

Human papillomavirus type 16 (HPV16), complete genome

GenBank: K02718.1

[FASTA Graphics](#)

[Go to:](#)

LOCUS PPH16 7904 bp DNA circular VRL 18-MAR-1994
DEFINITION Human papillomavirus type 16 (HPV16), complete genome.
ACCESSION K02718
VERSION K02718.1
KEYWORDS circular; complete genome.
SOURCE Human papillomavirus type 16
ORGANISM [Human papillomavirus type 16](#)
Viruses; Monodnaviria; Shotokuvirae; Cossaviricota;
Papovaviricetes; Zurhausenvirales; Papillomaviridae;
Firstpapillomavirinae; Alphapapillomavirus.
REFERENCE 1 (bases 1 to 7904)
AUTHORS Seedorf,K., Krammer,G., Durst,M., Suhai,S. and Rowekamp,W.G.
TITLE Human papillomavirus type 16 DNA sequence
JOURNAL Virology 145 (1), 181-185 (1985)
PUBMED [2990099](#)
REFERENCE 2 (sites)
AUTHORS Kennedy,I.M., Haddow,J.K. and Clements,J.B.
TITLE A negative regulatory element in the human papillomavirus type 16
genome acts at the level of late mRNA stability
JOURNAL J. Virol. 65 (4), 2093-2097 (1991)
PUBMED [1848319](#)
COMMENT Original source text: Papilloma virus type 16 DNA, isolated from
a
human invasive cervical carcinoma.
The sense strand of this double-stranded circular genome is shown,
with a numbering system matching the first 60 bp of HPV16, HPV6b
and BPV1. The annotation of sites and features is solely based
upon
homology comparison with these other papillomaviruses. In addition
to the coding sequences reported below, the authors note open
reading frames which do not start with 'ATG', but which are found
in other papillomaviruses. In particular, a second portion of the
E1 gene may be located out to base 2813 (the E1 protein is thought
to be generally involved in DNA replication).
A potential 'CAT'-box region is found beginning at base 7895
below,
and 'TATA' boxes for early and late transcripts may be located at
17, 65 and 4289. Potential polyadenylation signals are at bases
4213 and 7260.
HPV16, in comparison to HPV types 6 and 11, is more often

associated with malignant genital cancers in humans.

```
FEATURES             Location/Qualifiers
source               1..7904
                   /organism="Human papillomavirus type 16"
                   /mol_type="genomic DNA"
                   /db_xref="taxon:333760"
regulatory          17..23
                   /regulatory_class="TATA_box"
regulatory          65..71
                   /regulatory_class="TATA_box"
gene                83..559
                   /gene="E6"
CDS                 83..559
                   /gene="E6"
                   /note="E6 ORF from 65 to 559; putative"
                   /codon_start=1
                   /product="transforming protein"
                   /protein_id="AAA46939.1"
                   /translation="MHQKRTAMFQDPQERPRKLPQLCTELQTTIHDIILECVYCKQQL
LRREVDFAFRDLCIVYRDGNPYAVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKP
LCDLLIRCINCQKPLCPEEKQRHLDDKKQRFHNIQRWWTGRCMSCCRSSRTRRETQL"
gene                562..858
                   /gene="E7"
CDS                 562..858
                   /gene="E7"
                   /note="E7 ORF from 544 to 858; putative"
                   /codon_start=1
                   /product="transforming protein"
                   /protein_id="AAA46940.1"
                   /translation="MHGDTPTLHEYMLDLQPETTDLYCYEQLNDSSEEDEIDGPAGQ
AEPDRAHYNIVTFCCKCDSTLRLCVQSTHVDIRTLEDLLMGTGLGIVCPICSQKP"
gene                join(865..1140,1140..2813)
                   /gene="E1"
CDS                 join(865..1140,1140..2813)
                   /gene="E1"
                   /note="E1 interrupted ORF from 859 to 2813; putative"
                   /codon_start=1
                   /product="replication protein"
                   /protein_id="AAA46936.1"
                   /translation="MADPAGTNGEEGTGCNGWFYVEAVVEKKTGDAISDDENENDSDT
GEDLVDFIVNDNDYLTQAETETAHALFTAQEAQHRDAVQVLKRKYLVSPISDISGCV
DNNISPRLKAICIEKQSRAAKRRLFESSEDSGYGNTEVETQQMLQVEGRHETETPCSQY
SGGSGGGCSQYSSGSGGEGVSRHTICQTPLTNILNLVTKTSNAKAAMLAKFKELYGVS
FSELVRPFKSNKSTCCDWCIAAFGLTPSIADSIKTLQYCYLYLHIQSLACSWGVMVL"
```

LLVRYKCGKNRETIEKLLSKLLCVSPMCMMEIEPPKLRSTAAALYWKYTGISNISEVYG
DTPEWIQRQTVLQHSFNDCTFELSQMVMQWAYDNDIVDDSEIAYKYAQLADTNSNASAF
LKSNSQAKIVKDCATMCRHYKRAEKKQMSMSQWIKYRCDRVDDGGDWKQIVMFLRYQG
VEFMSFLTALKRFLQGIKPKNCILLYGAANTGKSLFGMSLMKFLQGSVICFVNSKSHF
WLQPLADAKIGMLDDATVPCWNYIDDNLRNALDGNLVSMDVKHRPLVQLKCPPLLITS
NINAGTDSRWPYLHNRLVVFTFPNEFPFDENGPNVYELNDKNWKSFFSRTWSRSLHE
DEDKENDGDSLPTFKCVSGQNTNTL"

gene 2755..3852

/gene="E2"

CDS 2755..3852

/gene="E2"

/note="E2 ORF from 2725 to 3852; putative"

/codon_start=1

/product="regulatory protein"

/protein_id="[AAA46941.1](#)"

/translation="METLCQRLNVCQDKILTHYENDSTDLRDHIDYWKHMRLECAIYY
KAREMGFKHINHQVPTLAVSKNKALQAIELQLTLETIYNSQYSNEKWTLQDVSLEVY
LTAPTGCIKKHGYTVEVQFDGDICNTMHYTNWTHIYICEEASVTVEGQVDYGYGLYVY
HEGIRTYFVQFKDDAEKYSKNKVWEVHAGGQVILCPTSVFSSNEVSSPEIIRQHLANH
PAATHTKAVALGTEETQTTIQRPRSEPDGTGNPCHTTKLLHRDSVDSAPILTAFNSSHK
GRINCNSNTTPIVHLKGDANTLKCLRYRFKKHCTLYTAVSSTWHWTGHNVKHKSAIVT
LTYDSEWQRDQFLSQVKIPKTITVSTGFMSI"

gene 3332..3619

/gene="E4"

CDS <3332..3619

/gene="E4"

/note="E4 ORF from 3332 to 3619; putative"

/codon_start=1

/protein_id="[AAA46937.1](#)"

/translation="YYVLHLCLAATKYPLLKLLGSTWPTTPRPIPKPSPWAPKKHRR
LSSDQDQSQTPETPATPLSCCTETQWTVLQSSSLHLTAHTKDGLTVIVTLHP"

gene 3863..4099

/gene="E5"

CDS <3863..4099

/gene="E5"

/note="E5 ORF from 3863 to 4099; putative"

/codon_start=1

/protein_id="[AAA46938.1](#)"

/translation="YCIHNITGVLFALLCVLLCVCLLIRPLLSVSTYTSLIILVLLL
WITAASAFRCFIVYIIFVYIPLFLIHTHARFLIT"

regulatory 4213..4218

/regulatory_class="polyA_signal_sequence"

/note="putative"

gene 4235..5656

CDS /gene="L2"
4235..5656
/gene="L2"
/note="L2 ORF from 4133 to 5656; putative"
/codon_start=1
/product="minor capsid protein"
/protein_id="AAA46942.1"
/translation="MRHKRS AKRTRKASATQLYKTCQAGTCCPDIIPKVEGKTIAEQ
ILQYGSMGVFFGGLGIGTGS GTGGRTGYIPLGTRPPTATDTLAPVRPPLTVDPVGP
PSIVSLVEETS FIDAGAPTSVPSIPPVSGFSITTSTDTTPAILDINNVTVTTHNN
PTFTDPSVLQPPTPAETGGHFTLSSSTISTHNYEEIPMDTFIVSTNPNTVTSSTPIPG
SRPVARLGLYSRTTQQVKVVDPAFVTTPTKLITYDNPAYEGIDVDNTLYFSSNDNSIN
IAPDPDFLDIVALHRPALTSRRTGIRYSRIGNKQTLRTRSGKSIGAKVHYYYDLSTID
PAEEIELQTITPSTYTTTSHAASPTSINNGLYDIYADDFITDTSTTPVPSVPSTSLSG
YIPANTTIPFGGAYNIPLVSGPDI PINITDQAPSLIPIVPGSPQYTI IADAGDFYLHP
SYYMLRKRKRRLPYFFSDVSLAA"

regulatory 4289..4295
/regulatory_class="TATA_box"
/gene="L2"

gene 5559..7154
/gene="L1"

CDS 5559..7154
/gene="L1"
/note="L1 ORF from 5526 to 7154; putative"
/codon_start=1
/product="major capsid protein"
/protein_id="AAA46943.1"
/translation="MQVTFIYILVITCYENDVNVYHIFQMSLWLPSEATVYLPPVPV
SKVVSTDEYVARTNIYYHAGTSRLLAVGHYPYFPIKKNPNNKILVPKVSGLQYRVFRIH
LPDPNKFGFPDTSFYNPDTQRLVWACVGEVGRGQPLGVGISGHPLLNKLDLDDTENASA
YAANAGVDNRECISMDYKQTQLCLIGCKPPIGEHWGKGSPTNVAVNPDCPPLELIN
TVIQDGMVHTGFGAMDFTTLQANKSEVPLDICTSICKYPDYIKMVSEPYGDSLFFYL
RREQMFVRHLFN RAGTVGENVPDDLYIKGSGSTANLASSNYFPTPSGSMVTSDAQIFN
KPYWLQRAQGHNGICWGNQLFVTVDTRSTNMSLCAAISTSETTYKNTNFKEYLRH
GEEYDLQFIFQLCKITLTADVMTYIHSMNSTILEDWNFGLQPPPGGTLEDYRFVTQA
IACQKHTPPAPKEDDPLKKYTFWEVNLKEKFSADLDQFPLGRKFLQAGLKAKPKFTL
GKRKATPTTSSTSTTAKRKRKL"

regulatory 7260..7265
/regulatory_class="polyA_signal_sequence"

ORIGIN Unreported.
1 actacaataa ttcatgtata aaactaaggg cgtaaccgaa atcggttgaa ccgaaaccgg
61 ttagtataaa agcagacatt ttatgcacca aaagagaact gcaatgtttc aggaccaca
121 ggagcgaccc agaaagtac cacagttagt cacagagctg caaacaacta tacatgatat
181 aatattagaa tgtgtgtact gcaagcaaca gttactgcga cgtgaggtat atgactttgc

241 ttttcgggat ttatgcatag tatatagaga tgggaatcca tatgctgtat gtgataaatg
301 tttaaagttt tattctaaaa ttagtgagta tagacattat tgttatagtt tgtatggaac
361 aacattagaa cagcaataca acaaaccgtt gtgtgatttg ttaattaggt gtattaactg
421 tcaaaagcca ctgtgtcctg aagaaaagca aagacatctg gacaaaaagc aaagattcca
481 taatataagg ggtcgggtga ccggctgatg tatgtcttgt tgcagatcat caagaacacg
541 tagagaaacc cagctgtaat catgcatgga gatacaccta cattgcatga atatatgtta
601 gatttgcaac cagagacaac tgatctctac tgttatgagc aattaaatga cagctcagag
661 gaggaggatg aaatagatgg tccagctgga caagcagaac cggacagagc ccattacaat
721 attgtaacct tttgttgcaa gtgtgactct acgcttcggt tgtgctgaca aagcacacac
781 gtagacattc gtactttgga agacctgta atgggcacac taggaattgt gtgccccatc
841 tgttctcaga aaccataatc taccatggct gatcctgcag gtaccaatgg ggaagagggt
901 acgggatgta atggatggtt ttatgtagag gctgtagtgg aaaaaaaaaac aggggatgct
961 atatcagatg acgagaacga aaatgacagt gatacaggtg aagatttgggt agattttata
1021 gtaaatgata atgattattht aacacaggca gaaacagaga cagcacatgc gttgtttact
1081 gcacaggaag caaaacaaca tagagatgca gtacaggttc taaaacgaaa gtatttggtta
1141 gtccacttag tgatattagt ggatgtgtag acaataatat tagtcctaga ttaaaagcta
1201 tatgtataga aaaacaaagt agagctgcaa aaaggagatt atttgaaagc gaagacagcg
1261 ggtatggcaa tactgaagtg gaaactcagc agatgttaca ggtagaaggc cgccatgaga
1321 ctgaaacacc atgtagtcag tatagtgggtg gaagtggggg tggttgcagt cagtacagta
1381 gtggaagtgg gggagagggt gttagtghaa gacacactat atgccaaaca ccacttacia
1441 atatttttaa tgtactaaaa actagtaatg caaaggcagc aatgttagca aaattttaaag
1501 agttatacgg ggtgagtttt tcagaattag taagaccatt taaaagtaat aaatcaacgt
1561 gttgagattg gtgtattgct gcatttggac ttacaccag tatagctgac agtataaaaa
1621 cactattaca acaatattgt ttatattttac acattcaaag tttagcatgt tcatggggaa
1681 tggttgtggt actattagta agatataaat gtggaaaaaa tagagaaaca attgaaaaat
1741 tgctgtctaa actattatgt gtgtctccaa tgtgtatgat gatagagcct ccaaaattgc
1801 gtagtacagc agcagcatta tattgggtata aaacaggat atcaaatatt agtgaagtgt
1861 atggagacac gccagaatgg atacaaagac aaacagtatt acaacatagt tttaatgatt
1921 gtacatttga attatcacag atggtacaat gggcctacga taatgacata gtagacgata
1981 gtgaaattgc atataaatat gcacaattgg cagacactaa tagtaatgca agtgcctttc
2041 taaaaagtaa ttcacaggca aaaattgtaa aggattgtgc aacaatgtgt agacattata
2101 aacgagcaga aaaaaaaca atgagtatga gtcaatgat aaaaataga tgtgataggg
2161 tagatgatgg aggtgattgg aagcaaattg ttatgttttt aaggatcaa ggtgtagagt
2221 ttatgtcatt tttactgca ttaaaaagat ttttgcaagg catacctaaa aaaaattgca
2281 tattactata tgggtgcagct aacacaggta aatcattat tggtagagt ttaatgaaat
2341 ttctgcaagg gtctgtaata tgttttgtaa attctaaaag ccatttttgg ttacaacat
2401 tagcagatgc caaaataggt atggttagat atgctacagt gccctggttg aactacatag
2461 atgacaattht aagaaatgca ttggatggha atttagtttc tatggatgta aagcatagac
2521 cattggtaca actaaaatgc cctccattat taattacatc taacattaat gctggtacag
2581 attctaggtg gccttattta cataatagat tgggtggtgt tacatttctc aatgagtttc
2641 catttgacga aaacggaaat ccagtgtatg agcttaatga taagaactgg aaatcctttt
2701 tctcaaggac gtggtccaga ttaagtttgc acgaggacga ggacaaggaa aacgatggag
2761 actctttgcc aacgttttaa tgtgtgtcag gacaaaatac taacacatta tgaaaatgat
2821 agtacagacc tacgtgacca tatagactat tggaaacaca tgcgcctaga atgtgctatt

2881 tattacaagg ccagagaaat gggattttaa catattaacc accaagtggg gccaacactg
2941 gctgtatcaa agaataaagc attacaagca attgaactgc aactaacgtt agaaacaata
3001 tataactcac aatatagtaa tgaaaagtgg acattacaag acgttagcct tgaagtgtat
3061 ttaactgcac caacaggatg tataaaaaaa catggatata cagtgggaagt gcagtttgat
3121 ggagacatat gcaatacaat gcattataca aactggacac atatatatat ttgtgaagaa
3181 gcatcagtaa ctgtggtaga gggtaagtt gactattatg gtttatatta tgttcatgaa
3241 ggaatacгаа catattttgt gcagttttaa gatgatgcag aaaaatatag taaaaataaa
3301 gtatgggaag ttcatgcggg tggtcaggta atattatgtc ctacatctgt gtttagcagc
3361 aacgaagtat cctctcctga aattattagg cagcacttgg ccaaccaccc cgccgcgacc
3421 catacгаа cggtcgcctt gggcacgгаа gaaacacaga cgactatcca gcgaccaaga
3481 tcagagccag acaccgгаа ccctgccac accactaagt tgttgacag agactcagtg
3541 gacagtgctc caatcctcac tgcatttaac agctcacaca aaggacggat taactgtaat
3601 agtaacacta cacccatagt acatttaaaa ggtgatgcta atactttaa atgtttaaga
3661 tatagattta aaaagcattg tacattgtat actgcagtg cgtctacatg gcattggaca
3721 ggacataatg taaaacataa aagtgcatt gttacactta catatgatag tgaatggcaa
3781 cgtgaccaat ttttgtctca agttaaata ccaaaaacta ttacagtgtc tactggattt
3841 atgtctatat gacaaatctt gatactgcat ccacaacatt actggcgtgc ttttgcctt
3901 gctttgtgtg cttttgtgtg tctgcctatt aatacgtccg ctgcttttgt ctgtgtctac
3961 atacacatca ttaataatat tggattact attgtggata acagcagcct ctgcttttag
4021 gtgttttatt gtatatatta tatttgttta tatacatta ttttaatac atacacatgc
4081 acgcttttta attacataat gtatatgtac ataatgtaat tgttacatat aattgttgta
4141 taccataact tactattttt tcttttttat tttcatatat aatttttttt tttgtttgtt
4201 tgtttgtttt ttaataaact gttattactt aacaatgcga cacaaacgtt ctgcaaacg
4261 cacaaacgt gcacgggcta cccaacttta taaaacatgc aaacaggcag gtacatgtcc
4321 acctgacatt atacctaagg ttgaaggcaa aactattgct gaacaaatat tacaatatgy
4381 aagtatgggt gtattttttg gtgggttagg aattggaaca gggtcgggta caggcggagc
4441 cactgggtat attccattgg gaacaaggcc tcccacagct acagatacac ttgctcctgt
4501 aagacccctt ttaacagtag atcctgtggg cccttctgat ccttctatag tttctttagt
4561 ggaagaaact agttttattg atgctgggtc accaacatct gtacctcca tccccaga
4621 tgtatcagga ttagtatta ctacttcaac tgataccaca cctgctatat tagatattaa
4681 taatactggt actactgtta ctacacataa taatcccact ttcactgacc catctgtatt
4741 gcagcctcca acacctgcag aaactggagg gcattttaca ctttcatcat cactattag
4801 tacacataat tatgaagaaa ttcctatgga tacatttatt gttagcacia accctaacac
4861 agtaactagt agcacacca taccagggtc tcgcccagtg gcacgcctag gattatatag
4921 tcgcacaaca caacaggtta aagttgtaga ccctgctttt gtaaccactc ccaactaaact
4981 tattacatat gataatcctg catatgaagg tatagatgtg gataatacat tatatttttc
5041 tagtaatgat aatagtatta atatagctcc agatcctgac tttttggata tagttgcttt
5101 acataggcca gcattaacct ctaggcgtac tggcattagg tacagtagaa ttggtaataa
5161 acaaacta cgtactcgta gtggaaaatc tatagggtgct aaggtaacatt attattatga
5221 tttaagtact attgatcctg cagaagaaat agaattacia actataacac cttctacata
5281 tactaccact tcacatgcag cctcacctac ttctattaat aatggattat atgatattta
5341 tgcagatgac tttattacag atacttctac aaccccggtc ccatctgtac cctctacatc
5401 tttatcaggt tatattcctg caaatacaac aattcctttt ggtgggtgcat acaatattcc
5461 tttagtatca ggtcctgata taccattaa tataactgac caagctcctt cattaattcc

5521 tatagttcca gggctctccac aatatacaat tattgctgat gcaggtgact tttatttaca
5581 tcoctagttat tacatgttac gaaaacgacg taaacgttta ccatattttt tttcagatgt
5641 ctctttggct gcctagttag gccactgtct acttgccctcc tgtcccagta tctaaggttg
5701 taagcacgga tgaatatggt gcacgcacaa acatatatta tcatgcagga acatccagac
5761 tacttgacgt tggacatccc tattttccta ttaaaaaacc taacaataac aaaatattag
5821 ttocctaaagt atcaggatta caatacaggg tattttagaat acatttacct gacccaata
5881 agtttggttt tcoctgacacc tcattttata atccagatac acagcggctg gtttggcct
5941 gtgtaggtgt tgaggtaggt cgtggtcacg cattaggtgt gggcattagt ggccatcctt
6001 tattaataa attggatgac acagaaaatg ctagtgctta tgcagcaaat gcaggtgtgg
6061 ataatagaga atgtatatct atggattaca aacaaacaca attgtgttta attggttga
6121 aaccacctat aggggaacac tggggcaaaag gatccccatg taccaatgtt gcagtaaatc
6181 caggtgattg tccaccatta gagttaataa acacagtta tccaggtggt gatatggttc
6241 atactggctt tgggtgctatg gactttacta cattacaggc taacaaaagt gaagttccac
6301 tggatatttg tacatctatt tgcaaatatc cagattatat taaaatggtg tcagaacat
6361 atggcgacag cttatttttt tatttacgaa ggaacaaat gtttgttaga catttatta
6421 atagggctgg tactgttggg gaaaatgtac cagacgattt atacattaaa ggctctgggt
6481 ctactgcaa tttagccagt tcaaatatt ttccacacc tagtggttct atggttacct
6541 ctgatgccc aatattcaat aaaccttatt ggttacaacg agcacaggc cacaataatg
6601 gcatttggtg gggtaaccaa ctatttgta ctggttga tactacacgc agtacaata
6661 tgtcattatg tgctgccata tctacttcag aaactacata taaaataact aactttaagg
6721 agtacctacg acatggggag gaatatgatt tacagtttat ttttcaactg tgcaataa
6781 cottaactgc agacgttatg acatacatac attctatgaa ttccactat ttggaggact
6841 ggaattttgg tctacaacct ccccaggag gcacactaga agatacttat aggtttgtaa
6901 cccaggcaat tgcttgtcaa aaacatacac ctccagcacc taaagaagat gatcccccta
6961 aaaaatacac tttttggaa gtaaattdaa aggaaaagt ttctgcagac ctagatcagt
7021 ttoccttagg acgcaaat tttactacaag caggattgaa ggccaaacca aaatttacat
7081 taggaaaacg aaaagctaca cccaccacct catctacctc tacaactgct aaacgcaaaa
7141 aacgtaagct gtaagtattg tatgtatggt gaattaggtg tgtttgtgtg gtatatgttt
7201 gtatgtgctt gtatgtgctt gtaaatatta agttgtatgt gtgtttgtat gtatggtata
7261 ataaacacgt gtgtatgtgt ttttaaatgc ttgtgtaact attgtgtcat gcaacataaa
7321 taaacttatt gtttcaacac ctactaattg tgttgtggtt attcattgta tataaactat
7381 atttgctaca tcoctgtttt gttttatata tactatattt tgtagcgcca ggcccatttt
7441 gtagcttcaa ccgaattcgg ttgcatgctt tttggcacia aatgtgtttt tttaaatagt
7501 tctatgtcag caactatggt ttaaacttgt acgtttcctg cttgccatgc gtgcaaatc
7561 cctgttttcc tgacctgcac tgcttgccaa ccattccatt gttttttaca ctgcactatg
7621 tgcaactact gaatcactat gtacattgtg tcatataaaa taaatcacta tgcgccaacg
7681 ccttacatac cgctgttagg cacatatttt tggcttgttt taactaacct aattgcatat
7741 ttggcataag gtttaactt ctaaggccaa ctaaatgtca ccoctagtta tacatgaact
7801 gtgtaaaggt tagtcataca ttgttcattt gtaaaactgc acatgggtgt gtgcaaacg
7861 attttgggtt acacatttac aagcaactta tataataata ctaa

//