Group G Sequences

HPV1   **HPV4**
HPV41   **HPV63**
**HPV65**

INTRODUCTION

Group G consists of the human papillomaviruses HPV-1, HPV-4, HPV-41, HPV-63 and HPV-65, a group primarily associated with the benign cutaneous lesions, commonly seen in the general population.

HPV-1 and HPV-4, in conjunction with HPV-2, a group F virus, are the major etiological agents of benign cutaneous papillomas in the general population. HPV-1 is primarily associated with deep palmoplantar warts, while HPV-4 has been correlated with common warts and keratotic flat lesions on the hands and feet [1,2]. HPV-4 is also frequently present in hand warts of meat handlers [3]. HPV-41, like HPV-4, has been linked to flat warts, which are mainly found on the face and feet. Unlike the others in this group, HPV-41 has been detected in skin squamous cell carcinomas and its precursor lesions [4]. Also unique to HPV-41 is the absence of typical E2 binding sites in the LCR. However, modified E2 sites, as reported for BPV-1, have been located near the E6 gene [5]. HPV-65 has been linked to pigmented common warts and keratotic flat lesions, while HPV-63 is associated with multiple punctate keratotic lesions of the foot [2].

The viruses HPV-1, HPV-4, HPV-63 and HPV-65 produce intracytoplasmic inclusion bodies in most infected epidermal cells. The inclusion bodies primarily contain E4 proteins that can be used to histologically identify these viruses. HPV-4 like HPV-65, is associated with a homogeneous type of ICB. HPV-63 is associated with a filamentous type of ICB (Fl-ICB) and HPV-1 presents a granular type (Gr-ICB) [2].

The primary target tissue of the group G viruses is the epithelium, however rare mucosal infection has been reported for HPV-1, HPV-4 and HPV-41. Both HPV-1 and HPV-41 have been identified in benign anogenital warts [4, 6 ,7]. Conversely, HPV-4 has been identified in isolated cases of both normal and malignant oral lesions [8].

The members of Group G were placed together primarily on the basis of their clinical associations. Within the group are three subsets, each of which is at least as distant from the other subsets of Group G as any intergroup comparison of Groups A–F. For the purposes of analysis "by group" we chose to create two distinct subgroups, one consisting of HPV-1a and HPV-63, the other of HPV-4 and HPV-65. The latter pair of sequences are nearly close enough in terms of non synonymous base changes to be considered "close types". HPV-41 is closer to the former subgroup than the latter, but sufficiently distant that it merits separation when analysis is performed. The members of Group G are generally closer to non primate animal papillomaviruses than to other human papillomaviruses.
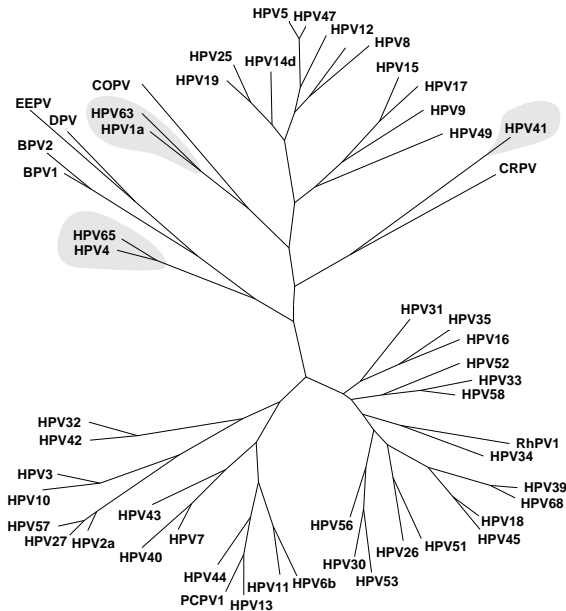
[1] Danos,O., Katinka,M., and Yaniv,M. Human papillomavirus 1a complete DNA sequence: a novel type of genome organization among Papovaviridae. *EMBO* **1**: 231–236 (1982)

[2] Egawa, K., Delius,H., Matsukura,T., Kawashima,M., and de Villiers,E.M. Two novel types of human papillomavirus, HPV 63 and HPV 65: comparisons of their clinical and histological features and DNA sequences to other HPV types. *Virology* **194**: 789–99 (1993)

[3] Melchers,W., de Mare,S., Kuitert,E., Galama,J., Walboomers,J., van den Brule,A.J. Human papillomavirus and cutaneous warts in meat handlers. *J Clin Microbiol* **31**: 2547–9 (1993)

[4] Grimmel,M., de Villers, E.M., Neumann,C., Pawlita,M., and zur Hausen, H. Characterization of a new human papillomavirus (HPV 41) from disseminated warts and detection of its DNA in some skin carcinomas. *Int. J. Cancer* **41**: 5–9 (1988)

[5] Hirt,L., Hirsch-Behnam,A., de Villiers,E.M. Nucleotide sequence of human papillomavirus (HPV) type 41: an unusual HPV type without a typical E2 binding site consensus sequence. *Virus Res* **18**: 179–89 (1991)

[6] Krzyzek,R.A., Watts,S.L., Anderson,D.L., Faras,A.J., and Pass,F. Anogenital warts contain several distinct species of human papillomavirus. *J Virol* **36**: 236–44 (1980)

[7] Gissmann,L., deVilliers,E.M., and zur Hausen,H. Analysis of human genital warts (condylomata acuminata) and other genital tumors for human papillomavirus type 6 DNA. *Int J Cancer* **29**: 143–6 (1982)

[8] Yeudall,W.A., and Campo,M.S. Human papillomavirus DNA in biopsies of oral tissues. *J Gen Virol* **72**: 173–6 (1991)

```
LOCUS       HPV1a        7815 bp    ds-DNA   Circular  VRL        06-JUL-1989
DEFINITION  Human papillomavirus 1a (HPV-1a), complete genome.
ACCESSION   V01116 X03321
KEYWORDS    circular; genome; origin of replication.
SOURCE      Human papillomavirus type 1a DNA.
REFERENCE   1  (bases 1 to 7811)
  AUTHORS   Danos,O., Katinka,M. and Yaniv,M.
  TITLE     Human papillomavirus 1a complete DNA sequence: a novel type of
            genome organization among papovaviridae
  JOURNAL   EMBO J. 1, 231-236 (1982)
REFERENCE   2  (bases 1 to 7815)
  AUTHORS   Danos,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JAN-1985) Danos O.
REFERENCE   3  (bases 7574, 7691, 7692 and sites; revision)
  AUTHORS   Palermo-dilts,D., Broker,T., and Chow,L.
  TITLE     Human papillomavirus type 1 produces redundant as well as
            polycistronic mRNAs in plantar warts
  JOURNAL   J. Virology 64, 3144-3149 (1990)
COMMENT     HPV-1a has a strong preferential association with deep plantar
            warts, characterized by a highly thickened corneal layer
            (hyperkeratosis).  Generally, they are present in a single
            location.  Skin warts are transmitted by direct contact with
            infected tissue or with contaminated objects.  A majority of warts
            regress, spontaneously within two years.  This is thought to be the
            result of a cell-mediated immune response.  HPV-1 DNA has also been
            detected in anogenital tissue, however rarely.

            The 7815 bp genome of HPV-1a is the embl corrected version of a
            previously published sequence.  In addition to the embl corrections
            this sequence has been updated as noted in reference 3:  a C has been
            changed to a T at nt 7574 and a CC has been changed to a GG at nts
            7691-7692.

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

mRNA species	Splice donor/acceptor pair	Coding potentials
a	827/3200	E1^E4, E5a
b	1231/3200	E1M, E2C, E5a
c	827/2545	E2, E5a
e	827/3200	E1^E4, L1
	3592/5431	
f	7710/3200	L1
	3592/5431	
g	7710/5431	L1
h	7710/3200	E5a, E5, L2

HPV1a

```
BASE COUNT     2396 a   1482 c   1666 g   2271 t
ORIGIN
        1 gttaactacc atcattcatt attctagtta caacaagaac cTAGgagtta tatgccagaa
                                                      E8 orf start ->
       61 gtaagccTAT AAAAtacaca ggTAAgactc tgcacaggac cagATGgcga caccaatccg
             signal ->     E6 orf start ->                E6 cds ->
      121 gaccgtcaga cagctttccg aaagcctctg tatcccatat attgATGttt tattgccttg
                                                          E8 cds ->
      181 taatttttgt aattattttt tgtctaatgc tgagaagctg ctttttgatc attttgattt
      241 gcatcttgtc tggagagaca atttggtgtt tggatgctgt caagggtgtg ctagaactgt
      301 tagcctattg gagtttgttt tatattatca ggagtcttat gaggtaccgg aaaTAGaaga
                                                                  <- E8 end
      361 aattttggac agacctttat tgcaaattga actccgttgt gttacatgca taaaaaaact
      421 gagtgttgct gaaaaattgg aggttgtgtc aaacggagaa agagtgcata gagttagaaa
      481 cagacttaaa gcaaagtgTA Gtttgtgtcg cttgtatgcT ATATAAcaAT Ggtgggcgaa
                 E7 orf start ->                              E7 cds ->
                                                           <- E6 end
                                   signal ->
      541 atgccagcac taaaggacct ggttcttcaa cttgaaccaa gcgtcctaga tttagatctt
      601 tattgttacg aggaggtgcc tcctgatgac atagaggagg agttagtgtc gcctcagcaa
      661 ccttatgctg tcgttgcttc ctgtgcctat tgcgagaaac tggttcgatt gaccgtcctc
      721 gcggatcaca cgcgccattag acagctggag gaactccttc tgcgatcttT GAacatcgtg
                                                          E1 orf start ->
      781 tgcccactgt gcaccctaca gcgacagTAA aATGgcagat AATAAAgGTa ctgaaaacga
                                E1 cds ->        5' sj /\
                                  <- E7 end      -> signal
      841 ttggtttttg gtggaggcga cagattgtga ggaaacgtta gaggaaACCT CACTTGGTga
                                                             -> E2 bind
      901 cctagataat gtttcttgtg ttagcgactt atctgattta ttagacgagg cgccgcaaag
      961 ccaggggaat tccctggaat tgttccacaa gcaagaatcg ctggaaagcg aacaggaact
     1021 taatgcttta aaacgaaagt tactttacag tcctcaggcg agaagcgcgc acgaaacaga
     1081 cattgctagc attagtccta gattagaaac tatttctatt acaaagcaag acaaaaaaag
     1141 gtatcgaagg caactgtttt ctcaggatga tagtggttta gagctatcgc tgcttcagga
     1201 tgaaactgaa aatattgatg aatcgacaca gGTagatcaa cagcagaaag aacatactgg
                                        5' sj /\
     1261 ggaagttggg gccgctgggg tgaacatttt gaaagctagt aatatccgcg ccgcattatt
     1321 aagcagattt aaagatacgg ctggcgtcag ttttacagAC CTGACGCGGT cgtacaagag
                                                      -> E2 bind
     1381 caacaaaacc tgttgtggag attgggtttt ggcagtttgg ggtgtccgtg aaaatttaat
     1441 tgacagtgta aaagaattat tgcaaaccca ttgtgtgtat attcaattgg aacatgcagt
     1501 aactgaaaaa aatagatttt tattttttatt ggtacgattt aaagcccaga aaagtagaga
     1561 gactgtgata aaacttataa ccacaattct tccagttgat gctagctata ttttgtctga
     1621 gcctccaaaa tcaagaagtg tggctgctgc attattttgg TATAAAagat ctatgtcttc
                                                      signal ->
     1681 aactgttttt acatggggta caactttgga gtggattgca cagcaaaccc ttattaatca
     1741 tcagttagat tccgaaagtc cctttgagct ttgtaaaatg gttcagtggg cctatgataa
     1801 tggacataca gaagagtgta aaattgcata ttattatgct gtttttagcag atgaggatga
     1861 aaaatgcaagg gcatttctaa gctctaattc acaggcaaaa tatgtgaaag actgtgcaca
     1921 aatggtaaga cactatttac gtgctgagat ggcacaaatg tctatgtcag agtggatttt
     1981 tagaaaacta gataatgtag aaggttctgg taattggaaa gaaattgtaa gattttttaag
     2041 atttcaagaa gttgaattta taagctttat gattgcattt aaagatttgt tatgtggtaa
     2101 gccaaagaaa aactgtttgt taatatttgg acctccaaat acaggaaaat caatgttttg
     2161 tacaagttta ttaaagttgt taggagggaa agtgatttca tactgtaaca gtaaaagtca
     2221 gttttggttg cagcctctgg ctgatgctaa gatagggcta ttagatgatg caacaaagcc
     2281 atgttgggat tatatggaca tttatatgag aaatgcattg gatggtaaca ctatttgtat
     2341 tgatttaaaa catagagctc tcaacaaat taaatgccca cctttactta ttactagtaa
     2401 tattgatgtt aaaatcagata cctgttggat gtatttgcat agtagaatat cagctttttaa
     2461 atttgctcat gagtttccat ttaaagacaa tggtgatcca ggattttcct taacagacga
     2521 aaaattggaaa tctttctttg aaAGgttttg gcaacagtta gaatTAAgtg accaagaaga
                                /\ 3' sj    E2 orf start ->
     2581 cgagggaaac gATGgaaaac ctcagcagtc gcttagactt actgcaagag cagctaatga
                 E2 cds ->
     2641 acctataTGA acaggacagt aaattgatag aagatcaaat taagcagtgg aatctaatta
                 <- E1 end
```

I-G-4
SEP 94

```
2701 gacaagaaca agttcttttc catttcgcca gaaaaaatgg ggtaatgaga attggattgc
2761 aggcagttcc atctttagcg tcctcacagg agaaggcaaa gacagctatt gaaatggtgt
2821 tacatttaga gtctttaaag gactcacctt atggcacaga ggattggtca cttcaagaca
2881 ctagcagaga gctgtttttg gcacccccag ctggcacctt caagaagagt ggcagcacac
2941 ttgaggttac ctatgacaat aaccctgata atcagacaag gcacacaatt ggaatcatg
3001 tgtattatca aaatggggac gatgtatgga gaaaagtatc cagtggtgtt gatgctgtag
3061 gagtgtacta tttagaacac gatggcTATA AAaattatta tgtgttattt gctgaggagg
                           signal ->
3121 cctctaagta cagcacaaca ggacaatatg ctgTAAatta caggggtaaa aggtttacaa
                                E4 orf start ->
3181 ATGttatgtc ttccactAGc tccccaaggg ctgctggggc tcctgcagta cactccgact
E4 cds ->                   /\ 3' sj
3241 acccaaccct atccgagagT GAcaccgccc agcaatcgac gtccatcgac tacaccgaac
                E3 orf start ->
3301 tcccaggaca gggggagacc tcgcaggtcc gacaaagaca gcagaaaaca cctgtacgca
3361 gacggcctta cggacggcga agatccagaa gtcccagagg tggaggacga agagaaggag
3421 aatcaacgcc ctcTAGgaca cccggatctg tcccttctgc gcgagacgtt ggaagtatac
                 <- E3 orf end
3481 acacaacgcc tcaaaaggga cattcttcaa gacttagacg acttctgcag gaagcttggg
3541 atccacccgt ggtctgtgTA Aaaggggtg ccaatcagct taagtgtctc agGTacagac
                 <- E4 end                          5' sj /\
3601 ttaaagcatc tactcaagtt gactttgaca gcataagcac cacatggcat tggacagata
3661 gaaaaaacac cgagaggata ggtagtgcta gaatgttagt aaagtttatt gatgaggctc
3721 aacgagagaa gtttcttgag agagttgctt gcccagatc agtgtctgtg tttttgggac
3781 agtttaatgg gtctTAAaat taatggaagt tgattttgct ggacgtgtg tacatagtcc
                 <- E2 end
3841 ctgTATATAT tcccctccta cccccacata cctTGAagct tgcaacattg TAAcaaATGt
 signal ->                          E5 orf start ->    L2 orf start ->
                                                            L2 cds ->
3901 atcgcctacg tagaaaacgc gctgcccccca aagaTATATA ccctcatgc aaaatatcaa
                                 signal ->
3961 acacctgccc acctgacatt caaAATAAAa ttgagcatac aacaattgct gataaaatat
                 signal ->
4021 tgcaatatgg cagtctggga gttttttttgg gaggtttggg cattggaaca gccagaggct
4081 ctggaggaag aattggttat actccctcg gtgagggtgg tggggttaga gttgctactc
4141 gtccaactcc agTAAggcct acaataccctg tggaaacagt aggccccagt gaaattttcc
                 <- E5 orf end
4201 ccatagatgt tgtagatcct acaggccctg ctgttattcc cctacaagat ttaggtagag
4261 acttcccaat accaactgtg caggttattg cagaaattca ccctatttct gacataccaa
4321 acattgttgc atcttcaaca aatgaaggag aatctgccat attagatgtg ttacgaggga
4381 atgcaaccat acgcactgtt tcaagaacac aatacaataa cccctctttc actgttgcat
4441 ctacatctaa tataagtgct ggagaagcat caacatcaga tattgtattt gttagcaatg
4501 gttcaggtga cagggtggtg ggcgaggata tccccttggt agaattaaac ttaggccttg
```

```
     4561 aaacagacac atcttctgtt gtacaagaaa cagcattttc cagcagcaca CCAATtgctg
                                                               signal ->
     4621 aaagaccctc ttttaggccc tcaagattcT ATAATAGgcg tctatatgaa caggtgcaag
                                          signal ->
     4681 tacaagaccc taggttcgtt gagcagccac agtcaatggt cactttttgat aatccagcat
     4741 ttgagccaga gcttgatgag gtgtctatta tcttccaaag agacttagat gctcttgctc
     4801 agacaccagt gcctgaattt agagatgtag tttatctgag caagcccaca ttttcgcggg
     4861 aaccaggggg acggttaagg gttagccgcc ttggcaaaag ttcaactatt cgtacacgcc
     4921 tgggcacagc aattggcgcc agaacccact tttttctatga tttaagttct attgctccag
     4981 aagactcaat tgaattattg cctttaggtg agcatagtca aacaacagtc attagttcca
     5041 acttaggtga cacagcattt atacaaggtg agacagcaga ggatgactta gaagttatct
     5101 ctttagaaac accacaatta tattcagaag aagagctttt agacacaaac gaaagtgtgg
     5161 gcgaaaattt gcaacttact attactaact cagagggtga ggtttctata ctagatttaa
     5221 cacaaagcag agtcaggcca ccttttggca ctgaagatac tagcttgcat gtatattacc
     5281 caaattcttc taaagggact ccaataatta atcctgaaga atcatttaca cctttggtta
     5341 ttaTAGctct taacaactca acaggggatt ttgagttaca tcctagtctt agaaagcgtc
L1 orf start ->
     5401 gtaaaagagc ttATGtaTAA tgttttttcAG atggctgtct ggttaccagc gcagaataag
                    L1 cds ->    <- L2 end   /\ 3' sj
     5461 ttctatcttc ctccccagcc catcactaga atcctgtcca ctgatgaata tgtaaccaga
     5521 accaatctct ctaccatgc aacatctgaa cgtctactgc tggtcggaca tcctttgttt
     5581 gagatctcca gtaatcaaac tgtaactata ccaaaagtgt caccaaatgc atttagagtt
     5641 tttagggtgc gttttgctga tccaaataga tttgcatttg gggataaggc aatttttaat
     5701 ccagaaacag aaagattagt ttggggccta gagggatag agataggtag aggccagcct
     5761 ttaggtatag gaataacggg ccaccctctt ttaaataagt tagatgatgc agaaaatcca
     5821 acaaattata ttaatactca tgcaaatgga gattctagac aaaatactgc ttttgatgca
     5881 aaacagacac aaatgttcct cgtcggctgt actcctgctt caggtgaaca ctggacaagt
     5941 agtcgttgcc caggggaaca agtgaaactt ggggactgcc ccaggggtgca aatgatagag
     6001 tctgtcatag aagatggtga catgatggat attggttttg gggctatgga ttttgctgct
     6061 ttacagcaag acaagtctga tgtcccttta gatgttgttc aagcaacatg caaatatcct
     6121 gattatatca gaatgaacca tgaagcctat ggcaactcta tgtttttttt tgcacgtcgc
     6181 gagcaaatgt ataccaggca cttttttact cgcgggggtt cggtgggtga taaggaggca
     6241 gtcccacaaa gcctgtattt aacagcagat gctgaaccaa gaacaacttt agcaacaaca
     6301 aattatgtag gcacaccaag tggctctatg gtttcatctg atgtccaatt gtttaataga
     6361 tcttactggc ttcagcgatg tcaaggccag aataatggca tttgctggag aaaccagtta
     6421 tttattacag ttggagataa taccagagga acaagtttat ctatcagtat gaaaaacaat
     6481 gcaagtacta catattccaa tgctaatttt aatgattttc taagacatac tgaagaattt
     6541 gatctttctt ttatagttca gctttgtaaa gtaaagttaa ctcccgaaaa tctagcctac
     6601 attcatacaa tggaccctaa tattttagag gattggcaac tatctgtatc tcaaccacct
     6661 accaatcctc tagaagatca atataggttt ttagggtctt ccttggcagc aaaatgtcca
     6721 gaacaggcgc ctcctgagcc ccagactgat ccttatagtc aaTATAAATt ctgggaagtc
                                                         signal ->
     6781 gatctcacag aaaggatgtc cgaacaatta gaccaatttc cactaggaag gaaatttcta
     6841 tatcaaagtg gcatgacaca acgtactgct actagttcca ccacaaagcg caaaacagtg
     6901 cgtgtatcta cgtcagccaa gcgcaggcgt aaggctTAGT ATATAtTATA TATAactata
                                              <- L1 end
                                            signal ->
     6961 tttattagta gattatttat TATATAtttt tatatttttta tactttttat acttgtttag
                       signal ->
     7021 ttctaaatag acatgtaaga tttacattag tataagtagg catgtattta cataaaatag
     7081 tcttggaaac cttttattag tgaaccatca tttacaatag tgacatcata gttcatctgc
     7141 aattgctatt ccatcgttct tcacatattc tacagtagtg ttctctagat tgtattgcta
     7201 ttttcctgtt aggcaaacaa caacatctgt acatggacca aacaacccac tttcatttta
     7261 ttgtgctgca tatattccag attgttgagg atttatttgt ttagactccg gtgcattata
     7321 cacaagtgtg catttttgt gttctctgat tgattgtgtg ttattttcct gcaatatgcA
                                                               signal ->
     7381 ATAAAgtga gctgtccttt cttttgtta atccctccct actccAATAA Aaaatcccta
                                                    signal ->
     7441 cccctaaaat ctgtttgtgc tggtttttatt aataattgcg ctctttTATA TAAtaagtac
                                                    signal ->
                                                          |-> mRNA start site
                                                              from P(7490)
                                                              promoter
     7501 tattaacACC GCACCCGTTg tggctaatcc cttatggtat ttaaaagact acacctacag
```

```
                 -> E2 bind
7561 gatgtattgt ctttattgtt tatggtttac cgcgctCCAA AGACGGTTTG CCCAAAGACG
                                         ->    repeat region start
7621 GTTTGCcaac cgcggttagg acttgtttca atttgctgcc aaacttatct ggtcgtgctc
         <- repeat region end
7681 caacgggttt ggtgccaagc acctaaaacg GTaggtgtgt actcttttca agaattaaca
                                      5' sj /\
                          -> E2 bind
7741 aaggagattt ctcccgccaa attagtttcg agcgACCGAA TTCGGTcgta aaaatctaaa
                                         -> E2 bind
7801 gtgatgattg ttgtt
```

```
LOCUS       HPV4          7353 bp ds-DNA          VRL         04-OCT-1993
DEFINITION  Human papillomavirus type 4 (HPV-4), complete genome.
ACCESSION   X70827
SOURCE      Human papillomavirus type 4 DNA.
REFERENCE   1  (bases 1 to 7353)
  AUTHORS   Egawa,K., Delius,H., Matsukura,T., Kawashima,M. and
            De Villiers,E.M.
  TITLE     Two novel types of human papillomavirus, HPV63 and HPV65
            comparisons of their distinct clinical and histological features
            and their DNA sequences to other HPV types
  JOURNAL   Virology 194, 789-799 (1993)
COMMENT     Submitted (27-JAN-1993) on tape to the EMBL Data Library by:  H.
            Delius, Deutsches Krebsforschungszentrum, Abt ATV, Im Neuenheimer
            Feld 506, W -6900 Heidelberg, FRG.

            HPV-4 has a strong preferential association with common warts,
            characterized by a dome shape with multiple conical projections
            (papillomatosis).  Most often they are multiple and are found on
            the hands.  Skin warts are transmitted by direct contact with
            infected tissue or with contaminated objects.  A majority of warts
            regress spontaneously within two years.  This is thought to be the
            result of a cell-mediated immune response.

            Infection from HPV-4 produces intracytoplasmic inclusion bodies
            (ICB) in most cells of the epidermis.  These inclusions have been
            previously shown to contain mainly E4 proteins for HPV-1.  To
            date, three morphologically different types of ICBs have been
            distinguished, and certain types of HPV are preferentially found
            in one type of ICB or another.  HPV-4, like HPV-65, is associated
            with a homogenous type of ICB (Hg-ICB).  The author notes that
            histologically, these Hg-ICBs were usually present as a single
            inclusion in each cell, surounding an extremely vacuolated
            nucleus.  HPV-4 and HPV-65 have both been detected in common
            warts and keratotic flat lesions on the hands and feet.

BASE COUNT     2363 a   1290 c   1543 g   2157 t
ORIGIN
        1 gtctgtaatg atagttggca acaatcatta cttatagcta tataTAAccg gaagagatac
                                            E6 orf start ->
       61 atataaaaag ggacagtgca tttctactaa atcctgtcca gATGgcagat ggcagacctg
                                            E6 cds ->
      121 caaccttgga cgacttctgc agacgattcg acatttcctt ttttgatttg cgccttactt
      181 gtattttttg ttctcatact gtcgatcttg cggatcttgc tttattctat cttaagaaac
      241 ttagtttagt atttagagga aattgttatt atgcatgttg ttctgaatgc ttaagattaa
      301 gtgcactgtt tgaacaagag aattattttc aatgttctat taaagctgta catttggagg
      361 aaattgctca gaaaaagatt aaggaaattt gcattagatg catttgctgc cttagattac
      421 ttgatattgt tgagaaatta gatttattat actctgacga gacttgctat ttaataaggg
      481 gtttgtggag gggctattgc agaaattgta tTAGgaaaca ATGAgaggag cagcgcccac
                                                    E7 cds -><- E6 end
                              E7 orf start ->
      541 ggttgcagat cttaatttag aactaaatga cttagtgtta ccagcaaacc tgctgagtga
      601 ggaggtcttg caatcttcag atgatgagta tgagattaca gaggaggagt cggtggttcc
      661 atttagaata gacacctgtt gctatagatg tgaagttgct gtaagaatta cattgtatgc
      721 tgctgagctc ggactacgga ccttggaaca acttcttgta gaaggaaagc TGAcgttttg
                                                        E1 orf start ->
      781 ctgcaccgct tgtgcaagaa gtcttaacag aaATGGcaga TAAaggtaca gacaattttg
                              E1 cds ->          <- E7 end
      841 acttagaagg gaataattgg tatattgtcc atgaagcaga atgcactgac agtatagata
      901 cgttggatga tttatgcgac gaaagtaatg acgattcaaa catttctaac ttaattgatg
      961 acgatgtcgt tgatcagggg aattcccttg cgctgtacaa tgcacaaata aatgaggatt
     1021 gtgacaatgc actagcacac ctaaaacgaa agtataacaa aagtccagag caggcagtcg
     1081 ctgaattgag tccgcagttg caggctgtga aaataactcc tgaaagacac agcaaaagga
     1141 gattatttca ggacagtggg attttcgaag atgaagctga aaattctctt acacaggtag
     1201 aatccgagag ccaggctgga cctttctagcc aagatggcgg cggagatatt aatttgttgt
     1261 tgttacaaag tagtaacagg agggcaacaa tgctagcaaa gtttaaagaa tggtatgggg
     1321 tctcatacaa tgaaataaca agaattttata aaagtgataa atcttgtagt gataattggg
```

```
1381 taatagttat ttttagagct gctgttgaag tattagaaag ttcaaagatt gttttaaagc
1441 agcattgtac atatattcaa gttaagatct ttggattttc agctttatat ttagtacagt
1501 ttaaaagtgc gaaaagtaga gaaactgtac aaaagttgat gtgttctata ttaaatatcc
1561 aagaatatca aatgttatgt gatcctccaa aattacgaag tgtacccaca gcattatact
1621 tttataagca tgctatgtta acagagagtt ctgttttttgg acaaacaccg gattggatcg
1681 caaaacaaac tctcgtaagt catcaagcag caactactgc agagactttt gagttatcta
1741 gaatggttca gtgggcatac gataataatt atgtggatga atgtgacatt gcttatcact
1801 atgcaatgta cgcagaggag gatgcaaatg ctgctgctta tttaaaaagt aataatcaag
1861 taaagcatgt acgagattgt agtacaatgg tcaggatgta taaaagatat gaaatgagag
1921 atatgtcaat gtcagaatgg atttataaat gttgtgatga atgttctgaa gaaggagatt
1981 ggaagccaat ctcacagttt ttaaaatatc aagtgttaa tatattatcc tttcttatag
2041 tgcttaaatc attttttaaaa ggtattccaa aaaaaaactg tatagttatt catggtccac
2101 cagatacagg aaaatcatta ttttgttatt cttttataaa attttttaaaa ggaaaagtag
2161 tttcatatgt aaatagaagt agccattttt ggttgcagcc tctgatggat tgcaaggtag
2221 gatttatgga tgatgctacc tatgtgtgct ggacatatat agatcaaaat ttaaggaatg
2281 cattagatgg taatccaatg tgtattgacg ctaaacacag agcaccacaa caattaaaat
2341 taccaccaat gctaataacg tcaaatattg atattaaaca ggaacaatct ttaatgtatt
2401 tacacagtag aatacagtgt tttaattttc ctaacaaaat gcctatttta gatgatggta
2461 gtcctatgta tacatttact gacggtactt ggaaatcttt tttccaaaag cttggcagac
2521 aattagaatT AAcagatcct gaagaggaaa acaATGgagt ccctagtcgc acgtttcgat
        E2 orf start ->              E2 cds ->
2581 gcacttcaag aagcaattct gactcatatT GAgtcacagg agagcacttt ggaatcccaa
                                  <- E1 end
2641 atccaatatt gggaaaatat cagaaaagaa aatgctataa tgcattatgc tcgaaaacaa
2701 ggcctaACCA AATTAGGTct acaaccactt cctacactag cagtaactga atacaatgca
            -> E2 bind
2761 aagcaagcta ttcagataca tttaactttg caatcattgt taaaatctcc ctttgcatct
2821 gaacggtgga cattgacaga tgttagtgca gaactgataa atacctctcc acaaaactgt
2881 ttaaaaaagg gaggttatga tgttgctgtg tggtttgata tgatagaca gaatgcaatg
2941 ctgtacacaa attgggactt tttatattat caagataTGA ATGaacagtg gcacaaagtt
                             E4 orf and cds start ->
3001 aaaggtgaag tggattatga tggcttatac tttacagacc atacgggaga aagagcttat
3061 tttacattat ttagctctga tgctcaaaga tttagcagaa ctggactgtg gactgtgcat
3121 tttaaaaccc aagttatttc ctcccctatt gttagctcta catactcctc ctccttcgac
3181 actgaggaac aacagttacc cgggccctcc accagctact ccgaagttac cgagcaggcg
3241 agccctactc gaaggaggaa accgaggaaa tccgacgcga cctccaccac gtcccctgaa
3301 accgagggag tacgactacg acgaagacga cgagaaggaa aatcagggcc cgggtcagga
3361 gaaacccccc gcaaaagaag aagaggagga ggaagaggag gaggagagac cgaattggga
3421 tctgcaccat ctcctgcaga agtggggagc agacatcgac aagttgaaag acaaggtctg
3481 tcgcgacttg gactcttaca agcagaagct agggatccgc ctaTGAtatt gttaaagggc
                                             <- E4 end
3541 acagcaaatt ctttgaaatg ttggagatat agaaaagtta actcaaattg ctgcaacttc
3601 ttattcatga gtactgtttg gaactgggtt ggagattgct cacataatca tagtcgcatg
3661 cttattgcat ttgatagcac tgaccaaaga gacgcttttg taaaacacaa cctttttcct
3721 aaactgtgta catatacctag cggctcatTG AatagtttaT AAaATGcaaa gcttgagtag
                                             <- E2 end
                    L2 orf start ->   L2 cds ->
3781 aaggaaaaga gattcagttc caaatctttta tgcaaaatgt caactgtctg gcaattgcct
3841 acctgatgta aaaAATAAAg tagaagctga tactcttgct gatcgtttgc tgagatggtt
                -> signal
3901 gggaagtgta atatacctag gaggcttggg tattggtact gggagaggta gtgggggggtc
3961 aactgggtat aatccaattg gagctccaag tagagtcaca cctagtggta ctttagtaag
4021 gcctacagtg cctgtggaaa gtttgggacc ctcagaaata atcccaatag atgcaataga
4081 cccaacaaca tcttctgttg tgccattaga ggatctgacc atcccagatg tcacagtaga
4141 tagtggagat acaagaggaa taggggagac tactcttcag cctgcacaag tagatatttc
4201 aacatcacat gacccatata cagatgtcac tggtgctagc agccacccta caatcatatc
4261 tggcgaggat aacgccattg cagtgttaga tgtgtcccct atagaacctc ccacaaaacg
4321 gatagcattg gcaactaggg gagcctcagc aactccacat gtaagtgtca tatctggcac
4381 aACCGAATTC GGTcagtcat ctgatctgaa tgtatttgtg aatgccacat tttcaggcga
        -> E2 bind
4441 ttccattggt tatacagaag aaattccatt agaaccgttg aacccctttc aagaattcga
4501 aatagaaagc cctccaaaaa ctagtacacc acgtgacgtt ttaaatcgtg caataggaag
4561 agcacgggat ttatataata gaaggggttca gcaaatacct actaggaacc cagctttact
4621 gacacagcct tcccgcgcaa tagtatttgg atttgaaaat cccgcctttg atgctgacat
```

```
4681 cactcaaaca tttgagcggg atttagaaca ggttgcagca gctccagatg ctgactttgc
4741 agacatagtc actatagggc gtccaaggtt ttcagagaca gatgctggtc aaattagagt
4801 tagcaggctt ggacgccgag gcacaataaa aactagaagt ggtgtgcaaa ttgggcaggc
4861 ggttcatttt tattacgacc taagtacaat agatactgct gatgctattg aattatctac
4921 tttaggtcaa cattcaggag aacaaagcat tgttgatgct atgatagaaa gcagcttaat
4981 agatcctttt gaaatgcccg atcctacttt tacagaagaa caacagcttt tagatccact
5041 tacagaagat tttagtcagt cacacttggt gcttactagt agcagacgtg gacatcatt
5101 tactatacct acaataccac ctggattagg tcttagcatt tatgtagatg atgtaggttc
5161 tgatttattt gtttcctatc cagaatctag agtaatacct gctggaggtt taccaactga
5221 gccatttgtt cctctagaAC CAGCTTTGTT atctgatata tttagtacgg attttgtata
                        -> E2 bind
5281 tcgtcctagt ttatatcgca agaaacggaa acgattagaa atgtttTAAt tgttttgcag
                                                              <- L2 end
                                           L1 orf start ->
5341 gaacATGtcg agttggttat ctacaacggg taaagtctac ttacctccag ctcaacctgt
   L1 cds ->
5401 ggcaagagtt ttggaaactg acgaatatat cactggaaca tctctgtatt ccacgctgg
5461 tacagaaagg cttttaactg taggccatcc ttattttcca gtgaaagatg tacaggaacc
5521 tcacaaagta ttagttccta aggtttcagg aagtcaattt agagtgttta gattcaattt
5581 gccagaccca aacagatttg ctttaattga taatggcttt tatgattctg atcatgaacg
5641 cctagtatgg aaactgaggg aatagaaat aggaagagga ggaccgcttg gtataggtac
5701 tacaggtcat cctttatata ataagtttgg agacacagaa aatcctaatg ctacaaaaa
5761 gcaatcagat gataaataga aggatgtctc tttagaccca aaacaaacac agatgtttat
5821 tataggttgc actcctgcaa taggtgaaca ttgggataaa gctgaacctt gtcccagccc
5881 tgctccgcaa cagggagatt gcccaccaat agagcttgta aattcataca ttcaagatgg
5941 agatatgtgt gacattggat ttggggcttt caattttaaa gctttgcagg ctgataaatc
6001 tagtgctcct ttggatgtca ttgccacgat ttgtaaatgg ccagatttt taaaaatggg
6061 gaaagatatc tatggagata gcttgttttt ctttggagga agagaacaac tatatgccag
6121 acatttcttt gtcagagcag gcaccatggg agatgctcta ccagaacctt ttgaagctac
6181 ctcagattat tttattggtg ctcaaaacca acaagatcag tacactttag gacctcatat
6241 ttatgtaggg acccctagtg gctctttagt atccagtgaa tcccagttgt ttaatcgACC
                                                              E2 bind ->
6301 GTATTGGTTa aacagagctc agggtacaaa taatggaatt tgttgggata tcagttgtt
6361 tgttactctt gtagataaca ctcataatac aaactttaca atttctgtga agtcagatgg
6421 tgctaatgac aattatcagt ataaagctag tgattttaaa cagtacctca gacatataga
6481 ggagtttgaa atggaattta tatttcaact ttgtaaagtt cctctaactg cagatgttat
6541 ggctcattta aatgtaatga atcctaatat tttggataat tggcagttaa attttgttcc
6601 accacctccc tctggaattg aggatcaata tagatttttg caatctagag ctacaagatg
6661 ccctacacag acccctgcaa ctgaaaaaga agatccatat aaagatttgt ctttttgggt
6721 tgttgattta agtgaaagat tttccagtga attgagccaa ttttccttag gcaggcggtt
6781 tttatatcaa agtggtttaa ttaatggttc tctaaaacgt aaaagaataa taagttcttc
6841 tcatgcacaa actaatacca aacgttctgc caaacgaaaa cggtctctga aaTAAcaatg
                                                              <- L1 end
6901 tgaactcttc tggaatgttt tattctgcca ggaaaacctt caactgagcc aaattattat
6961 ataatcgttc ttaatctcaa aattgagcta attatataag atttgcaaac gtgtatgtat
7021 ctgtttttgt gaactatagt gaaataaaact gccacatact tgccagtgtc cagtctctct
7081 gagtcatttg gtcaacatgc gtccgcaccc caataattat ttgcatacac agatcagtag
7141 gagaggcgcc aagacggaca tatcctcttc aaatttcctt aaaattattg aatttaacaa
7201 ctgtaagcta caaaagACCG TTATCGTTtc ctctaacctt gggaaaaagg tgagtgaaag
                            -> E2 bind
7261 ttttattgca ccttttgtga gtcaatttgt ctggcggcgc tgaacgaatt tggctgtcag
7321 cctttgcACC GGGAGTGGTg gaaaatagtt tct
            -> E2 bind
```

```
LOCUS       HPV41        7614 bp ds-DNA              VRL       04-JAN-1993
DEFINITION  Human papillomavirus type 41 (HPV-41), complete genome.
ACCESSION   X56147
KEYWORDS    papillomavirus.
SOURCE      Human papillomavirus type 41 DNA.
REFERENCE   1  (bases 1 to 7614)
  AUTHORS   De Villiers,E.M., Hirsch-Behnam,A. and Hirt,L.
  TITLE     Nucleotide sequence of human papillomavirus (HPV) type 41: an
            unusual HPV type without a typical E2 binding site consensus
            sequence
  JOURNAL   Virus Res. 18, 179-190 (1990)
COMMENT     *source: tissue=facial wart; *source: is_proviral=N; From EMBL
            entry PAP41CG;  dated 14-JUN-1991.

            HPV-41 was originally isolated from biopsies taken from a
            15-year-old female patient with facial, peri-anal and foot warts
            (Grimmel et al. Int. J. Cancer 41, 5-9).  The patient had a
            history of dermatitis atopica since early childhood, but no
            clinical or histological symptoms of epidermodysplasia
            verruciformis or any type of immunodeficiency at the time of
            treatment.  This suggests that HPV-41 caused the lesions.

            Subsequent screenings of tissues taken from 106 biopsies from
            benign and malignant skin lesions as well as 71 malignant tumors
            from non-cutaneous tissues indicated the presence of HPV-41 DNA in
            two out of ten squamous-cell carcinomas and one out of three
            arsenic keratoses (the latter being regarded as a precursor lesion
            to the former). HPV-41 DNA was not detected in any non-cutaneous
            tissues.

            Sequence analysis of HPV-41 reveals it to be highly divergent from
            all other known types. Overall nucleotide similarity to other
            sequenced HPV types is less than 50%, and the highly conserved
            E2 protein binding motif (ACCN6GGT) is found only in a modified form
            in the upstream regulatory region of the genome. (These modified
            sequences have been demonstrated to bind to the E2 protein in the
            BPV-1 genome.)

            The author observes the presence of three short ORFs (X, Y, and Z)
            in the HPV41 genome.  At the time of publication, they were no
            observed analogs in other sequenced papillomavirus genomes.
            Because of this, these ORFs are probably not functional. The author
            notes that these ORFs have the following characteristics: X ORF
            922-1212 with first methionine at 979, Y ORF 2811-3050 with first
            methionine at 2832, and Z ORF 5631-5882 with first methionine at
            5652.
BASE COUNT     2101 a   1665 c   1908 g   1940 t
ORIGIN
        1 aCAATCATaa tcatcgccct ttcgtgttat ttcttgtAAC GAATTCGTTa caaaacacac
     signal ->                                    -> E2-bind (modified)
       61 acacagTATA TAAgatagag gaacggattg gtacaccaca gATGgcatca acaagcggtg
          signal ->                                E6 cds ->
      121 tgggatccgt cgggcctgca agctgttgcg agacgcagaa gccacatacc atacgggagt
      181 tgtgtttggc gcagcagata acttatccat gcatacagct ctgctgccat tattgctata
      241 agatccttag cgtattggat atttacgctt cgaccagag ctgtctgtac ttatcctggg
      301 gagaagggg gccaacgggt atttgttctc agtgtactag agtgcttgca aggctggagt
      361 tcactgcacg gcacgaagtg tcttgtgcag ccagccgtct gccgcacttt ataggacaga
      421 gcctcagcga ccttgaggtg aggtgtgtga ggtgcctagc tcttctacaa tctgtggaaa
      481 aggattacat attgcgggaa gacttgtctg tgcaTAGaat tggcgggatc tggaggggaa
                                       E7 orf start ->
      541 cttgtgttcg ATGtatggta ggactgtatT AGctgtgaga ctaatatact gtttgctgta
             E7 cds ->                     <- E6 end
      601 ttgtattgct gtaatcgtgc gtaaattgct ataccctgta ataatgagag ggaatagtgt
      661 tgacctgcaa gaaattgtgc ttgttcagca gggggaggta cctgagaatg ctgcagtgca
      721 ttctcaggtag cattctgatg atgaggttga gagcgaggag gaggagcggg aacaggtgca
      781 gcaagtcccc acacccagga gaacattata cctggtagag agtcagtgtc cattttgcca
      841 ggctatcata cgatttgtat gcgTAGcaag caacactggg atacggaatc tacaggcact
```

```
                           E1 orf start ->
  901 cctggtcaac agtcacctTG Acctcgcttg tcacgcctgt gtcgagcaga ATGgcgtcca
             "X" ORF start ->                              E1 cds ->
  961 gggtctcaga caccggcaat ggcaaTGAaa acaaagagaa tgaagGTaca Gtggcatctg
                                <- E7 end     5' sj /\
                                           (putative E1,E4 fusion site)
 1021 atcattctga ggcgcgttgt agctatatat tatttgaggc tgaatgtagc gatggcgggg
 1081 acgatgagga aagtatggag gatagcttgg tggaagacct tgtggatgat gcttctgtgc
 1141 atcagggaaa ttccttgtcg ctgtttcatg cccaaactgt cgaggaatac gagggagaga
 1201 tccagagccT AAaacgaaag tttatcctga gtcccttgca tagggatgtg gcagaactaa
                 <- "X" ORF end
 1261 gcccgcgtct ggcgggtgtt tccctggaag aaaaccgtgg gaaaaaggct cgcaaatctc
 1321 tgttccacga tgacagtggc atagacagca gcgcagtgga agtctcccag ctatctagta
 1381 cgccatcagc tccagggcca gacatccggc tgcctaaacc ctcagatata gatctagagc
 1441 cactgttcca aagccgccag cgctgtacgc atatgtatag caaatttaaa gctgtgtacg
 1501 gggttagctt tacagatata accaggccat tcaaaagcga caaaacaaca tcacagcatt
 1561 gggttgtggc cgcctactat ttagcttttg atagtgagat aagtgctatg gaggtttgc
 1621 tgcgacaaca atgccaattt ttatacattg acaacaatga tggcattata ctgttcttcc
 1681 tggaatacaa cgtgcagaaa tctaggacta cagtgtacaa ttggttcaca gccaatttcc
 1741 attataatga aaatagaatg ctagctaatc cgccaaggac acgaaacatg cctgctgctt
 1801 tattcttcta tcatagattt atgggtacag ggggtataaa acatggcgca atgccagaaa
 1861 taattgtaaa ccagtgcgtg gtgtctaatc agcagacaga caccttggaa ttatcacgta
 1921 tggtacagtg ggcactggac aacgatctgc aagatgaaca tatgttagct ttagagtatg
 1981 ctttgcttgc tgaaagtgat ggcaatgcgc gggctttttt aaagcagaat aatcagccaa
 2041 tgatagtgaa gaattgtagc ataatggtta gacactacaa gacagcgctg gtcgcaaaaa
 2101 tgtctatttc acagtatgtg aataagcggt gtctggacca tggggaagct gatgaaaaca
 2161 gctggcgggg aattgtgcat tttctgaggt atcaaggtca ggaattcctg cccttcatgt
 2221 gtaaaatgca caatttccta caccatagac caaagaaatc aacacttgta ttatgtggac
 2281 cgtcggacac aggcaaatca tattttgcca atggtcttaa caaatttttg gatggacacg
 2341 tgctgagctt tgtcagcaat gggtcacatt tttggttatc accattacgt ggggcacggt
 2401 gctgtctaat agacgatgcg accctcacgt tttggaggta cgcggaccaa aacatgaggg
 2461 cactgctaga tggatatgag atttccattg atgcaaaaca cagaaaccca atgcaaacta
 2521 gagcaccacc attaataata accacaaatg aggacattat gcgattagat gaattcaaat
 2581 atctgcaaac cagaacaatg tatgtgtact ttaacaagcc atttcctctt aaaggaaatg
 2641 ggcaaccgtt atattacatt gatggttata catggaactc ttttttttagg aaattttggc
 2701 gtcacctaaa tcTAAaagac cctgaggATG agtcagatgg agagactcct ggaacgatta
           E2 orf start ->    E2 cds ->
 2761 gactatatac aagagcagat actgacacta taTGAgaaag atagtgtTGA cctagaggat
                                           <- E1 end
                                             "Y" orf start ->
 2821 catataaggc tATGgaatct gctaaggagg gaaaatgcaa tctggtatgt actcagacag
 2881 gaaggacacg caagggtcgg cggcagagcg gtgccggcaa tgacggtatc ggaagccaat
 2941 gccaaattcg caatagaaat gcagataaag ctagaatcac taaaggccag tccctatgcg
```

```
3001 gccgagggct ggtcattgca agaaaccacc aaggaacggt acttggcTGA accgtctcgg
                                                        <- "Y" end
3061 acatttaaga aattagggca gccagttacc ctaatgtttg acaatgatcc cgaaaacctt
3121 acagaagttg tattgtggaa atgggtttat tatattacac caacagatga atggtataaa
3181 gctagaggtg gcattgatga cactggtata tactacattg accacgagtc tgttaaaatg
3241 tactatgtga gatttgacat ggaagcggag aactttagcg agacaggcac tgtcacctac
3301 cggctaggca gcgccctggt aaatgtacct gaacctgTAA ctgttaccga cagctcctcc
                                      E4 orf start ->
                                      NH2 terminus unknown
3361 acgAGggaga gaaccccaaa ggtactacga ccgcaggggt cgagacgacg cagaaacgag
          /\ 3' sj
(putative E1,E4 fusion product)
3421 gaaacggggg agccggtcgc cccagcccct aagcgaagac gaggagctta cggacgcaga
3481 tcctccccga aggcccaacg caggaccgcg gcgtcgcctg tttctagagg aaacggagga
3541 tcgtctgact tcacttctgg agagtctgac gaaggacatc gagtcagaca tagagcactt
3601 cgaaagaaaa ctgcgggtgt tgctccagca gaaggacact atcTAGttgg cgccaaaggt
                                                        <- E4 end
3661 ccagtgaata gcctgcggtg cttaaggtac aaatggaaaa acaagtatag cggtgacaTA
3721 Atgtatctgg ggactacttt cacatggacg gagtctgacg gacagaacg gtgtgggtcg
E5 orf ->
 start
3781 gggcgctttt tttgtgcttt ctctaATGaa acaaaaagag aaaagttcct caaatctgtc
                          E5 cds ->
3841 aagattccta aaaacattgg gctgtttcgc gcacacgcag aaaagctgTG Acctgtgtat
                                                           <- E2 end
3901 catTAAacaA TGcttgctag gcaaagggtt aaacgcgcta atcctgaaca actgtataag
        L2 cds ->
L2 orf start ->
3961 acatgcaaag caacgggggg cgattgtcca cccgatgttA TTAAAcgcta tgagcaaact
                                             signal ->
4021 acacctgctg atagtatATT AAAgtatggg agtgtagggg ttttctttgg cggtctgggc
                signal ->   <- E5 end
4081 attggcacag gacgtggtgg cggtggcaca gtgcttgggg ctggggcagt tgggggacgc
4141 ccgtccatat ccagtggtgc aattggtccc cggatatttt gccaattga atcagggggg
4201 ccttcactgg cagaggaaat acctctgctt cccatggcac cccgtgtgcc aaggcctaca
4261 gatcccttta ggccgtcagt gctggaagag ccttttatta taaggcctcc tgaacgccca
4321 aacattttgc atgagcagcg tttccctaca gacgctgcac catttgacaa tggcaacaca
4381 gaaatcacaa ccattcctag ccaatatgat gttagtgggg gaggggttga cattcagata
4441 attgaactcc ctagtgtgaa tgaccccggt ccctcggttg ttacccgcac acaatacaac
4501 aatccaacgt ttgaggtgga ggtgtccact gacattagtg agaaacctc atcaacggac
4561 aacattattg taggagctga aagcggtggc acatccgtag gtgacaatgc tgaactgata
4621 cctttgctag atatatcccg gggggacaca attgacacaa caatacttgc ccctggcgag
4681 gaggagactg cctttgtgac cagcactcct gaacgtgtgc ctatacagga gcgattacct
4741 attaggcccc tatggcagaca gtatcagcaa gtgcgagtta ccgaccctga attttagac
4801 agcgctgcag tacttgtctc tttagagaat ccagtgtttg atgcagacat tactctcacg
4861 tttgaggatg atctgcagca ggcactacgt agtgacacag acctgcggga cgtgcgtcgc
4921 ctcagtagac cttattacca gaggcgcact actggccttg tgttagtcg cctggggcaa
4981 cgtcggggta ctatatccac gcgctctggt gttcaggtag gctccgctgc tcattttttc
5041 caggacatta gtccaatcgg ccaggctatt gagccaattg atgcaattga actagatgta
5101 ctgggtgagc aatccggtga ggggactatt gtgagaggag accctacgcc ttctattgag
5161 caagacatag gactaaccgc tttgggggac aacattgaaa atgaattgca ggaaatagat
5221 ttattaactg cggatggtga agaagaccag gagggcagag acctgcagtt ggtattttcc
5281 actggcaatg atgaggtggt tgatattaTG Actataccta tacgtgcagg cggggATGac
                          L1 orf start ->                L1 cds ->
5341 aggccttcag tatttatttt tagcgatgat ggcactcaca ttgtctatcc tactagcaca
5401 acagccacca ccccactcgt gcctgcacag cccagcgatg tgccctacat tgttgttgac
5461 ttgtatagtg gaagtatgga ttatgatata catcctagcc tgttgcgcag aaaacgtaaa
5521 aaacgcaaac gtgtttattt ttcagatggc cgtgtggctt ccaggcccaa aTAGatttta
                                                        <- L2 end
5581 cttacccct caacctatac aacggacatt gaacacagag gaatacgTGA gacgcaccag
                                      "Z" orf start ->
5641 tactttcctc cATGctgcca ctgaccgttt gcttactgtt ggacatccat tttacaatat
5701 tactaatgcg gatggcaaag aggtggtccc taaagtttcc tctaatcagt tcagggcctt
5761 ccgtgtccgt ttcccaaatc ccaatacctt tgcattttgt gataagtccc ttttttaaccc
```

```
5821 tgacaaggag cgtctggtct ggggtattcg tgggattgag gtttctaggg gacagccctT
5881 AGgtattggt gtaacaggga accctttttt taataagttt gatgatgctg aaaatcccta
        <- "Z" end
5941 caatggtata aacaaaaata acattactga ccaaggttca gactcaaggt tgagcattgc
6001 atttgaccct aagcaaacac agctgctgat agtaggtgct aaacctgcaa agggtgagta
6061 ctgggacgtt gctgcaacat gtgaaaaccc tccactgacc aaagcagatg acaaatgtcc
6121 tgctctagag cttaagtcct catacattga ggatgcagac atgagtgaca taggcctggg
6181 aaacttgaat ttttctacac tgcagagaaa caaatccgat gccccattag atattgtgga
6241 ttctatctgc aaatatcctg actacctgca aatgatagaa gaactatatg gagaccacat
6301 gtttttctat gtgcggcgtg aagctctgta tgctaggcat ataatgcaac acgcgggcaa
6361 gatggatgct gagcaatttc ccacttctct gtacatagac tcctctgtag aaggtgagaa
6421 attaaattcc ttgcagcgca ctgataggta tttcatgaca cccagcggct ccctggtagc
6481 tactgagcag cagctgttta acaggccctt ttggctgcag agatcccagg gccataacaa
6541 tggcatactg tggcacaacg aggcctttgt aacattggtt gacactacca ggggaactaa
6601 ctttaccatc agtgttcctg aggggggatgc ttcttcatat aacaattcta agttttttga
6661 gttttttaagg cacaccgagg agtttcagct tgcctttatt ctacagctgt gtaaggtaga
6721 ccttaccccct gagaatttgg cttacataca cacaatggat ccatccatta ttgaagactg
6781 gcatttagct gtcacttcac ctcccaattc tgtactggag gatcattata ggtacatact
6841 gtccattgca actaaatgtc cctctaagga tgcagatgat acctccactg acccatacaa
6901 agatcttaag ttttgggagg ttgatctacg ggatcgtatg acagagcaat tggaccagac
6961 tccccttggc aggaagtttt tgtttcaaac tggtatcact cagtcatcat caaataagcg
7021 ggtgtccacg cagtctactg cccttactac ctacaggcgg cctactaagc gccgccggaa
7081 ggctTAAacg aattgctggt attgtggtgc ggtgtcctcg acggtccatg tgtcatctta
            <- L1 end
7141 taatcacttg gtcagtccag ggtacaccac tccattatct atttacttcg catgtatttc
7201 tctgttatgt tcctgtatgg gttatgaatg tgttAATAAA atatgttggt aacgctgtgc
                                     signal ->
7261 acgggtttgt tcacgttcat gtctcatgat TTGGCAcccc tgtattcccg ccgccgcccg
                                     NF-1 bind ->
7321 ggggatcgca gatataatcc ccaaACCCAA AGCGTTccaa caTTGGCAaa cgtctctggc
                             -> E2-bind          -> NF-1 bind
                              (modified)
7381 cccgatacAA CTGAAACGGT ctgtctTGCC AAtagcccca tctggcgggg attcAACTGA
                -> E2-bind           -> NF-1 bind                   -> E2-bind
                 (modified)                                          (modified)
7441 AACGGTgtgt acTGCCAAgt aacatttttg ttattggAAC GCCTCCGGTg ctggcggaag
                   -> NF-1 bind              -> E2-bind
                                               (modified)
7501 cgcaaggatt taggcgcgaa gacagtttta tTGCCAAaac cttttggttg cTGCCAAtag
                            NF-1 bind ->              NF-1 bind ->
7561 caggcgtggt ctcAACGAAT TCGTTgcggc aaTAGgtatg taccatggtt atga
                      -> E2-bind        E6 orf ->
                       (modified)        start
```

```
LOCUS       HPV63        7348 bp ds-DNA           VRL       04-OCT-1993
DEFINITION  Human papillomavirus type 63 (HPV-63), complete genome.
ACCESSION   X70828
SOURCE      Human papillomavirus type 63 DNA recovered from punctate keratotic
            lesions of the foot.
REFERENCE   1  (bases 1 to 7348)
  AUTHORS   Egawa,K., Delius,H., Matsukura,T., Kawashima,M. and
            De Villiers,E.M.
    TITLE   Two novel types of human papillomavirus, HPV63 and HPV65
            comparisons of their distinct clinical and histological features
            and their DNA sequences to other HPV types
  JOURNAL   Virology 194, 789-799 (1993)
COMMENT     Submitted (27-JAN-1993) on tape to the EMBL Data Library by:  H.
            Delius, Deutsches Krebsforschungszentrum, Abt ATV, Im Neuenheimer
            Feld 506, W -6900 Heidelberg, FRG.

            The 7348 bp genome of HPV-63 was isolated from punctate keratotic
            lesions of the foot.  Infection from HPV-63 produces
            intracytoplasmic inclusion bodies (ICB) in most cells of the
            epidermis.  These inclusions have been previously shown to contain
            mainly E4 proteins for HPV-1.  Up to date three morphologically
            different types of ICBs have been distinguished, and certain types
            of HPV are preferentially found in one type of ICB or another.
            HPV-63 is associated with  a filamentous type of ICB (FI-ICB).  The
            author notes that histologically, these FI-ICBs were seen as
            heavily stained keratohyalin-like substances with filamentous
            brush-like structures within vacuolated cells, seen either as
            single cells or clumped in nests.

BASE COUNT     2239 a   1359 c   1612 g   2138 t
ORIGIN
        1 gtTAAcaact atcaggcgat tctctagttc taacacgaac gtttacggtc gttgccagct
  E6 orf start ->
       61 ttttccttat aaaactctgg tgggaatttc tcttgggaca gATGgacctg acatctgtac
                                                        E6 cds ->
      121 attcggttcg ggatctgagt tctgctctcc gtatcccatt tattgatttg gttgttcctt
      181 gcaatttttg cttgaaattt cttacaaatg ctgaaaaatt gctgtttgat tattttgact
      241 tgcatcttat ctggcgagat aatttcgtgt ttgcttgttg tcagtgctgt gctaggcatg
      301 ttagtctgct tgagtttatg ctttattatc aggagtcttt cgaggtatct gaagtagaag
      361 aattacttaa tcaacctctt gtaaatattg gtttaaggtg tgttacatgc acaaaaaaac
      421 tgactgtttc agaaaagtta gctgttgttt ctgctggaga aagagttcat aaagtaagga
      481 acaaattcaa agcaaagtgc agtttgtgca gactctacat tataTAGttt gtgcagactc
                                                              <- E6 end
      541 tataTAAtta acaATGgtgg gagagcagcc aaatataggt gatttggtga gtcaagaaga
  E7 orf start ->      -> E7 cds
      601 accaagcgtc ctagatctaa attgttatga ggatatacct gctgaggagg aggagtctga
      661 atatccatat gcaattgtgc ttccttgtgg tttgtgcgat cagctgttaa ggctgacctg
      721 cgtttctgac ctgtctactc ttacgcgtct ggaggagctg ctgttaggct cacTGAggat
                                                              E1 orf start ->
      781 cgtgtgtccc ctgtgtgcca ttcgacacca acgacacTAA gATGACCGAC AGAGGTacaa
                                                  E1 cds -> -> E2 bind
                                                        <- E7 end
      841 ataatgatga ttggtatatt gtggatgagg cagaatgtcg ggatgatgat gagagcgaat
      901 tggaggattt ggaggacacc tataattcat gtttaatag atctgaaagt gacatatcag
      961 atctattaga cgatacgcag caaagtcagg gaaattccct ggaactgttc cacttacagg
     1021 agcacttgca gaacgagcag gacctaaata ccctaaaacg aaagtactta aacagtcctc
     1081 cgcaggcaag tgccacagag actgcctgca atagcctcag tcccagattg gaatctataa
     1141 caatttcgca gagggaaaaa aaggcaagaa agcaactatt tacacaaaat gacagtggca
     1201 tagagttatc gctatgccag gatgaagttg acaatattaa cgaagcgctt caggagcagg
     1261 tagacatcgt acagtctctg ggaggtgggg tgcgtgactg tataggagtg gacattttga
     1321 aatgcagtaa tacaagatct gctctacttg ccaaatttaa agacacagta ggtgtcagtt
     1381 ttactgacct caccagagca tacaaaaaca caagacatg ctgtagttac tgggtcatag
     1441 cagtggggg agtaacatct acgtctgtgg acgttctgaa aactgtattc caagttcagt
     1501 gtaattatat gcatgtagaa cattgtttaa ctgaaaaaaa taagtttcta attgtattag
     1561 ctggctttaa agctcaaaaa agtagagaaa cagtgttaaa tctcgtaact agcagtttga
```

```
1621 atgtgcaaag taattacata atggctgaac caccaaaaaa tagaagtatg gcggcagcgt
1681 tatattggta taggagatct atgtctccag ctgtatatac ctggggagaa atgccagatt
1741 ggatggcgca gcagacattg ttgaatcatc aattagcatc agaaaagcat tttgaattgt
1801 cacaaatggt acaatgggct tatgataatg ctatacaga tgaaagtgat attgcatact
1861 attatgctat tttagcagaa gaagatgaaa atgcaaaagc attcttggct tctaatgcac
1921 aagcaaaata tgttaaggac tgtgctagaa tggttagtca ttacaaaagg gcagaaatga
1981 gtagtatgtc tatgtcagca tggatttata aaaagactgga ggaagttgaa aatggtggtg
2041 actggaaaca tattgtaaag ttcttgaggt ttcaagaagt agaatttata agtttcatga
2101 tagcatttaa ggaattgtta agtggtaaac caaagaaaaa ttgtcttgta atatatggtc
2161 caccaaatac tggtaaatct atgttttgta tgagtttgtt gagagtatta aaaggaaaag
2221 taatatctta tgtaaatagc aaaagtcaat tttggttgca accactagct agcactaaaa
2281 tagcattatt agatgatgca acaaaaccag catgggatta tattgattta tttttgagaa
2341 atgctttaga tgggaatcct atttgtgtag atctgaaaca taaggcacca caacaaataa
2401 aatgcctcc acttatgata acttctaata taaatgttaa ggctgatgta tgttggatgt
2461 atttacatag taggataaca tgttttgaat ttaaacaacc ttttccattt gatgaaaatg
2521 gtcaaccggc attttcctta acagacatca attggaaatc ttttttttgaa aggttttgga
2581 gccagttaga ctTAAgtgac caagaagacg aggagagtgA TGgaaagcct caacaaccgc
                E2 orf start ->                    E2 cds ->
2641 ttagactggc tacaagagca gcttctaact ctataTGAga aagacagtaa agatattgaa
                                                <- E1 end
2701 gatcagataa tgcagtggaa tctacttaga caggaacaag tgttattcca ctatgcccga
2761 aaaaaggaa taatgcgact tggcctgcaa gttgtgcctt cccttgcagc ttcccaggat
2821 aaagcaaaaa cagctataga aatgactctt tatcttagtg gcctcagaga ctcacaatat
2881 ggttctgaac agtggtcttt acaagatact agcagagaaa tctttttagc accaccagat
2941 catacattca aaaagggagg gcaaacaatt gaggtaatct atgatgagga tcccaataat
3001 agcaccagac atactgtatg gcgccatata tattatcaaa acggtgataa cagatggaga
3061 aaagcagcta gtgatgtaga tgttcatggt gtgtttttatt TAGaatATGa tggtgtcaaa
                                                E4 cds ->
                                              E4 orf start ->
3121 aactactatg ttgactttca agaagaggcc aatcgatata gcaaaacagg tcgatatact
3181 gttcaatatg agggtaaaag gttcacaaat gttatgtctc ctgtcaatag ctcccccacta
3241 cggacttctg ggtctcctac agacaccaac ccagccaccc aaggacaatc cacccaaact
3301 gccagaaaag cagagacgaa ggggtcgaga caccacccga aatcgccggc tgttcgcaag
3361 cgacggcct acggacgaag aaggtccaga agtcccagag ataccaccct cagacgagga
3421 gaaggagaat cggccagagc ctctgccggt agtggagaac gggtggcatt catttctccg
3481 ggagacgttg gaacatcaac taggtcgcct ccaaagggag gtcaatcaag acttcgaaga
3541 cttatacagg aggctcggga tccacccaTA Atttgtctga aggggggccc taatcaactt
                                                <- E4 end
3601 aagtgcttaa ggtataggat taaagcttca aattcatctg actttgaaag tatcagtact
3661 acatggcatt gggtacataa taaatgcaca gatagagtag gtcatgcacg tatgctggtg
3721 cgttttatat caacagaaca acgtgaccga tttttagata aggtggtggt gcctaaatct
3781 gtttctgtta ttttaggggc attTGAcggt tccTAAgggt gggtgttggg gtatattttg
                L2 orf start ->         <- E2 end
```

```
3841 taatcATGtt aagagtacgt aaacgacgag ctgctccaca agatatttat cctgcttgta
    L2 cds ->
3901 aggttgcaaa caattgcccc cctgatatac aaAATAAAat tgaacaaaca acagttgctg
                                signal ->
3961 acaagatttt acaatatggg agtttgggaa tattcctggg aggtttgggt attggtactg
4021 gcaagggtgg gggtggccgg tatggttata cacctctagg ggacagtggt gcggtgcgag
4081 ttggtggcag aagtacacct gtaagaccaa cagtacctgt ggagactgta ggaccaaggg
4141 atatattacc tatagattca ttggatcctt tagggccctc agtcattgaa ctagaagata
4201 ttccagccac aacagtggaa gtagtggctg aagtgcatcc catatctgat actccacaaa
4261 taccggcacc tactactgat gaatctagtt cagctgttct tcatattcca caagaaagtc
4321 ctgctgcacg tacaatcaca cgttcccaat acaataatcc tttattcagg atcacagctca
4381 gtgcagacat agcatcaggt gaagcttcag catctgataa tattttttata gatgtagata
4441 cgccgggtca aatagtagga caagaaatac cactagttaa ttttgatatg ggacctatat
4501 ctactgaagg tgagcttgaa actgagttca caactagtac accaagaacc acacaagtac
4561 aggaaaggcc tacacgtttc tataatagac gctattatga acaagtgcca gttactgcac
4621 ctgaatttat cacaaggcct gcttccttag ttacttttga gaatcctgca tttgaaagga
4681 gtgtttcttt gattttttgaa caagatttag aagatatttt aaatgctcct gatcaggatt
4741 ttagagacat tgtttatttta agcagaccaa catacagtcg tgcccctgat ggccgcatgc
4801 gcctaagccg cctgggacgc agagccacta taagtACCAG AAGTGGTgtt actataggtg
                                            -> E2 bind
4861 ctcaatcaca cttttatatg gatattagct ctatctcctc aaatgatggc attgagttac
4921 aaacactggg tgaagcttct ggcgagactg tggtgcaaag ttctcttgct gcatcggatc
4981 ctattgaagc agaacattca ttcattgaac cagcaccatc tatagatagt tatgatattg
5041 tttcacttca gtctgagact tattcagatg aacatttgtt agatatgtat gaacctgtag
5101 gttcttcctt gcaattacaa atatcagacg tcagaggtcg gccaactgtt attgatattc
5161 cctttagacc ccgcaggcct ccattaggtc ctataaatgc tggtgttgat atctatagtc
5221 caactgctag tgttggatca cctactataa atcctactga tcttgacatt ccattaatta
5281 ttatacattt agataattca acaggggatt atgatttaca tccaagtttg cgtaaacgtc
5341 gcaaatTAGt tcatattTGA tattttacag ATGgctgttt ggcttcctgc ccagaataag
    L1 orf start ->        <- L2 end     -> L1 cds
5401 tttttaccttc ctacccaacc gatcaccaag attctaagca gcgatgatta tgtgtctcgc
5461 accaacatct tctatcacgc taccagtgat cgactgctca ttgtgggaca cccgctctat
5521 gaggttaccc gtgcaaatga taacactatg actgtgccta agtttctcc aaatcagtat
5581 agagtctttc gtgttagatt tccagatcct aaccgatttg cctttggaga taaggatatt
5641 tttgacccag aaactgagag actagtttgg ggtcttagag gcatagaaat cggtaggggt
5701 caaccattag gtgtgggttat ttcaggcaat ccattataa ataggtttga tgatgctgaa
5761 aatcctagca gatataataa tacacatgca actggtgata ataggcaaaa tgttgctttt
5821 gatgcaaaac aaacccaaat gtttctaatt ggctgtacac cagccactgg ggaacactgg
5881 tcaatagctc gacgctgtgc aggaacacag tttcagcttg gagattgtcc tcctatagaa
5941 ttagttaaca cagttattga ggatggtgat atgtttgaca taggtctagg tgctatggac
6001 tttggttctt tgcaagcaaa caaagcagat gctccttgg atattgcagg cactgtctgc
6061 aaatatccag attatattaa aatgggacag gaagtacatg gtaattctct gttttttcttt
6121 gctcgcagag aacaaatgta tttaaggcat gtatttacac atgctggaat tgttagtgaa
6181 aaagagaaag tccctaccag tgcatatatt gctgctaaag ccgagcaacc ccaaaatact
6241 attgctacag ataattattt tgtagctccc agtggatctt tagtgtcctc tgatgtgcaa
6301 atttttaata ggccctattg gttacaacgt tctcaaggac agaacaatgg tatctgttgg
6361 agaaatgagt tatttgtaac tgtagctgat aataccagag gaaccacgat gaatataaat
6421 gttcttaaca aagcaacccc tgagacttat gatagcgcag attataatga gtatactcgc
6481 catgtggagg aatatgagtt atcctttata gttcagcttt gtaaggtaaa actaacacct
6541 gaaaatttag cattttttgca aatatggat ccaacaatta tcgattcctg gcagttaaca
6601 gtttctcaac ctcctgcaaa tgctatagag gacaagtata gatttattga atcattagca
6661 acaaaatgtc ctgataacgt gcccccaccc actcctactg atcccttacaa agatttacgt
6721 ttttgggatg tagacctcag tgagcgaatg tcggagcagc ttgatcaatt tcctttaggc
6781 cgcaaatttt tgtatcaaag tggtcttgca cagcgttctg ttccaaaaac tgtgaatttc
6841 agaaaacgta gatcctccaa tactactgtg gccaaacgga ggcgacgggc cTGAatatac
                                                            <- L1 end
6901 atgtgaatgt tgaatTATAT AAtgtgaatt gtgaattctt gactttggca cttgcactttt
                signal ->
6961 attcttggca tactgatact tgaaacttgt tcaatgcttg aaggttacac acctgtacag
7021 tattgttaat aaacgtttat gctgctgtca tttacctgtc ttcgagtcat tattgcctag
7081 tcatatagcc tcatgacttg gcatgcaatt ggtatgtggc agatacttca aacaggatac
7141 tggtatcctt tttggcgcgc gcgcgaattt tgaagttacc actgttccaa cttgttctga
7201 gacgtctgga tctgatcccg ACCGCTGTCG TTactgccaa agacgaaagt ggtaggcgcg
                        -> E2 bind
```

```
7261 aACCGTTTGT GGTttccctg gggctagcag aaactcttta ggttgcgACC GTTTTCGGTc
        -> E2 bind                                        -> E2 bind
7321 gggccaataa tctctttcga tcgttgtt
```

```
LOCUS       HPV65        7308 bp ds-DNA          VRL        04-OCT-1993
DEFINITION  Human papillomavirus type 65 (HPV-65), complete genome.
ACCESSION   X70829
SOURCE      Human papillomavirus type 65 DNA recovered from a black verrucous
            lesion.
REFERENCE   1  (bases 1 to 7308)
  AUTHORS   Egawa,K., Delius,H., Matsukura,T., Kawashima,M. and
            De Villiers,E.M.
    TITLE   Two novel types of human papillomavirus, HPV63 and HPV65
            comparisons of their distinct clinical and histological features
            and their DNA sequences to other HPV types
  JOURNAL   Virology 194, 789-799 (1993)
COMMENT     Submitted (27-JAN-1993) on tape to the EMBL Data Library by:  H.
            Delius, Deutsches Krebsforschungszentrum, Abt ATV, Im Neuenheimer
            Feld 506, W -6900 Heidelberg, FRG

            The 7308 bp genome of HPV-65 was isolated from a black verrucous
            lesion.  Lesions of this type are mainly distributed on the
            palmoplantar or lateral surfaces of the hands and feet.  Infection
            from HPV-65 produces intracytoplasmic inclusion bodies (ICB) in
            most cells of the epidermis.  These inclusions have been previously
            shown to contain mainly E4 proteins for HPV-1.  Up to date three
            morphologically different types of ICBs have been distinguished,
            and certain types of HPV are preferentially found in one type of
            ICB or another.  HPV-65, like HPV-4, is associated with a homogenous
            type of ICB (Hg-ICB).  Egawa et al. note that histologically, these
            Hg-ICBs were usually present as a single inclusion in each cell,
            surounding an extremely vacuolated nucleus.  HPV-4 and HPV-65 have
            both been detected in common warts and keratotic flat lesions on
            the hands and feet, as well as in two biopsies from cutaneous horns.
            All of the lesions in this study were pigmented.

BASE COUNT     2280 a   1306 c   1575 g   2147 t
ORIGIN
        1 attattaatg atagttggca acaaccatca tctaaagtaa tataTAAccg gaagagatac
                                              E6 orf start ->
       61 atataaaaag acacattgta ttcctgataa atcccatcca gATGgcagat ggcagacctg
                                              E6 cds ->
      121 cagctttgga cgacttctgc agacgatttg atatttcttt ttttgatttg catcttactt
      181 gtattttttg ttctcatact gttgatctgc aagatcttgc ttcgttttat cttaagaagc
      241 ttagtttagt tttcaggggg ggttgctact atgcatgctg ttctgaatgc ttacgtttaa
      301 gtgcaagatt tgagcaagag aattattttc agtgttctat aaaaagctgtt aatttggaag
      361 aagtagctca aaggaaaatt aaggagattt gcatcagatg tatatgttgt ttgagactac
      421 tagacattgt tgagaaatta gacttgctgt actctgacca ggcctgctac ttaataaggg
      481 gtttgtggag gggctattgc agaaattgca tTAGgaaaca ATGAgaggag cagcacccag
                          E7 orf start ->      -> E7 cds
                                                       <- E6 end
      541 ggttgcagat cttaatttag aattaaatga cttggtgttg cctataaacc tgctgagtga
      601 ggaggtcttg caaccttcag atgatgagtc tgaggctcca gaggaggagc ttttttccttt
      661 tagaatagac acctgttgct atagatgtga agttaatgta aggattactc tgtttgctgt
      721 ggaatttgga cttcgagcgt tggaacaact catagtggac ggaaagctga cgttttgctg
      781 tactacttgt gcaagaacct TAAgaaATGg cagaTAAagg tacagaaaat tttgacttag
                      E1 orf start -> -> E1 cds <- E7 end
      841 aagggagtag ttggtatatt gtgcatgaag cagaatgcac agacagtata gacactttgg
      901 aggatttgtg cgacgaaagc gattcaaatg tttccaattt aattgatgac gatgtggttg
      961 atcagggaa ttccctggca ctgtacaatg caaaaataac tgatgattgt gataatgcaa
     1021 tagcacacct aaaacgaaag tataacaaaa gtccagagca ggcagttgct gaattgagcc
     1081 ctcagttgca ggctgtgaaa attactcctg aaagaaacag caaaaggaga ttatttcagg
     1141 aggacagtgg gatatttgaa gatgaagctg aaaattctct tacacaggta gaatctaaca
     1201 gccagactgg tggcaacagc caagatggcg gaggagatat taacttactg ttgttacaaa
     1261 ctagtaatag gcgtgctaca atgcttgcaa aatttaaaga ttggtatggg gtttcatata
     1321 atgaaataac tagagtttat aaaagtgata agtcttgtag tgataattgg gttatagtta
     1381 tctttcgagc tgctgttgaa gttttggaaa gctctaaaat tgttttgcaa cagcactgta
     1441 catatattca agttaaaata tttggatttt cagctttata tttattacaa ttcaaaagtg
     1501 cgaaaagcag agaaacggta caaaagttaa tgtgttctat gttgaatatt caagagtttc
```

```
1561 aaatattgac tgaccctcct aagttacgaa gtgtgcctac agctttatac ttctataagc
1621 aagccatgtt aacagagagc tttgtttttcg gacagacacc agattggatt gcaaagcaaa
1681 ccttagtaag ccatcaagca gcaactactg cagaaacatt cgaattatca aaaatggttc
1741 agtgggctta tgataataat ctgctggaag aatgtgatat tgcgtatcat tatgctatgt
1801 atgctgatga agatgctaat gctgcagcct atttaaagag taataatcaa gtaaaacatg
1861 ttagagattg cagtacaatg gttagaatgt ataaaagata tgaaatgagg gatatgtcaa
1921 tgtcagaatg gatttataaa tgttgtgatg aatgcACCGA AGAAGGTgat tggaaaccta
                                                -> E2 bind
1981 tttctcaatt tttaaaatat caaggtgtca atattctttc ttttcttata gttctaaagt
2041 cttttttaaa aggtattcct aagaaaaatt gcatagtcat tcatggccca ccagatactg
2101 gaaaatctct gttttgttat tctctagtga aatttcttaa aggtaaagta gtatcctatg
2161 tgaatcgaag tagccacttc tggttgcagc ctttaatgga ttgtaaggta ggatttatgg
2221 atgatgcaac ctatgtatgc tggacataca tagatcagaa tttgaggaat gcattagatg
2281 gaaatccaat gtgtatagat gcaaaacata gagctccaca gcaattaaaa ttACCTCCTA
                                                               -> E2 bind
2341 TGTTaataac atcaaatatt gatgtgaaac aggaacaatc attaatgtat ttgcatagta
2401 gggtccagtg ttttagtttt cctaataaaa tgccttttttt agatgatggt tctcccatgt
2461 atacatttac tgacgcaacg tggaaatctt ttttccaaaa gcttggcaga caattagagc
2521 TAAcagatcc tgaagaggaa agcaATGgag tccctagtcg cgcgtttcga tgcacttcaa
 E2 orf start ->              E2 cds ->
2581 gaagcaattc tgactcatat TGAgtctcag gacgatactt tggaatccca aattcgatat
                        <- E1 end
2641 tgggaaaata tcagaaaaga aaatgcaata atgcattttg cacgaaaaca aggcctaaca
2701 aaattaggtc tgcaaccact tcccacatta gcagtaactg aatataatgc aaaacaagca
2761 atacaaatac atttgacttt gcagtcatta ttaaagtctc cttatggatc agagcgttgg
2821 actctgccag aagttagtgc agaactgatt aatactgctc cacagaactg tcTAAgaaa
                                                           E4 orf start ->
2881 ggaggttATG atgtgtctgt atggtttgat aatgatagat ataatgctat ggtgtataca
     E4 cds ->
2941 aattgggatt atctatatta tcaagatgtc aatgaaatat ggcataaagt taaaggtgaa
3001 gtggattatg atggcttata ctttacagac catacaggag aacgtgcata ctttactctg
3061 tttagcacag atgctcacag atttagcaga actggactat ggactgtgca ttttaaaaca
3121 caagttattt cctcctctgt tgtcagctca acaaacaccc cctccttcga ctttgaggaa
3181 caacaactac ccgggccctc aacacccacc tacaccgagc ttacccaggc gagcccttgt
3241 ggtaggggga atcgaggga atctcaaccg acctccacaa cgtcccccga aacctcgggg
3301 ctacgagtac gacgaggacg acgacaaaga aaatcagggc ccgggccagg agagaccccc
3361 agcaaaagaa gaagaggagg agggagagga ggaggagaga ccagattgga gtctgcgcca
3421 tctcctgggg aagtgggaat cagacataga acagttgaaa gacaaggtct gtcgcgactt
3481 ggacaactac aagctgaagc tagggatcca cccaTGAtat tgttaaaggg cacagcaaat
                                       <- E4 end
3541 tcgttgaaat gttggagata taggaaacaa aattccagta actgtgggtt tctctttatg
3601 agcactgttt ggaactgggt tggtgatgtg tcagaaaatc acagtcgcat gttaattgct
3661 tttaaaagtc ctggtcagag agactcattt gttaaacaca atctattccc aaaactctgt
```

```
3721 acatatacct atggttcttT AAatagctta TAAaATGcaa gcctcacgca gaacaaaaag
                L2 orf start ->    L2 cds ->
                                   <- E2 end
3781 agactcaata ccaaacctct atgcaaaatg tcaattatct ggcaactgtc ttcccgatgt
3841 aaaaAATAAA gtagaagctg atacccttgc tgatcgtttg ctcagatggt tgggaagtgt
  signal ->
3901 catatactta ggaggtttgg gaattgggac gggaagaggc agtggagggt cttctggtta
3961 taatccactg ggagcccota gcagagtaac acctagtgga actgttataa gaccaacagt
4021 ccctgtagaa ggtctggggc ctagtgaaat catccctgta gatgttgtaa ACCCTGGCAG
                                                           -> E2 bind
4081 TTcttctgtg gttcctttag aggatttaac agttccagaa gtcaccatag atagcggggga
4141 ggtgggagga ggtggtcttc atccttctga gatagatgtt gtcacttcct cagaccctat
4201 ttcagatgtc actggtacaa gtagccaccc tacaataata tctggcgaag ataacgccat
4261 tgcagtccta gatgtttctc caacagagcc ccctaccaaa cgcatagctt taggaactag
4321 gggcgcaaca tcaacaccac atataagtgt aatatcaggc acgACCGAAT TCGGTcagtc
                                                           -> E2 bind
4381 ttcagaccta aatgtgtttg ttaatgccac atttttctgga gattctattg gatatacaga
4441 agaaattcct ttagaagagt tgaatactat ccagcaattt gaaatagaaa cccctccaaa
4501 aactagcaca ccacgtgaga ctattggacg tgctttggaa agagcgcgag atctttataa
4561 cagaagagtg cagcagatag ccactagaaa tccagcaatg ctaggacagc cttcccgcgc
4621 aatagttttt ggatttgaaa atcccgcctt tgacgctgac atcactcaag tatttgagcg
4681 ggacttggaa caggttgcag cagctccaga tgctgatttt gctgatatag tcagaatagg
4741 gcgtccaaga ttttctcaga cagacactgg gcaaattaga attagcagac ttgggcgccg
4801 aggaactatt aaaactagaa gtggtttgca aattggtcag gctgtgcatt tttattatga
4861 cttgagtaca atagacactg ctgatgcaat tgagctatct cacttggtc agcattcagg
4921 agaacaaagc attgtagatg caatgataga gagcagtttc gtagatccct ttgaaactcc
4981 tgatcctaca tacacagaag agcaacagct tctagatccc cttacagagg atttttagcaa
5041 ctctcacttg gtttttaacta gcagtagacg tggatcttca ttcagtatcc ccactattcc
5101 tcctggttta ggcttgagaa tttatgttga tgatgtaggc tccgatttat ttgtatctta
5161 cccagaaaca agagttatac ctgcgggagg cttacctact gagcctttta ccccgttaga
5221 gcctccattt ttttcagagt tttacagttc tgactttgta tatcgtccTA Gtttgtatcg
                                                L1 orf start ->
5281 caagaaacga aaacgatcag atatatttTA Attttttgca ggaacATGGc gagttggtta
                                    <- L2 end          -> L1 cds
5341 tctgcaaagg gtaaagtgta cctacctccc gctcaacctg tggcaagagt tttggaaact
5401 gacgaatata tcactggaac atctctttat ttccacgctg gaacagaaag gcttttaact
5461 gtaggccatc cttattttcc agtgaaagat gtgcaggcat aacacaaggt tttagttcct
5521 aaagtttctg gaagtcagta tagagttttc agattctatt taccagatcc taataggttt
5581 gcattgattg ataatggttt ctatgattct gaccatgaac gtcttgtttg gagactcaga
5641 ggaatagaaa taggtagggg tggtccccctt gggattggaa caacaggcca tcctttatat
5701 aataagtttg gagattcaga aaatcctaat ggatacagga aacaatcaga tgacaacaga
5761 caggatgttt ctttagatcc aaaacagaca caaatgtttta tcataggttg cactcctgct
5821 attggggaac attgggataa ggctgagcct tgtgcaagtc ctgttcctca accaggagat
5881 tgtcctccta tagagcttgt gaacacatat attgaggatg gggatatgtg tgatattggt
5941 tttggagctt ttaattttaa agctttgcag caggataaat ctagcgctcc attggatgta
6001 gttgctacaa tgtgcaaatg gccagatttt ctaaaaatga gtaaagatgt ctatggagat
6061 agcttattct tttatggtag gcgagaacag ttatatgcca ggcacttttt tgttagagca
6121 ggagcaatgg gagatgcact accagaACCT TTCGAGGTta aaactgatta ctggattcct
                                   -> E2 bind
6181 gctcaagaag gtcaggatca gaatacttta ggtccacata tatatatagg cactcctagt
6241 ggatctttag tttcaagtga gtcccagtta tttaatcgAC CGTATTGGTT aaacagagct
                                              -> E2 bind
6301 cagggtacaa ataatggaat atgctgggat aatcaattat ttgtgacact tgttgacaat
6361 actcataata ctaattttac tatttctgta aaaacagaag cagctgatga atcctataaa
6421 tataaagcag gtgattttaa acagtattta agacacattg aggagtttga aatggaattt
6481 attttttcagc tttgcacagt tcctcttaca gccgatgtca tggctcactt aaatgtaatg
6541 aatcctaata tcttagacaa ttggcaatta aattttgtgc ccctcctcc ttcaggaatt
6601 gaggatcaat atagatttat acaatccaga gctacgagat gtcctactca gtctcccagt
6661 actgaaaagg aagacccttaa taaagattta tcattttggg gtgtagattt aacagaaaga
6721 ttttccagtg aacttagcca atttttccttaa gggagacgct ttttatatca aagtggttta
6781 attaatggaa cacttaaaag aaaacgtact attaattctc aagctccaac tagtattaaa
6841 cgttctgcca aacgaaaacg gtctattaaa cagTAGatta atccttataa gaatgcattt
                                        <- L1 end
6901 gttctgccaa tagaaactta aactgagcca attatTATAT AAtctttgga aacgtgtatg
```

```
                                        signal ->      <-
     6961 tatttgcttt ttctgtgaga aacttgagta aAATAAActg ccacattctt gccactgtcc
                                          signal ->
     7021 atattgtctg actcattggt caacatgtgt gtccgcaccc cagtaattac ctggatcgcc
     7081 tccacatccg tcggatgcgc gccaaaagag gaacgtcccc ttctattttt ctaaaaaatt
     7141 accgtttctg cagctgcaaa attgtgtaag ACCGTTATCG TTccctgtaa ccttggcaca
                                          -> E2 bind
     7201 aaaggtgagt gaaagtttta tagtaacgtt tatgagtcaa tttgtctggc ggcgctgaac
     7261 gaatttggct gtcagccttt gcACCGGGAG TGGTagaaaa tagtttct
                                          -> E2 bind
```