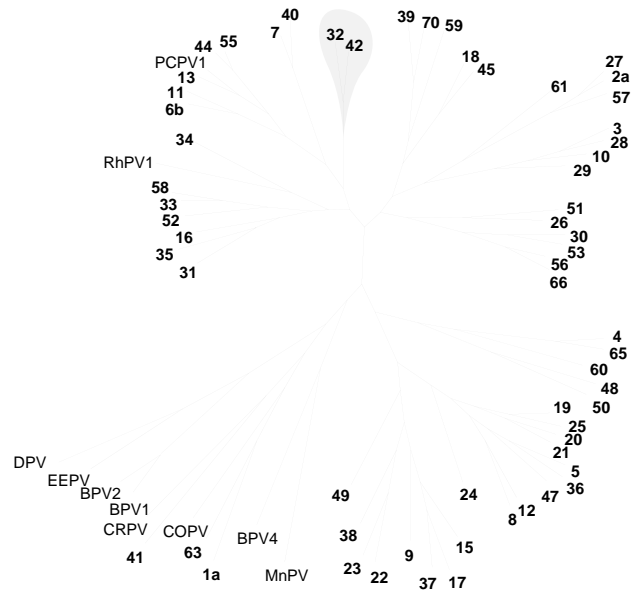# Group A1 Sequences

HPV32   HPV42

INTRODUCTION

The members of group A1 formed a subset of Group F in the old classification. This small group is composed of viruses HPV-32 and HPV-42, which are primarily associated with benign orogenital lesions. HPV-32 was originally isolated from an oral focal epithelial lesion and further detected in several oral papillomas [1]. When 113 benign and malignant tumors of the ororespiratory system were analyzed for the presence of HPV-32 DNA, 9% of the oral papillomas were positive for HPV-32 [2]. In contrast, HPV-42 is predominantly found in genital lesions. Lorincz et al. have classified HPV-42 as a "low-risk" virus [3]. This placement is justified by the infrequent association of HPV-42 with dysplastic lesions and the absence of HPV-42 in malignant lesions [4,5].

What's new?

No new sequences in Group A1 were released during 1995. The sequences of members of this group were published in *Human Papillomaviruses 1994* pp. I-F-26, and I-F-34.
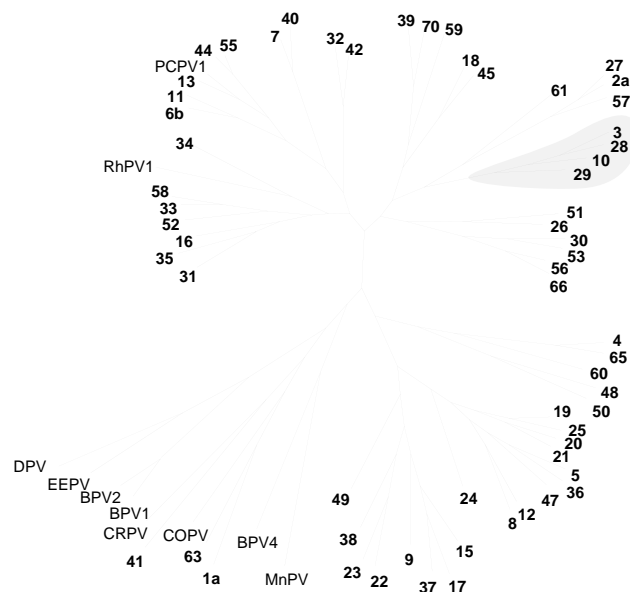

References

[1] de Villiers,E.M. Human pathogenic papillomavirus types: an update. in *Human Pathogenic Papillomaviruses*, edited by Harald zur Hausen, Springer-Verlag, Heidelberg, pp 1–12 (1994)

[2] de Villiers,E.M., Weidauer,H., Le,J.Y., Neumann,C., and zur Hausen,H. Papilloma viruses in benign and malignant tumors of the mouth and upper respiratory tract. *Laryngol Rhinol Otol (Stuttg)* **65**: 177–9 (1986)

[3] Lorincz,A.T, Quinn,A.P., Goldsborough,M.D., Schmidt,B.J., and Temple, G.F. Cloning and partial DNA sequencing of two new human papillomavirus types associated with condylomas and low-grade cervical neoplasia. *J Virol* **63**: 2829–34 (1989)

[4] Beaudenon,S., Kremsdorf,D., Obalek,S., Jablonska,S., Pehau-Arnaudet,G. Croissant,O., and Orth, G. Plurality of genital human papillomaviruses: characterization of two new types with distinct biological properties. *Virology* **161**: 374–84 (1987)

[5] Philipp,W., Honore,N., Sapp,M., Cole,S.T., and Streeck,R.E. Human papillomavirus type 42: new sequences, conserved genome organization. *Virology* **186**: 331–4 (1992)

Group A2 Sequences

HPV3 **HPV10**
HPV28 **HPV29**
HPV77

INTRODUCTION

The members of group A2, HPV-3, HPV-10, HPV-28 and HPV-29, formed a subset of Group F in the old classification. Group A2 consists of viruses that are associated with cutaneous benign lesions. HPV-3, HPV-10, and HPV-28 commonly cause benign flat warts in patients with EV, immunosuppressed patients, and in the general population [1,2,3]. HPV-28 has also been found in intermediate warts, a type which exhibits characteristics of both flat and common warts [3]. HPV-29 was isolated from a skin wart. Prevalence data indicates that HPV-29 occurs very rarely [4]. HPV-77 (isolate VS-93-1-G) was isolated from cutaneous lesions of renal transplant patients; identical or near-identical sequence was obtained from 10 isolates (4 common warts, 3 dysplastic warts and 3 SCCs) drawn from three or more patients [5]. Also relatively frequently found in this same study was HPV-27. The isolation of these types and HPV-28, which come from phylogenetically related groups, in immuno-suppressed patients is suggestive.

What's new?

Last year we published only MY09-MY11 fragments of HPV-28 and HPV-29. This year, on the following pages, we publish the complete genomes of these two viruses. We also publish here the novel L1 fragment of HPV-77.

References

[1] de Villiers,E.M. Human pathogenic papillomavirus types: an update. in *Human pathogenic papillomaviruses*, edited by Harald zur Hausen, Springer-Verlag, Heidelberg, pp 1–12 (1994)

[2] Kremsdorf,D., Jablonska,S., Favre,M., and Orth,G. Human papillomaviruses associated with epidermodysplasia verruciformis. II. Molecular cloning and biochemical characterization of human papillomavirus 3a, 8, 10, and 12 genomes. *J Virol* **48**: 340–51 (1983)

[3] Favre,M., Obalek,S., Jablonska,S., and Orth,G. Human papillomavirus type 28 (HPV-28), an HPV-3-related type associated with skin warts. *J Virol* **63**:4905 (1989)

[4] Favre,M., Croissant,O., Orth,G. Human papillomavirus type 29 (HPV-29), an HPV type cross-hybridizing with HPV-2 and with HPV-3-related types. *J Virol* **63**: 4906 (1989)

[5] Shamanin,V., Glover,M., Rausch,C., Proby,C., Leigh,I.M., zur Hausen,H. and de Villiers,E.M. Specific types of human papillomavirus found in benign proliferations and carcinomas of the skin in immunosuppressed patients. *Cancer Research* **54**:4610–3 (1994)

```
LOCUS       HPV28        7959 bp    DNA              VRL        18-JUL-1995
DEFINITION  Human papillomavirus type 28, complete genome.
ACCESSION   U31783
KEYWORDS    .
SOURCE      Human papillomavirus type 28.
REFERENCE   1  (bases 1 to 7959)
  AUTHORS   Delius,H.
  JOURNAL   Unpublished, Sequenced by Hajo Delius, Deutsches
            Deutsches Krebsforschungzentrum, Angewandte Tumorvirologie,
            I.N.F. 506, W-6900 Heidelberg, Germany
REFERENCE   2  (bases 1 to 7959)
  AUTHORS   Farmer,A.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-1995) Andrew D. Farmer, HIV Sequence Database,
            Los Alamos National Labroatory, T-10, Mail Stop K710, Los Alamos,
            NM 87501, USA
COMMENT     HPV-28 was originally isolated from butchers' warts (Favre et al.,
            J. Virol. 63, 4905).  Subsequent tests revealed it to be present in
            7 of 130 butchers suffering from warts, in 3 of 66 wart patients
            and in 1 of 55 immunosuppressed patients. It has been detected in
            flat warts and intermediate warts.  It is most closely related to
            HPV-3 and HPV-10.
BASE COUNT     2161 a   1677 c   1994 g   2127 t
ORIGIN      102 bp upstream from beginning of E6 cds
        1 taaataataa tgaacagaag caaaagTAGg gagggACCGA AAACGGTacg ACCGAATGGG
                          E6 orf start ->
                                    E2 binding ->    E2 binding ->
       61 GTacaTATAA AAagaggctt acattgaatg gcagaagctc aATGgatgac caacggccta
          signal ->                                   E6 cds ->
      121 aaaacatatt cctattgtgc agagacagtg gaatatcatt tgatgacctc cgtttgcact
      181 gcatattctg cgccaaagtg ctgaccacag cggagctatc agcatttgca ttaagagaac
      241 taaatgtggt gtggagaacg ggagcgccct atggtgcctg tgcccggtgt ttactactgc
      301 agggcattgt gcgacgcctg aagcactggg actattcatt gtatgtggaa ggcgtggaag
      361 aggagacaaa acaatctata gatacacagc aagtgagatg ctacatgtgt cacaaaccac
      421 tggtTAAaga agaaaaggac agacacagga cgaacggcg acgactgcac tgcatagctg
  E7 orf start ->
      481 gctactggag agggagctgt caatactgct ggttaagATG cacggtccgc atcccacagT
                                       E7 cds ->
      541 AAaagatata gaattgagtc ttgcaccaga ggacgtccca gtgcaatgca atgtgcaatt
          <- E6 end
      601 agatgaagaa gattatacaa atgtggagga accagcacaa caagcgtata gggtggtaac
      661 actgtgtcca aagtgtagtt caccactgcg actggtggta gagtgcagcc acgcagatat
      721 aagggcattg gagcagctcc tgttgggcac actgacggtc gtgtgccctc gctgcgtgTA
      781 ActggacATG gatgatactt caggtacaga gggggacgaa tgttcagagt tggaacgggc
          <- E7 end
          E1 cds ->
  E1 orf ->
  start
      841 tgggggatgg tttatggtag aggccatagt agacaggcgg acgggcgata agccatctag
      901 tgatgaggat gaggacgagg atgcagacga aggggaagat tttgtagatt ttatagatga
      961 caggcctgta ggggatggac aggaagtggc acaggaactg ttgctgcagc aagcagctgc
     1021 agatgatgat gtagcagttc aggctgtaaa acgaaagttt gctcccagtc cgtatttcag
     1081 tcctgtgtgt atgcaaccca gcatagaaaa tgaactcagt ccgaggctag atgccataaa
     1141 gctgggggaga cagtcgggga cagccaaacg ccggctattc cagctaccgg acagtgggta
     1201 tggccaaaca caggtggata cggaatcggg ACCACTGCAG GTacaggaca tttgtgagac
                          E2 binding ->
     1261 ggggacacaa gatggccgcc aggacgcgga tgaggggagc ggtaggaatg tggggggaaa
     1321 tggcggccag gaggaggagc gtgcaggagg ggatggggag gaatcgcaga ctcagggtgt
     1381 acaaacagac aaggcagcct gtggagtgtt ggcaatatta gagctagca atcaaaaagc
```

```
1441 tacgctactc ggtaagttca aagaacaatt tgggttggga tttaatgaac ttatccgaca
1501 cttcaaaagt agcaaaaccg tttgcctaga ttgggtggta tgcgtgtttg gggtatactg
1561 tacattggca gagggcatta aaacattgat acagcctcag tgtgactatg cacacataca
1621 ggtgctgtcc tgccaatggg ggatgacagt gctgatgctg gtacggtaca aacgtgccaa
1681 aaacagagaa actgtggcta aaggattaag cacattacta aatgtgccag agagccacat
1741 gttgatagaa ccacctaaac taagaagcgg cccagcagcg ctgtattggt ataagacagc
1801 catgtcaaac tgtagtgatg tgtatgggga aacaccagag tggatagtga acaaacgat
1861 ggttggacat gcactagagg aagcacagtt tagtctgtcg gagatggtac agtatgcata
1921 cgaccatgac attacagatg agagcatgct ggcatttgaa tatgcactac tagcagatac
1981 agatgccaat gcagccgcgt tcctaagtag caattgtcag gcaaaatatg taaaggacgc
2041 ttgcacaatg tgcagacact acaaaagagg tgaacaggcg cgcatgaaca tgtcagaatg
2101 gatatggttc agaggtgata aagtacaggg cgatggggac tggaaaccca tagtacaatt
2161 tttaagatat cacgatgtgg aatttatacc attcctgtgt gcctttaaaa ccttcctgca
2221 aggaattcca aaaaaaagtt gtttagtgtt ctatggtcca gcagatacag gaaagtcata
2281 tttctgcatg agtctattaa gatttctagg tggggtggta atatcttatg ccaattcaaa
2341 cagccatttt tggttgcaac cgttagctga tgccaaaata gggttgctag acgatgccac
2401 cagccagtgc tggtgctaca tagatacgta tttaagaaat gcattagatg gaaaccaggt
2461 ttgcatagac aggaagcata gagcactgct acagctgaaa tgtcctcctc tattaataac
2521 aacaaatata aatccactag aggatgacag atggaagtat ctgcgcagca gagtacaact
2581 gtttacattc aaaaacaaat ttccattaac aacacaagga gaaccactgt atacacttaa
2641 tgatcaaaat tggaaatgct ttttttcgaag gttgtgggca cgttTAAgcc ttaccgatcc
                                                E2 orf start ->
2701 tgacgacgag gaggaaaATG gaaaccctag cgaaccgttt agatgtgtgc caggacaaaa
                      E2 cds ->
2761 tgctagaact ttaTGAaaag gatagcaaca aacttgagga ccagataatg cattggcaat
                   <- E1 end
2821 tgatgcgtgt agaaaatgct ttgctgtaca aagcaaggga atgtggattg acacatattg
2881 gccatcaggt ggtgccacct cttagtgtaa ctaaagctaa ggcacgcagt gccattgaag
2941 tgcatgtagc tttgctacaa ttacaagaaa gtgcatatgc acaagactca tggacactgc
3001 gagacacatc acgagaaatg tgggacacag tacccaagaa gtgctggaaa aaaagaggtg
3061 taaccgtgga agttagatat gatggagacg aaaccaaatc catgtgttat gtacactgga
3121 gagacatatt tactcagaac tatagtgatg ataaatgggt aaaggtggca ggacacgtgt
3181 catatgaggg actgtattac atacacgaag gagaacaaac attttatgtc aaattcaaag
3241 atgacgcgta cgtgtatggg gaaacaggaa aatgggaggt acatgtggga ggcaaagTAA
                                                E4 orf start ->
                                                NH₂ terminus unknown
3301 ttcaccacca tgcatttgac cccgtatcta gcacacgaga aatacccgct gctggacctt
3361 tgtgcaccgg tgacaccacc aaggcgtcga ccgaaacctc ggtgggcgcg accgaaggac
3421 cgcagcaaaa acgacagcga ctcgagacac tcaactggga gcagcagcag cgacagtacc
3481 cccaaacccc ctccacgcaa accactgaac gagcaagtca accactggac gttacaagga
3541 ccagcgactg tgacactaca tgtccataca ccgtcgggca cccaagtgac cctgactgtg
3601 cacctgTAGt acacctaaaa ggtgatccta attgcttgaa atgtttttaga tataggttac
             <- E4 end
3661 acaaagggaa aagaaaactg tattgtaaaa cctcttccac atggcggtgg agctgtgaat
3721 cagaaaatca agcagctttt gtaactattt ggtatacaag ttattcacag cgtaatgaat
3781 tcttgtccac tgtaaaggta ccacctggta tacaggtaat attgggacat atgtcaatgt
3841 ttgtaTAAtt atcgaaccgt gtggtgatac agtgtgcaga ctggatactt atttgtgcag
             <- E2 end
3901 gctacatttt tgggcttatt ttgtgctgct gctgtgttcg ttttggctgt gtgtgcttcc
3961 tgcgctacct tgctatctgg caattgtgct gtgtttgtac ctgggcctgg tagcattata
4021 tttgcgaatt gtatcacata ttgtacaaat caatacataa gccttgctat gtatcctttg
4081 gtgctaaggg acaacaatgg cgatcatact gtcctgtttc tggagccaga tgacgtgtat
4141 ctattgttgt tgtttatatt ttttgttata cttgcattgt ttatcttgta tagacatttg
4201 ggttttttgt aactgtgTAG ttgtgtgtta gttggtcacc acaatgtatt tttaagtact
             L2 orf start ->
4261 gtgtatatcc cagtgtgcta ttgagccact gccttttcta cacttgtttt tttttttac
4321 agttcAATAA AgcaaccATG gtggcacatc gtgcaaggcg tcgcaagcgt gcatccgcca
      signal ->    L2 cds ->
4381 cacagcttta taggacctgc aaggccgcag gcacatgccc ccctgatgtt attccaaaag
```

```
4441 tggagggcac tactctggca gatcgtattt tgcagtgggg cgggctgggt atttatttgg
4501 ggggtctggg gattggtacc ggatctggga caggggggtcg cacgggatat gttcctctta
4561 gcacacggcc tgggactgtt gttgatgtta gtgtccctgc taggcctcct gtggtaattg
4621 aacctgtagg accatctgac ccctccattg taaacttatt ggaggactcc agtattatca
4681 attcgggttc caccgttcca acttttctg gtactggtgg ctttgaggtg acttcttcag
4741 ccacaactac ccctgctgtg ttggacatca cccctgccac tgacaatgtg gtcattagta
4801 gttccaactt taccaaccct gcctttactg aaccctcgtt gttagaggtt ccccaaaatg
4861 gtgaagtgtc gggacacatt cttgttagta cccccacggc tggtacacat agttatgaag
4921 aaatacctat ggagacattt gcttcgccag gtactggcaa tgaacctatt agtagcaccc
4981 ctgtacctgg tgtaagtagg atcgctggac cccgtctata tgccaaagct gtcacacagg
5041 ttaaagtaac agatcctgct ttcttgtctc gtcccacctc attggttact tttgacaatc
5101 ctgcatttga gccgggggat gaaactatta tatttgaacg cccataccct ccctcacaag
5161 tgcctgaccc agattttatg gacatcatac gtttgcatag gcctgcctta acatctcgca
5221 ggggtactgt gcgtttcagt cgcctcggta caaaacttag catgcacact cgcagtggca
5281 agggtatagg agctagggta cattattatc aggacctaag ccctataggc cccacggaag
5341 acattgaaat ggagcctttg cttgctcctg ctgaaaacgc tgctggggat tctatttatg
5401 atgtgtttgc ggacgtggag gacgctgaca tagcttttac tggtagaagt cgcagcgcca
5461 cgtcatctcg gggatacact acggtgtccc cactgtcttc cacactaact accaagtatg
5521 gcaatgtcac cattcccttt gtgtctcctg tggatgtgca cttacacccct gggcctgata
5581 ttataacgcc tgcgtctacg cagtggccgt ttgtgccctt ggttcctgct gacacaaccc
5641 attatgtgta caTAGATGgc ggggattttt atttatggcc tgttaccctg tttgtacccc
first 'ATG' of L1 orf ->
        L1 orf start ->
5701 gacgtcgtcg ccgtaaacgt ctctcatatt ttcttgcAGA TGgcactgtg gcgctcTAGt
                        probable L1 cds start ->                  <- L2 end
                                                /\ 3' sj
5761 gacaacctgg tgtacctgcc tcccacccct gtttccaaag ttctcagcac ggacgactat
5821 gtgacacgca ccaatatttta ttattatgca ggcacttctc ggttgctgac cgtgggtcat
5881 ccttattttc ccattcctaa atcatccact aacaaagcag atgtgcccaa agtgtccgcc
5941 tttcagtata gggtattccg ggtgcggttg cctgacccaa ataagtttgg cctgccagat
6001 gcacgcatat ataaccctga tgcagaaaga ctggtctggg cttgcactgg ggttgaggta
6061 ggccgcgggc agcctttggg tgtcggcctc agtggacatc ctctctataa caagctagat
6121 gacactgaaa actctaacat agcacatggg gacattggac aggattcccg agacaacata
6181 tctcttgata taaagcagac gcagctatgt attgttggtt gcacccctcc aatgggggag
6241 cattgggggca aggggacccc atgtcgtcaa aatacacaac agggagattg ccctccgttg
6301 gagcttataa ctgcccctat tcaggatggt gatatggtgg cactggctt tggtgcaatg
6361 gactttgctg tattacaggc caataaatcg gacgtgccct tagatatttg ccagtccact
6421 tgcaaatatc ctgattattt aggcatggcc gcggagccgt atggcgacag catgtttttt
6481 tacttgcgaa aggaacaact tttcgctcgg catttttta atagagctgg tgttgctggg
6541 gacaccattc ctgaaacatt gtatattaaa ggtcagggca acgggaggga tgtgattggt
6601 agtgctgtgt atagtcctac acctagtgga tctatgatct cctctgaggc ccagctttc
6661 aataagccat attggctgcg gcgtgctcag ggacacaata tggtatctg ttgggccaac
6721 caattgtttg taactgtagt ggatactaca cgcagtacaa acatgacgtt gtgtgtttct
6781 actgactctt cagctacgta cgatgctagt aaatttaagg aatacttaag gcacggggag
6841 gagtacgatt tgcagtttat attccagttg tgtaaagtaa ccttgacccc tgatattatg
6901 gcatatttac ataccatgaa caatagttta ttggaggact ggaactttgg gttgactttta
6961 ccaccatcca ctagcttgga ggacacgtat aggttcatat cttcctctgc cattacctgt
7021 caaaaggatg cttcccccac taccaaggaa gaccccttacg ctaaactaaa cttttgggaa
7081 gtggatctta aggatcgctt ttctcttgat ctatcgcaat ccctctgggg aaggaaattt
7141 ttaatgcagt tgggtgtagg tgcccgctcc agtgtctccg ttcgcaaacg tccagcgtcc
7201 accacgagag gatcgtctgc tgcaaaacga aaacgcgcca aaaagTAGca ctactgtgtt
                                                        <- L1 end
7261 ttgtatgttt gcctgttgtt gtcctgtgtg tcagtgtatg ggcgtccctg atatgtgttt
7321 gtacaaagta tggtgtatgt atttgtgtat gcatgtaatg tggttataac tgtgttttgt
7381 atgtttgcct gttgttgtcc tgtgtgtcag tgtgtgggcg tcctgctatg tgcttgtaca
7441 aagtatggtg tatgtatttg tgtatgcatg taatgtggtt ataagttatc atAATAAAct
                                                        signal ->
7501 gtgtcacata gtttttatatt ttttaatttc tgtaattgct gatcctgtga gtaagaaagg
```

```
7561 tggttccagg tcaagcgACC GATATCGGTa tccaaaatgg ccgcctatgc aggtgtgttc
                E2 binding ->
7621 caggtgtgct taataccaat tagtcataca gatccatatc ctgcaaacct gccctgtctc
7681 gcaagttttt ggctgtgata ctatctttct atagtttatt tcatttgctg tatcattctt
7741 tctggcacgt caaaatgtct ccattgcaaa ttaactgctc ctgggcacca acttattatg
7801 actactttca cataattact atcttggctg ggttttatag gttgccttgc caatagaatg
7861 tgcttccaaa tctccaccaa gacacACCTA ATCCGGTcgc tgcttgcttt ctagccataa
                     E2 binding ->
7921 attaagcagt tgctacacgt ttctttccaa ctataatta
```

```
//
```

```
LOCUS       HPV29          7916 bp    DNA              VRL        17-JUL-1999
DEFINITION  Human papillomavirus type 29, complete genome.
ACCESSION   U31784
KEYWORDS    .
SOURCE      Human papillomavirus type 29.
REFERENCE   1  (bases 1 to 7916)
  AUTHORS   Delius,H.
  JOURNAL   Unpublished, Sequenced by Hajo Delius, Deutsches
            Deutsches Krebsforschungzentrum, Angewandte Tumorvirologie,
            I.N.F. 506, W-6900 Heidelberg, Germany
REFERENCE   2  (bases 1 to 7916)
  AUTHORS   Farmer,A.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-1995) Andrew D. Farmer, HIV Sequence Database,
            Los Alamos National Labroatory, T-10, Mail Stop K710, Los Alamos,
            NM 87501, USA
COMMENT     HPV-29 was originally isolated from a skin wart. Subsequent tests
            failed to detect it in cutaneous wart specimens from 119 patients,
            including both butchers and immunosuppressed patients. These latter
            were included in the study on the basis of HPV-29's similarity to
            members of the group including HPV-10 and HPV-3, which are
            frequently detected in such patients. It thus seems to be a
            relatively uncommon member of this group (Favre et al.  J. Virol.
            63, 4906).
BASE COUNT     2177 a   1665 c   1955 g   2119 t
ORIGIN        102 bp upstream from beginning of E6 cds
        1 taTAAactat catcttcata ataaaaagta gggagggACC GAAAACGGTa cgACCGAATG
  E6 orf start ->                           E2 binding ->    E2 binding ->
       61 GGGTacaTAT AAAAagacat cactgcagcg tggcagaagc cATGtccaga ggtgatggct
          signal ->                                   E6 cds ->
      121 atccaaaaaa tatattcctg ttgtgcagag acagtggagt accatttgag gaccttcgcc
      181 tacagtgtgt tttctgcacg aaagagctaa ccagcccaga actggcagca ttttgcattc
      241 gggaattaaa tgtggtgtgg aaaagtggag ctccgtacgg tgcatgtgca cgctgcttat
      301 tgttTGAagg cataaagcgg cgcctaaaat actggcagta ttcttgtttt gtggaaggcg
  E7 orf start ->
      361 tggaagcgga gacaaacgag tccatatata cacagctaat tcgctgctac atgtgccaca
      421 agccacttgt cagagaggaa aaagacaaac accgaaacga aaagcgaaga ctacacaaaa
      481 tttctggata ctggagaggg agttgcctgt attgttggtc acgATGcatg ggccaaagcc
                                                  E7 cds ->
      541 cacggTAAaa gatattgaat tggatcttgc accagaggcc gtacctttag tatgcaatga
              <- E6 end
      601 gcaattagac agctcagatg aagaagattg tatagatgtt gtggaaccag cacaacaggc
      661 gtatagggtg gtaactttgt gtacaaagt tagtacaaca ctgcgactgg tggtagagag
      721 cagcgaagca gatataaggg cattccagga gctcctacta cgcacattga agatcgtgtg
      781 tcctcgctgt gcgTAActgg acATGgccga taactcaggt acagaggggg aggaggagga
                        <- E7 end
                        E1 cds ->
                 E1 orf start ->
      841 ctgttctgag gcggaacggg ctggaggatg gttcatggta gaggctatag tagacagacg
      901 gacagggac acaatatcca gtgacgagga tgaggaggat gagggtgaag acatggtaga
      961 ctttatagat gatagaccta taggggacgg acaggaagta gcacaggaac tgttgctgca
     1021 gcaagcagct gcggatgacg atgaagcagt gcacactgta aaacgaaagt ttgctcccag
     1081 tccctatttc agccctgtgt gtgtgcccag catagaacat gagctaagtc ccaggctaga
     1141 cgccataaag ctgggacggc agtcctctaa agccaaacgg aggctattcc aactaccgga
     1201 cagtgggtat ggccaaacac aggtggatac ggacacgggA CCAAGCCAGG Tacaagatgg
                                       E2 binding ->
     1261 ttgcgagacg ggtgatcaaa atggccgaca gcagtataag gaggggagtg gtacaaagga
     1321 tggggaaaat ggcagccaag aggaggagcg tgcaggaggg gatggggagg aatcgcaacc
     1381 tctgagtaca gaaacagaga aaggagcatg tggtgtgttg tctatactga aagctagtaa
```

```
1441 tcagaaagca ACCCTACTAG GTaagtttaa agaacaattt ggacttggat ataatgaatt
     E2 binding ->
1501 ggttaggcat tttaaaagta gtaggacagc atgtgtggat tgggtagtgt gtgtgtttgg
1561 ggtgtactgc actgtggccg agggcataaa acagttgata cagccactat gtgagtatgc
1621 acatatacaa gtgctaccct gtcaatgggg aatgacagtg ttaatgctgg tgcggtacaa
1681 acgtgccaag aatagggaga cagtagcaaa aggtcttagt actttattaa atgtaccaga
1741 aagccatatg ttaattgagc cacctaaact aagaagtagt ccagcagcat tgtattggta
1801 caaaactagt atgtccaata ttagtgatgt gtatggcgag acacctgaat ggatagtaag
1861 acagacaatg gtaggtcacg cattacaaga agtacagttc agtttatctg aaatggtaca
1921 atgggcatat gatcatgata taacagatga aggtaccttg catacgagt atgcattgat
1981 agcagatgta gactctaatg ctgcagcttt tcttgccagc aattgtcaag ctaaatatgt
2041 aaaggatgct tgcacaatgt gcagacatta caaacggggt gagcaggcac gaatgtccat
2101 gtctgagtgg atacggttta gaagcaacaa agtacaggga gaggggact ggaaaccaat
2161 agtacacttt ttaagatacc aaaatgtaga atttatacca tttctgtgtg ccttttaagtt
2221 attcctacaa ggcatacccca agaaaagctg tttagtgttt tatggacctg cagacacagg
2281 gaagtcatat ttttgcatga gtctgctaaa atttatgggc ggtgttgtaa tttcatatgc
2341 gaattcacac agccatttt ggctgcagcc attgtctgaa gctaaaatgg gtctgctaga
2401 cgatgcaaca agccaatgtt ggagttatgt agacacatat ttaagaaatg cattggatgg
2461 gaacgtaatg tgcatagata gaaaacacag gtccctacta caactcaaat gccctccact
2521 actaataact accaatgtga atccgttgga ggatgacaga tggaagtatt tgcgcagcag
2581 actgcaggta ttcacattca gcaatccatg tccattaaca agtaaaggag agccagttta
2641 tacactaaat gatcaaaatt ggaaatcatt ttttcaaagg ttatgggcac gttTAAgcct
                                                      E2 orf start ->
2701 taccgaccct gacgacgagg aggaaaATGg agaacctagc gaaccgttta gatgcgtgcc
                      E2 cds ->
2761 aggacaaaat actagaactg taTGAaagag atagcgacaa acttgaggac cagatcacgc
                       <- E1 end
2821 attggtatct tatgcgtgta gagagtgcgt tgtattataa agcaagagaa tgtggaatga
2881 cacgtatagg ccaccaggtg gtgccaacac ttagtgtagc taaagctaaa gcatgcagtg
2941 ctattgaaat gcatgtagct ttacaacaat tgcaacaaag tgcatatgga aaggaaccat
3001 ggacacttcg ggacacttca cgagaaatgt gggacgcagt accaaagagg tgctggaaaa
3061 aaagaggagt gactgtggaa gttagatatg atggagacga gactaaagca atgtgccatg
3121 tactgtggaa ggacataatt gtacaaaacc ttagtgatga ccagtgggtt aaagttaaag
3181 gtcaagtctc atatgagggg ctatattatg tgcacgaaga cgtaaaagtg tttttacgtga
3241 aattccataa agacgcacgt gtgtatgggg aaacaggcat atgggaggtg catgtgggag
3301 gcaaagTAAt tcatcacaat gcatttgacc ctgtatctag cacacaagaa gtacccgcta
 E4 orf start ->
 NH2 terminus unknown
3361 ctggacctct atacgcctcc cacaacacca cccgctcgcc cacccaagcc ccgttggggc
3421 ctgaggaggg acaggaacgg aaacgacgca ggcttgaagc agtcgggcct gggccacagc
3481 agcagcagca gcagcagcac cagcagcagc agcagcaaca gaccccgacc cacaccccct
3541 ccacgcaagc ctgtgcacga acgggtggac cagtggacag taacaggacc cgggactgtg
3601 actctacaag tcaaaacccc taccggcacc caagtgatcc tgactgtgca cctgTAAtac
                                                            <- E4 end
3661 acttacgagg tgacccaaac agtttaaaat gttttagata taggttacaa aacggaaaaa
3721 aagggttgta ctgtaaagca tcgtccacgt ggcggtggtc ctgtgaacca gaaaatcaat
3781 cagcatttgt aacaatatgg tacacaagtg ttacacagcg agccgaattt ttggctaatg
3841 ttaaaatacc accaggtatg caggccattt taggccatat gtctgtgttt TGActactgt
                                                           <- E2 end
3901 gccacaacgt gtagcagcct ggatttttat ctgtgtcgac tgtctctgtg ggtgtatttt
3961 gtgttgcttc tgtgtctttt ctggctatct gtgcttcctg cgcttacttg ctacttggcc
4021 attgtgttgt gtttatacct aggattggtg gcactatatt tacaagttgt gcagcacatt
4081 gcacgaaaca cttaggctat catgtatcct ataataatta tagatgggta tggggatcgt
4141 actgtattgc tgtttgagcc aagggacgtg tatgtgttgg gattgttaat actaatggta
4201 tgcctgttgt tatttatagt ttatagacat ttgggattat tataacctgT ATATAAcctg
                                                          signal ->
4261 tatttgtaca tatacatgta ttttatatgt ggctgtggta atacgtgtat tgtatacatg
4321 gccatacaat tgtgcatgtg ttttTAAagt tctaccactt tttttgtttt ttttgttgtt
                    L2 orf start ->
```

```
4381 cctttgtttt tacagttcAA TAAAgcaacc ATGgtggcac atcgtgcaag gcgtcgcaag
                         signal ->     L2 cds ->
4441 cgtgcatccg ccacagagct ttataaaacc tgcaaagttg caggcacatg cccccctgat
4501 gttattccaa aagttgaggg caccacactg gccgacagga tattgcaatg gggcagtcta
4561 ggtgtctatt tgggtgggtt aggtatcggt actgggtctg gcactggagg tcgcacaggt
4621 tatgtccctg tcggcactcg gccaggcact gttgtggatg ttagtattcc tacgcggcct
4681 cctgtggtta ttgagcctgt gggcccttct gatccttcta ttgttaccct gttagaagaa
4741 tccagtgtaa ttaattcggg tgctaccata cccACCTTTA CTGGTacatc cgggtttgag
                                     E2 binding ->
4801 ataacatcat ctgccacaac taccccggct gtgttagata taACCCCTGC TGGTgacaat
                                         E2 binding ->
4861 gtagtcatta ctagcacaaa ctttaataat cctttattca ccgagccttc actcttggaa
4921 attccacaaa ctggagaaac ttctggacgt gtcctggtgg gcacacccac ctcgggtgtc
4981 cacgggtatg aagaaatacc catggacacg tttgccacct ctggaactgg gttagagcct
5041 attagcagca ctcccgtccc tggtgtcagc agggttgcag gtccccgcct ctatggcaag
5101 gccctaacac aggttagggt gtctgatcct gcgtttttga ctcagccttc ttcgtttgta
5161 acctttgata atcctgtgta tgatcctgag gatgaaacta ttattttga gcgtccttct
5221 cccggcactc gtgtgcctga tcccgatttt atggatattg ttaagctgca taggcccgca
5281 ttaacatctc gcaggggcac ggtgcgcttc agtcgcgttg gtcagaagtt tagcatgcgc
5341 actcgcagtg gcacaaacat aggtgccagg gttcactatt atcatgacct gagtcccata
5401 cttcccaccg aggacataga gttggaacca ctgctccccc ctgcagatcc cactgctgag
5461 gagtctctgt atgatatata tgctgatgtg gacgaggctg acatggcttt tacaggcggt
5521 ggtcgcggcg ccaccactta cggggggtcgc attactccat ctgtatttttc ctccacactg
5581 tctacgaggt atggcaatgt cactattcca ttcgtgtcgc cagttgatgt gcctttacac
5641 acgggccctg atattattct gccctcctct gcacaatggc cttttgttcc tgTAGcaccc
                                                       L1 orf start ->
5701 gcagacacga cacattatgt gtacattgAT Ggaggggatt attttttgtg gcctgttacc
                 first 'ATG' of L1 cds ->
5761 tttcctgtgt cccgaaaacg tcgccgtaaa cgtctttcat attttcttgc AGATGgcttt
                                         probable L1 cds start ->
                                                             /\ 3' sj
5821 gtggcgctcT AGtgacaacc tggtgtacct gcctcccacc ccagtctcaa aagttatcag
                 <- L2 end
5881 cacggacgac tatgtgacac gcacaaatat ttattattat gcaggcagtt ctcgcctgct
5941 cactgtgggt catccacatt attcaattcc caaatcctct ggtaataagg tagatgtgcc
6001 taaggtgtct gcatttcagt acaggggtttt ccgtgtgcgt ttgcctgacc ctaataagtt
6061 tggtttgccc gatgcccgca tatataaccc tgaggcagaa cgtttggtgt gggcctgcac
6121 tggtgtggag gtaggtcgag ggcaacctct cggtgtcggg ttgagtggac accctctgta
6181 taacaaactg aatgacacag aaaactctaa tattgcacat gctgaaaatg gtcaggattc
6241 cagggacaac attgctgttg actataagca aacacaactg tgcattctgg ctgtacgcc
6301 tcccatgggc gaacactggg gtaagggcac tgtgtgtgca cgcactagtt ccgctgctgg
6361 tgattgcccc cccctggagt taatgaccac acatattgag gatggcgata tggtggatAC
                                                             E2 binding ->
6421 CGGGTACGGT gccatggact ttgctgctct gcaagttaat aagtctgatg tgcccccttga
6481 tatttgccag tctacgtgta aatatcctga ctacttaggc atggctgctg accctatgg
6541 cgacagcatg tttttttttc tgcgtaggga caactgtttt gccaggcact ctctttaatcg
6601 tgctggtgta gtaggggaca aaatcccaga ttccttgtac ttaaagggta caacgggcg
6661 agaaactcct ggcagtgcca tatacagtcc cacacctagt gggtccatgg taacgtctga
6721 ggctcaaata tttaataagc cttactggct acagcaggcc cagggacaca acaatggtat
6781 atgctgggcc aatcaggtat ttttaactgt ggtggacacc acacgcagca ccaatatgtc
6841 gttgtgtgct accacagagt ctcaaccgtt gaccacttat gatgctacca agattaaaga
6901 atatttgaga catgggaagg aaatatgatt tgcagtttatt ttccagttgt gtaaagttac
6961 attgacacct gaaattatgg cttaccttca tactatgaac agtgccttac ttgaagactg
7021 gaatttggga ttgacattgc caccttccac tagcttggaa gacacgtata ggtttgtaac
7081 atcctctgcc ataacttgtc aaaaagattt ggccctaca gaaaagcagg atccgtatgc
7141 aaagctaaat ttctgggatg tagatttaaa ggatagattt accctggatt tgtcacagtt
7201 tccccctggga cgtaaatttt tattacagat cggtgcgcgc cggcgttcag tagtcccctc
7261 cagaaagcgc cgaacgacca ccacggcccc cacccctgca aagcgaaaac gctcgaaaaa
```

```
7321 gTAAccccag tgttgttgtg tgctgtatgt tgtgtaatgt aatgtgtgta tgtatttatt
          <- L1 end
7381 accatatgtg tttgtatgtc tgtatgtctg tacaatgtat gtatgtatac ttcaactatg
7441 tatgtgtgga tgtataAATA AAgtatgtca catagtttta tattttatac atataattgt
               signal ->
7501 ttgctgagta agaagttaag gtataggtca ggggACCGAT TTCGGTctaa aatggccgcc
                                 E2 binding ->
7561 ggtgcaggtg tgcacaccac taattactca tattattcaa tttcctgcga catgccgtct
7621 cacgcacagt tttggcagca attttttgct ttccactgtt tattttactg ctgtatcatt
7681 cttcttggca agtttgcaca tatacattgc aaattcgctg cttctgggca ccaacttatt
7741 atgactactt tcacataatt actgtcttgg cccagttttc taagttatct tgccaataaa
7801 acgtgtttgc aaatctccac cttaacaatg tgtttccatg acacACCTAA TCCGGTcgct
                                      E2 binding ->
7861 gcttgctttc taaccttaat taatgcagct gccacacctg tctttctaac tataat
```

//

```
LOCUS       HPV77L1      647 bp     DNA             VRL       15-NOV-1994
DEFINITION  Human Papillomavirus type 77 DNA, part of L1 ORF.
ACCESSION   X79947
KEYWORDS    capsid protein; L1 gene.
SOURCE      Human papillomavirus.
REFERENCE   1  (bases 1 to 647)
  AUTHORS   Shamanin,V.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUN-1994) to the EMBL/GenBank/DDBJ databases. V.A.
            Shamanin, Deutsches Krebsforschungszentrum, Im Neuenheimre Feld
            242, 69120 Heidelberg, FRG
REFERENCE   2  (bases 1 to 647)
  AUTHORS   Shamanin,V., Glover,M., Rausch,C., Proby,C., Leigh,I.M., zur
            Hause,H. and Villiers,E.M.
  JOURNAL   Unpublished
COMMENT     HPV77 (VS93) was isolated from a squamous cell carcinoma biopsy
            of a renal allograft patient, and appears likely to constitute
            a new type.  It was found in a search for HPV DNA in 118 skin
            lesions taken from 46 renal allograft patients [1].  The same
            sequence was obtained from lesions from three or more
            patients.  HPV29 appears to be the most closely related known
            type.
            NCBI gi: 562322
BASE COUNT     167 a    157 c    157 g    166 t
ORIGIN
        1 ggtagaggac agccattagg cgtggggtta agtggacacc ctctgtataa caaactgaat
  L1 cds ->
       61 gacactgaaa actccaacat tgcacatgct gacaatagtc ctgactcccg ggacaacatt
      121 tctgttgact gtaagcaaac acaactgtgc atactgggct gtacgccccc catggggggaa
      181 tactggggta agggtacccc ttgtgcacgt actaatacta ccccaggaga ctgtcctccc
      241 ttggagttaa tgacatctta tattcaggat ggcgacatgg tggataccgg gtatggtgcc
      301 atggacttta ctgccctgca atttaataag tctgacgtgc cccttgatat ttgccagtct
      361 atttgcaaat atcccgatta tttgggcatg gctgccgacc cgtatggcga tagcatgttc
      421 ttttttcctcc gtcgggaaca actgtttgcc agacacttttt tcaatcgtgc gggtgatgtt
      481 ggagacaaaa ttccagaatc tttgtacctc aaagggagta gcgggcgtga gactcccggc
      541 agtgctatat acagccccac acccagtggg tctatggtga cctctgaggc acaaatattc
      601 aataagtctt actggctaca gcaagctcaa ggccaaaata acggtat
                                                  L1 cds ->
//
```

Group A3 Sequences

|	
HPV61	**HPV62**
HPVCP4173	**HPVCP6108**
HPVCP8304	**HPVLVX100**
HPVMM8	**HPVX06**

INTRODUCTION

Group A3 comprises HPV sequences CP4173, CP6108, CP8304, LVX100, MM8, HPVX06, and the types HPV-61 and HPV-62. These viruses formed group E in the old classification. CP4173 and LVX100 are believed to be representatives of viral type HPV-72. HPVX06, an L1 fragment, is very similar, but not identical to HPV61 over this region. Viruses in this group primarily infect anogenital tissue and have relatively unknown oncogenic potential and prevalence.

The previously characterized HPV types HPV-61 and HPV-62 have been isolated from tissues with at least some degree of dysplasia. Both HPV-61 and HPV-62 have been derived from vulvar intraepithelial neoplasias [4].

CP4173, CP6108 and CP8304 were obtained through clinical studies conducted in the state of New Mexico among a tri-ethnic population [1]. LVX100 and LVX82 were isolated from an Amazonian Indian population [2]. MM8 was identified through studies conducted in the state of California. Initial prevalence data for MM8 are similar to that obtained for characterized "intermediate risk" viruses [3]. All samples were obtained from cervical lavages or genital swabs (CP6108 and CP4173 were isolated from normal cervices).

With the exception of HPV61, the complete genome of which is presented on the following pages, the members of Group A3 have been sequenced only over the My09-My11 fragment of L1. Phylogenetic analysis of the L1 region categorizes the group A3 viruses as a distinct group.
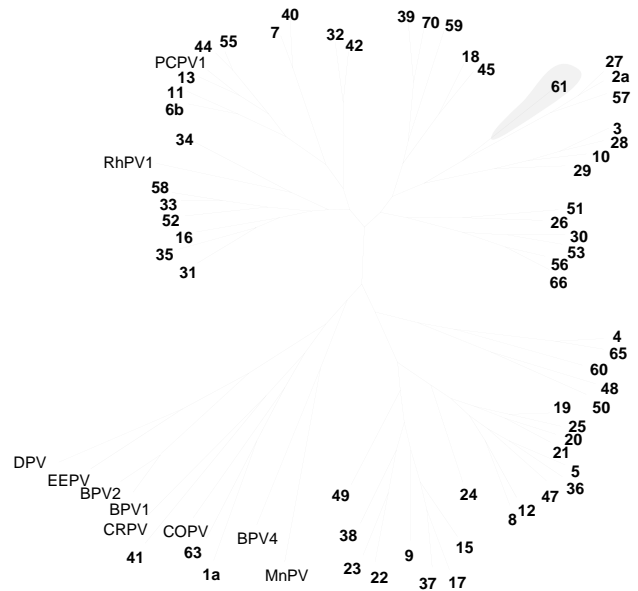
What's new?

HPV61 presented on the following pages is the only new complete sequence in Group A3 released during 1995. HPVX06, an L1 fragment, is also presented here. The sequences of the other members of this group were published in *Human Papillomaviruses 1994* pp. I-E-2 et seq.

References

[1] Peyton,C.L. and Wheeler,C.M. Identification of five novel human papillomaviruses in the New Mexico triethnic population. *J. Infect. Dis*. **170**:1089–92 (1994)

[2] Ong,C.-K., Bernard,H.-U. and Villa,L.L. Identification of genomic sequences of three novel human papillomaviruses in cervical smears of Amazonian Indians. *J. Infect. Dis*. **170**:1086–8 (1994)

[3] Manos,M.M., Waldman,J., Zhang,T. Greer,C., Eichinger, G.,Schiffmann,M., and Wheeler, C. Epidemiology and partial nucleotide sequence of four novel genital human papillomaviruses. *J Infect Dis* **170**:1096–99 (1994)

[4] de Villiers,E.M. Human pathogenic papillomavirus types: an update. in *Human pathogenic papillomaviruses*, edited by Harald zur Hausen, Springer-Verlag, Heidelberg, pp 1–12 (1994)

```
LOCUS       HPV61          7989 bp    DNA              VRL        18-JUL-1995
DEFINITION  Human papillomavirus type 61, complete genome.
ACCESSION   U31793
KEYWORDS    .
SOURCE      Human papillomavirus type 61.
REFERENCE   1  (bases 1 to 7989)
  AUTHORS   Delius,H.
  JOURNAL   Unpublished, Sequenced by Hajo Delius, Deutsches
            Deutsches Krebsforschungzentrum, Angewandte Tumorvirologie,
            I.N.F. 506, W-6900 Heidelberg, Germany
REFERENCE   2  (bases 1 to 7989)
  AUTHORS   Farmer,A.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-1995) Andrew D. Farmer, HIV Sequence Database,
            Los Alamos National Labroatory, T-10, Mail Stop K710, Los Alamos,
            NM 87501, USA
COMMENT     HPV-61 has recently been characterized and isolated from a vulvar
            intraepithelial neoplasia by Dr. T. Matsukura (unpublished data). A
            variant of HPV-61 (isolate AE4) was identified in a study conducted
            to screen cervical lavages from over 500 women for novel HPV types
            (Tachezy et al, Biochem. Biophys. Res. Comm. 204, 820-827).
BASE COUNT     2113 a   1726 c   1973 g   2177 t
ORIGIN      102 bp upstream from beginning of E6 cds
        1 taacaatcat atatgtaaaa agtaggttgg cACCGAAAAC GGTcggACCG ACCTGGGTac
                                           E2 binding ->   E2 binding ->
       61 aTATATAAag aaaccgtagg gtcagcaaag cacactcatc tATGggaccg tgcaatccaa
     E6 orf start ->                                    E6 cds ->
     signal ->
      121 ctaatatttt cctgctctgc aaggactacg aggtggactt tgaggattta aggctgacat
      181 gtgtattttg caaaaatgaa ttaacaacag aagaattgct ggcgtttgca ctaaaggagc
      241 taagcattgt gtggagacat aactggccat ttggagtatg cgcaccatgc ttggcacgtg
      301 aagtaaaagt gagggagctg cgacattggg accattcctg ttacggaccc actgtggaac
      361 agacaacagg acggtcatta gctgaactat atatacggtg ccatgcatgc agcaaaccgt
      421 taagtataca ggaaaaggag catcaggtac aggcatacat ccacttccac tatatagctg
      481 gacagtggac gggaaggtgt tgccagtgTA GagggccATG cacggccagg tggcaaccaT
                                           E7 orf start ->
                                                      E7 cds ->
      541 AAaggacata gtccttgaag agcgtcctga ggtggttgac ctacattgca atgagcagtt
          <- E6 end
      601 attagacagc tcagagtcag aggaggagga tagtgtgcgt gagcaacttg ttgaacaagc
      661 acagcaggcc tacagggtgg ttactacctg tggcatttgt aagtgtccag ttaggctggt
      721 ggtgcagtgc ggagacgcag acctgaaggt gctacaTGAa ctactgctgg gcgacttgtc
                                          E1 orf start ->
      781 catagtgtgt cctggctgcg caTAAgggac ATGgctgaca gtgaaggtac agaaagcggg
                                 <- E7 end
                                 E1 cds ->
      841 gatgggaccg aggccgcgga acgcgcaggg gggtggtttc tggtagaagc cgtggtagac
      901 cgcaccacag gataccaggt gtccagtgat gaggaggaca atagcattga cacaggggaa
      961 gacctagtag acttcataga tacaaggcgc cccgggggtg ggcaggaagt gccgcttgcg
     1021 ttgttcgttc aacaaaatgc acaggatgac gctgcaacgg tgcaggcact aaaacgaaag
     1081 tatacatgta gccctgcaag cagcacctgt gtgtccttgg tggacagtga attaagtccc
     1141 cggctggacg ccatacggat acaccgggga caggacaggg ctaggagaag gctgtttgag
     1201 caagatagtg gctatggcca tacgcaagtg gaaattggag catcagaaag tcaggtaccg
     1261 ggggatgcgc aacatgaggg gggggggaa tccgtgcagg aagcagagga ggagcgtggg
     1321 ggggggacg gagaggccga ggccacaggt aaccaggaaa cgcaagcgca ggagcaggcg
     1381 gcagacatat tagaggtgtt taaggttagt aatttaaaag caaaattact gtacaaattc
     1441 aaggacctat ttggactagc atttgggggag ctggtaagaa atttttaaaag tgataagtca
     1501 atatgtgggg actgggtaat atgtgcgttt ggtgtatacc atgctgtggc agaggctgta
     1561 aaaaccttaa tacaacctat atgtgtgtat gcacacatac aaatacagac atgccagtgg
```

```
1621 ggaatggtaa tattaatgct tgtgcgatac aaatgtggga agagcaggga aacagtagca
1681 cacagcatgg gaaaactgtt aaacataccg gaaagacaga tgctaattga accaccaaag
1741 attagaagcg caccgtgcgc actatattgg tatagaacag ccatgggaaa cgccagcgag
1801 gtgtatggcg aaacacctga atggatagta agacaaacag tggtaggaca tgcaatgcaa
1861 gaagcacagt ttagtttgtc catgttagtg caatgggcat atgacaatga cataacagat
1921 gaaagtgtgc tggcctatga atatgcattg ctgggaaatg aggacccaaa tgcggcggcc
1981 ttcttggcca gtaactgcca agcaaagtat attaaagatg caattacaat gtgtaagcat
2041 tatagacgtg cagagcaggc aaaaatgaca atggcacagt ggattacaca tagaggccgc
2101 aaggtggcag acacagggga ttggaaggct attgtaaagt atttaagata tcaacaggta
2161 gaatttgtac catttattag tgcccttaaa ttgtttctaa aggggtacc aaaaaaaagc
2221 tgcatggtat tttacggacc aagcgacaca gggaagtcgc tattttgcat gagcctacta
2281 aattttctgg gaggagctgt gatttcatat gtaaattcaa gcagccattt ttggttatca
2341 cctttagcag atactaaagt gggactactg gacgatgcca cctatcagtg ctggcaatac
2401 atagatacat atttacggac agtattagat ggcaatgcca taagcataga taggaaacac
2461 agaaatttaa cacagctgaa gtgtcctcca ttaatgataa caacaaatat aaatccacta
2521 gaggatccta cgtttaaata cctgcacagt agaatagtgg tgtttcagtt tttacataag
2581 tgtccactaa acagtaatgg ggacccagta tatccctga ataatgaaaa ttggaaatca
2641 ttttttccgaa ggtcctgggc acgcaTAGag ggatctgatc aacaggagga ggaggaggag
                    E2 orf start ->
2701 gaggATGagg atggagtcac tagccgaccg tttagatgcg tgccaggaga aattactaga
   E2 cds ->
2761 cctctaTGAa aaagatagca acaagcttga ggaccaaata ttgcattggc actatgtgcg
            <- E1 end
2821 tttggaaaat gcaatgttat ttaaggcacg acaagcagga ctaacgcgtg tgggccacca
2881 gatggtgcca acacttagtg ttacaaaagg caaagcacat aaagcaattg aagtacatct
2941 gtcattgcaa gggttacaaa caagtgcata tgcgcatgaa ccctggacat tgcaaaccac
3001 ctctttggaa atgtggaata cgcagccaca acgatgctgg aagaaaaaag acgcagact
3061 aacagttaaa tttgatggcg aggaccacaa agcagtggaa tacgttagtt gggggtatat
3121 ttatgtacaa agcacagaaa ctgatctgtg gtataaggtt cctggaaaag tgtcatataa
3181 gggcctatat tatgaaatgg agggacagga acactattat gtgacatttg cacaggaggc
3241 ccaaaaatat ggggaaacag gcaaatggga ggtacatatg ggaaacacag TGAtttatga
                                E4 orf start ->
                                NH2 terminus unknown
3301 accctgcgcc tctgtatcta gtacccagga cgccgtgcga gaagtatcca ctgctgaaac
3361 tgctggacac ctgcgggaca acaccacaca gaccaccacc accccacgt gcgtgggccc
3421 cacccagaca tccacctcgg tgcagacgcc gcctcataag cgacagcgac tccacagaga
3481 cagagagcag cagcccgaca caacacaaaa agacaaccac aagcgggtgg acagtactga
3541 ccagtggatc aacggtcacc gtaacagcac agaaacaggg gacaactgtg acagttacag
3601 ttcacctgTA Atacacctaa aaggtgatcc aaataaatta aagtgttttc ggtataggtt
            <- E4 end
3661 acaacattca gtacctgagc tgtttgacaa agcatcatcc acatggcatt gggccggtgg
3721 acaaagtaca accagagcag catttgtaac attgtggtat gttaatgtgg aacaacgaaa
3781 acaatttta aatcgtgtaa ccattccaaa gggtatacaa gccactgcag ctatatgtc
3841 aatgtgtata TAAgtgtgtg tatacaaaaa cataaatact gtgttgcaag tattgagcca
            <- E2 end
3901 gctacagcta tttacatacc acacctgcat tattgtgTAA accataaaca atcctggacg
                                E5 orf start ->
                                NH2 terminus unknown
3961 agcttgtgtg tttgatagat ggatggtgta cgatactcct gctgcttata cttttctggc
4021 tttcctatct ttcctcacta agcgcatttt tggtgttttt ctgtgccata tatgtaggat
4081 tgttgtgcat atacatgcaa gtgttgtggt atatggaata actgtggcgc caccagccat
4141 acactgctgc tcatatgtaa atagtgttcc ttgtgtctgt tgttcgtttg caatatgtat
4201 ccccttgTAG tTAGttcccc tcaggggttt gcattactgg ttgtggatca gcaaaactgc
            <- E5 end
        L2 orf start ->
4261 ggattgttat ttttgttttt gctattactg cttattgcaa tactgttgta caggctttta
4321 cattaggtgg tgtaccttgt gtgtttttccc tccccacagt tttttacttt gttatatttt
4381 tatatgttta tggcctAATA AAcattATGg ctcttaaacg tcgcaagcgt gcatctgcaa
            signal ->  L2 cds ->
```

```
    4441 ctgacctcta tcgtacatgt aaacaatcag gcacatgtcc ccctgatgtt gttccaaagg
    4501 tggaagggga tacccttgct gaccgcattt taaagtgggc tagtttgggg gtttttctttg
    4561 ggggattggg tataggcACC GGCTCTGGTa ctgggggtcg tactgggtat gtgcctattg
              E2 binding ->
    4621 gcacccggcc tcctactgta gtggatattg gccctgtgtc ccgcccccct gtagttattg
    4681 atccagtggg tgctgccgat ccctctatag ttacccttgt agaggaatcc agcgtcattg
    4741 aggctggggc tacagtccct acattttccg ggtccggggg cttcaatgtc acgtcgtctt
    4801 ctactaccac cccggctgtg ttggacatta ctccatcggg tgggtctgtg caggttagca
    4861 gcaccagttt tattaatcct ttattcactg agccttctat tattgaacct cctcaggcag
    4921 gggaccttgc cggccatgtc atttccagta cccctaccgc tgggtcccac agttttgagg
    4981 aaatcccaat gcacaccttt gctACCTCTG AGGGTcctgg aagtagtacc ccgctgcctg
                  E2 binding ->
    5041 gcatccgtcg ccttgcacgg ccccgcctga atttatacag taaggctaac caacaaatta
    5101 aggttgctaa ccctacattt atgtctgatc cagcctcctt aattacatat gacaatccta
    5161 tatttgatcc tgaggaaact attatatttg aacatcctag tatatataca cctcctgacc
    5221 cggactttt agatattgtg tccttacaca ggcccgcgct cacatcccgc cagggaactg
    5281 tacggtttag tcgccttggt caacgtgcca ccttgcgtAC CGTAGTGGT aggcgcattg
                                   E2 binding ->
    5341 gcgctcgggt acactttttac catgacatta gtcccattcc atcggatgct gtggagttac
    5401 agcctttggt gccgtcctcg tcccctagta taacatatga tatatatgct gaccctgagg
    5461 tgttggatct tcctgcacaa catacacaac ccacacttac agtacagggc ccttccctct
    5521 ctgctgcatc tgcatctacc aaggtacaca atgtcactgt tccactggcc acagggcttg
    5581 atacgcctgT GAcatcggc cctgATGttg attttgcaca tgctcctgcg cctgtaccag
          L1 orf start ->
             first 'ATG' of L1 cds ->
    5641 ctgtaccata tgtacctgcc acccatccac attctatata tattcagggc tctgatttttt
    5701 atttgttgcc tgcttatgtg ttttttccta aacgtcgtaa acgtgtaccc tattctttttt
    5761 cAGATGgctt tgtggcggcc tggTGAcggc aaggtatacc tgcctcccac tcctgtgtct
probable L1 ->                        <- L2 end
cds start  /\ 3' sj
    5821 aaagttatca gtactgaccg ttatgtgcaa cgcaccaact tattttatta tggtggcagt
    5881 tcccgtctgc ttactgtagg acatccctat tgtagtttgc agcttgatgg gctgcagggc
    5941 aagaaaaaca ctatccccaa ggtgtctggc tatcaatata gggtgtttag ggtacagctt
    6001 cctgacccta ataaatttgc tttgcctgat ggcaccttgt ataatcctga tactgagcgc
    6061 atggtgtggg cctgcagggg cattgaggtt ggtagggggc agccattggg ggttggtact
    6121 agtggtcacc cgctgtacaa tcgcctggat gacactgaaa acaccacatt acttgttgca
    6181 gagtcctctg acagtcgtga caatgtgtct gttgattata agcagaccca gctacttatt
    6241 gtaggttgca aaccccctat tggggaacac tggacaaagg gtactgcttg tgctaacccc
    6301 gcgccccggc ccactgactg tcctccattg gaatttacaa ataccaccat acaggatggc
    6361 gatatggtgg aaacaggcta tggggccatt gattttgctg cattgcagga aaataaatct
    6421 gaagtgccat tggatatttg caccactata tgcaagtatc ctgattattt acaaatggct
    6481 gctgaaccat atggagactg tatgttttttt tgtttacgca gggaacaaat gtttgctcgc
    6541 cactttttca atcgccaggg tgtcatgggt gaggcactgc ctgactctta ttatcttaag
    6601 ggtgcgaatg acaaggcggc ccctggtagt tatatttatt cccctacgcc tagtggctct
    6661 atggtgtcct ctgattccca attatttaat aagccatatt ggctgcaacg tgcccagggc
    6721 cacaacaatg gtatttgttg gtttaatgaa ttgtttgtaa ccgttgtgga taccacccgc
    6781 agtactaatt taaccatttg tactgctaca tccccccctg tatctgaata taaagccaca
    6841 agctttaggg aatatttgcg ccatacagag gagtttgatt tgcaatttat ttttcagtta
    6901 tgtaaaatac atttaacccc tgaaattatg gcctacctac ataatatgaa taaggccttg
    6961 ttggatgact ggaactttgg tgtggtacca ccaccctcta ccagtttaga agacacatat
    7021 aggtttttgc agtccagagc tattacatgt cagaagggtg ctgctgcccc gccgcccaag
    7081 gaggatcgct atgccaagtt atcctttttgg actgttgatt tacgagacaa gttttccact
    7141 gatttggatc agtttccttt ggggcgcaag ttttttattgc aggccggtcc ccgctcagta
    7201 tctgtgtcac gcaaacgtgc tgcccccagc tctaccccta catcctcccc tgctacaaaa
    7261 cgcaaaaagc gtaaacagTA Gtgatgtttg tctgtatgct gttgttgtgc atgtgttgtt
                          <- L1 end
    7321 ttgtattgtg ctgtactggc cttgtgcatg tgctgtttgt atggtgcctg gtttgtatga
    7381 ctgtatttgc tatggttgtg tattgtgtga ccctgtactg acccttgtcc cctgtgttgt
```

```
7441 gtatctgtgg aatgtgtgtt acactgtgtg tgcattAATA AActgtgtgt catgtgtgtg
                                          signal ->
7501 tggtgtccgg tgcaccctgt gagtaagtgt gtttattgtt gcacgcctta ttttgtgcag
7561 cttccccatt tctgtgttcc ttgtgttcgg gtagccattt tgtgttgccg ACCGTTTTCG
                                                            E2 binding ->
7621 GTctcccgcc tttttttgttg ccgACCGGTT TCGGTcgccc gcctttttgg ttgtggcact
                E2 binding ->
7681 atgccactgt acacatacca catcctgttt aaacaaactt tatgtgttgc acctgggtgt
7741 gtttgacagg ttgctaacac acacagtttg gaacaatgct ccatctatct tttttgacat
7801 gttgtgcaat tactgtgctt ttaggcacat attttgtgct gactactttc ccttaattgc
7861 tggtttggca ttgcttgtat actttttctag ccaggaaact gtcttgcaac tttgggcatg
7921 ttttacacac attatgttac aaagccgatc cggtcgctcc ctatgtctca ttctttttat
7981 gatagttta
```

//

```
LOCUS       HPVX06        211 bp   DNA             VRL        09-NOV-1992
DEFINITION  L1 orf [human papillomavirus HPV, X06 isolate, Genomic, 211 nt].
ACCESSION   S42984
SOURCE      Human papillomavirus HPV X06 isolate.
REFERENCE   1  (bases 1 to 211)
  AUTHORS   Iwamoto,A., Yoshikawa,H., Kitagawa,K., Igarashi,H., Kawana,T. and
            Yoshikura,H.
  TITLE     Short sequence in L1 region of human papillomaviruses correlates
            with clinical pictures and grouping by cross-hybridization
  JOURNAL   Jpn. J. Cancer Res. 83, 315-319 (1992)
COMMENT     GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 111372] from the original journal article.

            This sequence comes from Fig. 1.

            NCBI gi: 253864
BASE COUNT      52 a     55 c     50 g     54 t
ORIGIN
        1 gcagatggct ttgtggcggc ctggtgacgg caaggtatac ctgcctccca ctcctgtgtc
       61 caaagtttta agtacggacc gctatgtcca gcgtaccaac ctgtactatt atggaggcag
      121 ctcccgcctt attacagttg gccacccta ttatactgta caggtcaatg gtgctaacaa
      181 aaaggccaac atacctaagg tatcagggta t
//
```
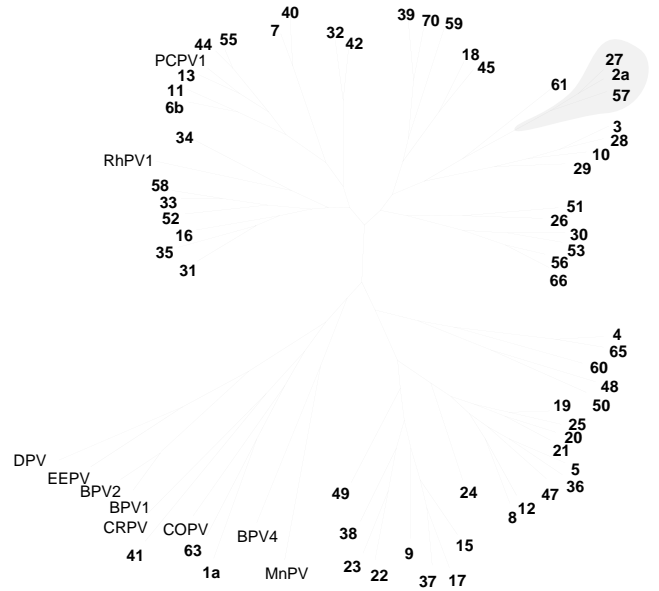
Group A4 Sequences

HPV2a   **HPV27**
HPV57

INTRODUCTION

The members of group A4 formed a subset of Group F in the old classification. HPV-2a, HPV-27, and HPV-57 infect both cutaneous and mucosal tissue. HPV-57 seems to preferentially infect the mucosa, and only rarely infects cutaneous tissues. Conversely, the primary site of infection for HPV-2a is cutaneous tissue. Due to the rare detection of HPV-27, its tissue preference in the normal population is unknown [1]. HPV-57 has been isolated from a number of oral, upper respiratory and genital lesions, but in only two cases of verrucae vulgares of immunosuppressed patients [2]. HPV-2a frequently causes common and filiform warts, infecting EV patients, immunosuppressed patients and the general population [3]. Hybridization under conditions of low stringency to a tongue carcinoma and cross-hybridization to HPV-18 indicated dual tissue infectivity [4]. Subsequently, HPV-2 has been detected in numerous oral lesions of differing severity and in an anogenital wart [1]. HPV-27 was isolated from a common hand wart from a renal transplant patient and subsequently from cases of anogenital lesions [1,6]; it has also recently been isolated from multiple skin lesions of additional renal transplant patients [7]. Recently, an isolate originally thought to constitute a subtype of HPV-2, HPV-2c, was sequenced over the LCR, E6, and L1 region and was found to be identical to the sequence of HPV-27. HPV-27 and HPV-2a differ by 55 nucleotides, but qualify as "close types" [5].

What's new?

No new sequences in Group A4 were released during 1995. The sequences of members of this group were published in *Human Papillomaviruses 1994* pp. I-F-4, I-F-20, and I-F-41.

References

[1] de Villiers,E.M. Human pathogenic papillomavirus types: an update. in *Human pathogenic papillomaviruses*, edited by Harald zur Hausen, Springer-Verlag, Heidelberg, pp 1–12 (1994)

[2] de Villiers,E.M., Hirsch-Behnam,A., von Knebel-Doeberitz,C., Neumann,C., and zur Hausen, H. Two newly identified human papillomavirus types (HPV 40 and 57) isolated from mucosal lesions. *Virology* **171**: 248–53 (1989)

[3] Corley,E., Pueyo,S., Goc,B., Diaz,A., and Zorzopulos,J. Papillomaviruses in human skin warts and their incidence in an Argentine population. *Diagn Microbiol Infect Dis* **10**: 93–101 (1988)

[4] Hirsch-Behnam,A., Delius,H., and de Villiers,E.M. A comparative sequence analysis of two human papillomavirus (HPV) types 2a and 57. *Virus Res* **18**: 81–97 (1990)

[5] Chan,S.Y., Tan,C.H., Delius,H. and Bernard,H.U. Human papillomavirus type 2c is identical to human papillomavirus type 27. *Virology* **201**: 397–8 (1994)

[6] Ostrow,R.S., Zachow,K.R., Shaver,M.K., and Faras,A.J. Human papillomavirus type 27: detection of a novel human papillomavirus in common warts of a renal transplant recipient. *J Virol* **63**: 4904 (1989)

[7] Shamanin,V., Glover,M., Rausch,C., Proby,C., Leigh,I.M., zur Hausen,H. and de Villiers,E.M. Specific types of human papillomavirus found in benign proliferations and carcinomas of the skin in immunosuppressed patients. *Cancer Research* **54**:4610–3 (1994)

Group A5 Sequences

HPV26   **HPV51**
**HPV69**   **HPVIS039/L1AE2**
HPVMM4

INTRODUCTION

The members of group A5 formed one of two subsets of group D in the old classification. Group A5 consists of human papillomavirus types 26, 51, 69, and the novel viruses IS039/AE2, and MM4, a group primarily associated with anogenital lesions, some with considerable oncogenic potential. Lorincz et al. classified HPV-51 as an "intermediate-risk" virus [1].

The group A5 viruses exclusively infect the anogenital area with the exception of HPV-26 which has also been detected in cutaneous lesions. HPV-26 was first identified in numerous cutaneous lesions of an immunosuppressed patient [2]; it has rarely been detected in genital sampless. However, HPV-26 was identified in 4 out of over 1000 cervical cancers in the IBSCC study, a survey of HPV prevalence in a worldwide sampling of cervical cancers [3]. HPV-69 was isolated and characterized from a cervical intraepithelial neoplasia [2].

The novel virus IS039 was identified through international screening efforts. HPVIS039 was isolated from an invasive cervical carcinoma biopsy and is closely related to HPVMM4 (90.8% over the My09-11 segment) [4]. MM4 was found in conjunction with HPV-53 in a normal cervix [5]. Initial prevalence data for MM4 is similar to that obtained for characterized "intermediate-risk" viruses [6]. The virus AE2 was detected in a cervical lavage sample obtained in a Bronx, N.Y. clinical trial [6]; the sequence is nearly identical to that of IS039. Incidence rates and risk assessments have not been published for the recently identified IS039/AE2.

In the IBSCC study, group A5 viruses were detected as follows: HPV-26 (0.4%) HPV-51 (0.8%) IS039 (0.1%) MM4 (0.6%).

Of the members of Group A5, complete genomic sequences are available for HPV-26, and HPV-51. HPV-69 and the novel sequences HPVIS039, HPVMM4, and AE2 have been sequenced only over the My09-My11 region of L1. The sequences of IS039 and AE2 are virtually identical.
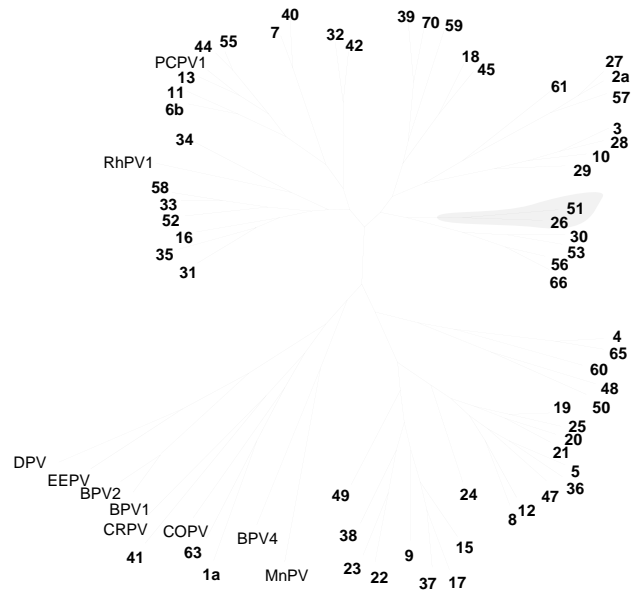
What's new?

No new sequences in Group A5 were released during 1995. The sequences of members of this group were published in *Human Papillomaviruses 1994* pp. I-D-3, I-D-11, I-D-25, 26, 27, and 28.

References

[1] Lorincz,A.T., Reid,R., Jenson,A.B., Greenberg,M.D., Lancaster,WD, and Kurman,R.J. Human papillomavirus infection of the cervix: relative risk associations of 15 common anogenital types. *Obestet Gynecol* **79**: 328–337

[2] de Villiers,E.M. Human pathogenic papillomavirus types: an update. in *Human pathogenic papillomaviruses*, edited by Harald zur Hausen, Springer-Verlag, Heidelberg, pp 1–12 (1994)

[3] Bosch, F.X., Manos, M.M., Munoz, N., Sherman, M., Jansen, A.M., Peto, J., Schiffman, M.H., Moreno, V., Kurman, R., Shah, K.V., International Biological Study on Cervical Cancer (IBSCC) Study Group. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. *J Natl Cancer Inst* **87**:796–801 (1995)

[4] Peyton,C.L., Jansen,A.M., Wheeler,C.M., Stewart,A.-C., Peto,J., Bosch,F.X., Munoz,N., Teyssie, A.R., Torroella, M., Wabinga, H.R., Sarjadi, Ngelangel,C., and Manos,M.M. A novel human papillomavirus sequence from an international cervical cancer study. *J Infect Dis* **170**:1089–92 (1994)

[5] Manos,M.M., Waldman,J., Zhang,T. Greer,C., Eichinger,G., Schiffmann,M., and Wheeler, C. Epidemiology and partial nucleotide sequence of four novel genital human papillomaviruses. *J Infect Dis* **170**:1096–99 (1994)

[6] Tachezy,R., Van Ranst,M.A., Cruz,Y. and Burk,R.D. Analysis of short novel human papillomavirus sequences. *Biochem Biophys Res Commun* **204**:820–7 (1994)
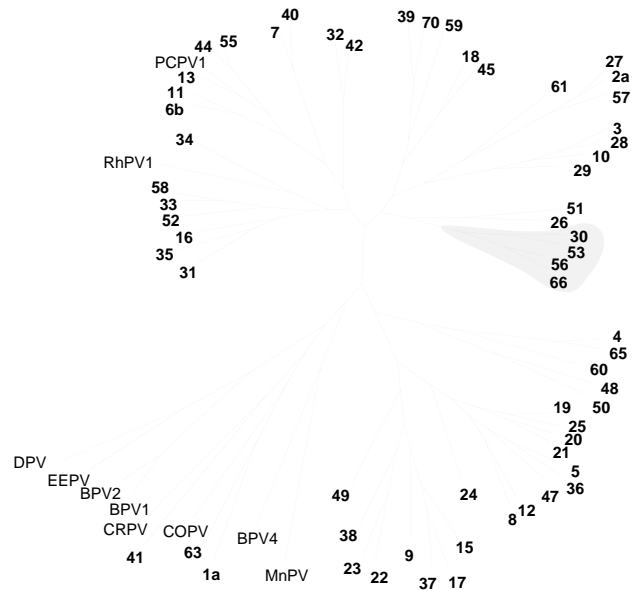
Group A6 Sequences

HPV30	**HPV53**
HPV56	**HPV66**

INTRODUCTION

The members of group A6 formed one of two subsets of Group D in the old classification. Group A6 consists of human papillomavirus types 30, 53, 56, and 66, a group primarily associated with anogenital lesions, some with considerable oncogenic potential. Lorincz et al. classified HPV-56 as a "high-risk" virus [1], while Bergeron et al. placed HPV-30 and HPV-56 in the intermediate-risk category [1].

The Group A6 viruses have exclusively been detected in the anogenital area with the exception of HPV-30 which has also been detected in respiratory lesions. HPV-53 and HPV-30 are rarely detected in the genital tract [2]. HPV-30 was initially isolated from a laryngeal carcinoma which suggested it might be the etiological agent of this form of cancer. An extensive effort was undertaken to probe laryngeal lesions for HPV-30 DNA. Unexpectedly, in 116 additional laryngeal squamous cell carcinomas, HPV-30 DNA was not detected [4]. Infrequent identification in genital lesions has been reported [3,5]. HPV-53, originally isolated from a normal cervix, is also rarely found in the genital tract. When 189 asymptomatic women were examined for infection with HPV-53, six positive samples were identified [6]. HPV-66 has been reported in cases of multiple HPV infection. For example, it was accompanied by HPV-16 and HPV-45 when originally isolated from an invasive carcinoma [7]. Additionally, it was detected in three out of four cervical intraepithelial cases accompanied by another HPV type [7]. In the IBSCC, group A6 viruses were detected as follows: HPV-56 (1.7%) HPV-66 (0.1%) [8]

Complete genomic sequences are available for all known members of Group A6.

What's new?

The complete genome of HPV-66 is the only new sequence in Group A6 released during 1995. The sequences of other members of this group were published in *Human Papillomaviruses 1994* pp. I-D-7, I-D-15, and I-D-1.

References

[1] Lorincz,A.T., Reid,R., Jenson,A.B., Greenberg,M.D., Lancaster,WD, and Kurman,R.J. Human papillomavirus infection of the cervix: relative risk associations of 15 common anogenital types. *Obestet Gynecol* **79**: 328–337

[2] Bergeron,C., Barrasso,R., Beaudenon,S., Flamant,P., Croissant,O., and Orth,G. Human papillomaviruses associated with cervical intraepithelial neoplasia. Great diversity and distinct distribution in low- and high-grade lesions. *Am J Surg Pathol* **16**:641–649

[3] de Villiers,E.M. Human pathogenic papillomavirus types:an update. in *Human pathogenic papillomaviruses*, edited by Harald zur Hausen, Springer-Verlag, Heidelberg, pp 1–12 (1994)

[4] Syrjanen,S., Syrjanen,K., Mantyjarvi,R., Collan,Y., and Karja,J. Human papillomavirus DNA in squamous cell carcinomas of the larynx demonstrated by in situ DNA hybridization. *ORL J Otorhinolaryngol Relat Spec* **49**:175–86

[5] Kahn,T., Schwarz,E. and zur Hausen,H. Molecular cloning and characterization of the DNA of a new human papillomavirus (HPV 30) from a laryngeal carcinoma. *Int J Cancer* **37**:61–5 (1986)

[6] Gallahan,D., Muller,M., Schneider,A., Delius,H., Kahn,T., de Villiers, E.M. and Gissmann,L. Human papillomavirus type 53. *J Virol* **63**:4911–2 (1989)

[7] Tawheed,A.R., Beaudenon,S., Favre,M., and Orth,G. Characterization of human papillomavirus type 66 from an invasive carcinoma of the uterine cervix. *J Clin Microbiol* **29**:2656–60 (1991)

[8] Bosch, F.X., Manos, M.M., Munoz, N., Sherman, M., Jansen, A.M., Peto, J., Schiffman, M.H., Moreno, V., Kurman, R., Shah, K.V., International Biological Study on Cervical Cancer (IBSCC) Study Group. Prevalence of human papillomavirus in cervical cancer:a worldwide perspective. *J Natl Cancer Inst* **87**:796–801 (1995)

```
LOCUS       HPV66         7824 bp   DNA              VRL        18-JUL-1995
DEFINITION  Human papillomavirus type 66, complete genome.
ACCESSION   U31794
KEYWORDS    .
SOURCE      Human papillomavirus type 66.
REFERENCE   1  (bases 1 to 7824)
  AUTHORS   Delius,H.
  JOURNAL   Unpublished, Sequenced by Hajo Delius, Deutsches
            Deutsches Krebsforschungzentrum, Angewandte Tumorvirologie,
            I.N.F. 506, W-6900 Heidelberg, Germany
REFERENCE   2  (bases 1 to 7824)
  AUTHORS   Farmer,A.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-1995) Andrew D. Farmer, HIV Sequence Database,
            Los Alamos National Labroatory, T-10, Mail Stop K710, Los Alamos,
            NM 87501, USA
COMMENT     HPV66 was originally isolated from a biopsy of a 38-year-old
            patient with a stage I invasive squamous-cell carcinoma of the
            uterine cervix. The sample was coinfected with HPV16, HPV45 and the
            new HPV66.  The HPV16 was present in an integrated form; HPV45 and
            HPV66 were present in episomal (circular) form. The biopsy specimen
            showed a low-grade intraepithelial neoplasia which continues as a
            poorly differentiated invasive carcinoma. HPV66, but not HPV45 nor
            HPV16, was detectable by in-situ hybridization in the nuclei of
            terminally differentiating cells of the CIN; Tawheed et al (J.
            Clin. Micro. 29, 2656-2660) took this to suggest that HPV66 has
            oncogenic potential.  Screening of 160 anogenital biopsies revealed
            the presence of HPV66 in three cases of high-grade CIN and one PIN.
            HPV16, HPV51, and HPV58 were also each detected in one of these
            HPV66 positive lesions. Putative splice sites suggest that HPV66,
            like many other genital carcinoma-associated HPV types, encodes
            both a full-length and a truncated form of the E6 protein.  The
            LXCXE motif in HPV16 E7 associated with binding of p105-RB is
            preserved.  HPV66 hybridized with high stringency with HPV56, and
            more weakly with HPV30 and HPV53. Phylogenetic analysis supports
            this grouping.  No conserved restriction sites between HPV56 and
            HPV66 were found. HPV-66 was independently isolated on two separate
            occasions by independent researchers.  Dr. M. Manos isolated the
            same strain from a patient with normal cytology and designated it
            PAP88.  It was later discovered that the PAP88 sequence was
            identical to the sequence of HPV-66 (M. Van Ranst personal
            communication).  Tachezy et al (Biochem. Biophys. Res. Comm. 204,
            820-827) isolated a variant of HPV-66, designated AE3 (U01533).
BASE COUNT     2471 a   1355 c    1653 g    2345 t
ORIGIN         102 bp upstream from beginning of E6 cds
        1 gaaagtttca atcatacttt attatattgg gagtaACCGA AATGGGTtta ggACCGAAAA
                                           E2 binding ->      E2 binding ->
       61 CGGTacaTAT AAAAggcagc ctgttgtgcc tgTAGatatc cATGgattcc atattcagca
          signal ->                E6 orf start ->
                                                     E6 cds ->
      121 atacacagga acgtccacga agcctgcacc atctgagcga ggtattacaa atacctttac
      181 ttgatcttag attatcatgt gtatactgca aaaaggaact tacaagttta gagctatata
      241 ggtttgcatg tattgagtta aaactagtat atagaaacaa ttggccatat gcagtatgta
      301 gggtatgttt attgtttttat agtaaggtta gaaaatatag gtactataaa tattcagtgt
      361 atggggcaac attagaaagt ataactaaaa aacagttatc tgatttatca ataaggtgct
      421 accgatgtca atgtccgtta acaccggagg aaaaacaatt gcactgtgaa cataaaagac
      481 gatttcatta tatagcatat gcatggaccg ggtcatgttt gcagtgttgg agacatacga
```

```
     541 gTAGacaagc tacagaatct acagtaTAAC CATGCATGGT aaagtaccaa cgttgcaaga
E7 orf start ->                                <- E6 end
                                        E7 cds ->
                          E2 binding ->
     601 ggttatatta gaacttgcac cgcaaacgga aattgaccta caatgcaatg agcaattgga
     661 cagctcagag gatgaggatg aggatgaaat agaccatttg ctggagcggc cacagcaagc
     721 tagacaagct gaacaacata agtgttacct aattcacgta ccttgttgta agtgtgagtt
     781 ggtggtgcag ttggacattc agagtaccaa agaggagcta cgtgtggtac aacagctgct
     841 tatgggtgcg ttaacagtaa cgtgcccact ctgcgcatca tcTAAaTAAc tgcaATGgca
                                            E1 orf start -> E1 cds ->
                                                            <- E7 end
     901 tcacctgaag gtacagatgg ggaggggatg ggatgttgtg gatggtttca ggtagaagca
     961 attgtagaaa gaaaaacggg ggatacaata tcagatgatg aaagcgagga ggagaatgaa
    1021 acagatacag atgtagatgg atttatagac aatacactta taaacaatac acaggaagac
    1081 agggagacag ctcaacaatt attgcaagta caaacagcac atgcagatgc acagacgttg
    1141 caaaaactaa aacgaaagta tataggtagt cccttaagtg atattagtaa tcagcaaact
    1201 gtgtaccgag aggaagtaaa acgaaggcta atattatcag aagacagcgg gtatggcaat
    1261 acattggaaa cattggaaac atcacaacag gtagaatacg aaaagggaaa tgggtgcggg
    1321 agctcacaaa atggaggctc gcaaaacagt aattgtagtg agcactcggt atcaaatatg
    1381 gatatagata caaatatgga aacaccaaca caccaattgc aggaactatt aaaaagtagt
    1441 aacgtacaag gaagattaca ttttaaattt aaagaagtgt atggagtgcc atatacagag
    1501 ttggtgcgaa catttaaaag cgatagtaca tgttgtaacg attggatatg tgcaatattt
    1561 ggcgttaatg aaacattagc agaggcgtta aaaactatac taaaaccaca atgtgtgtac
    1621 tatcatatgc aatgcttaac atgttcatgg ggagtaattg taatgatgct aattagatat
    1681 atatgtggaa aaaatagaaa aacaattaca aaatcgctaa gctcaatttt aaatgtacca
    1741 caagagcaaa tgttaattca accaccaaaa ctacgaagtc ctgctgtagc attatatttt
    1801 tataaaacag caatgtcaaa tattagtgag gtgtatgggg aaacaccaga atggatacaa
    1861 agacagacac aattgcaaca cagtttacaa gacaatcaat ttgaattgtc taaaatggta
    1921 cagtgggcat ttgataatga gtaacagat gatagccaaa ttgccttttt atatgcacaa
    1981 ctagcagaca tagatagcaa tgcacaagca ttttttaaaa gtaatatgca agcaaaatat
    2041 gtaaaggatt gtggaataat gtgtagacat tacaaaaggg cacagcaaca gcaaatgaat
    2101 atgtgccagt ggataaagca tatatgtagt aaagtagatg aaggggggtga ttggaaaccc
    2161 attgtgcaat ttttacgata tcaaggggtc gacttcattt catttttaag ttattttaaa
    2221 ttatttttac aaggaacgcc taaacataat tgtttggtac tgtgtggacc accaaataca
    2281 ggtaaatcat gttttgctat gagccttata aattttttcc aagggtcagt catttcattt
    2341 gttaattcac aaagccactt ttggttacag ccactagaca atgccaaatt aggtttgctg
    2401 gatgatgcaa cagatacgtg ttggagatac atagatgatt atctaagaaa tttattagat
    2461 gggaatccca taagtttaga taggaaacat aaacaattag tacaaataaa atgtcctcca
    2521 gttattatta caactaatgt aaatcctatg caagatgcaa aattaagata tttacacagt
    2581 agaatttcag tgtttaagtt tgaaaatcca tttcattag ataacaatgg taatcctgtg
    2641 tatgaattaa gtaatgtaaa ttggaaatgt tttttttgaaa ggacatggtc cagatTAAat
                                                        E2 orf start ->
    2701 ttggataacg acgaggacaa agaaaacaAT Ggagactcta tcccaacgtt tagatgcgtg
                          E2 cds ->
    2761 ccagaacaaa atactagact gttaTGAaaa agatagtaaa tgcattatag atcacataga
                                  <- E1 end
    2821 ctattggaaa gctgtacgac atgaatatgt attatattat aaagcaagag aaaatgacat
    2881 taatgtacta aaccaccaga tggtgccctc tttacaagtg tgtaaagcaa aagcatgtag
    2941 tgcaatagaa ttacaaatag cactggaagc aataagtaac acaatatata aaaatgaaga
    3001 gtggacatta cgtgatacat gtgatgaact gtggcgcacg gagcctaaaa actgtttttaa
    3061 aaaagaagga caacacaTAG aagtgtggtt tgatggtaac aaaaataatt gtatggaata
              E4 orf start ->
              NH2 terminus unknown
    3121 tgtggtgtgg aaatttatat attataatgg agagtgtggg tggtgtaaag tgtcatcagg
    3181 ggtggattac agaggcatat attatatgca tgatggccac aaaacatatt acacagactt
    3241 tgaacaggag gccaaaaaat atgggtgtac aaacatatgg gaagtacata tggaaaccga
    3301 gagtatttac tgtcctgact ctgtgtctag tacctgtaga tacaacgtac ccctgttga
    3361 gactgttaac gaatacaaca accacaggac caccaccacc gcctccacct ttgtgggcgc
    3421 ccaagacgcc gcggtatccc acagaccagg aaaacgaccc agagcaagtg aatcagaacc
```

```
3481 tgactcctcc agagagtcct acgcacactg tgtcacaaca gacacagaca tcagtaacaa
3541 cgccaacagt agaagtccac gtatcaacac acaaagccac tgtggtgata aaactacgcc
3601 tgTAAtccat ttaaaaggtg aagctaatag attaaagtgt tgtagataca gatttcaaaa
          <- E4 end
3661 atataaaaca ttatttacag atgtaacaac aacatatcat tggacaagta cagataataa
3721 agacagtagt attattacaa tattatataa agatgaaaca caacgggaca ccttttttaaa
3781 tgttgtaaaa ataccaccta gtgtacaggt tattttggga caaatgagtt gtccaTAAag
                                                                  <- E2 end
3841 tgttgtatat attgtatata catatgtgtt attgtaacac tggtacaggt gaagtgtaat
3901 tgccatacat tgctgctaag catatatatt gcacccatta attgtatttg gtatattatg
3961 tgttattgta acactgggaa aggtaacgtg TAAtcgccat atattgcaac cattgatttt
                        E5 orf start ->
                        NH₂ terminus unknown
4021 tgtgtaattt gtgtgtttgc gctttgcttt tgtgtttgtc tgtgtgtgtg ccattttgtc
4081 ccgcttttgc tatctgcatc tttatttaca agttgtctta tactaattat tttattttgg
4141 tttgttgtgg ctacatcatt ttttTGAtact tttatactgt ttttactatt tttttatata
                  L2 orf start ->
4201 cctacactgt gtatatattg ccatgctttg tggttAATAA AccatttgTA Acagtagtaa
                                      signal ->              <- E5 end
4261 tttttgctac tATGgttgcc caccgtgcca cacgacgcaa acgcgcatct gccacacaat
          L2 cds ->
4321 tatataaaac atgcaaatta tctggtacat gtcctgagga tgttattaat aaggtggagc
4381 aaaaaacatg ggctgatagg atttacaat ggggaagttt atttacatat tttgggggggc
4441 ttggcattgg tactgggtct gggtcgggtg tcgggcggg ctatgttccc ttaggctcta
4501 ggccttctac tatagttgat gtcactcctg cacgaccACC TATTGTGGTg gagtcagttg
                        E2 binding ->
4561 ggcctacaga tccttctatt gttacactgg tagaagaatc tagtgttatt aactcagggg
4621 ctggtgttcc caatttttact gggtcagggg gatttgaagt tacatcctct tccacaacca
4681 cacctgctgt gttggatatt acacccacat ctagtactgt acatgtaagt agtactacta
4741 taacaaaccc actatatatt gatcctccag taattgaggc tccacaaact ggagaggtat
4801 ctggtaatat tttgattagc actcctacat ctggaataca tagctatgag gaaataaccta
4861 tgcaaacatt tgctatacac ggtactggca acgaacctat tagtagtacc cctattccag
4921 gttttagacg ccttgctgct cccaggttat atagtagggc ttttcagcag gttagggtca
4981 ctgacccagc atttttggac aaccccacaa cattaatatc tgctgataat cctgttttttg
5041 aaggtgctga cacaacgttg acctttttctc cctcgggtgt ggctcctgat cctgattttaa
5101 tggatatagt tgcattacat aggcctgcat ttactacacg tagaacaggt gtgcgttttaa
5161 gtaggctagg caaaaaggct accatgcaaa cacgtagggg tacgcaaata ggtgctcgtg
5221 tgcattatta ttatgatata agtcctattg cacaggctga tgaaattgaa atgcagccat
5281 tattgtctac agacaattca tttgatggcc tatatgatat ttatgcaaat attgatgatg
5341 aggcacccat ttcatttcgt cagtctggtg ctacaccttc tgcacaatta cctattaaac
5401 cttctacatt atcctttgct agtaacacag ctaatgttac tgccccttttg ggaaatgttt
5461 gggaaacacc attttattca ggtcctgata TAGtttttacc tacaggcccc agtacttggc
                        L1 orf start ->
5521 ccttcgtacc tcagtctcct tctgATGtta cacatgatgt atatatacag ggagctacat
          first 'ATG' of L1 cds ->
5581 ttgcactatg gcctgtatat tttttttaaac gtaggcgccg taaacgtatt ccctattttt
5641 ttgcAGATGg cgatgtggcg gccTAGtgac aataaggtgt acctacctcc aacacctgtt
probable L1 ->                <- L2 end
cds start /\ 3' sj
5701 tcaaaggttg tggcaacgga tacatatgta aaacgtacca gtatattttta tcatgcaggt
5761 agctctaggt tgcttgctgt tggccatcct tattactctg tttccaaatc tggtaccaaa
5821 acaaacatcc ctaaagttag tgcatatcag tatagagtgt ttaggtacg gttgcctgat
5881 cctaataagt ttggccttcc tgatccatct ttctataatc ctgaccagga acgtttggta
5941 tgggcctgtg taggtttgga ggtaggccga ggtcaacctt taggtgctgg gttaagtggt
6001 catccattat ttaataggct ggatgacact gaggtctcta atttagcagg taataatgtt
6061 atagaagata gccgggacaa tatatctgtt gattgtaaac aaacccagtt atgtattgtg
6121 ggatgtgcac cagcattagg ggaacattgg actaagggcg cggtgtgtaa gtctacacca
6181 ggtaatacag gggattgtcc acctcttgca ttagttaata ccccgataga ggacggtgac
```

```
6241 atggtggacA CCGGGTTTGG Tgcaatggac tttaagctat tacaggaatc aaaggctgag
       E2 binding ->
6301 gtgccattgg acattgtaca atctacatgt aaatatcctg attatttaaa aatgtctgca
6361 gatgcctatg gggattctat gtggtttttac ttacgcaggg aacaattgtt tgccagacat
6421 tactttaata gggcaggtaa tgttggggaa gccattccta cagatttgta ttggaagggt
6481 ggcaatggca gggacccctcc tcccagttct gtatatgttg ctactcctag tgggtccatg
6541 attacctctg aggcccaatt atttaataaa ccttattggt tgcaacgtgc acagggccat
6601 aataatggca tatgctgggg taatcaggta tttgttactg ttgtggatac taccagaagc
6661 accaacatga ctattaatgc agctaaaagc acattaacta aatatgatgc ccgtgaaatc
6721 aatcaatacc ttcgccatgt ggaggaatat gaactacagt ttgtgtttca actttgtaaa
6781 ataaccttaa ctgcagaagt tatggcatat ttgcataata tgaataatac tttattagac
6841 gattggaata ttggcttatc cccaccagtt gcaactagct tagaggataa atataggtat
6901 attaaaagca cagctattac atgtcagagg aacagcccc ctgcagaaaa gcaggatccc
6961 ctggctaaat ataagttttg ggaagttaat ttacaggaca gctttttctgc agacctggat
7021 cagtttcctt tgggtagaaa attttttaatg caactaggcc ctagaccccc tagacccaag
7081 gctagtgtat ctgcctctaa aaggcgggcg gctcctacct cttcctcttc ttcaccagct
7141 aaacgtaaaa aacgaTAGtt gtgtgttgtg tgttgtatgt attgtatggt tgtgcttgta
                    <- L1 end
7201 ctgtatgttt ttgtgtatgt ttatgtattt tataattgtg tatgtgctat gtgtatgtat
7261 gactgtatgt atgtgtaatg ttttgtgtgt atgtAATAAA catgcatggt tactttttacg
                    signal ->
7321 cgtggttgca taaactaagg tgcggtagta tccttgggca gtgtgtgtca ggttaggtgg
7381 tgttccttac tgtttaatgt tatattaaat aggttgtttg tatgcactat agtaacacac
7441 caaactccat tttagtgctg tacgccattt tatgcatgca ACCGAATTCG GTtgcctagc
                    E2 binding ->
7501 cttttgtcct tatttaaacc caaaacgact tttcagcaaa acagttaatc ctttggcata
7561 ttgccgtttc ctgttgtatg attcaggtat gtacactgcc ttaccctgta ttactcacct
7621 gtatttctgt gccaactatg cttttatctg catactttgg cgctgttggg catatgtttt
7681 tatgcaggtg tttgcaatat attttgttgg cgtgtagccc ttattgtata agccaagtat
7741 ctgtcttgca aatatgtaac catatactta ctcattttac aaaACCGTTT ACGGTcgtgc
                    E2 binding ->
7801 taaaacaggt ttcttttaat tgtt
//
```

Group A7 Sequences

| HPV18	HPV39
HPV45	HPV59
HPV68	HPV70

INTRODUCTION

Group A7 is equivalent to group C in the old classification. Consisting of human papillomavirus types 18, 39, 45, 59, 68, and 70, group A7 is primarily associated with anogenital lesions, some with considerable oncogenic potential. Lorincz et al. classified HPV-18 and HPV-45 as "high risk" viruses [1]. However, Bergeron et al., [2] placed HPV-45 in an "intermediate risk" category. HPV-16, a "high risk" group A9 virus, was originally thought to be the most lethal of the HPV types. Comparison studies of HPV-18 and HPV-16 suggest that HPV-18 may have the higher oncogenic potential of the two or, alternatively, that HPV-18 infection may progress to malignancy more rapidly. These proposals are based on research that measured the relative detection frequency of HPV-18 and HPV-16 in squamous cell-carcinomas compared to detection in CIN lesions (the ratio for HPV-16 was 1.2 as opposed to 2.3 for HPV-18) [3].

The viruses in group A7 primarily cause anogenital lesions. HPV-18 has been found in high prevalence in adenocarcinomas and in moderate prevalence in squamous cell carcinomas. Consistently, HPV-18 and HPV-16 are the most prevalent HPV types in adenocarcinomas and adenosquamous carcinomas [4]. Relative frequencies of the two types vary among studies. However, in most studies HPV-18 is detected at least as frequently as HPV-16 [4–7]. In contrast, HPV-18 is found less frequently than HPV-16 in squamous cell carcinomas of the genital tract, in some cases up to five times less [8, 9]. Other viruses in this group, HPV-39, HPV-68, HPV-45, and HPV-59, have been identified the majority of the time in anogenital tissues which exhibit some degree of dysplasia, if not full-blown in-situ carcinoma. HPV-39 was first isolated from penile Bowenoid papules and subsequently detected in a few cases of intraepithelial neoplasias and invasive cervical carcinomas [10]. HPV-45 was initially derived from a recurrent cervical lesion with mild to moderate dysplasia [11]. HPV-68 was originally isolated from a genital lesion [12]. Subsequently, a partial HPV genome was recovered from the cell line ME180, derived from a cervical carcinoma [13]. This sequence was more than 90% homologous to the original HPV-68 isolate [13]. And finally, HPV-59 was originally isolated from a vulvar intraepithelial neoplasm of the genital mucosa [14]. In the International Biological Study on Cervical Cancer (IBSCC) [15], group A7 viruses were detected in a worldwide sample of cervical cancer as follows: HPV-18 (13.7%), HPV-45 (8.4%), HPV-59 (1.6%), HPV-39 (1.5%), and HPV-68 (1.2%).

In addition to the infection of anogenital tissue, HPV-18 is highly associated with infection of the ororespiratory system. Carcinomas of the oral cavity, tongue, esophagus, sinonasal epithelium and lung have been positive for HPV-18 DNA [16–20]. In approximately 10% of all HPV-positive lung carcinomas, HPV-18 DNA has been detected [19]. In addition, HPV-59 has been isolated from a papilloma on the lip [12].

The viruses CP141, AE1 and LVX160, published last year, are now known to be representatives of HPV-70 [21–23], the full genome of which [24] was released this year.

Of the members of Group A7, complete genomic sequences are available for HPV-18, HPV-39, HPV-45, HPV-59, and HPV-70, the latter two sequences having been released in 1995. The
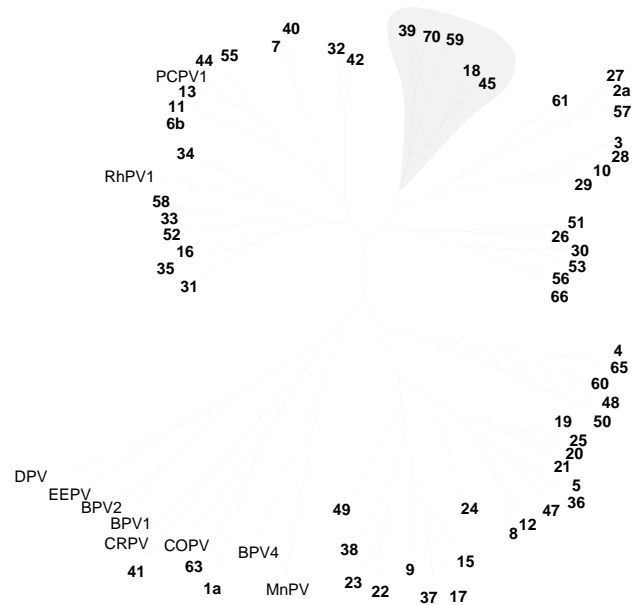
virus HPV68ME180, whose partial genome was published last year, was discovered integrated into chromosomal DNA of the ME180 cervical carcinoma cell line. Approximately 2 kbp of the viral genome in the E1-E2 region were missing.

What's new?

Last year we published only MY09-MY11 fragments of HPV-59 and HPV-70 (see *Human Papillomaviruses 1994* pp. I-C-16, I-C-20, I-C-22, and I-C-23. This year, on the following pages, we publish the complete genomes of HPV-59 and HPV-70 [24].

References

[1] Lorincz,A.T., Reid,R., Jenson,A.B., Greenberg,M.D., Lancaster,WD, and Kurman,R.J. Human papillomavirus infection of the cervix: relative risk associations of 15 common anogenital types. *Obestet Gynecol* **79**:328–337

[2] Bergeron,C., Barrasso,R., Beaudenon,S., Flamant,P., Croissant,O., and Orth,G. Human papillomaviruses associated with cervical intraepithelial neoplasia. Great diversity and distinct distribution in low- and high-grade lesions. *Am J Surg Pathol* **16**:641–649

[3] Arends,M.J., Donaldson,Y.K., Duvall,E., Wyllie,A.H., and Bird,C.C. Human papillomavirus type 18 associates with more advanced cervical neoplasia than human papillomavirus type 16. *Hum Pathol* **24**:432–7 (1993).

[4] Matsuo,N., Iwasaka,T., Hayashi,Y., Hara,K., Mvula,M., and Sugimori,H. Polymerase chain reaction analysis of human papillomavirus in adenocarcinoma and adenosquamous carcinoma of the uterine cervix. *Int J Gynaecol Obstet* **41**:251–6 (1993)

[5] Leminen,A., Paavonen,J., Vesterinen,E., Wahlstrom,T., Rantala,I., and Lehtinen,M. Human papillomavirus types 16 and 18 in adenocarcinoma of the uterine cervix. *Am J Clin Pathol* **95**:647–52 (1991)

[6] Johnson,T.L., Kim,W., Plieth,D.A., and Sarkar,F.H. Detection of HPV 16/18 DNA in cervical adenocarcinoma using polymerase chain reaction (PCR) methodology. *Mod Pathol* **5**:35–40 (1992)

[7] Lee,K.R., Howard,P., Heintz,N.H., and Collins,C.C. Low prevalence of human papillomavirus types 16 and 18 in cervical adenocarcinoma in situ, invasive adenocarcinoma, and glandular dysplasia by polymerase chain reaction. *Mod Pathol* **6**:433–7 (1993)

[8] Low,S.H., Thong,T.W., Ho,T.H., Lee,Y.S., Morita,T., Singh,M., Yap,E.H., Chan,Y.C. Prevalence of human papillomavirus types 16 and 18 in cervical carcinomas: a study by dot and Southern blot hybridization and the polymerase chain reaction. *Jpn J Cancer Res* **81**:1118–23 (1990)

[9] Arends,M.J., Donaldson,Y.K., Duvall,E., Wyllie,A.H., and Bird,C.C. HPV in full thickness cervical biopsies: high prevalence in CIN 2 and CIN 3 detected by a sensitive PCR method. *J Pathol* **165**:301–9 (1991)

[10] Beaudenon,S., Kremsdorf,D., Obalek,S., Jablonska,S., Pehau-Arnaudet,G., Croissant,O., and Orth,G. Plurality of genital human papillomaviruses: characterization of two new types with distinct biological properties. *Virology* **161**:374–84 (1987)

[11] Naghashfar,Z.S., Rosenshein,N.B., Lorincz,A.T., Buscema,J., and Shah,K.V. Characterization of human papillomavirus type 45, a new type 18-related virus of the genital tract. *J Gen Virol* **68**:3073–9 (1987)

[12] de Villiers,E.M. Human pathogenic papillomavirus types: an update. in *Human pathogenic papillomaviruses*, edited by Harald zur Hausen, Springer-Verlag, Heidelberg, pp 1–12 (1994)

[13] Reuter,S., Delius,H., Kahn,T., Hofmann,B., zur Hausen,H., and Schwarz,E. Characterization of a novel human papillomavirus DNA in the cervical carcinoma cell line ME180. *J Virol* **65**:5564–8 (1991)

[14] de Villiers, E.M. Heterogeneity of the human papillomavirus group. *J Virol 63*:4898–4903 (1989)

[15] Bosch, F.X., Manos, M.M., Munoz, N., Sherman, M., Jansen, A.M., Peto, J., Schiffman, M.H., Moreno, V., Kurman, R., Shah, K.V., International Biological Study on Cervical Cancer (IBSCC) Study Group. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. *J Natl Cancer Inst* **87**:796–801 (1995)

[16] Woods,K.V., Shillitoe,E.J., Spitz,M.R., Schantz,S.P., and Adler-Storthz,K. Analysis of human papillomavirus DNA in oral squamous cell carcinomas. *J Oral Pathol Med* **22**:101–8 (1993)

[17] Toh,Y., Kuwano,H., Tanaka,S., Baba,K., Matsuda,H., Sugimachi,K., and Mori,R. Detection of human papillomavirus DNA in esophageal carcinoma in Japan by polymerase chain reaction. *Cancer* **70**:2234–8 (1992)

[18] Kashima,H.K., Kessis,T., Hruban,R.H., Wu,T.C., Zinreich,S.J., Shah, K.V. Human papillomavirus in sinonasal papillomas and squamous cell carcinoma. *Laryngoscope* **102**:973–6 (1992)

[19] Shindoh,M., Sawada,Y., Kohgo,T., Amemiya,A., and Fujinaga,K. Detection of human papillomavirus DNA sequences in tongue squamous-cell carcinoma utilizing the polymerase chain reaction method. *Int J Cancer* **50**:167–71 (1992)

[20] 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 pathogenic papillomaviruses*, edited by Harald zur Hausen, Springer-Verlag, Heidelberg, pp 177–197 (1994)

[21] Peyton,C.L., and Wheeler,C.M. Identification of five novel human papillomaviruses in the New Mexico triethnic population. *J. Infect. Dis.* **170**:1089–92 (1994)

[22] Ong,C.-K., Bernard,H.-U., and Villa,L.L. Identification of genomic sequences of three novel human papillomaviruses in cervical smears of Amazonian Indians. *J. Infect. Dis.* **170**:1086–88 (1994)

[23] Tachezy,R., Van Ranst,M.A., Cruz,Y. and Burk,R.D. Consensus primer mediated PCR allows identification of novel human papillomavirus PCR-types in cervicovaginal lavages. *Biochem Biophys Res Com* **204**:820–7 (1994)

[24] Forslund O., Hansson, B.G. Cloning of the complete genome of a new genital human papillomavirus type candidate by the use of overlapping PCR products. Thirteenth International Papillomavirus Conference. Amsterdam, The Netherlands, Oct. 1994

```
LOCUS       HPV59       7896 bp    DNA            VRL       10-OCT-1994
DEFINITION  Human papilloma virus type 59, complete viral genome.
ACCESSION   X77858
SOURCE      Human papillomavirus type 59.
REFERENCE   1  (bases 1 to 7896)
  AUTHORS   Rho,J., Roy-Burman,A., Kim,H., De Villiers,E.M., Matsukura,T. and
            Choe,J.
  TITLE     Nucleotide sequence and phylogenetic classification of human
            papillomavirus type 59
  JOURNAL   Virology 203, 158-161 (1994)
REFERENCE   2  (bases 1 to 7896)
  AUTHORS   Choe,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-1994) to the EMBL/GenBank/DDBJ databases. J.
            Choe, C/O Hajo Delius, DKFZ - Abt. ATV, Im Neuenheimer Feld 506,
            69120 Heidelberg, FRG
COMMENT     HPV-59 was isolated and cloned from a vulvar intraepithelial
            neoplasia (Matsukura et al, unpublished data), and also detected
            in a lesion of the lip.  Subcloned fragments of the genome were
            sequenced.  The genome is similar in organization to other
            HPV genomes.  HPV-59 contains an intact E5 ORF.  There is
            an additional small ORF contained inside of the E1 ORF.
            Similarity and phylogenetic analysis (of the L1 and L2 ORFs)
            group HPV-59 most closely with HPV-18, HPV-45 and HPV-39.
            The HPV-59 LCR does not contain a consensus GRE binding site.
            E2 binding sites found in the LCR of related types are conserved;
            additional E2 binding sites are found within the L1 and L2 ORFs.
            Cys-X-X-Cys motifs are found in E6 and E7, as in other types;
            an additional such motif is found in E5.  A putative tissue
            specific motif in L2 has been noted: T-T-P-A-V/I-L/I-D/N-V/I;
            an extended motif uniquely associated with the high risk group
            containing HPV-59 is: T-T/D-P-A-V-L-D-D-I-T-P.

            A 149 bp segment in the E2 orf of HPV-59 (starting at nt 2970
            of the complete sequence) is dissimilar from all other
            papillomaviruses in an otherwise well-conserved region.  The
            reverse complement of this segment, however, is extremely similar
            to related papillomaviruses, suggesting that the segment may have
            resulted from either an error in assembling the sequence or an
            actual inversion in the cloned genome.  In light of this, it is
            interesting to note that the first six bases at the 5' end of the
            segment ("CTGCAG") are exactly repeated at the 3' end, and that
            these two sub-sequences are palindromic.  If the inversion was
            present in the genome of the original isolate, it is likely that
            the resultant E2 protein was defective, perhaps being a factor
            in the induction of cancer, since disruption of E2 appears to be
            a factor in oncogenesis in other types.  It should be noted that
            the inversion does not result in a premature stop in the E2
            protein.
            NCBI gi: 557236

BASE COUNT    2473 a   1457 c   1594 g   2372 t
ORIGIN
        1 gttaagACCG AAAACGGTgc atataaaggt agtTGAaaag aaaagggcaa cggcATGgca
          E2 bind ->                     E6 orf start ->         E6 cds ->
       61 cgctttgagg atcctacaca acgaccatac aaactgcctg atttgagcac aacattgaat
      121 attcctctgc atgatattcg catcaattgt gtgttttgca aaggggaact gcaagaaaga
      181 gaggtatttg aatttgcttt taatgactta tttatagtgt atagagactg tacaccgtat
      241 gcagcgtgtc tgaaatgcat ttcatttttat gcaagagtaa gagaattaag atattataga
```

```
 301 gattccgtgt atggagaaac attagaggct gaaaccaaga caccgttaca tgagctgctg
 361 atacgctgtt atagatgcct aaaacctcta tgtccaacag ataaattaaa gcatataact
 421 gaaaaaagaa gattccataa tatagctgga atatatacag gacagtgtcg tgggtgtcgg
 481 acccgagcaa gacaccTAAg acagcaacga caagcgcgta gtgaaacact ggtgTAAaac
            E7 orf start ->                                   <- E6 end
 541 aATGcatgga ccaaaagcaa cactttgtga cattgtttta gatttggaac cacaaaatta
E7 cds ->
 601 tgaggaagtt gaccttgtgt gctacgagca attacctgac tccgactccg agaatgaaaa
 661 agatgaacca gatggagtta atcatccttt gctactagct agacgagctg aaccacagcg
 721 tcacaacatt gtgtgtgtgt gttgtaagtg taataatcaa cttcagctag tagtagaaac
 781 ctcgcaagac ggattgcgag ccttacagca gctgtttatg gacacactat cctttgtgtg
 841 tcctttgtgt gcagcaaacc agTAAcctgc aATGgccgat cggaaggta cagatgggga
                             <- E7 end
                           E1 cds ->
                  E1 orf start ->
 901 agggacgggg tgcaatggat ggttttttgt gcaggcaata gtagataaaa aaacaggtga
 961 caaaatttca gatgacgagg atgaaaatgc aacagataca ggttcagact tggtagattt
1021 tattgatgat accacaacaa tttgtgtaca ggcagagcgc gagacagcac aggccttgtt
1081 taatgtgcag gaagcccaaa gggatgcacg ggaaatgcat gtttttaaaac gaaagtttgg
1141 gtgcagtata gaaaacagta gtgagaaagc ggcggcagga aaaaaagcta agtcaccatt
1201 acaagaaata tcagtaaatg ttaaccaccc aaaagtaaaa agaaggttaa taacagtgcc
1261 agacagcggc tatggctatt ctgaagtgga aatgctcgag actcaggtaa ccgtggagaa
1321 tactggaaat ggggatagca atggcagtgt tgtagcgac agtcaaatag actgtagcga
1381 cagcagtaac atggatgttg aaaacatagt tccaacatcc cccactaatc aattgttaca
1441 gttattacat agcaaaaata gaaagcagc tatgtatgca aaatttaaag aattgtatgg
1501 gttatcattt caagatttgg ttaggacatt taaaagtgac agaactacct gtagcgattg
1561 ggtaaccgcc attttggtg ttaatccaac tgtagcagaa ggatttaaaa cattaataca
1621 accctatgtg ctatatgcac atatacaatg cttagattgt gcatggggag tagtaatatt
1681 agcattatta agatataaat gtggaaaaaa tagaataaca gttgcaaaag acttagcac
1741 attactacat gtaccagata cgtgcatgtt aattgaacca cccaaattgc gtagtggtgt
1801 tgcagcacta tattggtaca gaacaggaat gtccaatatt agtgaagtta taggggaaac
1861 gcccgaatgg atacaaagac taacaattat acaacatgga gttgatgata gcgtgtttga
1921 cctgtcagaa atgatacaat gggcgtttga taatgaccta acagatgaaa gtgatattgc
1981 atatgaatat gcattaatag cagatagtaa tagtaacgcc gctgcatttt taaaaagcaa
2041 ctgccaggca aaatacctaa aagattgtgc agttatgtgt aggcattata aaagagcaca
2101 aaaaagacaa atgagtatgt cacagtggat aaaatggaga tgtgataaaa tagaagaggg
2161 gggagattgg aaacccatag tacaattttt aagatatcaa ggagtagaat ttataacgtt
2221 tttatgtgca ttaaaagatt tttaaaagg tacccccaaaa agaaattgca ttgtgctgtg
2281 tgggccagca aatacaggca agtcatactt tggaatgagc ctgctacatt ttttacaagg
2341 aactgtaatt tcacatgtaa attcaaatag tcacttttgg ctagaacctt aacagatcg
2401 taaattagct atgctagacg atgcaacaga tagttgttgg acatattttg atacatatat
2461 gcgaaatgct ttggatggca atcctataag tgtagataga aagcataggc acctagtaca
2521 aattaaatgt ccaccaatgc ttattacatc aaatacaaat ccagttacag ataacaggtg
2581 gccatattta aatagcagat taatggtatt taaatttcca aacaaattgc catttgacaa
2641 aaatagaaat ccagtatata caattaatga cagaaactgg aaatgttttt ttgaaaggac
2701 gtggtgcaga ttagattTGA acgaggaaga ggaagATGca gacagtgatg gacaccctt
            E2 orf start ->        E2 cds ->
2761 cgcagcgttt aagtgtgtta caggatcaaa tattagaaca ttaTGAaaac gatagtaaag
                                                      <- E1 end
2821 acattaatga acacataaac tattggaaac tggtgcgtat ggaaaatgta atttatttg
2881 cagcaagaga gaacaatata catacattaa accaccaggt ggtGCCAAcg tttttggtgt
                                          NF-1 bind ->
2941 ctaaaaacaa ggcatgtgaa gctattgaac tgcagtcaaa ccgtacttcc actgtaatgc
       beginning of 149 bp inversion ->
3001 cctgttttt aaaacatttt ttaggtgctg tttgccatag ttcttggcat gtttcttgca
3061 ttgtccattg ctcattttta aactcagttt gtGCCAAact ctctaacgcc atctgcagca
                             NF-1 bind ->                      <- end of
                                                              inversion
3121 aggaaaacac aatgcattac acaagctgga catttatata ttatgtaaat gatgtaggac
```

```
3181 agtggtgtaa aaccacagga aatgtggact tttgggggact atattataaa gtggaagagg
3241 aacaggtgta ctatgTAAaa tttatacATG atGCCAAaaa atatgggact acagacaagt
            E4 orf start -> E4 cds ->    -> NF-1 bind
3301 gggaagtgca ttataatggc aaggttattg attgttatga ctctatgtgc agtaccagtg
3361 acgagcaagt atccactgct ggatcttctg agcaactatc atacccctcc gcaacgcccc
3421 ccgaagccac gtacttgggc ccccaaacgt ggaaccgtca gacgaagact ggaaagcgac
3481 caagacagtg tggatacaca cagcaccctc agtctaccag cgtgtcagtg gactactgtg
3541 acaacccagt cgtccgtttg catccaggca acaacccgcg acggcacatc ccttgcagta
3601 acactacgcc taTAAtacac ttaaaaggtg acaaaaatgg ccttaagtgt ttaaggtata
                 <- E4 end
3661 gattaagaaa agtacactgg ttatttgaaa atatttcctc tacctggcat tggacaggaa
3721 acagaggatc aGCCAAaaca ggcattttaa cattaacata tacaagcgaa acacaacgca
        NF-1 bind ->
3781 atgaattttt agatactgta aaaattccta atagtgtaca aatacatgtt gggtatatga
3841 gtgtgTAAtg gttgttatgc aaatgTAAca caaGCCAAta ctgctgctat attgtatagc
            <- E2 end         NF-1 bind ->
                  E5 orf start ->
                  NH2 terminus unknown
3901 tgaggaaATG ataacccttg tatttgtgtg ttgtgtttgt gtttgcttgt gtgtgtgttg
      E5 cds ->
3961 caatgtcccg cttctgcaat ctgtctatat gtgtgcatat acatggttac tagtatttgt
4021 gtatattgtg gttatcacct cctcatatga gtgttttttta ctatatatat tgttttttat
4081 aattccactg ttactactat atgcccatgc aatactgtcc atacaaTAAt tgctgtatat
                                                    <- E5 end
4141 tgtaaattac attgcactgt attgtacagt atattttaaa cacattatta tttttgttag
4201 gtgttggttt tgttacattt ataaTAAaac ATGgtttccc atcgtgctgc tcgtcgtaaa
                  L2 cds ->
                  L2 orf start ->
4261 cgtgcctcag caacagactt atataaaact tgcaagcagg caggtacatg ccccttctgat
4321 gttattaata aagttgaagg tacaacttta gctgataaaa tattgcagtg gaccagccta
4381 ggaatatttt taggtggact aggtattggt actggatctg gtaccggtgg cagaacaggg
4441 tacatacctt tagggggggcg tacaaacact atagtagatg tatcgcctgc taaaccacca
4501 gtagttattg aacctgttgg acctacagat ccatcatag ttacattagt tgaggattct
4561 agtgttataa catctggagc ccctgcccca acatttacag gtacttcagg atttgaaata
4621 tctacctcta gtacaacaac accagctgtt ttggatataa ccccaacctc ttctgttcaa
4681 attagtagct ctagtttttat aaatcctgca tttacagacc cttctgtcat tgaggttccc
4741 caaacaggtg aaatttctgg taatatatta attagtaccc ctacctctgg tgcacatggc
4801 tatgaagaaa ttccaatgca aacgtttgct acggaaggta ctggtttgga acccattagc
4861 agtacccccca atccaacagt acgtcgtgtg ctggaccta gattgtacag tagggctaat
4921 caacaagttc gggtgtctaa cgctgacttt ttaacacgtc catccacatt tgttacatat
4981 gataaccctg cttatgatcc aattgatact acattaactt ttgaccccctc atcagaggtt
5041 ccagacccgg actttatgga tatagttcgt ttgcataggc ctgcattaac atccagacgc
5101 agcactgtaa ggtttagtag gctaggacaa cgggcaacca tgtttACCCG TAGTGGTaaa
                                                    E2 bind ->
5161 caaattgggg cccgtgtaca ttttttatcat gatataagcc ctataccaca tgctgaagat
5221 attgaattgc aacctcttgt ttcttcccag gctgctactg atgatatata tgatatatat
5281 gcagatatta cagatgaagc acctactagt actGCCAAca ctgcatttac aattcctaaa
                  NF-1 bind ->
5341 tcttctttttc aaagtttgtc attaacacgg tcggcatcta gcacctttttc aaatgtaact
5401 gttcctttgg ctactgcctg ggatgttcct gtaaatacag gacccgatat agttttacct
5461 aatactaata ttgttgaacc cacttattct actacacccct ttaccaccat acagtctatt
5521 aataTAGaag gcacaaatta tttttttatgg cctatatatt attttttacc tcgtaaacgt
L1 orf start ->
5581 aaacgtgttc cctatttttt tacagATGgc tctatggcgt tcTAGtgaca acaaggtgta
                  L1 cds ->                     <- L2 end
5641 tctacctcca ccttcggtag ctaaggttgt cagcactgat gagtatgtca cccgtaccag
5701 tattttctac cacgcaggca gttccagact tcttacagtt ggacatccat atttttaaagt
5761 acctaaaggt ggtaatggta gacaggatgt tcctaaggtg tctgcatatc aatacagagt
5821 atttagggtt aagttacctg atcccaataa atttggcctt ccagataaca cagtatatga
```

```
5881 tcctaactct caacgcttgg tctgggcctg tgtaggtgtt gaaatcggtc gggGCCAACC
                                                             E2 bind ->
                                               NF-1 bind ->
5941 TTTAGGGGTa ggactcagtg gtcatccatt atataataaa ttggatgaca ctgaaaactc
6001 tcatgtagca tctgctgttg ataccaaaga tacacgtgat aatgtatctg tggattataa
6061 acaaactcag ctgtgtatta ttggctgtgt acctgccatt ggagaacact ggacaaaggg
6121 cactgcttgt aagcctacta ctgtggttca gggcgattgt cctccactag aattaataaa
6181 tacaccaatt gaagatggtg atatggtaga cacaggatat ggggctatgg actttaaatt
6241 gttgcaggat aacaaaagtg aagtaccatt ggatatttgt cagtctattt gtaaatatcc
6301 tgattattta caaatgtcag cagatgctta tggagacagt atgttttttt gtttaaggcg
6361 agaacaggtt tttgccagac attttttggaa tagatctggt actatgggtg atcaacttcc
6421 tgaatcacta tatattaaag gtactgacat acgtGCCAAc ccaggcagtt atttatattc
                                       NF-1 bind ->
6481 cccttcccca agtgggtctg tggttacttc tgattcacaa ttatttaata aaccatattg
6541 gctgcacaag gctcagggtt taaacaatgg tatatgttgg cacaatcaat gttttttaac
6601 agttgtagat actactcgca gcaccaatct ttctgtgtgt gcttctacta cttcttctat
6661 tcctaatgta tacacaccta ccagttttaa agaatatgcc agacatgtgg aggaatttga
6721 tttgcagttt atatttcaac tgtgtaaaat aacattaact acagaggtaa tgtcatacat
6781 tcataatatg aataccacta ttttggagga ttggaatttt ggtgttacac cacctcctac
6841 tgctagttta gttgacacat accgttttgt tcaatctgct gctgtaactt gtcaaaagga
6901 caccgcaccg ccagttaaac aggacccta tgacaaacta aagttttggc ctgtagatct
6961 taaggaaagg ttttctgcag atcttgatca gtttcctttg ggacgtaaat ttttattgca
7021 attaggagct agacctaagc ccactatagg cccacgcaaa cgtgcagcgc ctgcccctac
7081 ctctacccca tcaccaaaac gtgttaagcg tcgcaagtct tccagaaaaT AGtgttgttt
                                                             <- L1 end
7141 gttatgtgtt tgtatgtgtg catgttgtat gttttgtatt gtttgcctgt ttgtatgttg
7201 tgtatatgta catgtttgtt tgtctgctgt atgtgtgtat ttgttttttgt acataataaa
7261 gtatgcatga cagtttcatg tgtggttgca cccaatgagt aaggtactgt ccctttattg
7321 tttctttgtc cttattacac attattacac attgccctac ttacataggt gtgtttgttc
7381 cttcattttg tcctgaatgt ccagttttgc atttgcacat tatatggcgt ccattttatc
7441 ctttaaatcc tccattttgc tgtgcaACCG TTTTCGGTta ccttggtttta accttacctt
                                  E2 bind ->
7501 tttgaacaat taatctgttt aaacatcagc aaaacagtta atccccatct tgtttcctcc
7561 tacacgccta gactactaac acaacttaca aacGCCAAat agttagtcat catcctgtcc
                                        NF-1 bind ->
7621 aggtgcactc taacaatact tgcataactt tggtggcgcc cttgttaata aaacagcttt
7681 taggcacata ttttcactgt ttttactact ttaattgcat aattggcttg caaaactact
7741 gtgcaatcca agaatgtgtc tataatttat tgtaaaaaac atgactaagg tttttgtcat
7801 tgttaagcaA CCGAAAAAGG Tcgggcaagt acatgcacac tttctactta ttacttttta
      E2 bind ->
7861 caatcatagt aataaaaaag ggtgtaaccg aaaacg
```

//

```
LOCUS       HPV70       7904 bp    DNA             VRL        28-MAR-1995
DEFINITION  Human papillomavirus type 70, complete genome.
ACCESSION   U21941
SOURCE      Human papillomavirus.
REFERENCE   1  (bases 1 to 7904)
  AUTHORS   Forslund,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-1995) Ola Forslund, Clinical Virology, Medical
            Microbiology, Malmoe University Hospital, Ing. 78 MAS, Malmoe,
            S-205 02, Sweden
REFERENCE   2  (bases 1 to 7904)
  AUTHORS   Forslund,O. and Hansson,B.G.
  JOURNAL   Unpublished
COMMENT     This sequence was isolated from a cervical brush sample from a
            patient diagnosed with cervical condyloma.  Initial attempts
            to clone the entire genome were unsuccessful. The genome was cloned
            in pieces from three overlapping PCR products. The original sample
            designated HPV70 was isolated by G. Orth et al. The sequence
            HPVCP141 (Accession U12476), a My09-My11 fragment of a type 70
            virus, differs by 2% (8 of 455 nts) from the corresponding region
            of the sequence in this entry.
            NCBI gi: 733077
BASE COUNT     2421 a   1495 c   1696 g   2292 t
ORIGIN
        1 cttataacat tttacaatca taatttaaaa aaagggaggc ACCGAAAACG GTcacgACCG
                                                     E2 bind ->        E2 bind ->
       61 AAAACGGTgt ataTAAaacc atgcaaaagt tgcttgccca tacggaATGg cgcgatttcc
             E6 orf start ->                          E6 cds ->
      121 caatcctgca gaacggccat acaaattgcc tgacctgtgc acggcgctgg acactacatt
      181 gcacgacatt acaatagact gtgtctattg taaaacacag ctacagcaaa cagaggtata
      241 tgaatttgca tttagtgatt tatttatagt atatagaaac ggggagccat atgctgcatG
                                                                  NF-1 binding ->
      301 CCAAaaatgt attaaatttc atgctaaagt aagggaacta cggcattatt cgaactcggt
      361 gtatgcaaca actttggaaa gcataactaa taccaagtta tataatttat caataaggtg
      421 catgagttgc ctgaaaccat tgtgtccagc agaaaaatta aggcatgtTA Ataccaaaag
                                                               E7 orf start ->
      481 aagatttcac caaatagcag gaagctatac aggacagtgc cgacactgct ggaccagcaa
      541 ccgggaggac cgcagacgta tacgaagaga aacacaagta TAAatataaa tATGcatgga
                                                           <- E6 end
                                                           E7 cds  ->
      601 ccacggccga cattgcaaga gattgtttta gatttatatc catacaatga aatacagccg
      661 gtcgaccttg tatgtcacga gcaattagaa gattcagaca atgaaacaga tgaacccgac
      721 catgtagtta atcaccaaca caactacta gccagacggg aagaaccaca gcgtcacaaa
      781 atacagtgta tgtgttgtaa gtgtaatact acactgcact agtagtaga agcctcacaa
      841 gagaacctgc gatctctact gcagctgttt atggagacac tgtcatttgt gtgtccctgg
      901 tgtgcatcgg gaacccagTA AcctgcaATG GCCAAttgtg aaggtacaga tggggatggg
                              <- E7 end   -> NF-1 binding
                    E1 orf start ->
                          E1 cds ->
      961 tcgggatgta acggatggtt cctagtacag gcaatagtag ataaacaaac gggcgacact
     1021 gtgtcagagg acgaggacga aaatgcaaca gatacaggtt cagacttggc agactttatt
     1081 gatgatacta cagatatttg tgtacaggca gagcgcgaga cagcacaggt actgtataat
     1141 atgcaagagg cccaaaggga tgcacaatca gtgcgtgcct aaaacgaaa gtatggaggg
     1201 agcaatctaa ataaaagtcc ttgtgcaaaa ccgccaggcg tacataggga caaaagggta
     1261 acactacaag agctcccggt aaacatatgc aataaacagg caagaacaaa cgtgtattca
     1321 gtaccagaca gcggctatgg caatatggaa gtggaaacag ctgaagtgga ggtaactgta
     1381 gtaaataata caaatgggga agaggaaggg gaaaatggcg gggaaaatgg cggcagcata
     1441 cgggaggagt gcagtagtgt agacagtgct attgatagtg agaatcaaga tccacagtca
     1501 cctactgcac agctaaaaac agtattacag gctaataacc aaaaagccat actactatca
```

```
1561 caatttaaac acacatatgg attagcattt aacgacctgg tacgtacatt taaaagtgat
1621 aaaaccatat gtactgactg ggtagcagca atatgtggag taaatcccac catagcagaa
1681 ggctttaaaa cactaattca gccatatgcg ttatatacac atatacagtg tttggatacc
1741 aaatatggag tgtatatact actattaatt agatatataat gtggaaaaaa caggataaca
1801 gtaggcaaag gattaagtaa attattacat gtgccagaaa gttgtatgct aattgaacca
1861 cctaaattgc gtagccctgt tgcagcactg tattggtata gaactggaat gtctaatata
1921 agtgaagtgt caggtactac gccagaatgg atacagcgat aacagtaat acagcatgga
1981 atagatgaca gtgtatttga cctgtctgat atggtacaat gggcatttga taatgatgta
2041 acagaagaca gtgacatagc atatggatat gcattattag cagatagtaa tagtaatgct
2101 gcagcatttt taaaaagtaa ctgccaggca aaatatgtac gcgactgtgc tacaatgtgc
2161 agacattata aaagggcaca aaaaaaacaa atgactatgg cgcaatggat taggtttaga
2221 tgtgataaat gtgacgatgg gggcgactgg cgaccaatag tgcaatttct aaggtatcaa
2281 ggggtagaat ttataacctt tttgtgtgca tttaaggagt ttttaaaggg caccccaaag
2341 aaaaattgca tagtaataca gggaccacca aacacaggca agtcatactt ttgtatgagt
2401 ttaatgcact ttttacaagg tacagtaatt tcatatgtaa attccactag tcattttgg
2461 ttagagccac ttgcagatgc aaaggtagca atgttggatg atgccacagg cacatgctgg
2521 tcatatttcg atacgtatat gagaaatgca ttagatggaa atcctataag ccttgacaga
2581 aaacatagac atttaataca aattaagtgt ccacccatat aataacatc caataccaat
2641 cctgtagagg aaaataggtg gccataccta actagcagac taacagtgtt tacatttcct
2701 aatgcattcc catttgacca aaacaggaat ccagtgtaca caatcagtaa taaaaactgg
2761 aaaagttttt tccaaaagac ttggtgcaaa tTAGacttgc agcaggacga ggATGaagga
                      E2 orf start ->              E2 cds ->
2821 gacaatgatg gaaacactat cccaacgttt aaatgcgtta caggagaaaa tactagaaca
2881 ttaTGAacag gacagtaaac taatatatga tcaaatcaat tattggaaat atgtgcgact
         <- E1 end
2941 ggaaaatgca atattttatg cagcacggga acgtggcatg catactatag accaccaggt
3001 ggtgccacca ggcactactt caaaagcaaa agcatatcaa gctattgaac tgcagatggc
3061 cctagagagc cttgcacaag ctgactttaa taaagaggag tggacattaa aggacacaag
3121 taatgaaatg tggcagacaa aGCCAAaaca atgtttaaa aaaaaggtg ttacagtgga
             NF-1 binding ->
3181 ggtgtggtac gatggaaaca aggacaattc tatgcattat gtagtgtggg gagcaatata
3241 ttataaaaca catacagaca tgtggtgtaa aacagaaggg tatgtggatt actggggtat
3301 atattatgtg cacgagcagc ataagacata ttatgaagtg tttaagcagg atgcacaaat
3361 gtatgggact agcggaaaat gggaagtgca ttgtaatggc aacaTAAttc attgtcctga
                                        E4 orf start ->
                                        NH₂ terminal unknown
3421 ctctatgtac agtaccagtg acgacacagt acccactact gagcttactg cagaactaca
3481 acacaccacc ccggcccata ccgccgcaac aaccccatgc accaaaaaaa ctaagtcggc
3541 gccgtcttgc aagtgtggag tctccagacc ctcagaaaca gacggagtgt cgtggacct
3601 tgttacaagt aaaggctgca caaacgacg gcaccagtgt tgtggtgaca ctacacctaT
3661 AGtgcattta aaaggtgaca aaaatggttt aaagtgtctt aggtatcgat tgcgaaaatt
         <- E4 end
3721 taattcattg tatgaaaata tttcatgtac ttggcattgg atagggggca agggaagtaa
3781 acatacaggt atactaactg taacatatac tactgaagca caacGCCAAa aattttttgga
                                        NF-1 binding ->
3841 aactgttaga attccaccta gtgtacatgt atctgtggga tataTGAcat tgTAAcagca
                                        E5 orf start ->      <- E2 end
3901 catgctgtAT Gtatattgta tacatatcaa tgattgcatt ggtgtttttg gtgtggtttg
         E5 cds ->
3961 ctgtatgctt atatatatgt tgcagtgtcc cgcttttgcc gtctgtgcat ttgtgtgcgt
4021 atatgtggct actttattt gtgtttattg ttgtacatac cacaccattg caaatgtttt
4081 gtatatattt actatttttt atattgccta tgtggtttt acacatcctt tcagtatatg
4141 ctTAAgttgt gttgctgcat agtgtattgt acattacttg tttttacatt tatattgtac
         <- E5 end
    L2 orf start ->
4201 caataaacAT Ggtttctagc cgtgcgtcca ggcgtaagcg tgcatctgca acagacatat
         L2 cds ->
4261 ataaaacctg caagcaatca ggcacatgtc cgcctgatgt tgttaataag gtggagggta
4321 ccacactggc tgataggttt ttacaatggg ctagtttagg tatttttttg ggtggtttgg
```

```
4381 gaatcggtac gggtactggt actggggggcc gcacagggta cattcctttg gggggtaggc
4441 ctagtacagt tgtagatgtt acccctgcac gtcctcctgt ggttatagaa cctgtaggac
4501 ctacagaacc ttctattgtt cagttggtag aggaatctag tgttgtttcc tctggtacac
4561 ccatccctac ttttacaggc acatctgggt ttgaaattac atcttctgca accacaacac
4621 ctgctgtatt agatattacc cctgcttctg ggtctgttca aattagtacc actagttata
4681 ccaatcctgc atttgctgat ccatcgttaa ttgaggttcc acaaacaggt gaggtgtcag
4741 gcaatatatt tgatactact ccaacatctg gaacacatgg atatgaagaa attcctatgc
4801 aggttttttgc ctcacatgga acaggcacag aacctattag tagtactcct gttcctggtg
4861 ttagtcgtgt ggcaggccca cgtttatata gtagggccta tcatcaggtt cgtgttaata
4921 attttgattt tgtaacccgc ccttcatctt ttgtaacatt tgacaatcca gcttttgagc
4981 ctggtgatac atccttaaca tttgaacctg ctgacacagc tcctgatcca gattttctgg
5041 acattgttcg tttacatcgg cctgctttaa cctcacgacg cggaacagta cgctttagta
5101 ggcttggtaa aaaggccaca atgtttACCC GGCGGGGTac acaaattggg gcacaggttc
                       E2 bind ->
5161 attattatca tgatgttagt aacattactg caacagaaga cattgagatg caacctttac
5221 ttacctctga atctacagat ggtttatatg atatatatgc agatgcagat atagataatg
5281 caatgttaca tactacttct catacaggtt ctacaggacc taggtcccat ctttcatttc
5341 cttctatacc ttctacagtg tctacaaaat atagtaatac aaccattcca tttactactt
5401 cttgggacat acctgtaacc actggccctg acatagtttt acctactgca tcccccaatt
5461 tgccctttgt ccctcctaca tctatagata ccacagttgc aatagccatt cagggctcca
5521 attattattt attgccttta ttatattatt ttcTAAagaa gcgtaaacgt attccctatt
                       L1 orf start ->
5581 tttttacagA TGgctttgtg gcggtcTAGt gacaacacgg tgtatttgcc accccccttct
       L1 cds ->                         <- L2 end
5641 gtggcgaagg ttgtcaatac agatgattat gtaacacgta caggcatata ttattatgct
5701 ggaagctctc gcttattaac agtagggcat ccttatttta aggtacctgt aaatggtggc
5761 cgcaagcagg aaatacctaa ggtgtctgca tatcagtata gggtatttag ggtatcccta
5821 cctgatccta ataagtttgg ccttccggat ccttcccttt ataatcctga cacacaacgc
5881 ctggtatggg cctgtatagg tgtggaaatt ggtagaggcc agccattggg cgttggcgtt
5941 agtggacatc ctttatataa tagattggat gatactgaaa attctcattt ttcctctgct
6001 gttagtacac aggacagtag ggacaatgtg tctgtggact ataagcaaac acagttatgt
6061 attataggct gtgttcctgc tatgggagag cactgggcta agggcaaggc ctgtaagtcc
6121 actcaacagg gcgattgtcc accattagaa ttagttaata ctgcaattga ggatggcgat
6181 atgatagata caggctatgg tgccatggac tttcgtacat tgcaggaaac caaaagtgag
6241 gtaccactag atatttGCCA Atccgtgtgt aaatatcctg attatttgca gatgtctgct
       NF-1 binding ->
6301 gatgtatatg gggacagtat gtttttttgt ttgcgcaagg aacagttgtt tgccaggcac
6361 ttttggaata gaggtggcat ggtgggcgac acaatacctt cagagttata tattaaaggc
6421 acggatatac gtgagcgtcc tggtactcat gtatattccc cttcccaag tggctctatg
6481 gtctcttctg attcccagtt gtttaataag ccctattggt tgcataaggc ccagggacac
6541 aataatggca tttgttggca taaccagttg tttattactg tggtggacac tacacgtagt
6601 actaatttta cattgtctgc ctgcaccgaa acggccatac ctgctgtata tagccctaca
6661 aagtttaagg aatatactag gcatgtggag gaatatgatt acaatttat atttcaattg
6721 tgtactatca cattaactgc tgacgttatg gcctacatcc atactatgaa tcctgcaatt
6781 ttggtcaatt ggaatatagg agttacccct ccaccatctg caagcttggt ggacacgtat
6841 aggtatttac aatcagcagc tatagcatgt caaaaggatg ctcctacacc tgaaaaaaag
6901 gatcccctatg acgatttaaa attttggaat gttgatttaa aggaaaagtt tagtacagaa
6961 ctagatcagt ttcctttggg gcgcaaattt ttactacagg tagggctcg cagacgtcct
7021 actataggcc ctcgcaaacg ccctgcgtca gctaaatcgt cttcctcagc ctctaaacac
7081 aaacggaaac gtgtgtccaa gTAAtgtatg tatgttgtat gctgtgtatt atgtactatt
                       <- L1 cds end
7141 acatatttgt gttttttatgt tgtatgcttg cacactgttt acatatttgt gtttgtatgt
7201 tgtatgcttg cacactgtac tgtatatgtt tgtcctggta catatttgtg gttgtatgtg
7261 tatatgttgc gtgctatgtg tatgtttttag aagtatgtgt gtatgtatgt ttttgttaat
7321 aaagtatgta tggaggtttc atttgtggtt gcacctgtg actaaggtgt tgtccctgtt
7381 ttacatataa taggagtgtg attaccaaca tttcctacat aattttatgc cctaccctaa
7441 ggtgtgtgta taccatttgt agtttataca tttatatttt atagtgggtt acctgtatac
7501 agcaacggcc attttgtgtg aaACCGTTTT CGGTtgcatt tggctttgta ccatcagtta
                       E2 bind ->
```

```
7561 cccttataaa ccttttgtat cagcaaaaac atgtcctgta acctaagttc acctacatac
7621 ttggcactac taacagtttt agtggcacac ctacacttag tcatcatcct gtccaggtgc
7681 actacaacaa tgctttggca accttatgca cctccaccct gtctaataaa gtgcttttag
7741 gcatgtattt tacctgtttt tacttaccta agagcatagt tggcctgtat aacagctttt
7801 acatccaaga atgtgtcgtt tggtgcaagt tatattttgt gactaatatt tttacagacc
7861 tgtgtgcacc cgaaataggt tgggcagaca ttcctatact ttta
```

//

Group A8 Sequences

HPV7   **HPV40**
HPV43

INTRODUCTION

Group A8 is equivalent to one of the four "clusters" that comprised the old group F. Viruses HPV-7, HPV-40 and HPV-43, which form Group A8 are predominantly associated with benign lesions. HPV-40 and HPV-43 infect the genital tract. HPV-7 infects both cutaneous and oral tissue, both with significant frequency. HPV-7 was first isolated from Butcher's warts, the benign hand warts of meat handlers; see Maitland et al for a recent review of this association [1]. Initially, it was almost exclusively detected in individuals with animal contact which suggested the possibility of cross-species transmission. This hypothesis was refuted when 37 bovine tumors were screened for the presence of HPV-7 DNA, and not one positive hybridization was reported [2]. Additionally, HPV-7 is frequently detected in common warts of individuals without known animal contact. HPV-7 has been detected in a significant proportion of oral papillomas (9%) and in one case of a tonsillar carcinoma [3,4]. HPV-40 is a rare HPV of the genital tract. It has been detected in benign lesions, low-grade neoplasias and in bowenoid lesions [5,6]. HPV-43 was classified by Lorincz et al. as a "low-risk" virus [7]; it was originally isolated from vulvar tissue [8].

What's new?

No new sequences in Group A8 were released during 1995. The sequences of members of this group were published in *Human Papillomaviruses 1994* pp. I-F-12, I-F-30, and I-F-38.

References

[1] Maitland,N.J., Keefe,M., al-Ghamdi,A., and Coggon,D. Human papillomavirus type 7 and the Butcher's Wart. *Papillomavirus Report* **6**: 33–7 (1995)

[2] Oltersdorf,T., Campo,M.S., Favre,M., Dartmann,K., and Gissmann,L. Molecular cloning and characterization of human papillomavirus type 7 DNA. *Virology* **149**: 247–50 (1986)

[3] de Villiers,E.M., Weidauer,H., Le,J.Y., Neumann,C., and zur Hausen,H. Papilloma viruses in benign and malignant tumors of the mouth and upper respiratory tract. *Laryngol Rhinol Otol (Stuttg)* **65**: 177–9 (1986)

[4] 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

[5] de Villiers,E.M. Human pathogenic papillomavirus types: an update. in *Human pathogenic papillomaviruses*, edited by Harald zur Hausen, Springer-Verlag, Heidelberg, pp 1–12 (1994)

[6] de Villiers,E.M., Hirsch-Behnam,A., von Knebel-Doeberitz,C., Neumann,C., and zur Hausen, H. Two newly identified human papillomavirus types (HPV 40 and 57) isolated from mucosal lesions. *Virology* **171**: 248–53 (1989)

[7] Lorincz,A.T., Reid,R., Jenson,A.B., Greenberg,M.D., Lancaster,WD, and Kurman,R.J. Human papillomavirus infection of the cervix: relative risk associations of 15 common anogenital types. *Obestet Gynecol* **79**: 328–337

[8] Lorincz,A.T, Quinn,A.P., Goldsborough,M.D., Schmidt,B.J., and Temple, G.F. Cloning and partial DNA sequencing of two new human papillomavirus types associated with condylomas and low-grade cervical neoplasia. *J Virol* **63**: 2829–34 (1989)

Group A9 Sequences

HPV16	**HPV31**
HPV33	**HPV35**
HPV52	**HPV58**
HPV67	**RhPV**
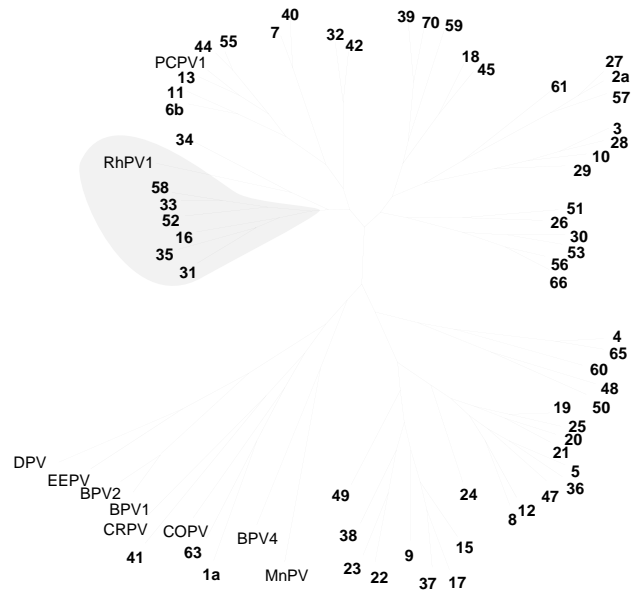
INTRODUCTION

Group A9 is equivalent to the old group A. It consists of human papillomavirus types 16, 31, 33, 35, 52, 58, and 67, a subset of the "high-risk and intermediate-risk" anogenital viruses as classified by Lorincz et al. [1] The Rhesus monkey papillomavirus (RhPV-1), which in the past we treated as an old group "I," or animal papillomavirus, is now included in Group A7. Phylogenetic analyses based on the L1 gene do indeed link RhPV with A7 viruses, however the link, when based on other genes, is weak. It may well be that RhPV would be better classified as the only known representative of a new group within the A supergroup.

The Lorincz risk assessment [1] used probes from 15 common anogenital HPV types to screen 2627 U.S. subjects by high-stringency hybridization. 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, 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. In a recent international study [3] of over 1000 invasive cervical cancers, Group A9 viruses were identified in approximately 64% of the tumors. The following overall type-specific prevalences in this worldwide study were noted as follows: HPV-16 (49.9%), HPV-31 (5.3%), HPV-33 (2.8%), HPV-52 (2.7%), HPV-58 (2.0%), HPV-35 (1.7%).

Group A9 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 [2,4]. Carcinomas of the larynx and lung have also been found to contain HPV-33 [4]. Bowenoid lesions of the finger are associated with HPV-35 [2]. Conversely, types 31, 52, 58, and 67 have been detected exclusively or almost exclusively in anogenital tissue (HPV-67 tissue restriction is based on limited data; one case of laryngeal carcinoma yielded a cell line containing HPV-31 [4]). Multiple site infection is common with all Group A9 viruses, but is most prevalent for HPV-16 [5]. 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 [5].

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 [6]. Transcription of the major HPV-16 RNA species begins at nucleotide 97, from the only known HPV-16 promoter, P_{97}. 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 [7]. 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 [5].

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) [8]. A splice site within the E6 gene produces truncated E6 products, E6*I, E6*II, and E6*III. After translation, proteins may be 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 [9]. 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 A9 except HPV-67, which has been sequenced only over the My09-My11 region of L1. Of the sequences in this group and HPV-33 and HPV-58 are "close types"—sequences that qualify to be distinct types under the criterion of ten percent dissimilarity at the nucleotide level, but between which most of the changes are "silent," causing no difference at the amino acid level (*Human Papillomaviruses 1994* Part III-C). HPV-35 and HPV-35h are sequences determined from the same clone; differences between the two sequences are likely to constitute amplification-induced mutations or sequencing errors.

> ## What's new?
> Although no new sequences in Group A1 were released during 1995, we have revised the sequence for HPV-16 and present it on the following pages. The sequences of other members of this group were published in *Human Papillomaviruses 1994* pp. I-A-1 et seq.

References

[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. *Obestet 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] Bosch, F.X., Manos, M.M., Munoz, N., Sherman, M., Jansen, A.M., Peto, J., Schiffman, M.H., Moreno, V., Kurman, R., Shah, K.V., International Biological Study on Cervical Cancer (IBSCC)

Study Group. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. *J Natl Cancer Inst* **87**:796–801 (1995)

[4] 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

[5] Shah,K. and Howley,P. Papillomaviruses. in Virology, second edition, edited by B.N. Fields, Raven Press, N.Y., pp 1651–1676

[6] 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

[7] McBride, A., Romanczuk,H., and Howley,P. The papillomavirus E2 regulatory proteins. *J. Biol. Chem.* **266**:18411–18414

[8] 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

[9] 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

HPV16R

```
LOCUS       HPV16R       7906 bp ds-DNA   circular  VRL       15-APR-1991
DEFINITION  Human papillomavirus type 16 (HPV16), complete genome.
ACCESSION   <not yet entered in GenBank>
KEYWORDS    circular; complete genome.
SOURCE      Papilloma virus type 16 DNA recovered from a human invasive
            cervical carcinoma.
REFERENCE   1  (bases 1 to 7906)
  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  (base 1139; revision)
  AUTHORS   Matsukura,T., Kanda,T., Furuno,A., Yoshikawa,H., Kawana,T.,
            and Yoshiike,K.
  TITLE     Cloning of monomeric human papillomavirus type 16 DNA
            integrated within cell DNA from a cervical carcinoma.
  JOURNAL   J. Virol. 58, 979-82 (1986)
REFERENCE   3  (base 1139; revision)
  AUTHORS   Baker,C.C., Phelps,W.C., Lindgren,V., Braun,M.J., Gonda,M.A.,
            and Howley,P.M.
  TITLE     Structural and transcriptional analysis of human papillomavirus
            type 16 sequences in cervical carcinoma cell lines.
  JOURNAL   J. Virol. 61, 962-71 (1987)
REFERENCE   4  (base 3907; 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, 243-6 (1988)
REFERENCE   5  (base 3907; revision)
  AUTHORS   Halbert,C.L., and Galloway,D.A.
  TITLE     Identification of the E5 open  reading frame of human
            papillomavirus type 16.
  JOURNAL   J. Virol. 62, 1071-5 (1988)
REFERENCE   6  (base 4365; revision)
  AUTHORS   Baker, C.C.
  TITLE     The Genomes of the Papillomaviruses
  JOURNAL   Genetic Maps: Locus Maps of
            Complex Genomes, 6th edition, Book 1: Viruses.
            S.J. O'Brien ed. Cold Spring Harbor Laboratory Press.
REFERENCE   7  (base 4365; revision)
  AUTHORS   Xi,S.Z., and Banks,L.M.
  TITLE     Baculovirus expression of the human papillomavirus
            type 16 capsid proteins: detection of L1-L2 protein complexes.
  JOURNAL   J. Gen. Virol. 72, 2981-8 (1991).
REFERENCE   8  (bases 6242 and 6903-6905, deletion of GAT at 6952; revision)
  AUTHORS   Kirnbauer,R., Taub,J., Greenstone,H., Roden,R., Durst,M.,
            Gissmann,L., Lowy,D.R., and Schiller,J.T.
  TITLE     Efficient self-assembly of human papillomavirus type 16
            L1 and L1-L2 into virus-like particles.
  JOURNAL   J. Virol. 67, 6929-36 (1993)
REFERENCE   9  (bases 6903-6905, deletion of GAT at 6952; revision)
  AUTHORS   Parton,A.
  TITLE     Nucleotide sequence of the HPV16 L1 open reading frame.
  JOURNAL   Nucl. Acids Res. 18, 3631 (1990)
REFERENCE   10 (base 7434; revision)
  AUTHORS   Bernard,H.-U.
  TITLE     Personal communication, 2-6-95.
  JOURNAL   Unpublished
REFERENCE   11 (base 7863; revision)
```

```
   AUTHORS   Bernard,H.-U.
   TITLE     Personal communication, 2-22-95.
   JOURNAL   Unpublished
REFERENCE   12 (base 7863; revision)
   AUTHORS   Romanczuk,H., Thierry,F., and Howley,P.M.
   TITLE     Mutational analysis of cis elements involved in E2 modulation of
             human papillomavirus type 16 P97 and type 18 P105 promoters.
   JOURNAL   J. Virol. 64, 2849-59 (1990)
COMMENT     In developing countries, cancer of the cervix is responsible for
             24% of all cancers in women.  In these areas, it is the most
             frequent female malignancy.  In developed countries, it ranks
             behind cancers of the breast, lung, uterus, and ovaries and
             accounts for 7% of all female cancers.  HPV-16 is most often found
             in lesions of the genital mucosa which may have a risk for
             malignant progression.  Estimates indicate that HPV-16 has been
             recovered from about 50-60% of all invasive cervical cancers.
             Studies show that the predominance of HPV-16 in high-grade
             anogenital lesions and invasive cancers is evident also for the
             vulva, the penis, and the anus.  HPV-16 has also been found in
             approximately 80% of the Bowenoid papulosis lesions (premalignant).

             In contrast to the extrachromosomal form of the viral genome found
             in skin carcinomas of EV patients, the viral genome of HPV16 is
             frequently found integrated into the cellular genome in genital
             tract cancers.  The integration of the genome requires
             linearization of the genome which most often occurs in the E1/E2
             region.  This results in the inactivation of the E2 repressor of
             the E6 and E7 genes.  With repression removed, the E6 and E7 genes
             are able to immortalize keratinocytes.

             The DNA of HPV-16 has been molecularly cloned from a genomic
             library of an invasive cervical carcinoma in bateriophage lambda.
             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 4291.  A
             potential CAT box precedes the TATA boxes near the E6 region, at bp
             7898.  Polyadenylation signals can be found at the end of the early
             and late gene regions, at bp 4215 and 7321.  Open reading
             frame E4 does not have a translation initiation codon, which
             is present in most other sequenced papilloma viruses.

             The originally published sequence of [1] contained several sequencing
             errors which have since been pointed out by other authors ([4]-[12]).
             These changes, which have been incorporated into the present entry,
             are as follows: insertion of 't' at bp 3907; substitution of 't' for
             'a' at bp 4365; insertion of 'cat' at bp 6903; deletion of 'gat' at
             bp 6952; substitution of 'cgg' for 'gc' at bp 7434; deletion of 'a'
             at bp 7863 (nucleotide positions given in terms of the corrected
             sequence presented herein).  In addition to these sequencing errors,
             the following differences between the originally sequenced clone and
             the wild-type HPV-16 DNA have been noted ([2] and [3]) and are
             incorporated into the present sequence: insertion of 'g' in wild-type
             at bp 1139; substitution of 'g' for 'c' in wild-type at bp 6242.

BASE COUNT    2599 a   1377 c   1511 g   2419 t
ORIGIN      83 bp upstream from beginning of E6 cds.
```

HPV16R

```
   1 actacaataa ttcatgTATA AAActaaggg cgtaACCGAA ATCGGTtgaA CCGAAACCGG
                 signal ->                      -> E2 bind      -> E2 bind
  61 TTAGTATAAA Agcagacatt ttatgcacca aaagagaact gcaATGtttc aggacccaca
 ori <-/                                               E6 cds ->
   signal ->                                          |-> mRNA start site from
         <- LCR                                       P97 promoter
 E6 orf start->
 121 ggagcgACCC AGAAAGTTac cacagttatg cacagagctg caaacaacta tacatgatat
             -> E2 bind
 181 aatattagaa tgtgtgtact gcaagcaaca gttactgcga cgtgagGTat atgactttgc
                                                 5' sj /\ (E6*transcripts)
 241 ttttcgggat ttatgcatag tatatagaga tgggaatcca tatgctgtat gtgataaatg
 301 tttaaagttt tattctaaaa ttagtgagta tagacattat tgttatagtt tgtatggaac
 361 aacattagaa cagcaataca acaaaccgtt gtgtgatttg ttaattAGgt gtattaactg
                                                           /\ 3' sj (E6*I)
 421 tcaaaagcca ctgtgtcctg aagaaaagca aagacatctg gacaaaaagc aaagattcca
 481 taatataagg ggtcggtgga ccggtcgatg tatgtcttgt tgcAGatcat caagaacacg
                                                           /\ 3' sj (E6*II)
 541 TAGagaaacc cagctgTAAt cATGcatgga gatacaccta cattgcatga atatatgtta
 E7 orf start ->       E6 end <-   -> E7 cds
 601 gatttgcaac cagagacaac tgatctctac tgttatgagc aattaaatga cagctcagag
 661 gaggaggatg aaatagatgg tccagctgga caagcagaac cggacagagc ccattacaat
 721 attgtaacct tttgttgcaa gtgtgactct acgcttcggt tgtgcgtaca agcacacac
 781 gtagacattc gtactttgga agacctgtta atgggcacac taggaattgt gtgccccatc
 841 tgttctcaga aaccaTAAtc taccATGgct gatcctgcag GTaccaatgg ggaagagggt
          E1 orf start ->     -> E1 cds  5' sj /\
                      <- E7 end
 901 acgggatgta atggatggtt ttatgtagag gctgtagtgg aaaaaaaaac aggggatgct
 961 atatcagatg acgagaacga aaatgacagt gatacaggtg aagatttggt agattttata
1021 gtaaatgata atgattattt aacacaggca gaaacagaga cagcacatgc gttgtttact
1081 gcacaggaag caaaacaaca tagagatgca gtacaggttc taaaacgaaa gtatttggGt
             additional 'g' in wild-type absent from original clone ^
1141 agtccactta gtgatattag tggatgtgta gacaataata ttagtcctag attaaaagct
1201 atatgtatag aaaaacaaag tagagctgca aaaggagat tatttgaaag cgaagacagc
1261 gggtatggca atactgaagt ggaaactcag cagatgttac agGTagaagg cgccatgag
                                                 5' sj /\
1321 actgaaacac catgtagtca gtatagtggt ggaagtgggg gtggttgcag tcagtacagt
1381 agtggaagtg ggggagaggg tgttagtgaa agacacacta tatgccaaac accacttaca
1441 aatattttaa atgtactaaa aactagtaat gcaaaggcag caatgttagc aaaatttaaa
1501 gagttatacg gggtgagttt ttcagaatta gtaagaccat ttaaaagtaa taaatcaacg
1561 tgttgcgatt ggtgtattgc tgcatttgga cttacaccca gtatagctga cagtataaaa
1621 acactattac aacaatattg tttatattta cacattcaaa gtttagcatg ttcatgggga
1681 atggttgtgt tactattagt aagatataaa tgtggaaaaa atagagaaac aattgaaaaa
1741 ttgctgtcta aactattatg tgtgtctcca atgtgtatga tgatagagcc tccaaaattg
1801 cgtagtacag cagcagcatt atattggtat aaaacaggta tcaaaatat tagtgaagtg
1861 tatggagaca cgccagaatg gatacaaaga caaacagtat tacaacatag ttttaatgat
1921 tgtacatttg aattatcaca gatggtacaa tgggcctacg ataatgacat agtagacgat
1981 agtgaaattg catataaata tgcacaattg gcagacacta tagtaatgc aagtgccttt
2041 ctaaaaagta attcacaggc aaaaattgta aaggattgtg caacaatgtg tagacattat
2101 aaacgagcag aaaaaaaaca aatgagtatg agtcaatgga taaaatatag atgtgatagg
2161 gtagatgatg gaggtgattg gaagcaaatt gttatgtttt taaggtatca aggtgtagag
2221 tttatgtcat ttttaactgc attaaaaaga ttttttgcaag gcatacctaa aaaaaattgc
2281 atattactat atggtgcagc taacacaggt aaatcattat ttggtatgag tttaatgaaa
2341 tttctgcaag ggtctgtaat atgtttttgta aattctaaaa gccattttg gttacaacca
2401 ttagcagatg ccaaaatagg tatgttagat gatgctacag tgccctgttg gaactacata
2461 gatgacaatt taagaaatgc attggatgga aatttagttt ctatggatgt aaagcataga
2521 ccattggtac aactaaaatg ccctccatta ttaattacat ctaacattaa tgctggtacA
```

```
2581 Gattctaggt ggccttattt acataataga ttggtggtgt ttacatttcc taatgagttt
         /\ 3' sj
2641 ccatttgacg aaaacggaaa tccagtgtat gagcttaatg ataagaactg gaaatccttt
2701 ttctcaagGA cgtggtccag atTAAgtttg cacgaggacg aggacaagga aaacgATGga
            /\ 3' sj (E2,E5)
                       E2 orf start ->                        E2 cds ->
2761 gactctttgc caacgtttaa atgtgtgtca ggacaaaata ctaacacatt aTGAaaatga
                                                                <- E1 end
2821 tagtacagac ctacgtgacc atatagacta ttggaaacac atgcgcctag aatgtgctat
2881 ttattacaag gccagagaaa tgggatttaa acatattaac caccaagtgg tgccaacact
2941 ggctgtatca aagaataaag cattacaagc aattgaactg caactaacgt tagaaacaat
3001 atataactca caatatagta atgaaaagtg gacattacaa gacgttagcc ttgaagtgta
3061 tttaactgca ccaacaggat gtataaaaaa acatggatat acagtggaag tgcagtttga
3121 tggagacata tgcaatacaa tgcattatac aaactggaca catatatata tttgtgaaga
3181 agcatcagta actgtggtag agggtcaagt tgactattat ggtttatatt atgttcatga
3241 aggaatacga acatattttg tgcagtttaa agatgatgca gaaaaatata gtaaaaataa
3301 agtatgggaa gttcatgcgg gtggtcaggT AAtattatgt cctacatctg tgtttAGcag
                       E4 orf start ->                        /\ 3' sj
                       NH2 terminus unknown
3361 caacgaagta tcctctcctg aaattattag gcagcacttg gccaaccacc ccgccgcgac
3421 ccataccaaa gccgtcgcct tgggcaccga agaaacacag acgactatcc agcgaccaag
3481 atcagagcca gacaccggaa acccctgcca caccactaag ttgttgcaca gagactcagt
3541 ggacagtgct ccaatcctca ctgcatttaa cagctcacac aaaggacgga ttaactgtaa
3601 tagtaacact acacccaTAG tacatttaaa agGTgatgct aatactttaa aatgtttaag
                    <- E4    5' sj /\
                            end
3661 atatagattt aaaaagcatt gtacattgta tactgcagtg tcgtctacat ggcattggac
3721 aggacataat gtaaaacata aaagtgcaat tgttacactt acatatgata gtgaatggca
3781 acgtgaccaa tttttgtctc aagtTAAaat accaaaaact attacagtgt ctactggatt
                    E5 orf start ->
                    NH2 terminus unknown
3841 tatgtctatA TGAcaaatct tgatactgca tccacaacat tactggcgtg ctttttgctt
            -> E5 cds
                  <- E2 end
3901 tgcttttTgtg tgcttttgtg tgtctgccta ttaatacgtc cgctgctttt gtctgtgtct
            âdditional 't'
3961 acatacacat cattaataat attggtatta ctattgtgga taacagcagc ctctgcgttt
4021 aggtgtttta ttgtatatat tatatttgtt tatataccat tattttaat acatacacat
4081 gcacgctttt taattacaTA Atgtatatgt acataatgta attgttacat aTAAttgttg
                       <- E5 end               L2 orf start ->
4141 tataccataa cttactattt tttcttttt attttcatat ataatttttt ttttgtttg
4201 tttgtttgtt ttttAATAAA ctgttattac ttaacaATGc gacacaaacg ttctgcaaaa
early poly-A signal ->                 L2 cds ->
4261 cgcacaaaac gtgcatcggc tacccaactt TATAAAAcat gcaaacaggc aggtacatgt
                         signal ->
4321 ccacctgaca ttatacctaa ggttgaaggc aaaactattg ctgaTcaaat attacaatat
                                          ^ 'a' replaced by 't'
4381 ggaagtatgg gtgtattttt tggtgggtta ggaattggaa cagggtcggg tacaggcgga
4441 cgcactgggt atattccatt gggaacaagg cctcccacag ctacagatac acttgctcct
4501 gtaagacccc ctttaacagt agatcctgtg ggcccttctg atccttctat agtttcttta
4561 gtggaagaaa ctagtttat tgatgctggt gcaccaacat ctgtaccttc cattcccca
4621 gatgtatcag gatttagtat tactacttca actgatacca cacctgctat attagatatt
4681 aataatactg ttactactgt tactacacat aataatccca ctttcactga cccatctgta
4741 ttgcagcctc caacacctgc agaaactgga gggcatttta cactttcatc atccactatt
4801 agtacacata attatgaaga aattcctatg gatacattta ttgttagcac aaaccctaac
4861 acagtaacta gtagcacacc cataccaggg tctcgcccag tggcacgcct aggattatat
4921 agtcgcacaa cacaacaggt aaagttgta gaccctgctt ttgtaaccac tcccactaaa
4981 cttattacat atgataatcc tgcatatgaa ggtatagatg tggataatac attatatttt
5041 tctagtaatg ataatagtat taatatagct ccagatcctg acttttttgga tatagttgct
```

```
5101 ttacataggc cagcattaac ctctaggcgt actggcatta ggtacagtag aattggtaat
5161 aaacaaacac tacgtactcg tagtggaaaa tctataggtg ctaaggtaca ttattattat
5221 gatttaagta ctattgatcc tgcagaagaa atagaattac aaactataac accttctaca
5281 tatactacca cttcacatgc agcctcacct acttctatta ataatggatt atatgatatt
5341 tatgcagatg actttattac agatacttct caaccccgg taccatctgt accctctaca
5401 tctttatcag gttatattcc tgcaaataca acaattcctt ttggtggtgc atacaatatt
5461 ccttttagtat caggtcctga tatacccatt aatataactg accaagctcc ttcattaatt
5521 cctaTAGttc cagggtctcc acaatataca attattgctg ATGcaggtga ctttttattta
     L1 orf start ->                                   L1 cds ->
5581 catcctagtt attacatgtt acgaaaacga cgtaaacgtt taccatattt tttttcAGat
                                                                /\ 3' sj
5641 gtctctttgg ctgccTAGtg aggccactgt ctacttgcct cctgtcccag tatctaaggt
                         <- L2 end
5701 tgtaagcacg gatgaatatg ttgcacgcac aaacatatat tatcatgcag gaacatccag
5761 actacttgca gttggacatc cctattttcc tattaaaaaa cctaacaata acaaaatatt
5821 agttcctaaa gtatcaggat tacaatacag ggtatttaga atacatttac ctgaccccaa
5881 taagtttggt tttcctgaca cctcatttta taatccagat acacagcggc tggtttgggc
5941 ctgtgtaggt gttgaggtag gtcgtggtca gccattaggt gtgggcatta gtggccatcc
6001 tttattaaat aaaattggatg acacagaaaa tgctagtgct tatgcagcaa atgcaggtgt
6061 ggataataga gaatgtatat ctatggatta caaacaaaca caattgtgtt taattggttg
6121 caaaccacct ataggggaac actggggcaa aggatcccca tgtaccaatg ttgcagtaaa
6181 tccaggtgat tgtccACCAT TAGAGTTaat aaacacagtt attcaggatg gtgatatggt
                        -> E2 bind
6241 tGatactggc tttggtgcta tggactttac tacattacag gctaacaaaa gtgaagttcc
        ^ wild-type has 'g', original clone had 'c'
6301 actggatatt tgtacatcta tttgcaaata tccagattat attaaaatgg tgtcagaacc
6361 atatggcgac agcttatttt tttatttacg aagggaacaa atgtttgtta gacatttatt
6421 taatagggct ggtactgttg gtgaaaatgt accagacgat ttatacatta aaggctctgg
6481 gtctactgca aatttagcca gttcaaatta ttttcctaca cctagtggtt ctatggttac
6541 ctctgatgcc caaatattca ataaACCTTA TTGGTTacaa cgagcacagg gccacaataa
                                  -> E2 bind
6601 tggcatttgt tggggtaacc aactatttgt tactgttgtt gatactacac gcagtacaaa
6661 tatgtcatta tgtgctgcca tatctacttc agaaactaca TATAAAata ctaactttaa
                                                 signal ->
6721 ggagtaccta cgacatgggg aggaatatga tttacagttt attttttcaac tgtgcaaaat
6781 aaccttaact gcagacgtta tgacatacat acattctatg aattccacta ttttggagga
6841 ctggaatttt ggtctacaac ctcccccagg aggcacacta gaagatactt ataggtttgt
6901 aaCATcccag gcaattgctt gtcaaaaaca tacacctcca gcacctaaag aagatcccct
        ^ ^ 'cat' added                                      /\ 'gat' deleted
6961 taaaaaatac actttttggg aagtaaattt aaaggaaaag ttttctgcag acctagatca
7021 gtttcccttta ggacgcaaat ttttactaca agcaggattg aaggccaaac caaaatttac
7081 attaggaaaa cgaaaagcta cacccaccac ctcatctacc tctacaactg ctaaacgcaa
                                                      /-> inhibitory element
                            nuclear localization signal (NLS) ->
7141 aaaacgtaag ctgTAAgtat tgtatgtatg ttgaattagt gttgtttgtt gtgtatatgt
         NLS  <-       <- L1 end
                      -> LCR
7201 ttgtatgtgc ttgtatgtgc ttgtaaatat taagttgtat gtgtgtttgt atgtatggta
            inhibitory element <-/
7261 taataaacac gtgtgtatgt gtttttaaat gcttgtgtaa ctattgtgtc atgcaacata
7321 AATAAActta ttgtttcaac acctactaat tgtgttgtgg ttattcattg tatataaact
poly-A ->
7381 atatttgcta catcctgttt ttgtttttata tatactatat tttgtagcgc cagCGGccat
                                              'gc' replaced by 'cgg' ^ ^
7441 tttgtagctt caACCGAATT CGGTtgcatg cttttttggca caaaatgtgt tttttttaaat
                     -> E2 bind
7501 agttctatgt cagcaactat ggtttaaact tgtacgtttc ctgcttgcca tgcgtgccaa
7561 atccctgttt tcctgacctg cactgcttgc caaccattcc attgtttttt acactgcact
```

```
7621 atgtgcaact actgaatcac tatgtacatt gtgtcaTATA AAAtaaatca ctatgcgcca
                                       signal ->
7681 acgccttaca taccgctgtt aggcacatat ttttggcttg ttttaactaa cctaattgca
7741 tatttggcat aaggtttaaa ctTCTAAGGc caactaaatg tcaccctagt tcatacatga
                        ^ ^ ^ ^ sequence of [1] correct
7801 actgtgtaaa ggttagtcat acattgttca tttgtaaaac tgcacatggg tgtgtgcaaa
                                                              /-> ori
7861 ccgttttggg ttacacattt acaagcaact taTATAATAa tactaa
        /\ 'a' deleted           signal ->
```