Molecular Cloning and Nucleotide Sequence of Deer Papillomavirus

DENNIS E. GROFF¹* AND WAYNE D. LANCASTER^{1,2}

Departments of Obstetrics and Gynecology¹ and Pathology,² Georgetown University Medical Center,
Washington, DC 20007

Received 22 April 1985/Accepted 5 July 1985

The genome of deer papillomavirus (DPV) isolated from American white-tailed deer was cloned into pBR322, and the entire nucleotide sequence of 8,374 base pairs was determined. The overall genetic organization of the DPV genome was similar to that of other papillomaviruses. All significant open reading frames were located on one strand, and the locations of putative promoters and polyadenylation signals were similar to those identified in the closely related bovine papillomavirus type 1 (BPV-1) genome. The DPV genome was approximately colinear with BPV-1 except for a noncoding region separating the early and late regions. The regions of highest nucleotide sequence homology between DPV and BPV-1 were found in the E1 open reading frame coding for BPV-1 DNA replication function and in the L1 open reading frame, which encodes the major capsid protein of BPV-1.

Deer papillomavirus (DPV) represents an unusual class of papillomavirus. DPV isolated from American white-tailed or mule deer (previously designated deer fibromavirus) primarily causes a fibroblastic proliferation (9, 10, 32). The virus matures only in the overlying, keratinizing epithelial layers; however, proliferation of these layers is uncommon (32). This attribute of DPV is in marked contrast to the majority of papillomaviruses which routinely cause epithelial neoplasms, either alone like human papillomaviruses, or in conjunction with fibroblastic proliferations like bovine papillomavirus (BPV) types 1 and 2 (for reviews, see 15, 17, 25).

Physically, DPV exhibits a higher degree of relationship with BPV-1 than has been found between any other papillomaviruses isolated from different species. Antisera against these viruses show cross-reactivity in hemagglutination inhibition assays and the viral DNAs cross-hybridize to a measurable level under stringent hybridization conditions (18). However, DPV DNA is significantly larger than BPV-1 (8.3 versus 8.0 kilobases).

The oncogenic potential of DPV provides a model system for papillomavirus-induced carcinogenesis that is complementary to the well-characterized BPV-1. DPV exhibits not only biologic similarities to BPV-1 but also striking differences. Both DPV and BPV-1 cause tumors in hamsters; however the latent period of DPV tumor formation is shorter than with BPV-1 (16). Although both DPV and BPV-1 transform NIH 3T3 mouse cells in culture (5, 10, 20), DPV, unlike BPV-1, does not transform the C127 mouse cell line. This also distinguishes DPV from BPV-2 (5), European elk papillomavirus (33), and possibly cottontail rabbit papillomavirus and human papillomavirus type 5 (34).

These biological differences may be reflected in the potential coding sequences of DPV gene products. To examine this possibility, we molecularly cloned the genome of DPV from the American white-tailed deer (*Odocoileus virginianus*) into pBR322 and determined the nucleotide sequence. Analysis of the genetic organization and regions of DNA sequence homology with BPV-1 indicate similarities between potential papillomavirus functions and invite speculation on the molecular mechanisms of papillomavirus-induced cell transformation.

MATERIALS AND METHODS

DNA cloning. DPV DNA was isolated from purified virions as described by Lancaster and Sundberg (18). DPV DNA was linearized at the unique *Eco*RI site and cloned into *Eco*RI-digested, dephosphorylated pBR322 DNA by standard procedures (22). Ampicillin-resistant colonies of *Escherichia coli* HB101 were screened with ³²P-labeled virion DPV DNA as described by Grunstein and Hogness (11). One clone containing DPV DNA, designated pDEG98, was selected for subsequent sequencing.

Nucleotide sequencing. Nucleotide sequencing reactions were done by the dideoxy chain termination method of Sanger et al. (28). The nucleotide sequence flanking the DPV *Eco*RI site was also determined by the chemical cleavage method of Maxam and Gilbert (23).

Sequencing strategies. Four approaches were used to generate DNA templates for dideoxy chain termination sequencing reactions. (i) Single-stranded templates were derived from small restriction fragments randomly cloned into M13 vectors as described by Gingeras et al. (8). (ii) Progressive deletions of large restriction fragments in M13 vectors were prepared as described by Hong (14). (iii) Plasmid templates were obtained by partial restriction endonuclease digestion of pDEG98 and sequenced with synthetic pBR322 primers (New England BioLabs, Inc., Beverly, Mass.) as described by Guo et al. (12). (iv) Progressive deletion mutants of pDEG98 DNA were generated for sequencing by BAL 31 nuclease digestion by a strategy similar to that described by Guo et al. (12).

To generate deletion mutants with BAL 31 nuclease, we linearized plasmid DNA by *Hpa*I digestion and purified it by phenol extraction and ethanol precipitation. Linear DNA (5 µg/ml) was digested for various periods of time with BAL 31 nuclease (4 U/ml) in 0.2 M NaCl–12 mM CaCl₂–12 mM MgCl₂–20 mM Tris hydrochloride (pH 8.0) at 30°C. Under these conditions, the rate of degradation was 60 to 80 base pairs (bp) per minute per end. Nuclease digestions were terminated by addition of EDTA to a final concentration of 20 mM and extraction with phenol, and DNA was collected by ethanol precipitation. BAL 31-digested plasmid DNA was cleaved at the unique *Sal*I site, the ends were repaired with *E. coli* DNA polymerase I (large fragment), and it was recircularized with T4 DNA ligase and used to transform *E.*

* Corresponding author.

1 GTTAAACAATA ACCAGACCTT AGCCSTTTTT GGGTGAAGCG GAAAGATGSG TTACAGGTTT TATAAAGCA CCACACGCA CAAGGTTGCT ATCACTTGTG
 201 TGTCTGTGACA GACCCCTTCT TCTGCATGTC TGTCTGATTAC TATGAACTAT TATCTGTGTG ATTTTGTCTAC TGTGTCTGTG TGTGTCTGTG
 301 ATGCGCCTCA ACCTGTATAT CAGGCGCTCCA TTGAAGAGCC TGATCTTTTC ATTTCAAAAG CCGTGCATAG ATGCTATGAG TGTGTGTGTG
 401 TGACGAAAGG GACAGGCGACA GATATTTTGA AGAGCTTTAC GTGATATTTCA GGAATCAGGT TCTTGTGAGA TGTGTGTGTG TGTGTGTGTG
 501 TGCGCCTCTT ACCGGGCGAAG CGCTACCGCG TGATGAATCA CTTTGTCTTAA CATTGATTTT GAGGCGCAGT TCGGGAAGAG CAGCAAGAA
 601 GTGCTGTGTG ATAGCTGTGG AAAACCGCGC CTTAAACGCG GTATAAATGTG ACTGTTTCTCT ACTGTTTCTCT TCTTCTGTGT TGTGTGTGTG
 701 CTGTCAAAAC TACCTGTGAG ACAATACTCA CCTGTGACAG ACTCTTGACA GAGGACCTGT ATTTCTGTGT TCTTGTGTGT TGTGTGTGTG
 801 AAAATGAGAG TAGCTGTGGG GTTGGGGGGG ATTTCTTTAT CTTCTTTGAG GCAGAATGCT CAGATACAGA TTTCTGAATCA TGTGAATGAG
 901 TGATGAGGAT TTACTAGATA ATGCCACTGC GTTCCGGGGA AACCACTGTG AGCTCTTCCA AACTCAGGAA AAGAGGCGG GAGAAAGAGA GATTTCATTT
 1001 TTGAAAGAGAA AACTGTGTTT AAGCCCTTGC TCTGCTGACT CTGAGGTGGA GCAAGCTCAAG TCTTGGGCTT GCTGTATATA ATATGAGGCT GGTGAAGGCA
 1101 ATCCGCTGTT TAGACGCGAG CTTTTCGAAA GAGGTGATCC AAGGCGGTGT /ACACACCTG TGAACCATGA AGCTGACAA TTTTCTGCTT CAGGACTGCA
 1201 GGTACAGTCT GGGGAAAATA GGTGAGGCA GAAAGAGGGA AAAGGGGAG TTTCCGCGCT GCTAGCTCA GCTGAGGCAA ATATGAGGCT CTGATACAG
 1301 AAATTTGTTCA AGACTCTCTA CATCGCTCC CATGGGAGCA TCACTGTGTG ATTTCAAAGT AATAAGACTG TTAACCATCA CTGGGTGATT CTGGCATATG
 1401 GAGTTAGTGA GGTGTGTAT TCTGCTAGCT TTGATCTCTT TGTAAACAG TGTAACTGCC TGCAAACTG TGAAGAGGTT CATGAAAAAG GAGATATTTT
 1501 TGTTCACGTG TGTATGTTCA ATGTTGCCAA AAGTGAAGAT ACAATGAGT AATTAATGAC CACAATTCTG AATGTTACCG CCGGCAACCT CCTCTACAG
 1601 CCCCCTAAAA TCAGAGGTCT CCGGCTGTCT CTATTCTGTG TTAAGCTCAC ACTGTCACT GCTACCTTAA CCAATGATAC CCAATGATAC TGTGTGTGTG
 1701 AGGCACTAA TGTGTGCGAG AATCTGTGAG AGGCGGCTAA ATTTGATTTA GGAACATATG TGCAGTGGG TTAATGACAT GTTTTCAAGG AGGATGTGAA
 1801 AATTGCTTAT GAATATGCTG GAGCGCTGAG GAGCGCTGAG GAGCGCTGAG CAATGTGAGG TATATATAA AGGCGGTGTG ATCAAACTGC AGGAAGTGG AGCTGGCTCT
 1901 ATGGGTGAGAC ATTACCTGCG TGTGTGAGTA CAGGCGCTGA CAATGTGAGG TATATATAA AGGCGGTGTG ATCAAACTGC AGGAAGTGG AGCTGGCTCT
 2001 CTATCTAGAA TTTGCTTAAA TATCATGGGA TAGAACATAT AATGCTTAA AGCCTTGGTT AAAAGGCATT CCAAAATATA CCAAAATATA
 2101 AATTGTTGGC CCGCTAACA GTGGGAAGTC ACTCTTTGAG AACTCTTTGA TAGCATTTCT TGGCGGCAAG GTGCTTACAT TTGCAACAGA CCACGACGCA
 2201 TTTTGGCTCG CACCTCTAGC GGAATGCGCA GTTGTCTTAA TGTGATGAGT TACCACACT TGTGAGAGT ACTTTGACAC ACACCTGACAT ATTTGTGTGG
 2301 ATGGCTATCC ATTTGGTATT GATAGAAAA ACATACCGG TGTTCATATG AAAAGCGCTC CCTCTTATG AACCAGTAA TTTGATGTGG ATGCAAGAGA
 2401 AAGGTATTTT TATTGCGACA GCGAGTAA TCCGCTTTAC TTCAAGGAGC GTCCGCTGCT TCCAGACAA GTTGAAGCTA TGTTCATATG AACTGATGCT
 2501 GATTGGAAC ATTTCTTTGA AAGGCTATGG GAGGCTTAA ACCTGAGCG TCAAGAGGAC GAGGTTGAGC ACATGAGTG CAGCAAGAGA CAGGTTACTT
 2601 GCACGCGAAG AAGGCAAAAT GACATTAAT AAAAAGATA GAAAGATCAC ATAGACTCTT GAGGCTCGGT GAGGCTCGGT CAGGCTCGGT
 2701 TTTATGCTGC CAGACACAAA GCTTGTATT GGTCTGCTTT GAACCTGTG CTGAGAGTG CTAGAGGCT CTAAGAGCT CCGCAAGCAA TTGAGATGCA
 2801 GCTTCTGCGG AACAGCTTAA AGGAGAGGCC ATGAGTGAAT GAGGCTGATG CACTGTGTGA CTTAAGCTGG CTAAGCTGCT AAGCGCTATC AGCAAACT
 2901 TTGAAAGAGG GCGCGAGCT GGTGAAAGT GAGTATGAT GAGGCTCCAC TAATAAACT TGTATACCG CTGGAATTC ATTTGATGCT CCGCAAGCT
 3001 ATGAGGAGGG CTGGGAGGCG GCGACTGCTG GTGCGAGGCG AGGCGCTCTG TGTCTACTA CCAATGTGCG TACAAGGCT TACTTATGCT AAGCAAACT
 3101 AGATGCGAGC AGATACAGCA CTACAGGAGC TTGAGCTGTG AGGATGAAGC ATGCTACTTA ATGCTACTTA TCACTGACAT TCTGCGCTCT CCACTCTAG AGAGGACTC
 3201 GAGGAGGCTGT GGAATCTGCG GGGCGGTGAA AGGAGGAGC CCACCACTCT GCGCGAGGCT GCGGCTGCTC ACACCTCTCC TGGAGGAGC ACCGTTCAAG
 3301 GTTCCGCTGAG AGCTTGTGAA AACCGGGTCT GGTCTTAA CCGCGGAGCT CCGTACAGCA CAGGAGGCT CCGGAGGAGT CCGGAGGAGT CCGGAGGAGT
 3401 CTTCCGCTGAG CCGAGGCGGG TACGCGAGAA CCGCGGTGCT GTATCTCTAC GTGCTCTTCA GGTGCGCTCT GGTGCGCTCT CCGGAGGAGT CCGGAGGAGT
 3501 CCGCGGTGAG TCCCAACCC CAGCGCGGAA GAGCGGCTGT TTAGCTTCTA TGGCTCTTCA GGTGCGCTCT GGTGCGCTCT CCGGAGGAGT CCGGAGGAGT
 3601 AAGTCAAGTG CTATTCTCTC CCGGTGAAAG GATGGCATAG GAGCAAGAT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT
 3701 ACCAGGAGT GCGCAGAGTA TGTGCTGCTT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT
 3801 GTGACTATGA CTGTTGACTT TTAGGCTGTG TGTGCTACTG GCTGCTACTG GCTGCTACTG GCTGCTACTG GCTGCTACTG GCTGCTACTG GCTGCTACTG
 3901 CACTGTGTGA AGCATGAAC ATGCTGGCTG TTTCTGTGTT CTGGGAGTGT GCTGCTACTG GCTGCTACTG GCTGCTACTG GCTGCTACTG GCTGCTACTG
 4001 GTATGCTGAG ATGAGTTTGG GTGTAGTGT GAAAGCTTCC ACATGTAAAT CCGCTTCTG CTGAGGAGG GATGAGGAGT TCAAGTATGG TGTGCTGTG
 4101 GATGAATAA ATCATTTTAT CCAAGCTTCT GAAATAGCA TTAAAGGAGC ATTAATTTACT GCGGAGTGTG GCGGAGTGTG GCGGAGTGTG GCGGAGTGTG
 4201 AAGGAGGCTT CTCTCTCTCT CTAGTCTCT CAGGTGTGAC AGGATTAATCT TCCCTCTCTT ATTTATTTT GATATTTTCT GATATTTTCT GATATTTTCT
 4301 TATTATTATC CAGGCTGAG TACTGTTATC CAAGGTGAGC AAGGATGAGT CAGGATGAGT CAGGATGAGT CAGGATGAGT CAGGATGAGT CAGGATGAGT
 4401 ATGAGCTATG GACTAAAACT TAAGTGTGCTG TGCAGGATC ACCAGGAGGCT CAGGATGAGT CAGGATGAGT CAGGATGAGT CAGGATGAGT CAGGATGAGT
 4501 ATGAGCTATG GACTAAAACT TAAGTGTGCTG TGCAGGATC ACCAGGAGGCT CAGGATGAGT CAGGATGAGT CAGGATGAGT CAGGATGAGT CAGGATGAGT
 4601 GCTGCTATTA TTTATAGGCT ATTGAGGAGC AGTTTGGTGT ATATTGTTGT TGTCTGTGCT CCAATGAGCT ATTAAGAGCT ATTAAGAGCT ATTAAGAGCT
 4701 TGTGCTGCTG GTTTCAGCTT CCAATGAGCT GTTTAAGTGT CAGGATGAGT CAGGATGAGT CAGGATGAGT CAGGATGAGT CAGGATGAGT CAGGATGAGT
 4801 TGTGCTGCTG GTTTCAGCTT CCAATGAGCT GTTTAAGTGT CAGGATGAGT CAGGATGAGT CAGGATGAGT CAGGATGAGT CAGGATGAGT CAGGATGAGT
 4901 ATGAGCTATG GACTAAAACT TAAGTGTGCTG TGCAGGATC ACCAGGAGGCT CAGGATGAGT CAGGATGAGT CAGGATGAGT CAGGATGAGT CAGGATGAGT
 5001 GCGAGGAGGAG TGGGTCTCTC ACATCTATT TTAGGAGGCT TGTGAGGAGT TGTGAGGAGT TGTGAGGAGT TGTGAGGAGT TGTGAGGAGT TGTGAGGAGT
 5101 GAGGAGGAGT GCTGCGAGCT TGTGAGGAGT TGTGAGGAGT TGTGAGGAGT TGTGAGGAGT TGTGAGGAGT TGTGAGGAGT TGTGAGGAGT TGTGAGGAGT
 5201 GTGCAAGGAG AAGTCAACCA AGAAGGCTCT ATCAGCTTTT GTGCAAGGAGT TGTGAGGAGT TGTGAGGAGT TGTGAGGAGT TGTGAGGAGT TGTGAGGAGT
 5301 AACCAACCTT GATTTCAGCT TCAAGGAGCT CCAAGGCTTT GTTTCAGGAGT TGTGAGGAGT TGTGAGGAGT TGTGAGGAGT TGTGAGGAGT TGTGAGGAGT
 5401 ATTTGAGGAG GTTGGGAGGAG TGGGTGAGC CACTGAGGAG GAGATGAGT TCAAGGAGT TGTGAGGAGT TGTGAGGAGT TGTGAGGAGT TGTGAGGAGT
 5501 GCGGAGGAGT GCAATTTCAA CTGAGTCAAC AGAATGAGT CAGGATGAGT TGTGAGGAGT TGTGAGGAGT TGTGAGGAGT TGTGAGGAGT TGTGAGGAGT
 5601 AGAATGAGT ATGAGTATCA AAGGAGTCA AAGTATGAGT CAGGATGAGT TGTGAGGAGT TGTGAGGAGT TGTGAGGAGT TGTGAGGAGT TGTGAGGAGT
 5701 GCTACTTTGA GAGGAGTCA GCGGAGTCA GCGGAGTCA GCGGAGTCA GCGGAGTCA GCGGAGTCA GCGGAGTCA GCGGAGTCA GCGGAGTCA
 5801 CTGCGTCACT GGTGAGGAGT TATTGATGAG CTTGAGGAGT TATTGATGAG CTTGAGGAGT TATTGATGAG CTTGAGGAGT TATTGATGAG CTTGAGGAGT
 5901 TGTGAGGAGT TATTGATGAG CTTGAGGAGT TATTGATGAG CTTGAGGAGT TATTGATGAG CTTGAGGAGT TATTGATGAG CTTGAGGAGT TATTGATGAG
 6001 AGGACAAATA TTCTGTGCA TGCAGGAGG CCGAGTGGGT TGTGCTCTAG GAGATACAT TATGATGAGT TGTGCTCTAG GAGATACAT TATGATGAGT
 6101 ATGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT
 6201 GTCTGTGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT
 6301 CTGTGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT
 6401 TCTGTGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT
 6501 CCGTATGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT
 6601 GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT
 6701 ACAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT
 6801 CCTCTATGAT AATTAAGATG AAGCTTATAT GAGGATGAGT GAGGATGAGT GAGGATGAGT GAGGATGAGT GAGGATGAGT GAGGATGAGT GAGGATGAGT
 6901 ATCTACCTCT GAGGAGGAGT AAGCTTATAT GAGGATGAGT GAGGATGAGT GAGGATGAGT GAGGATGAGT GAGGATGAGT GAGGATGAGT GAGGATGAGT
 7001 GAGGAGGAGT TATGTAAGAT AACTTATGAG CAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT GAGGAGGAGT
 7101 AAAATGCCAA GTTGTCTTCT TGAAGTCCA AGTGGAGGAGT TATGATGAGT TATGATGAGT TATGATGAGT TATGATGAGT TATGATGAGT TATGATGAGT
 7201 TGAGGAGGAGT TATGATGAGT TATGATGAGT TATGATGAGT TATGATGAGT TATGATGAGT TATGATGAGT TATGATGAGT TATGATGAGT TATGATGAGT
 7301 TTTGAGGAGT TATGATGAGT TATGATGAGT TATGATGAGT TATGATGAGT TATGATGAGT TATGATGAGT TATGATGAGT TATGATGAGT TATGATGAGT
 7401 CTGAGGAGT TCTCTAGTCT CCAAGGAGTCA ATGCGGAGTCA ATGCGGAGTCA ATGCGGAGTCA ATGCGGAGTCA ATGCGGAGTCA ATGCGGAGTCA
 7501 ATGCGGAGT TCTCTAGTCT CCAAGGAGTCA ATGCGGAGTCA ATGCGGAGTCA ATGCGGAGTCA ATGCGGAGTCA ATGCGGAGTCA ATGCGGAGTCA
 7601 AAAAGAGGAG AATTTCTTAT ATTTGAGGAGT GTTTCCTTCA GTTTCCTTCA GTTTCCTTCA GTTTCCTTCA GTTTCCTTCA GTTTCCTTCA
 7701 TTTGAGGAGT CTTGAGGAGT GTTTCCTTCA GTTTCCTTCA GTTTCCTTCA GTTTCCTTCA GTTTCCTTCA GTTTCCTTCA GTTTCCTTCA GTTTCCTTCA
 7801 CTTGAGGAGT TTTGAGGAGT GTTTCCTTCA GTTTCCTTCA GTTTCCTTCA GTTTCCTTCA GTTTCCTTCA GTTTCCTTCA GTTTCCTTCA GTTTCCTTCA
 7901 AGTGTATGAT ACCTTGTGAG CAGGAGTGTCA GCAACTTTTA ATTTGAGGAGT GTTTCCTTCA GTTTCCTTCA GTTTCCTTCA GTTTCCTTCA GTTTCCTTCA
 8001 TTAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT
 8101 AGGCGGAGT TTAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT
 8201 AAGGCGGAGT TTAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT CAGGAGGAGT
 8301 AATGATGAGT AAGTACTTTG CAGGAGTGTG TCTGAGTCA TGAAGTCTT GATTTCCTTCA GTTTCCTTCA GTTTCCTTCA GTTTCCTTCA GTTTCCTTCA

FIG. 1. Nucleotide sequence of DPV DNA. The sense strand sequence is shown. Nucleotide 1 is the G residue on the *HpaI* recognition sequence (5'-GTTAAC-3') originally placed at 0.65 map units (18). This numbering system corresponds to those reported for BPV-1 (2), human papillomavirus type 1a (4), and human papillomavirus type 6b (31).

coli HB101 to ampicillin resistance. Plasmid DNAs from deletion mutants varying in size by about 150 bp were used for the sequencing reactions as described above.

DNA sequence data. Computer programs developed by Pustell and Kafatos (27, 28) were used for analysis of the DPV sequence data and for comparison with the BPV-1 sequence. The BPV-1 DNA sequence (1, 2) contained mod-

ifications of Danos et al. (4) and Stenlund et al. (see reference 31).

RESULTS

Molecular cloning of DPV DNA. DPV was purified from the epithelial layers of a pool of naturally occurring fibromas obtained from American white-tailed deer. Viral DNA was

FIG. 2. Positions of translation termination codons in DPV DNA. Groups of 10 codons were analyzed in each reading frame, and a vertical bar was drawn if at least 1 termination codon existed. The sense strand is shown at the top; the antisense strand is at the bottom. ORFs in the sense strand encoding putative polypeptides of greater than 90 amino acid residues are designated E1 to E7 and L1 to L3 in correspondence to papillomavirus nomenclature (2). For orientation, the positions of some restriction sites are indicated (B, *Bgl*II; C, *Hinc*II; D, *Hind*III; E, *Eco*RI; H, *Hpa*I; K, *Kpn*I; P, *Pvu*I).

extracted, linearized with *Eco*RI and cloned into pBR322. DPV-containing clones were identified by hybridization with 32 P-labeled DPV virion DNA. One positive clone, designated pDEG98, contained an 8.3-kilobase insert after digestion with *Eco*RI and hybridized with 32 P-labeled BPV-1 DNA under stringent hybridization conditions (data not shown), as expected for DPV DNA (18). Restriction enzyme digestion of pDEG98 DNA produced fragments whose sizes were consistent with a DPV-pBR322 recombinant. In addition to restriction sites previously mapped (18), cleavage sites in DPV DNA for *Pvu*I (one site) and *Kpn*I (four sites) were identified and mapped by double digestion of pDEG98 DNA.

DNA sequence of DPV. The DPV DNA sequence contained 8,374 nucleotides (Fig. 1). This was in good agreement with the reported size of 5.25×10^6 daltons or 8.3 kilobases (18). Numbering began with the first G of the *Hpa*I recognition sequence at 0.65 map units and proceeded toward map unit 1.0 (18). This alignment placed the *Eco*RI site at position 2,975 and allowed direct comparison with other papillomavirus sequences (2, 4, 22). The G+C content of DPV DNA was 47.5%; however, the G+C content was not evenly distributed throughout the genome. For example, G+C-rich regions were found in the regions bp 3,215 to 3,536 (65% G+C), bp 4,883 to 4,957 (60% G+C), and bp 8,047 to 8,123 (61% G+C). Dinucleotide frequency analysis revealed a rare occurrence of CG (2.9%) and an abundance of TG (8.4%). One striking feature of the nucleotide distribution of DPV DNA was a large, T-rich region in the middle of the genome. The 730-bp region between bp 3,974 and 4,703 contained 38% T residues compared with 24% A, 20% G, and 18% C residues.

The distribution of termination codons in all six reading frames of DPV DNA was determined (Fig. 2). Similar to other papillomaviruses, one strand contained most of the significant open reading frames (ORFs) capable of encoding polypeptides larger than 90 amino acids in length. The ORFs comprised 85% of the DPV genome and were arranged in two clusters of overlapping segments separated by intergenic regions with no ORFs. The DPV ORF clusters have been designated E and L to correspond to the BPV-1 designations (2). E refers to the region transcribed in BPV-1-transformed cells, and L indicates the region transcribed only in productive, virus-producing warts (6). The location and potential coding capacity for each of the DPV ORFs are presented in Table 1. The organization of DPV ORFs was similar to BPV-1 with exceptions being the absence of the BPV-1 E5 and E8 ORFs and the presence of a third ORF (L3) in the L region of DPV which is not found in BPV-1 DNA. Intergenic noncoding regions (NCRs) accounted for about 15% of the genome. NCR-1, which corresponds to the BPV-1 regulatory region, contained 646 bp between the 3' end of the L1 ORF and the 5' end of the E6 ORF. NCR-2, which is absent in BPV-1, contained 602 bp between the 3' end of the E2 ORF and the 5' end of the L2 ORF.

The presence of potential transcription elements was located in DPV DNA. Potential promoter sequences were found in four positions. Three of these corresponded to similar sequences in BPV-1 DNA; one (CTATAAAA) was located near the 5' end of the E ORFs (bp 61), the second (ATATAAAA) was located within the E1 ORF (bp 1,954), and the third (ATATATAG) was at the 3' end of the L1 ORF (bp 7,721). The fourth sequence (GTATAAAC) was found in the NCR-2 intergenic region (bp 4,291); no corresponding sequence is found in BPV-1. Four potential polyadenylation

TABLE 1. ORFs in DPV DNA

Open reading frame	Nucleotide position			Translation capacity	
	First nucleotide	First ATG	Termination codon	Total	From first methionine
E1	749	791	2,630	69,741	68,058
E2	2,547	2,574	3,822	47,337	46,296
E4	3,142		3,508	13,364	
E6	66	126	531	17,839	15,623
E7	445	490	796	12,745	11,183
L1	6,210	6,255	7,794	59,741	58,163
L2	4,424	4,763	6,242	66,551	53,725
L3	5,376	5,592	5,697	12,433	4,239

TABLE 2. Potential transcriptional elements in DPV DNA

Positions of polyadenylation signal sequences (bp)		Distance between sequences (bp)
AATAAA	CA	
2,952	2,980	22
4,751	4,767, 4,770	11, 14
7,795	7,826	26
8,022	8,040	13

TABLE 3. Comparison of DPV and BPV-1 splice donor and acceptor sequences

Splice sequence type (genome)	Sequence	Position (splice site [bp])
Donor		
DPV	CAG/GGGT	323/4
BPV-1	CAG/GTGA	304/5
DPV	CAG/GTAG	809/10
	AAG/GTGA	890/1
BPV-1	AAG/GTAG	864/5
Acceptor		
DPV	CATTGATTTTGGAG/CCAG	564/4
BPV-1	TGATTCACCTGCAG/GACC	527/8
DPV	CCCTCCCCTCTAG/AGAG	3,190/1
BPV-1	TCTTCTGATTTTAG/AGAT	3,223/4

signals (AATAAA) were found in DPV DNA (Table 2). Two of these signals (bp 4,751 and 8,022) preceded a CA dinucleotide by 11 to 14 nucleotides, a distance considered optimal for mRNA polyadenylation (7). The positions of these two polyadenylation signals 3' of the E ORFs and 3' of the L ORFs correlated with mRNA polyadenylation sites used on the BPV-1 genome (6).

Several potential splice donor and acceptor sites were found in the DPV genome corresponding to known sites used in BPV-1-transformed mouse cells (35) (Table 3). Use of these RNA processing signals would allow DPV to direct the synthesis of fusion proteins similar to the predicted BPV-1 E6-E7 and E6-E4 proteins (35). The presence of a splice acceptor at the 5' end of E4 is noteworthy because DPV does not contain an AUG initiation codon in the E4 ORF. Therefore, unlike all other papillomavirus genomes sequenced to date, DPV would not be able to direct synthesis of an unfused E4 polypeptide.

DPV DNA sequence homology to BPV-1. The sequence of DPV DNA was compared with the BPV-1 sequence by using a matrix program (26). In this analysis, the degree of homology was assigned a letter code (A, 100%, B, 98 to 99%, C, 96 to 97%, etc.), and a minimum value of 55% was plotted (Fig. 3). The presence of a 45° diagonal indicated that regions of the DPV and BPV-1 genomes representing the E and L ORF clusters are approximately colinear. With the exception of sequences adjacent to the E region, no apparent homology was observed between DPV NCRs and BPV-1 DNA.

To observe the amount of homology between DPV and BPV-1 DNA as a function of genomic position, we converted a similar analysis to numerical output and graphically presented it (Fig. 4). Gaps were introduced in each sequence to maximize the alignment of the regions of homology. Although regions of measurable homology were scattered throughout the genome, they were, in general, confined to ORF clusters. The highest homologies (greater than 80%) were located in the 3' end of the E1 ORF and throughout the L1 ORF. This corresponded to two regions of homology in DPV DNA (bp 1,993 to 2,566 and 5,582 to 7,982) which were mapped by filter hybridization of BPV-1 DNA to restriction fragments of DPV DNA under stringent hybridization conditions (data not shown).

Significant nucleotide divergence from BPV-1 was found

in the E6, E7, and L2 ORFs. In addition, the middle of the E2 ORF (containing the E4 ORF) showed greater divergence than either the 5' or 3' ends. The two NCRs failed to show measurable sequence homology even though the minimum value plotted was reduced to 40%. The lack of nucleotide sequence homology in the E6 and E7 ORFs was also reflected in the amino acid sequences of the predicted polypeptides, which showed only 36 and 29% overall homology with the BPV-1 E6 and E7 ORF products, respectively. The Cys-X-X-Cys motif reported for other papillomavirus sequences (4) was also present in the DPV genome. The amino acid sequence of the E6 ORF contained four Cys-X-X-Cys patterns separated by 29, 35, and 29 amino acid residues, respectively. The E7 polypeptide contained two Cys-X-X-Cys repeat units separated by 29 amino acids. Thus, despite the low amino acid homology to BPV-1, structural features encoded in the E6 and E7 ORFs of the DPV genome appear to be highly conserved.

DISCUSSION

The genetic organization of DPV exhibits a typical papillomavirus architecture in that the sense strand contains two clusters of overlapping ORFs. Although the organization of these individual DPV ORFs within the E and L regions are colinear with BPV-1, a major difference in the E region was the lack of E5 and E8 ORFs in DPV. Because these ORFs do not appear to be conserved among the papillomaviruses (31), the putative E5 and E8 gene products may not be essential for papillomavirus DNA replication or for virus production. The DPV genome contains a unique ORF (L3) in the L region. This ORF has not been found in other papillomaviruses, and thus it seems probable that DPV L3 is not a functional coding sequence.

The regions of measurable sequence homology between DPV and BPV-1 were most prominent in the two coding regions. The greatest degree of homology (greater than 90%) was found in the 3' half of the E1 ORF. The BPV-1 E1 ORF encodes a *trans*-acting DNA replication function that maintains extrachromosomal viral DNA in BPV-1-transformed mouse cells (21). Since DPV DNA sequences remain episomal in virus-transformed mouse cells (19), it is likely that the DPV E1 ORF functions in autonomous viral DNA replication. Furthermore, the 3' half of the BPV-1 E1 ORF exhibits striking homologies to nucleotide-binding domains of the simian virus 40 large T antigen (3), which suggests that the gene products of both the BPV-1 and DPV E1 ORFs exhibit functions similar to each other and to that of the simian virus 40 large T antigen, based on the high degree of amino acid homology observed. The L1 ORF, which encodes at least a part of the major capsid protein (6), also shows a high degree of sequence homology to the BPV-1 L1 ORF. The high nucleotide and amino acid sequence homology in this region of the genomes may reflect structural constraints of the major capsid protein necessary for protein-protein interactions that maintain the integrity of the virus capsid. Additionally, this homology could reflect tissue receptor specificity, since both viruses infect fibroblasts in the natural host.

The transformation potential of BPV-1 DNA in mouse cells has been extensively investigated. Sequences at the 3' end of the E ORF cluster, including the E2, E4, and E5 ORFs, are capable of transforming both NIH 3T3 and C127 mouse cells in culture (24, 29). Recently, a second class of BPV-1 transforming genes has been identified in the BPV-1 E6 ORF (30, 35). The activated E6 transforming function is

FIG. 3. Matrix plotting of DNA sequence homologies between DPV and BPV-1. DPV is on the horizontal axis and BPV-1 is on the vertical axis. The genetic organization of each virus is shown for orientation. We compared sequences by using the forward hash matrix program described by Pustell and Kafatos (27). Range, 20 (41-bp window); compression, 75; minimum value plotted, 55%.

restricted to C127 cells and does not transform NIH 3T3 mouse cells. Therefore, it appears that C127 and NIH 3T3 cells can respond to different transforming functions. In considering DPV transformation, the putative transforming ORFs (E2, E4, and E6) showed more nucleotide divergence with BPV-1 than was found in the 3' half of the E1 ORF and the L1 ORF. Because DPV transforms NIH 3T3 mouse cells but not C127 mouse cells, the putative DPV transforming gene products may function specifically in NIH 3T3 cells. Another possibility is that the DPV transforming functions represent a distinct third class of papillomavirus transforming genes. This seems unlikely, however, since the DPV genome does not contain a unique ORF (except L3) which would be capable of encoding a transforming protein which has no counterpart in BPV-1.

DPV varied significantly from BPV-1 in the intergenic NCR regions. The DPV intergenic NCR-1 region was considerably smaller than the corresponding BPV-1 region (646 versus 901 bp). Conversely, the DPV NCR-2 region was large (602 bp) and contained no identifiable ORF. The corresponding region in BPV-1 is significantly smaller (349 bp) and contains the 3' half of the E5 ORF. This BPV-1 E2-L2 intergenic region appears to play a necessary role in cell transformation, since a BPV-1 mutant in the E5 ORF (4-bp deletion) has been constructed which fails to transform either C127 or NIH 3T3 cells (D. E. Groff and W. D. Lancaster, unpublished observations). It is not known whether the BPV-1 E5 ORF is translated. An RNA species containing E5 ORF sequences as part of the 5'-proximal coding region has not been identified in transformed cells (6,

FIG. 4. Distribution of DNA sequence homologies between DPV and BPV-1. The DPV and BPV-1 genomes are aligned, and the amount of homology is plotted for each successive 60 bp. Gaps were introduced in each sequence to maximize the homologies. Sequences were compared by using the forward hash matrix program of Pustell and Kafatos (27) with the output converted to numbers for plotting. Range, 30 (61-bp window); compression, 60; minimum value plotted, 40%.

13). However, it is possible that this RNA is a rare transcript and has escaped detection. Alternatively, it could be that the E5 ORF is translated from known viral transcripts and that a small E5 polypeptide is the product of a functionally polycistronic eucaryotic mRNA. If NCR-2 is not translated, a *cis*-acting role of this region must be considered. Since the DPV NCR-2 region is T rich, it could function by stabilizing mRNA necessary for expression of the transforming function, possibly by allowing duplex formation with the 3' polyadenylic acid. It is also conceivable that the sequences of specific sites in this region play a role in the correct processing and transport of viral mRNA from the nucleus to the cytoplasm. Determining the roles of these sequences at the 3' end of the early transcription region may be essential in understanding the regulation and function of papillomavirus genes in cellular transformation.

ACKNOWLEDGMENTS

We thank Judith Abraham for her help with the Maxam-Gilbert sequencing technique, and Jackie Zusi for technical assistance.

This work was supported by Public Health Service grants CA32603 and CA32638 from the National Institutes of Health.

LITERATURE CITED

- Ahola, H., A. Stenlund, J. Moreno-Lopez, and U. Pettersson. 1983. Sequences of bovine papillomavirus type 1 DNA—functional and evolutionary implications. *Nucleic Acids Res.* **11**:2639–2650.
- Chen, E. Y., P. M. Howley, A. D. Levinson, and P. H. Seeberg. 1982. The primary structure and genetic organization of the bovine papillomavirus type 1 genome. *Nature (London)* **299**:529–534.
- Clertant, P., and I. Seif. 1984. A common function for polyoma virus large-T and papillomavirus E1 proteins? *Nature (London)* **311**:276–279.
- Danos, O., L. W. Engel, E. Y. Chen, M. Yaniv, and P. M. Howley. 1983. Comparative analysis of the human type 1a and bovine type 1 papillomavirus genomes. *J. Virol.* **46**:557–566.
- Dvoretzky, I., R. Shober, S. K. Chattopadhyay, and D. R. Lowy. 1980. A quantitative *in vitro* focus assay for bovine papilloma virus. *Virology* **103**:369–375.
- Engel, L. W., C. A. Heilman, and P. M. Howley. 1983. Transcriptional organization of bovine papillomavirus type 1. *J. Virol.* **47**:516–528.
- Fitzgerald, M., and T. Shenk. 1981. The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. *Cell* **24**:251–260.
- Gingras, T. R., D. Sciaky, R. E. Gelinas, J. Bing-Dang, C. Yen, M. Kelly, P. Bullock, B. Parsons, K. O'Neill, and R. J. Roberts. 1982. Nucleotide sequences from the adenovirus-2 genome. *J. Biol. Chem.* **257**:13475–13491.
- Groff, D. E., and W. D. Lancaster. 1984. Evidence that integration of virus DNA may not be necessary for maintenance of cell transformation. *Prog. Med. Virol.* **29**:218–230.
- Groff, D. E., J. P. Sundberg, and W. D. Lancaster. 1983. Extrachromosomal deer fibromavirus in deer fibromas and virus-transformed mouse cells. *Virology* **131**:546–550.
- Grunstein, M., and D. Hogness. 1975. Colony hybridization: a method to isolate cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci. U.S.A.* **72**:3961–3965.
- Guo, L.-H., R. C. A. Yang, and R. Wu. 1983. An improved strategy for rapid direct sequencing of both strands of long DNA molecules cloned in a plasmid. *Nucleic Acids Res.* **11**:5521–5540.
- Heilman, C. A., L. Engel, D. R. Lowy, and P. M. Howley. 1982. Virus-specific transcription in bovine papillomavirus-transformed mouse cells. *Virology* **119**:22–34.
- Hong, G. F. 1982. A systematic DNA sequencing strategy. *J. Mol. Biol.* **158**:539–549.
- Jenson, A. B., R. J. Kurman, and W. D. Lancaster. 1984. Human papillomaviruses, p. 951–968. In R. B. Belsche (ed.), *Textbook of human virology*. PSG Publishing Co., Inc., Littleton, Mass.
- Koller, L. D., and C. Olson. 1972. Attempted transmission of warts from man, cattle, and horses and of deer fibroma, to selected hosts. *J. Invest. Dermatol.* **58**:366–368.
- Lancaster, W. D., and C. Olson. 1982. Animal papillomaviruses. *Microbiol. Rev.* **46**:191–207.
- Lancaster, W. D., and J. P. Sundberg. 1982. Characterization of papillomaviruses isolated from cutaneous fibromas of white-tailed deer and mule deer. *Virology* **123**:212–216.
- Law, M.-F., D. R. Lowy, I. Dvoretzky, and P. M. Howley. 1981. Mouse cells transformed by bovine papillomavirus contain only extrachromosomal viral DNA sequences. *Proc. Natl. Acad. Sci. U.S.A.* **78**:2727–2731.
- Lowy, D. R., I. Dvoretzky, R. Shober, M.-F. Law, and P. M. Howley. 1980. *In vitro* tumorigenic transformation by a defined sub-genomic fragment of bovine papilloma virus DNA. *Nature (London)* **287**:72–74.

21. Lusk, M., and M. R. Botchan. 1984. Characterization of the bovine papilloma virus plasmid maintenance sequences. *Cell* **36**:391-401.
22. Maniatis, T., E. F. Frisch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. Maxam, A., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**: 499-560.
24. Nakabayashi, Y., S. K. Chattopadhyay, and D. R. Lowy. 1983. The transforming function of bovine papillomavirus DNA. *Proc. Natl. Acad. Sci. U.S.A.* **80**:5832-5836.
25. Pfister, H. 1984. Biology and biochemistry of papillomaviruses. *Rev. Physiol. Biochem. Pharmacol.* **99**:111-181.
26. Pustell, J., and F. Kafatos. 1982. A high speed, high capacity homology matrix: zooming through SV40 and polyoma. *Nucleic Acids Res.* **10**:4765-4782.
27. Pustell, J., and F. Kafatos. 1984. A convenient and adaptable package of computer programs for DNA and protein sequence management, analysis, and homology determination. *Nucleic Acids Res.* **12**:643-655.
28. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **74**:5363-5367.
29. Sarver, N., M. S. Rabson, Y.-C. Yang, J. C. Byrne, and P. M. Howley. 1984. Localization and analysis of bovine papillomavirus type 1 transforming functions. *J. Virol.* **52**:377-388.
30. Schiller, J. T., W. C. Vass, and D. R. Lowy. 1984. Identification of a second transforming region in bovine papillomavirus DNA. *Proc. Natl. Acad. Sci. U.S.A.* **81**:7880-7884.
31. Schwarz, E., M. Durst, C. Demankowski, O. Lattermann, R. Zech, E. Wolfsperger, S. Suhai, and H. zur Hausen. 1983. DNA sequence and genome organization of genital human papillomavirus type 6b. *EMBO J.* **2**:2341-2348.
32. Shope, R. E., R. Mangold, L. G. MacNamara, and K. R. Dumbell. 1958. An infectious cutaneous fibroma of the virginia white-tailed deer (*Odocoileus virginianus*). *J. Exp. Med.* **108**: 797-802.
33. Stenlund, A., J. Moreno-Lopez, H. Ahola, and U. Pettersson. 1983. European elk papillomavirus: characterization of the genome, induction of tumors in animals, and transformation in vitro. *J. Virol.* **48**:370-376.
34. Watts, S. L., W. C. Phelps, R. S. Ostrow, K. R. Zachow, and A. J. Faras. 1984. Cellular transformation by human papillomavirus DNA in vitro. *Science* **225**:634-636.
35. Yang, Y.-C., H. Okayama, and P. M. Howley. 1985. Bovine papillomavirus contains multiple transforming genes. *Proc. Natl. Acad. Sci. U.S.A.* **82**:1030-1034.