatggtag tccgcgtgca aaacacattt
1201 gtgtaaatat acaggtgtgt ttaccaaacc gcaaacctatg tcacgtagaa gacagcgct
1261 atggcaatag tgaagtggaa gcgcagcaga tggcagacca ggtagacggg caaaatggcg
1321 actggcaaa taacagtagt caatcaagtg ggggtggggc tagtaattca gatgtaagtt
1381 gtactagtag agaggacaat gaggaaaata gtaatagaac gctaaaaagc atacaaaata
1441 ttatgtgcga aaatagcata aaacaactg tattatttaa atttaaagaa acatatgggtg
1501 ttagctttat ggaattagta agaccattta aaagtaatag aagtagttgt acagattgggt
1561 gtattatagg aatgggagta acaccatcag ttgcagaagg attaaaagta ttaatacagc
1621 cctatagcat atatgccat ttgcaatggt taacatgtga cagagggcgtg cttatactgc
1681 tgctaattag gtttaaatgt ggaaaaaaca gattaacagt gtccaaacta atgtcacagc
1741 tgttaaatat accagaaaca catatggtaa tagaaccacc aaaattacga agtgcacct
1801 gtgcattata ttggtataga acaggtttgt ctaatattag tgaggtatat ggtaccacc
1861 cagaatggat agaacaacaa acagatttac agcatagctt tgacaatagc atattcgatt
1921 ttggagaaat ggtgcaatgg gcatatgac atgatataac agatgatagt gacatgcat
1981 ataaatatgc acagttagca gatgtaaata gcaatgctgc agcattccta aaaagcaatt
2041 cgcaagcaaa aatagtaaag gactgtgcaa ccatgtgtag acattataaa cgggcagaaa
2101 gaaaacatat gaatattgga caatggatag agtatagatg tgatagaata gatgatgggtg
2161 gagattggag gcctatagta agatttttaa gatatcaaga catagaattt acagcctttt
2221 tagacgcatt taaaaaattt ttaaaaagta tacctaaaaa aaattgttta gtattatag
2281 gacctgcaaa cacaggaaaa tcatattttg gaatgagttt aattaggttc ttaagtggat
2341 gtgtaaatat ctatgtaaac tcaaaaagcc atttttggct acaaccatta acagatgcaa
2401 aagtgggtat gatagatgat gtaaacaccta tatgttggac atatatagat gattatatga
2461 gaaatgcact ggatggaat gatatatcag tagatgtaaa gcatagagcc ttagtaciaaa
2521 taaaatgccc accattaatt ttaacaacaa atacaatagc aggaacagat cctaggtggc
2581 catatttaca tagtagattg gttgtgtttc atttcaaaaa cccatttcca tttgatgaaa
2641 atggcaatcc tatatatgaa attaacaacg aaaattggaa atcctttttc tcaaggacgt
2701 ggtgcaaat agattTAAta caggaagagg acaaggaaaa cgATGgagtc gataccggca
    E2 orf start ->
    E2 cds ->
2761 cgtttaaagt cagtcagga aaaaatacta gatctatacg aagcTGAtag taatgaccta
                                     <- E1 end
2821 aacgcacaaa ttgaacattg gaaattgact cgaatggaat gtgttttgtt ttcaaaagca
2881 aaggaactgg gaataactca tataggccac caggtgggtg cccaatggc agtgtctaag
2941 gcaaaggcct gccaaagctat tgaactacaa ttggcattgg aggcattaaa caaaacacaa
3001 tatagcacag atggatggac attacaacaa acaagtctag aaatgtggcg tgcagaacca
3061 caaaaatact ttaaaaaaca tgggtataca ataacagtgc aatacgataa tgataaaaac
3121 aatactatgg attatacaaa ctggaaggaa atttatttac ttggtgagtg tgaatgtaca
3181 attgtagaag gacaagtaga ttactatggg ttatattatt ggtgtgatgg agaaaaaata
3241 tattttgtaa aathtagtaa cgatgcaaag caatattgtg taacaggagt atgggaagta
3301 catgtgggtg gtcaggTAAt tgtttgtcct gcatctgtat ctagtaacga agtatccact
    E4 orf start ->
3361 actgaaactg ctgtccact atgcaccgaa acctccaaga cctccgcagt gtcctgggtg
3421 gccaaagaca cacacctaca accaccacag aaacgacgac gaccagacgt cacagactcc
3481 agaaacacca agtaccocaa caaccttttg cggggacaac aatccgtgga cagtactaca
3541 cggggactcg tcaactgcaac tgagtgcaca aacaaaggac ggggtgcaca tacaacttgt
3601 actgcaccta TAAtacacct aaaaggatg cctaatagtt taaaatgttt aagatatagg
    <- E4 end
3661 gtaaaaacac ataaaagttt gtatgttcaa atttcatcta cctggcattg gaccagtaat

```

```

3721 gaatgtacaa ataataaact aggtattgta acaataacgt acagtgatga aacacaacgt
3781 caacaatttt taaaaactgt taaaatacca aatactgtgc aagttataca aggtgtcatg
3841 tcattgTGAT atttgtacat atgtatatat gtatatgtgt atggTAAaca cccaacacaa
      <- E2 end                               E5 orf start ->
3901 gccaatattg ctgctattgt gtatatataa caATGttagg attatttgta ttttgtttta
      E5 cds ->
3961 ttttgcttat ggtgttttgt gcagtgccta ggccgctctt gctatctata tcgggtgatg
4021 cgcagggtgt ggtgctgggt cttttgctat gggatctctat tgggtcacca tttaaagtgt
4081 ttttttgta cctactgttt ttatattttc caatgttttg tattcactgt catgcacagt
4141 atttggcaca actgcaaTAA ctgtacatgt agattggcta catgcatata tgcaaaaat
      <- E5 end
4201 actttttcac ttttgtagtt tgtctaataa ataactttat attttttaT AGcttgtcgc
      L2 cds start ->
4261 aATGagatac agacggctca cagggcacia acgtgcttct gcaacacagc tatatcaaac
L2 cds ->
4321 atgcaaagcc tctggcacct gcccccccga tgttattcct aaagtggaag gcacaactat
4381 tgcagatcaa cttttaaaat atggcagcct aggggtggtt tttggaggtt tgggtatagg
4441 tacaggtgca ggctctgggt gttagggcagg ctatgtgcca ttgtccactc gtccctcccac
4501 tagtagtatt accacgtcca ccattcgtcc cctgttaact gtagaaccca ttggctcctt
4561 agaaccatct atagtttcta tgatagaaga aacaacattt attgagtctg gcgcacctgc
4621 tccatctatt ccatcagcaa cagggtttga tgttacaaca tctgcaaata atactcctgc
4681 aataattaat gtaacatcta taggtgaatc atctgtacaa tcagtttcta cacatttaa
4741 tctacatctt actgaacct ctataataca gcccccgga cctgcagaag catctggcca
4801 tgtattgttt tctagtccaa ctatttagtac acacacctat gaagaaatcc ctatggatac
4861 atttgttacc tctactgaca gcagcagtggt aacaagtagt acacctattc cagggtctcg
4921 ccctacgaca cgccttgggt tatatagccg tgccacacaa cagggttaagg tagtcgacct
4981 tgcttttatg tcatcaccac agaaattagt aacatataac aatcctggtt ttgagggcgt
5041 tgatacagat gaaactataa tttttgatcg ttcacaactt ttacctgac cggatcctga
5101 ttttttagac attatagctt tgcataggcc tgcattaacc tctcgaagag gtactgttag
5161 gtttagcagg cttggtaata agggcaccct acgtacacgt agtggaanaa aaattggggc
5221 acgggtacat tattatcatg atatttagtc tatccagcct gctgaagttc aggaagacat
5281 agaattgcaa cttttattac cacagtcgtt gtccctctac actattaatg atggtttgta
5341 tgatgtgatg gcagattcct tgcagcaacc cacgtttcac ttaccttcca cactttctac
5401 ccataataat actttcactg tacctattaa tagtggatt gactttgtat atcaaccac
5461 tatgtccatt gagtcaggtc ctgacattcc attacctcgc ttaccacacac atactcctt
5521 tghtcctaTA Gcccctacag ctccatctac atctattatt gttgATGgta cagattttat
      L1 orf start ->                               L1 cds ->
5581 tttacatcct agttattttt tactacgctg caggcgtaaa cgttttccat atttttttac
5641 agatgtccgt gtggcggccT AGtgaggcca ctgtgtacct gcctcctgta cctgtctcta
      <- L2 end
5701 aggttgtaag cactgatgag tatgtgtctc gcacaagcat ctattattat gcaggcagtt
5761 ctcgattact aacagtagga catccctatt tttctattaa aacACCAGT AGTGGTaatg
      -> E2 bind <-
5821 gtaaaaaagt tttagttccc aaggtgtctg gctgcaata cagggtattt agaattaaat
5881 tgccggaccc taataaattt ggttttccag atacatcttt ttataaccca gaaacccaaa
5941 ggttggtgtg ggctgtaca ggcttgaaa ttggtagggg acagccttta ggtgtgggta
6001 ttagtgggca tcctttatta aacaagtttg atgatactga aaccagtaac aaatatgctg
6061 gtaaacctgg tatagataat agggaaatgt tatctatgga ttataagcag actcagttat
6121 gcattttagg atgcaaacct cctataggtg aacattgggg taagggaacc ccttgtaata
6181 ataattcag agaatcctgg gattgtcctc ccctacagct cattaacagt gtaatacagg
6241 atggggacat ggtagataca ggatttgggt gcatggattt taatacctg caagctagta
6301 aaagtgatgt gccattgatg atatgtagca gtgtatgtaa gtatccagat tatttgcaaa
6361 tggctagcca gccataggtg gacagtttgt tcttttttct tagacgtgag caaatgtttg
6421 ttgacactt ttttaatagg gccggtacct taggtgacct tgtgccaggt gatttatata
6481 tacaagggtc taactctggc aatactgcca ctgtacaaaag cagtgtttt tttcctactc
6541 ctagtggttc tatggtaacc tcagaatccc aattatttaa taaACCGTAC TGGTTacaac
      -> E2 bind
6601 gtgcgcaggg ccacaataat ggcataatgt ggggcaatca gttgtttgtc acagttgtgg
6661 ataccactcg tagcactaac atgactttat gtgctgaggt taaaaggaa agcacaTATA
      signal ->
6721 AAAatgaaaa ttttaaggaa taccttctgc atggcgagga atttgattta caatttattt
6781 ttcaattgtg caaaattaca ttaacagctg atgttatgac atacattcat aagatggatg
6841 caactatttt agaggactgg caatttggcc ttaccccacc accgtctgca tctttggagg
6901 acacatacag atttgtcact tctactgcta taactgtca aaaaaacaca ccacctaaag

```

HPV52

```
6961 gaaaggaaga tcctttaaag gactatatgt tttgggaggt ggatttaaaa gaaaagtttt
7021 ctgcagattt agatcagttt cctttaggta ggaagttttt gttacaggca gggctacagg
7081 ctaggcccaa actaaaacgc cctgcatcat cggccccacg tacctccaca aagaagaaaa
7141 aggttaaaaag gTAAccattg tctgttgggt aattgtctgt gtcatgtatg tgttgtgtat
      <- L1 end
7201 gtcaaacaca ggttaaaagg taaccattgt ttggtatgta attgttttgt gtgtgtactg
7261 tgttgtttgc atgttatgta tgtgtgtgca tgtttgttgt atttgcagt tcctgtatgt
7321 atgttttgtg tatgtattaa taaagtactg tatttactaa actatttata gtagtcttat
7381 gttatgttat ggttgcacc ccatgagtaa caatacagtt gctcctaatac tattgcatct
7441 cctgccctac cctgtgtccc ctgccctacc ctgtgtccta ctttgttaca ctactaatta
7501 gccttatact ctccattttg taccattttg tactatccac cattttaaata cctaACCGAA
      -> E2 bind
7561 TTCGGTtggt cttggcacia ctttggttgt ctttggcaca gtaacaacta tttttatata
7621 agtttcagca aactgcttaa tcctttgggt toctgcagtc cactggtcta cacttgttgt
7681 cccgcctaaa ctgacttctt gctgactcac aggtcctgca gtgcagctaa acaatacatt
7741 gcctaacatt gcatgtttta aactgctttt aggcacatat tttatttaa ctttcaatgc
7801 actaattaca gtgttggctt acacaagtac atcctacgcc aaatatgtct tgtaaaacat
7861 gattaaatac tgttactcac cagggtgtgca ctacacgACC GGTACGGTt accgtaccca
      -> E2 bind
7921 caaccacttt tttttataat ta
```

LOCUS HPV58 7824 bp ds-DNA VRL 24-JAN-1992
 DEFINITION Human papillomavirus type 58 (HPV-58), complete genome.
 ACCESSION D90400
 SOURCE Human papillomavirus type 58 DNA recovered from an invasive
 cervical carcinoma, patient specimen GN479.
 REFERENCE 1 (bases 1 to 7824)
 AUTHORS Kirii,Y., Iwamoto,S.-I. and Matsukura,T.
 TITLE Human papillomavirus type 58 DNA sequence
 JOURNAL Virology 185, 424-427 (1991)
 COMMENT Data kindly submitted in computer readable form by Yasuyuki Kirii,
 Kanebo Institute for Cancer Research, 1-5-90 Tomobuchi-cho,
 Miyakojima-ku, Osaka 534, Japan, Phone:06-921-1281,
 Fax:06-921-3883.

HPV-58 is considered to be a member of the "intermediate risk" anogenital group. All of the viruses in this group are most prevalent in high-grade intraepithelial lesions and less prevalent in invasive cancers. In a study which used 15 common anogenital probes to screen 2627 subjects, hybridization to the "intermediate risk" probes occurred in 26.4% of all high-grade intraepithelial neoplasias, 23.3% of all low-grade intraepithelial neoplasias, 13.7% of all invasive cancers, and in 3.3% of all normal samples. The relatively high percentage of high-grade lesions with respect to invasive cancers justifies the placement of type 58 in the "intermediate risk category. HPV58 is considered to be of moderate prevalence. Of all invasive cancer biopsies tested, 11% were positive for DNA in the "intermediate risk" group. HPV-58 was initially isolated and cloned from an invasive cervical carcinoma. It has also been detected in cervical intraepithelial neoplasias, other invasive cervical carcinomas but not condyloma. Some studies seem to indicate that the prevalence of HPV-58 in invasive carcinomas of the cervix is higher in Asia than elsewhere.

The 7824 nucleotide sequence of HPV-58 was aligned with HPV-16 and HPV-33, and was numbered accordingly. Common to all papillomavirus sequences, the HPV-58 genome contains a multimodal pattern of GC-rich regions, which are clustered around ORFs E7, E2, and L2. The E4 ORF lacks an initiation codon also familiar to types 16, 31, 33, and 39.

The long control region (LCR) of HPV-58 houses many features common among the papillomaviruses. It contains a glucocorticoid response element found at position 7246, a polyadenylation signal at position 7296, E2 binding sites at position 7487, 7780, 40, and 55, binding sites for nuclear factor 1 at position 7723 and 12, a binding site for AP-1 at position 7773, and a promoter element for TFII at position 72.

The E6 ORF has a splice acceptor/donor pair like those found in many oncogenic HPV types. Both the E6 and E7 ORFs contain the cysteine doublet motif Cys-X-X-Cys; E6 contains four sets and E7 contains two.

HPV58

```

BASE COUNT      2487 a   1388 c   1576 g   2373 t
ORIGIN
    1 ctaaactata aTGCCAAatc ttgtaaaaac taggggtgtaA CCGAAAACGG TctgACCGAA
      NF-1 bind ->
    61 ACCGGTgcat aTATAAagca gacatttttt ggtaggctac tgcaggactA TGTtccagga
      promoter for TF-II ->
      E6 orf start ->
    121 cgcagaggag aaaccacgga cattgcatga tttgtgtcag gcgttggaga catctgtgca
    181 tgaaatcgaa ttgaaatgcg ttgaatgcaa aaagactttg cagcgatctg aggtatatga
    241 ctttgtatth gcagatttaa gaatagtgta tagagatgga aatccatttg cagtatgtaa
    301 agtgtgctta cgattgctat ctaaaataag tgagtataga cattataatt attcgtata
    361 tggagacaca ttagaacaaa cactaaaaaa gtgtttaaht gaaatattaa tttagtgtat
    421 tatttgtcaa agaccattgt gtccacaaga aaaaaaaagg catgtggatt taaacaaaag
    481 gtttcataat atttcgggtc gtggacaggg gcgctgtgca gtgtgttggg gaccccgagc
    541 TAGacaaaaca caagtgTAAC ctgtaacaac gccATGagag gaaacaaccc aacgctaaga
E7 orf start ->
    601 gaatatatth tagatttaca tcctgaacca actgacctat tctgctatga gcaattatgt
    661 gacagctcag acgaggatga aataggcttg gacgggccag atggacaagc acaaccggcc
    721 acagctaatth actacattgt aacttgttgt tacacttgtg gcaccacggg tcgtttgtgt
    781 atcaacagta caacaaccga cgtacgaacc ctacagcagc tgcttatggg cacatgtacc
    841 attgtgtgcc cttagctgtgc acagcaaTAA acaccatctg caATGgatga ccctgaaggt
      <- E6 end   E7 cds ->
      E1 orf start ->
    901 acaaacgggg taggggcggg ctgtactggc tggtttgagg tagaagcggg aatagaacga
    961 agaacaggag ataataatthc agatgatgag gacgaaacag cagacgatag tggtagagat
  1021 ttaatagagt ttatagatga ttcagtacaa agtactacac aggcagaagc agaggcagcc
  1081 cgagcgttgt ttaatgtaca ggaaggggtg gacgatataa atgctgtgtg tgcactaaaa
  1141 cgaaggttthc cagcatgctc agaaagtgtc gtagaggact gtgtggaccg ggcctgcaaat
  1201 gtgtgtgtat cgtggaaata taaaaataaa gaatgcacac acagaaaacg aaaaattatth
  1261 gagctagaag acagcggata tggcaatact gaagtggaaa ctgagcagat ggcacaccag
  1321 gtagaagacc aaaaatggcga cgcagactta aatgactcgg agtctagtgg ggtgggggct
  1381 agttcagatg taagcagtga aacggatgta gacagttgta atactgttcc attacaaaat
  1441 attagtaata ttctacataa cagtaatact aaagcaacgc tattatataa attcaaaagaa
  1501 gcttatggag taagtttthc ggaattagtt agaccattta aaagtgataa aacaagctgt
  1561 acagattggt gtataacagg gtatggaata agtccctccg tagcagaaag tttaaaagta
  1621 ctaattaaaac agcacagtat atatacacac ctacaatgthc taactgtgtg cagaggaatt
  1681 atattatthc tgtaatttag attttaaattg agcaaaaata gattaactgt ggcaaaatta
  1741 atgagtaatt tactatcaat tcctgaaaca tgtatgatta tgcagccacc aaaattacga
  1801 agtcaagcat gtgccttata ttggtttaga acagcaatgt caaatataag tgatgtgcaa
  1861 gggacaacac cagaatggat agatagatta acagtgthc agcatagctt taatgatgat
  1921 atatttgatth taagtgaat gatacaatgg gcatatgata atgacattac agatgatagt
  1981 gacattgcat aaaaatgthc acagttagca gatgttaata gtaatgcagc agcattthta
  2041 agaagcaatg cacaagcaaa aatagtaaaa gactgtggcg ttatgtgagc acattataaa
  2101 agagcagaaa agcgtggtat gacaatggga caatggatac aaagttagtg tgaaaaaaca
  2161 aatgatggag gtaattggag accaatagta caattthta gatataaaa tattgaattt
  2221 acagcattthc tagttgcatth taaacagthc ttacaaggthc taccataaaa aagttgtatg
  2281 ttactgtgtg gccagcaaaa tacagggaaa tcataatthc gaatgagthc aatacattth
  2341 ttaaaaggat gcattatthc atatgtaaat tccaaaagthc atthttgthc gcagccatta
  2401 tcagatgctc aactaggtat gatagatgat gtaacagcca taagctggac atatatagat
  2461 gattatatga gaaatgcatt agatggtaac gacattthc tagatgtaaa acatagggca
  2521 ttagtacaat taaaatgthc accattaata attacctcaa atacaaatgc aggcaaaagat
  2581 tcacgatggc catatthcga cagtagacta acagatthc aatthtaaaa tccattthc
  2641 tttgatgcaa atggtaatcc agtgtataaa ataaatgatg aaaaatggaa atcctthtthc

```

```

2701 tcaaggacgt ggtgcaaatt aggcttaaTA Gaggaagagg acaaggaaaa cgATGgagga
      E2 orf start ->                E2 cds ->
2761 aatatcagca cgtttaagtg cagtgccagga caaaatccta gacatatatcg aagcTGATAa
      <- E1 end
2821 aaatgattta acatcacaaa ttgaacattg gaaactaata cgcattggagt gtgctataat
2881 gtatacagcc agacaaatgg gaatatcaca tttgtgccac caggtgggtgc cgtcattggt
2941 agcatcaaag actaaagcgt ttcaagtaat tgaactgcaa atggcattag agacattaaa
3001 tgcatacaca TATAAAAcag atgaatggac attgcaaca acaagcttag aagtgtggtt
      signal ->
3061 atcagagcca caaaaatgct ttaaaaaaaaa aggcataaca gtaactgtac aatattgacaa
3121 tgataaagca aacacaatgg attatacaaaa ttggagtga atatatatta ttgaggaaac
3181 aacatgtact ttggtagcag gagaagtga ctatgtgggg ttgtattata tacatggcaa
3241 tgaaaagacg tattttaaat attttaaaga ggatgcaaaa aagtactcta aaacacaatt
3301 atgggaggta catgtgggta gtcgggTAAt tgtatgtcct acatctatac ctatgtgatca
      E4 orf start ->
3361 aatatccact actgaaactg ctgacccaaa gaccaccgag gccaccaaca acgaaagtac
3421 acaggggaca aagcgacgac gactcgattt accagactcc agagacaaca cccagtactc
3481 cacaaagtat acagactgag ccgtggacag tagaccacga ggaggaggac tacacagtac
3541 aactaactgt acatacaaaag ggcggaacgt gtgtagttct aaagtttcac ctatcgtgca
3601 ttTAAaaggT gacccaataa gttttaaagt ttaagatat agattaaac catttaaaga
      <- E4 end
3661 cttatactgt aatatgtcat ccacatggca ttggaccagt gatgacaaa gtgacaaagt
3721 aggaattggt actgtaacat acacaacgga aacacaacga caactgtttt taaacactgt
3781 taaaatacca cccactgtgc aaataagtac tgggtgtatg tcattgTAAAt tgtattgtac
      <- E2 end
3841 aattactgta tgTAAaccac aagccaatat gtgctgctaa gtgtatatac aATGatatta
      E5 orf start ->                E5 cds ->
3901 cctatttttg ttgtttgttt tatactgttt ttatgcttgt gcattttttt gcggccattg
3961 gtgctatcta tttctatata tgcttgggtg ctgggtgttg tttgtgctct ttgggtgtct
4021 gtggggtcgg ctctacgaat tttttctgt tacttaatat ttttatata accaatgatg
4081 tgtattaatt ttcattgaca atacttaacc caacaagacT AActgtatac tggttctgca
      <- E5 end
4141 catgggtgga ttggtattgta aatatttact gttgtgtgtg ttgtttttat tatttttata
4201 catttactaa taaatacttt tatattttTA Gcactgtcct attATGagac acaaacggtc
      L2 orf start ->                L2 cds ->
4261 tacaaggcgc aagcgtgcat ctgctacaca actttacca acatgcaagg cctcaggcac
4321 ctgcccacct gatgttatac ccaaagtga aggactact atagcagatc aaatattacg
4381 atatggtagc ttagggtgtg tttttggagg tttaggcatt ggtacagggt cgggtacagg
4441 tggcaggact ggatattgtc cccttggtag taccaccagc tctgaggcta tacctttaca
4501 gccatacagt cccccagtta ccgttgatac tgtggggcct ttggattcct ctattgtatc
4561 ttaataagag gaatctagtt ttatagacgc cggtgcaaca gccccatcaa tcccactcc
4621 atctggtttt gatattacca cctctgcaga tactacacct gcaataacta atgtttctc
4681 tattggagaa tcatctatac aaactgtttc tacacattta aatocctcct ttactgagcc
4741 atccgtactc cgccctcctg cacctgcaga ggcctctgga catttaatat tttcctctcc
4801 tactgttagc acacatagtt atgaaaacat accaatggat acctttgtta tttctactga
4861 cagtggcaat gtcacgtcta gcacacccat tccagggtct cgccctgtgg cacgccttgg
4921 ttatacagc cgcaacACCC AACAAGTTaa ggtgtgtgac cctgcttttt taacatctcc
      E2 bind ->
4981 tcatagactt gtaacatag ataatccagc atttgaaggc ttaaccctg aggacacatt
5041 gcagtttcaa catagtgaca tategcctgc tcctgatcct gattttctag atattgtgc
5101 attacacaga cctgcattaa cctctcgcag gggtagctga cgttatagta gggttgggca
5161 aaaggctaca cttcgtactc gcagtggaag gcaaataggg gctaaagtac attactacca
5221 agacttaagt cccatacagc ctgtccagga acaggtacaa cagcagcaac aatttgatt
5281 acaatcttta aatacttctg tttctcccta tagtattaat gatggacttt atgatattta
5341 tgctgacgat gctgatacta tacatgattt tcagagtcct ctgcactcac atacgtcctt
5401 tgccaccaca cgtaccagta atgtgtccat accattaaat actggatttg acactcctct
5461 tgtgtcattg gaacctggtc cagacattgc atcttctgta acatctatgt ctagtccatt
5521 tattcctata tctccactaa ctctttttaa taccaTAAAt gtggATGgtg ctgattttat
      L1 orf start ->                -> L1 cds
5581 gttgcaccct agctattttta ttttcgctcg cagacgtaaa cgttttccat atttttttgc
5641 agatgtccgt gtggcgccT AGtgaggcca ctgtgtacct gcctcctgtg cctgtgtcta
      <- L2 end
5701 aggttgaag cactgatgaa tatgtgtcac gcacaagcat ttattattat gctggcagtt
5761 ccagactttt ggctgttggc aatccatatt tttccatcaa aagtcccaat aacaataaaa

```

HPV58

```

5821 aagtattagt tcccaaggta tcaggcttac agtatagggg ctttaggggtg cgtttacctg
5881 atcccaataa atttggtttt cctgatacat ctttttataa ccctgataca caacgtttgg
5941 tctgggcatg tgtaggcctt gaaataggta ggggacagcc attgggtgtt ggcgtaagtg
6001 gtcacacctta tttaaataaa tttgatgaca ctgaaaccag taacagatat cccgcacagc
6061 cagggtctga taacagggaa tgcttatcta tggattataa acaaacacaa ttatgtttaa
6121 ttggctgtaa acctcccact ggtgagcatt ggggtaaagg tgttgcctgt aacaataatg
6181 cagctgctac tgattgtcct ccattggaac tttttaattc tattattgag gatggtgaca
6241 tggtagatac agggtttggg tgcattggact ttggtacatt gcaggtaat aaaagtgatg
6301 tgcctattga tatttgaac agtacatgca aatatccaga ttatttaaaa atggccagtg
6361 aaccttatgg ggatagtttg ttcttttttc ttagacgtga gcagatgttt gttagacact
6421 tttttaatag ggctggaana ctggcgagg ctgtcccgga tgacctttat attaaagggt
6481 ccgtaatac tgcaattatc caaagtagtg cttttttcc aactcctagt ggctctatag
6541 ttacctcaga atcacaatta ttaataaagc cttattggct acagcgtgca caaggtcata
6601 acaatggcat ttgctggggc aatcagttat ttgtaacctg ggttgatacc actcgtagca
6661 ctaatatgac attatgcact gaagtaacta aggaaggtag aTATAAAAat gataatttta
                                     signal ->
6721 aggaatatgt acgtcatgtt gaagaatag acttacagtt tgtttttcag ctttgcaaaa
6781 ttacactaac tgcaagagata atgacatata tacatactat ggattccaat attttggagg
6841 actggcaatt tggtttaaca cctcctcctg ctgccagttt acaggacaca tatagatttg
6901 ttacctccca ggctattact tgccaaaaaa cagcaccccc taaagaaaag gaagatccat
6961 taaataaata tactttttgg gaggtttaact taaaggaaaa gttttctgca gatctagatc
7021 agtttccctt gggacgaaa tttttattac aatcagccct taaagcaaaag cccagactaa
7081 aacgttcggc ccctactacc cgtgcacat ccaccaaacg caaaaagggt aaaaaaTAAT
                                     <- L1 end
7141 tgttgggta cttacactat tttattatac atgtttggtt gttttatgta tgtgtgtct
7201 gtttgtttat gtttgtgtat atgttgtatg tgttatgtgt catgtTTGTG TACATGTTCT
                                     glucocorticoid response element ->
7261 atgtccttgt cagtttctctg tttctgtata tatgtAATAA Actattgtgt gtattgtaaa
                                     signal ->
7321 ctatttgtat tgtttgggtg tatctatgag taagggtgctg tccctaaatt gcctacacct
7381 gccctgccta ttatgcatac ctatgtaata gtatttgtat gatatgtatt ttatagtttt
7441 taacagtact gcctccattt tacttttaact ccattttgtg catgtaACCG ATTTCGGTtg
                                     -> E2 bind
7501 ctggcacaaa cgtgtttttt ttaaaactaca atttaacaaa tacagttaat cttttccctt
7561 cctgcactgc ttttgcctat acttgcatac gtgactcata tatacatgca gtgcagttgc
7621 aaaatgttta attatactca tagtttaaac atgcttatag gcacatattt taacttactt
7681 tcaatgctta agtgcagttt tggcttgcac aatagtttgt taTGCCAaac tatgtcttgt
                                     NF-1 bind ->
7741 aaaagtgact cactaacatt tattgccagg tgTGGACTAA CCGTTTTGGG Tcacattgtt
                                     AP-1 bind -> <-
                                     -> E2 bind
7801 catgtttcaa cattttatat aata

```


LOCUS HPV67MY911 449 bp ds-DNA VRL 16-OCT-1994
 DEFINITION Human papillomavirus type 67 (HPV-67), partial L1 cds, MY09/MY11 region.
 ACCESSION U12492
 SOURCE Human papillomavirus type 67 DNA recovered from a patient with vulvar intraepithelial neoplasia (VaIN).
 REFERENCE 1 (bases 1 to 449)
 AUTHORS Bernard,H.-U., Chan,S.-Y., Manos,M.M., Ong,C.-K., Villa,L.L., Delius,H., Peyton,C.L., Bauer,H.M., and Wheeler,C.M.
 TITLE Identification and assessment of known and novel human papillomaviruses by PCR amplification, restriction fragment length polymorphisms, nucleotide sequence, and phylogenetic algorithms
 JOURNAL J. Infect. Dis. (1994) In press
 COMMENT HPV-67 has recently been isolated from a vulvar intraepithelial neoplasia by Dr. T. Matsukura. The cloned DNA was subsequently sequenced by Dr. H. Delius over the L1 MY09/MY11 segment. HPV-67 and the several other types sequenced by Dr. Delius over this region were used as type-specific probes to screen DNA for novel genital HPV types. The screened DNA was obtained from four recent epidemiological studies.

BASE COUNT 156 a 87 c 79 g 127 t

ORIGIN

```

1 gccacaggac ataacaatgg tatatgctgg ggtaatcaaa tatttgttac tgttgtagac
L1 cds ->
61 actacacgta gtaccaacat gactttatat tctgagggaa aatcagaggc tacatacaaa
121 aatgaaaact ttaaggaata ccttagacat gtggaagaat atgatttgca gtttatattt
181 cagctgtgca aaatatccct tactgcaaat gttatgcaat acatacacac catgaatcca
241 gatatattag aggactggca atttggcctt acaccacctc cttcaggtaa ttacaggac
301 acatatagat ttgttacctc gcaggctatt acctgtcaaa aacatcccc tccaacagca
361 aaggaagatc ctcttaaaaa gtacagtttt tgggaaatca atttaaagga aaaattttct
421 gcagatttag atcagtatcc cttggacg
L1 cds ->

```