Human papillomavirus type 33, complete genome

GenBank: M12732.1

FASTA Graphics

Go to:

```
LOCUS       PPH33CG              7909 bp    DNA     circular VRL 21-MAR-1994
DEFINITION  Human papillomavirus type 33, complete genome.
ACCESSION   M12732
VERSION     M12732.1
KEYWORDS    complete genome.
SOURCE      human papillomavirus 33
  ORGANISM  human papillomavirus 33
            Viruses; Monodnaviria; Shotokuvirae; Cossaviricota;
            Papovaviricetes; Zurhausenvirales; Papillomaviridae;
            Firstpapillomavirinae; Alphapapillomavirus.
REFERENCE   1  (bases 1 to 7909)
  AUTHORS   Cole,S.T. and Streeck,R.E.
  TITLE     Genome organization and nucleotide sequence of human
papillomavirus
            type 33, which is associated with cervical cancer
  JOURNAL   J. Virol. 58 (3), 991-995 (1986)
   PUBMED   3009902
COMMENT     Original source text: Human papillomavirus type 33 DNA isolated
            from a human invasive cervical carcinoma, clone p15-5.
            Draft entry and computer-readable copy of the sequence were kindly
            provided by S.T.Cole, 04-AUG-1986.
FEATURES             Location/Qualifiers
     source          1..7909
                     /organism="human papillomavirus 33"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10586"
     regulatory      7..10
                     /regulatory_class="TATA_box"
                     /note="putative"
     protein_bind    39..50
                     /note="putative"
                     /bound_moiety="E2"
                     /function="gene transcription"
     protein_bind    54..65
                     /note="putative"
                     /bound_moiety="E2"
                     /function="gene transcription"
     regulatory      71..75
                     /regulatory_class="TATA_box"
                     /note="putative"
```

gene 109..558
/gene="E6"
CDS 109..558
/gene="E6"
/note="E6 ORF from 76 to 558; putative"
/codon_start=1
/product="transforming protein"
/protein_id="AAA46958.1"
/translation="MFQDTEEKPRTLHDLCQALETTIHNIELQCVECKKPLQRSEVYD
FAFADLTVVYREGNPFGICKLCLRFLSKISEYRHYNYSVYGNTLEQTVKKPLNEILIR
CIICQRPLCPQEKKRHVDLNKRFHNISGRWAGRCAACWRSRRRETAL"
gene 573..866
/gene="E7"
CDS 573..866
/gene="E7"
/note="E7 ORF from bp 543 to 866; putative"
/codon_start=1
/product="transforming protein"
/protein_id="AAA46959.1"
/translation="MRGHKPTLKEYVLDLYPEPTDLYCYEQLSDSSDEDEGLDRPDGQ
AQPATADYYIVTCCHTCNTTVRLCVNSTASDLRTIQQLLMGTVNIVCPTCAQQ"
gene 879..2813
/gene="E1"
CDS 879..2813
/gene="E1"
/note="E1 ORF from bp 867 to 2813; putative"
/codon_start=1
/product="replication protein"
/protein_id="AAA46960.1"
/translation="MADPEGTNGAGMGCTGWFEVEAVIERRTGDNISEDEDETADDSG
TDLLEFIDDSMENSIQADTEAARALFNIQEGEDDLNAVCALKRKFAACSQSAAEDVVD
RAANPCRTSINKNKECTYRKRKIDELEDSGYGNTEVETQQMVQQVESQNGDTNLNDLE
SSGVGDDSEVSCETNVDSCENVTLQEISNVLHSSNTKANILYKFKEAYGISFMELVRP
FKSDKTSCTDWCITGYGISPSVAESLKVLIKQHSLYTHLQCLTCDRGIIILLLIRFRC
SKNRLTVAKLMSNLLSIPETCMVIEPPKLRSQTCALYWFRTAMSNISDVQGTTPEWID
RLTVLQHSFNDNIFDLSEMVQWAYDNELTDDSDIAYYYAQLADSNSNAAAFLKSNSQA
KIVKDCGIMCRHYKKAEKRKMSIGQWIQSRCEKTNDGGNWRPIVQLLRYQNIEFTAFL
GAFKKFLKGIPKKSCMLICGPANTGKSYFGMSLIQFLKGCVISCVNSKSHFWLQPLSD
AKIGMIDDVTPISWTYIDDYMRNALDGNEISIDVKHRALVQLKCPPLLLTSNTNAGTD
SRWPYLHSRLTVFEFKNPFPFDENGNPVYAINDENWKSFFSRTWCKLDLIEEEDKENH
GGNISTFKCSAGENTRSLRS"
gene 2749..3810
/gene="E2"
CDS 2749..3810

/gene="E2"
/note="E2 ORF from bp 2728 to 3810; putative"
/codon_start=1
/product="regulatory protein"
/protein_id="AAA46961.1"
/translation="MEEISARLNAVQEKILDLYEADKTDLPSQIEHWKLIRMECALLY
TAKQMGFSHLCHQVVPSLLASKTKAFQVIELQMALETLSKSQYSTSQWTLQQTSLEVW
LCEPPKCFKKQGETVTVQYDNDKKNTMDYTNWGEIYIIEEDTCTMVTGKVDYIGMYYI
HNCEKVYFKYFKEDAAKYSKTQMWEVHVGGQVIVCPTSISSNQISTTETADIQTDNDN
RPPQAAAKRRRPADTTDTAQPLTKLFCADPALDNRTARTATNCTNKQRTVCSSNVAPI
VHLKGESNSLKCLRYRLKPYKELYSSMSSTWHWTSDNKNSKNGIVTVTFVTEQQQQMF
LGTVKIPPTVQISTGFMTL"

gene 3326..3577
/gene="E4"

CDS <3326..3577
/gene="E4"
/note="E4 ORF from bp 3326 to 3577; putative"
/codon_start=1
/protein_id="AAA46957.1"
/translation="LFVLRLYLATKYPLLKLLTYRQTTITDHHKQRPNDDDLQTPQTP
PSPLQSCSVQTPPWTIEQHVLQLTAQTSSGLCVVLTLHL"

gene 3854..4081
/gene="E5"

CDS 3854..4081
/gene="E5"
/note="E5 ORF from bp 3842 to 4081; putative"
/codon_start=1
/protein_id="AAA46962.1"
/translation="MIFVFVLCFILFLCLSLLLRPLILSISTYAWLLVLVLLLWVFVG
SPLKIFFCYLLFLYLPMMCINFHAQHMTQQE"

gene 4210..5613
/gene="L2"

CDS 4210..5613
/gene="L2"
/note="L2 ORF from bp 4198 to 5613; putative"
/codon_start=1
/product="minor capsid protein"
/protein_id="AAA46963.1"
/translation="MRHKRSTRRKRASATQLYQTCKATGTCPPDVIPKVEGSTIADQI
LKYGSLGVFFGGLGIGTGSGSGGRTGYVPIGTDPPTAAIPLQPIRPPVTVDTVGPLDS
SIVSLIEETSFIEAGAPAPSIPTPSGFDVTTSADTTPAIINVSSVGESSIQTISTHLN
PTFTEPSVLHPPAPAEASGHFIFSSPTVSTQSYENIPMDTFVVSTDSSNVTSSTPIPG
SRPVARLGLYSRNTQQVKVVDPAFLTSPHKLITYDNPAFESFDPEDTLQFQHSDISPA
PDPDFLDIIALHRPAITSRRHTVRFSRVGQKATLKTRSGKQIGARIHYYQDLSPIVPL

```
                     DHTVPNEQYELQPLHDTSTSSYSINDGLYDVYADDVDNVHTPMQHSYSTFATTRTSNV
                     SIPLNTGFDTPVMSGPDIPSPLFPTSSPFVPISPFFPFDTIVVDGADFVLHPSYFILR
                     RRRKRFPYFFTDVRVAA"
     gene            5594..7093
                     /gene="L1"
     CDS             5594..7093
                     /gene="L1"
                     /note="L1 ORF from bp 5516 to 7093; putative"
                     /codon_start=1
                     /product="major capsid protein"
                     /protein_id="AAA46964.1"
                     /translation="MSVWRPSEATVYLPPVPVSKVVSTDEYVSRTSIYYYAGSSRLLA
                     VGHPYFSIKNPTNAKKLLVPKVSGLQYRVFRVRLPDPNKFGFPDTSFYNPDTQRLVWA
                     CVGLEIGRGQPLGVGISGHPLLNKFDDTETGNKYPGQPGADNRECLSMDYKQTQLCLL
                     GCKPPTGEHWGKGVACTNAAPANDCPPLELINTIIEDGDMVDTGFGCMDFKTLQANKS
                     DVPIDICGSTCKYPDYLKMTSEPYGDSLFFFLRREQMFVRHFFNRAGTLGEAVPDDLY
                     IKGSGTTASIQSSAFFPTPSGSMVTSESQLFNKPYWLQRAQGHNNGICWGNQVFVTVV
                     DTTRSTNMTLCTQVTSDSTYKNENFKEYIRHVEEYDLQFVFQLCKVTLTAEVMTYIHA
                     MNPDILEDWQFGLTPPPSASLQDTYRFVTSQAITCQKTVPPKEKEDPLGKYTFWEVDL
                     KEKFSADLDQFPLGRKFLLQAGLKAKPKLKRAAPTSTRTSSAKRKKVKK"
     repeat_region   7098..7323
                     /note="putative"
     protein_bind    7502..7513
                     /note="putative"
                     /bound_moiety="E2"
                     /function="gene transcription"
     repeat_region   7517..7574
     misc_structure  7520..7530
                     /standard_name="Z DNA structure"
                     /note="putative"
     misc_structure  7599..7609
                     /standard_name="Z DNA structure"
                     /note="putative"
     regulatory      7891..7894
                     /regulatory_class="TATA_box"
                     /note="putative"
     regulatory      7901..7904
                     /regulatory_class="TATA_box"
                     /note="putative"
ORIGIN
        1 gtaaactata atgccaagtt ttaaaaaagt agggtgtaac cgaaagcggt tcaaccgaaa
       61 acggtgcata tataaagcaa acattttgca gtaaggtact gcacgactat gtttcaagac
      121 actgaggaaa aaccacgaac attgcatgat ttgtgccaag cattggagac aactatacac
      181 aacattgaac tacagtgcgt ggaatgcaaa aaacctttgc aacgatctga ggtatatgat
```

241 tttgcatttg cagatttaac agttgtatat agagagggaa atccatttgg aatatgtaaa
301 ctgtgtttgc ggttcttatc taaaattagt gaatatagac attataatta ttctgtatat
361 ggaaatacat tagaacaaac agttaaaaaa ccttttaaatg aaatattaat taggtgtatt
421 atatgtcaaa gacctttgtg tcctcaagaa aaaaaacgac atgtggattt aaacaaacga
481 tttcataata tttcgggtcg ttgggcaggg cgctgtgcgg cgtgttggag gtcccgacgt
541 agagaaactg cactgtgacg tgtaaaaacg ccatgagagg acacaagcca acgttaaagg
601 aatatgtttt agatttatat cctgaaccaa ctgacctata ctgctatgag caattaagtg
661 acagctcaga tgaggatgaa ggcttggacc ggccagatgg acaagcacaa ccagccacag
721 ctgattacta cattgtaacc tgttgtcaca cttgtaacac cacagttcgt ttatgtgtca
781 acagtacagc aagtgaccta cgaaccatac agcaactact tatgggcaca gtgaatattg
841 tgtgccctac ctgtgcacaa caataaacat catctacaat ggccgatcct gaaggtacaa
901 atggggctgg gatggggtgt actggttggt ttgaggtaga agcagtcata gagagaagaa
961 caggagataa tatttcagaa gatgaggatg aaacagcaga tgacagtggc acggatttac
1021 tagagtttat agatgattct atggaaaata gtatacaggc agacacagag gcagcccggg
1081 cattgtttaa tatacaggaa ggggaggatg atttaaatgc tgtgtgtgca ctaaaacgaa
1141 agtttgccgc atgttcacaa agtgctgcgg aggacgttgt tgatcgtgct gcaaacccgt
1201 gtagaacgtc tattaataaa aataaagaat gcacatacag aaaacgaaaa atagatgagc
1261 tagaagacag cggatatggc aatactgaag tggaaactca gcagatggta caacaggtag
1321 aaagtcaaaa tggcgacaca aacttaaatg acttagaatc tagtggggtg ggggatgatt
1381 cagaagtaag ctgtgagaca aatgtagata gctgtgaaaa tgttacgttg caggaaatta
1441 gtaatgttct acatagtagt aatacaaaag caaatatatt atataaattt aaagaggcct
1501 atggaataag ttttatggaa ttagtaagac catttaaaag tgataaaaca agctgtacag
1561 attggtgtat aacaggatat ggaattagtc catcagtagc agaaagttta aaagtattaa
1621 ttaaacagca tagtttgtat actcatttac aatgtttaac ttgcgataga ggaataataa
1681 tattattgtt aattagattt aggtgtagca aaaacaggtt aacagtagca aaactaatga
1741 gtaatttatt atcaatacct gaaacatgta tggttataga gccaccaaaa ttacggagcc
1801 aaacatgtgc attgtattgg tttagaacag caatgtcaaa cattagtgat gtacaaggta
1861 caacacctga atggatagat agactaactg ttttacaaca tagctttaat gataatatat
1921 ttgatttaag tgaaatggta cagtgggcat atgataacga gttaacggac gatagtgaca
1981 ttgcatatta ttatgcacaa cttgcagatt caaatagtaa tgctgctgca tttttaaaaa
2041 gtaactcaca agcaaaaata gtaaaggact gtggaataat gtgtagacat tataaaaaag
2101 cagaaaaacg taaaatgtca ataggacaat ggatacaaag tagatgtgaa aaaacaaatg
2161 atggaggaaa ttggagacca atagtacagt tgttaagata tcaaaacatt gaatttacag
2221 catttttagg tgcatttaaa aagtttttaa aaggtatacc aaaaaaaagc tgtatgctaa
2281 tttgtggacc agcaaataca ggaaagtcat attttggaat gagtttaata cagtttttaa
2341 aagggtgtgt tatatcatgt gtaaattcta aaagtcactt ttggttgcag ccattatcag
2401 atgcaaaaat aggaatgata gatgatgtaa cgccaataag ttggacatat atagatgatt
2461 acatgagaaa tgcgttagat ggaaatgaaa tttcaataga tgtgaaacat agggcattag
2521 tgcaattaaa atgtccacca ctgcttctta cctcaaatac aaatgcaggc acagactcta
2581 gatggccata tttacatagt agattaacag tatttgaatt taaaaatcca ttcccatttg
2641 atgaaaatgg taacccagtg tatgcaataa atgatgaaaa ttggaaatcc tttttctcaa
2701 ggacgtggtg caaattagat ttaatagagg aagaggacaa ggaaaaccat ggaggaaata
2761 tcagcacgtt taaatgcagt gcaggagaaa atactagatc tttacgaagc tgataaaact
2821 gatttaccat cacaaattga acattggaaa ctgatacgca tggagtgtgc tttattgtat

2881 acagccaaac aaatgggatt ttcacattta tgccaccagg tggtgccttc tttgttagca
2941 tcaaagacca aagcatttca agtaattgaa ctacaaatgg cattagagac attaagtaaa
3001 tcacagtata gtacaagcca atggacattg caacaaacaa gcttagaggt gtggctttgt
3061 gaaccaccaa aatgttttaa aaaacaagga gaaacagtaa ctgtgcaata tgacaatgac
3121 aaaaaaaata caatggatta tacaaactgg ggtgaaatat atattataga ggaagataca
3181 tgtactatgg ttacagggaa agtagattat ataggtatgt attatataca taactgtgaa
3241 aaggtatatt ttaaatattt taaagaggat gctgcaaagt attctaaaac acaaatgtgg
3301 gaagtacatg tgggtggtca ggtaattgtt tgtcctacgt ctatatctag caaccaaata
3361 tccactactg aaactgctga catacagaca gacaacgata accgaccacc acaagcagcg
3421 gccaaacgac gacgacctgc agacaccaca gacaccgccc agccccttac aaagctgttc
3481 tgtgcagacc ccgccttgga caatagaaca gcacgtactg caactaactg cacaaacaag
3541 cagcggactg tgtgtagttc taacgttgca cctatagtgc atttaaaagg tgaatcaaat
3601 agtttaaaat gtttaagata cagattaaaa ccttataaag agttgtatag ttctatgtca
3661 tccacctggc attggaccag tgacaacaaa aatagtaaaa atggaattgt aactgtaaca
3721 tttgtaactg aacagcaaca acaaatgttt ttaggtaccg taaaaatacc acctactgtg
3781 caaataagta ctggatttat gacattataa gtgtacatca caagccaata tgtgctgcta
3841 attgtatata accatgatat ttgtttttgt attatgtttt atattgtttt tatgcttatc
3901 cttattatta cgtcctttaa tactttccat ttctacctat gcttggttgc tggtgttggt
3961 attgctgctt tgggtgtttg tgggatctcc tttaaaaatt tttttttgct atttgttgtt
4021 tttatattta ccaatgatgt gtattaattt tcatgcacag catatgacac aacaagagta
4081 atgtatatac atgtatatat tgtttgtata tatgtgcaca tggtggtgtt ttaacattgt
4141 tgttgttatt ttagtttttt ttttttgta ttactaataa ataccttat attttagcag
4201 tgtattatta tgagacacaa acgatctaca aggcgcaagc gtgcatctgc aacacaacta
4261 taccaaacat gcaaggccac aggcacctgc ccacccgatg ttattcctaa agtggaagga
4321 agtaccatag cagatcaaat tcttaaatat ggcagtttag gggtttttt tggtggttta
4381 ggtattggca caggctctgg ttcaggtgga aggactggct atgtacctat tggtactgac
4441 ccacctacag ctgcaatccc cttgcagcct atacgtcctc cggttactgt agacactgtt
4501 ggacctttag actcgtctat agtgtcatta atagaagaaa caagttttat agaggcaggt
4561 gcaccagccc catctattcc tacaccatca ggttttgatg ttactacatc tgcagatact
4621 acacctgcaa ttattaatgt ttcatctgtt ggggagtcat ctattcaaac tatttctaca
4681 catttaaatc ccacatttac tgaaccatct gtactacacc ctccagcgcc tgcagaagcc
4741 tctggacatt ttatattttc ttcccctact gttagcacac aaagttatga aaacatacca
4801 atggatacct ttgttgtttc cacagacagt agtaatgtaa catcaagcac gcccattcca
4861 gggtctcgcc ctgtggcacg ccttggttta tatagtcgca atacccaaca ggttaaggtt
4921 gttgaccctg ctttttaac atcgcctcat aaacttataa catatgataa tcctgcattt
4981 gaaagctttg accctgaaga cacattacaa tttcaacata gtgatatatc acctgctcct
5041 gatcctgact ttctagatat tattgcatta cataggcctg ctattacatc tcgtagacat
5101 actgtgcgtt ttagtagagt aggtcaaaaa gccacactta aaactcgcag tggtaaacaa
5161 attggagcta gaatacatta ttatcaggat ttaagtccta ttgtgccttt agaccacacc
5221 gtgccaaatg aacaatatga attacagcct ttacatgata cttctacatc gtcttatagt
5281 attaatgatg gtttgtatga tgtttatgct gacgatgtgg ataatgtaca caccccaatg
5341 caacactcat acagtacgtt tgcaacaaca cgtaccagca atgtgtctat acctttaaat
5401 acaggatttg atactcctgt tatgtctggc cctgatatac cttccccttt atttcccaca
5461 tctagcccat ttgttcctat ttcgcctttt tttccttttg acaccattgt tgtagacggt

```
5521 gctgactttg ttttacatcc tagttatttt attttacgtc gcaggcgtaa acgttttcca
5581 tatttttta cagatgtccg tgtggcggcc tagtgaggcc acagtgtacc tgcctcctgt
5641 acctgtatct aaagttgtca gcactgatga atatgtgtct cgcacaagca tttattatta
5701 tgctggtagt tccagacttc ttgctgttgg ccatccatat ttttctatta aaaatcctac
5761 taacgctaaa aaattattgg tacccaaagt atcaggcttg caatataggg tttttagggt
5821 ccgtttacca gatcctaata aatttggatt tcctgacacc tccttttata accctgatac
5881 acaacgatta gtatgggcat gtgtaggcct tgaaataggt agagggcagc cattaggcgt
5941 tggcataagt ggtcatcctt tattaaacaa atttgatgac actgaaaccg gtaacaagta
6001 tcctggacaa ccgggtgctg ataataggga atgtttatcc atggattata aacaaacaca
6061 gttatgttta cttggatgta agcctccaac aggggaacat tggggtaaag gtgttgcttg
6121 tactaatgca gcacctgcca atgattgtcc acctttagaa cttataaata ctattattga
6181 ggatggtgat atggtggaca caggatttgg ttgcatggat tttaaaacat tgcaggctaa
6241 taaaagtgat gttcctattg atatttgtgg cagtacatgc aaatatccag attatttaaa
6301 aatgactagt gagccttatg gtgatagttt atttttcttt cttcgacgtg aacaaatgtt
6361 tgtaagacac ttttttaata gggctggtac attaggagag gctgttcccg atgacctgta
6421 cattaaaggt tcaggaacta ctgcctctat tcaaagcagt gctttttttc ccactcctag
6481 tggatcaatg gttacttccg aatctcagtt atttaataag ccatattggc tacaacgtgc
6541 acaaggtcat aataatggta tttgttgggg caatcaggta tttgttactg tggtagatac
6601 cactcgcagt actaatatga ctttatgcac acaagtaact agtgacagta catataaaaa
6661 tgaaaatttt aaagaatata taagacatgt tgaagaatat gatctacagt ttgtttttca
6721 actatgcaaa gttaccttaa ctgcagaagt tatgacatat attcatgcta tgaatccaga
6781 tattttagaa gattggcaat ttggtttaac acctcctcca tctgctagtt tacaggatac
6841 ctataggttt gttacctctc aggctattac gtgtcaaaaa acagtacctc caaaggaaaa
6901 ggaagacccc ttaggtaaat atacattttg ggaagtggat ttaaaggaaa aattttcagc
6961 agatttagat cagtttcctt tgggacgcaa gtttttatta caggcaggtc ttaaagcaaa
7021 acctaaactt aaacgtgcag cccccacatc cacccgcaca tcgtctgcaa aacgcaaaaa
7081 ggttaaaaaa taacactttg tgtaattgtg ttatgttgtt gttttgttct gtctatgtac
7141 tttgtgttgt tgtgttgtgt tgttgtttgt tttttgtgta tgtgttacaa tgtatgttat
7201 gttgtatgtt actgtgtttg ttttatgtgt acttgtttgt gtgcatgttc tatgtacttg
7261 tcagtttcct gtttgtgtat atgttaataa aacattgtgt gtatttgtta aactatttgt
7321 atgtatgtta tgtatatggg tgtacctata tgagtaagga gttgtattgc ttgccctacc
7381 ctgcattgca atgtacctac ctttatttcc ctatatttgt agtacctaca tgtttagtat
7441 tgctttacct tttgacatac tagtgtccat attgtacaat ttcctccatt ttgtatgcct
7501 aaccgttttc ggttacttgg catacatacc ctatgacatt ggcagaacag ttaatccttt
7561 tctttcctgc actgtgtttg tctgtacttg ctgcattggc atacataccc tatgacattg
7621 gcagaacagt taatcctttt ctttcctgca ctgtgtttgt ctgtacttgc tgcattgact
7681 catatataca tgcagtgcaa ttgcaaaata cttaattgta ctaatagttt acacatgctt
7741 ttaggcacat atttttactt tactttcaaa ccttaagtgc agttttggct tacacaattg
7801 ctttgtatgc caaactatgc cttgtaaaag tgagtcacta cctgtttatt accaggtgtg
7861 gactaaccgt tttaggtcat attggtcatt tataatcttt tatataata
//
```