Biologic Properties and Nucleotide Sequence Analysis of Human Papillomavirus Type 51

OCTAVIAN LUNGU,¹ CHRISTOPHER P. CRUM,² AND SAUL SILVERSTEIN^{1*}

Department of Microbiology and Cancer Research Center, Columbia University, 701 West 168th Street, New York, New York 10032,¹ and Department of Pathology, Brigham and Womens Hospital, Boston, Massachusetts 02115²

Received 25 March 1991/Accepted 16 May 1991

Human papillomaviruses (HPVs) may be grouped according to the site from which they are isolated and the disease with which they are associated. We recently identified and cloned HPV type 51 (HPV-51) from a low-grade precancerous lesion (G. Nuovo, E. DeVilliers, R. Levine, S. Silverstein, and C. Crum. *J. Virol.* 62:1452-1455, 1988). Molecular epidemiologic analysis of cervical lesions, including condylomata and low- and high-grade precancers, revealed that HPV-51 was present in about 5% of the samples we examined. We have now determined the complete nucleotide sequence of this virus and compared it with other sequenced HPVs. Our analysis reveals that the 7,808-bp genome is composed of eight open reading frames which are encoded on the same strand and that this virus is most closely related to HPV-31. Sequence comparisons place this virus in the group of high-risk viruses (those with an increased risk of progressing to malignancy) along with HPV-16, -18, -31, and -33. Morphologic transformation experiments demonstrated that HPV-51 had transformation potential and that transformed cells contained RNAs homologous to E6 and E7.

Human papillomaviruses (HPVs) are small icosahedral viruses that contain a double-stranded DNA genome of approximately 8 kbp. The genomes of more than 60 independent HPVs have been isolated as molecular clones. A new isolate is designated a novel strain if it is $\leq 50\%$ homologous to any other known HPV as determined by liquid hybridization analysis (13). The best criteria for the classification of novel isolates are direct DNA sequence analysis and comparison of the amino acid coding potential of novel isolates with known oncogenic and benign strains. Members of this virus family are the causative agents for a variety of benign lesions that result from abnormal epithelial cell differentiation, and HPVs are frequently subgrouped according to their association with cutaneous or mucosal sites. Some of these viruses are associated with genital neoplasias which range in severity from mild dysplasia to frank carcinoma (7, 24, 27, 38, 70, 71). The ability of specific virus early gene products to interact with host proteins such as p53 and RB105 (25, 43, 54, 66) may be the molecular basis that differentiates viruses with the ability to cause malignant disease from those which cause benign papillomas.

We previously used Southern blot hybridization, in a molecular epidemiologic study, to correlate HPVs with their segregation patterns according to the pathology of the lesions at genital sites (18). In this study several HPV DNAs with novel restriction endonuclease digestion profiles were detected. One of these DNAs, originally identified in a condyloma as an ~ 8 kbp *HindIII* fragment, was cloned and used to screen a panel of virus DNAs. Molecular hybridization analysis revealed that this DNA genome represented a novel isolate which was designated HPV type 51 (HPV-51) (47). The prevalence of HPV-51 in a small sampling of biopsies derived from patients with abnormal Papanicolaou smears was examined, and it was shown to be present in 7 of 127 samples derived from patients with condylomata, cervical intraepithelial neoplasia, or carcinoma. A more recent prevalence study using polymerase chain reaction (PCR)

analysis suggested that this virus was present in a percentage ($\sim 3\%$) of the population similar to that of other virus types (e.g., HPV-6, -16, -18, -31, and -33) which were more frequently associated with genital lesions (4).

We have determined the complete nucleotide sequence of HPV-51 and compared the arrangement and homologies of its open reading frames (ORFs) with those of other genital isolates. Transfection of HPV-51 DNA into either the BALB 3T3 or the CREF rodent fibroblast cell line results in morphological transformation of these cells. The HPV-51-transformed lines express RNAs that have the potential to encode the E6 and E7 oncoproteins.

MATERIALS AND METHODS

Cloning and sequence analysis. The HPV-51 genome as described in Nuovo et al (47) was excised from vector DNA by digestion with *HindIII*, and the 7.6-kb virus band was purified from a 0.5% agarose gel by solubilizing the gel and binding the DNA to glass powder (64). The DNA was cleaved with *PstI*, and the three *PstI* and the two *HindIII*-*PstI* fragments were subcloned into M13mp18 and -19. Nucleotide sequence was determined by the dideoxy chain terminator method (53) both manually and by using an Applied Biosystems automated DNA sequencer. Where necessary the length of the insert was pared with exonuclease III, and the family of nested deletions that resulted was sequenced. In several instances oligonucleotides that flanked regions for which no sequence information was available were prepared and used to determine the intervening sequence. Sequences were compiled and analyzed by using a commercially available sequence analysis package, DNASIS, from LKB.

The 232-bp *HindIII* fragment which encodes the E6 ORF was cloned after amplification of DNA isolated from two biopsies containing HPV-51 by using primers that flanked the *HindIII* site. Isolated DNA was heat denatured in the presence of the following primers: 5'-AAGTATAGAAGAA CACCATGT, which is complementary to the anticoding strand at positions 80 to 99, and 5'-CTTTGACATCTATGA

* Corresponding author.

FIG. 1. Sequencing strategy and ORFs of HPV-51. The M13 clones that were used to determine the sequence of the HPV-51 genome are shown beneath a linear representation of the genome. Arrows without names refer to oligonucleotides that were used as primers to "walk" through regions of the sequence for which no clones were available. Fragments F1 to F6 represent *Pst*I-*Pst*I or *Pst*I-*Hind*III subclones that were used to generate the various subclones in M13. The letters above the line refer to restriction endonuclease sites for the following enzymes: *H*, *Hind*III; *Spe*, *Spe*I; *P*, *Pst*I; *Rv*, *Eco*Rv; *M*, *Mlu*I; *Pv*, *Pvu*II; *B*, *Bam*HI; and *Hc*, *Hinc*II. The circles represent the potential or verified E2 binding sites, and the squares represent potential polyadenylation sites. The lower part of the figure depicts the location and boundaries of the various ORFs of HPV-51. The distribution of termination codons (long vertical lines) and initiation codons (short vertical lines) is shown.

CACCT, which is complementary to the coding strand at nucleotides 419 to 400. This DNA was subjected to 35 cycles of denaturing, annealing, and extension at 95, 55, and 72°C, respectively, with *Taq* polymerase in a Perkin-Elmer-Cetus Thermocycler. The resulting product was gel purified and cloned into the filled-in *Bam*HI site of pIBI31. The sequence was determined from double-stranded plasmid DNA by using the primers synthesized for amplification and T7 DNA polymerase.

Morphologic Transformation. Continuous rat embryo fibroblasts (CREF) (26) were obtained from Paul Fisher, and BALB 3T3 cells were obtained from Mitchell Goldfarb. These cells were maintained at subconfluence in Dulbecco's modification of Eagle medium containing 5% fetal calf serum and plated at 7×10^5 cells per 5-cm dish prior to transfection. Each dish of cells was transfected with 3.3 μ g of linearized HPV-51 or -18 DNA and 6.7 μ g of salmon sperm carrier DNA or with only 10 μ g of salmon sperm DNA, as previously described (67). Two days posttransfection each dish of cells was split 1:3, and the cells were fed fresh medium every third day. After 21 days, foci of transformed cells were photographed, cloned, and counted.

PCR amplification and RNA analysis. DNAs from biopsies, plasmids, or transformed cells were tested for HPV DNA by PCR amplification of the L1 region with the MY09 and MY11 primers described by Manos et al. (37). These DNAs were examined for the presence of HPV DNA ho-

mologous to the E6 ORF by using primers complementary to the coding strand (RT primer) 5'GGGAATTCCTTCA CAGTCCATCGCCGTTG and anticoding strand (5' E6 and E7 primer) 5'GGGGGATCCAACACCATGTTTCGAA GACAAG, respectively. These oligonucleotides hybridize to HPV-51 DNA at nucleotides 887 to 866 and 91 to 111. RNA synthesized in transformed cells was analyzed by primer extending 25 μ g of RNA with the RT primer and subsequently isolating the cDNA product and subjecting it to 35 cycles of amplification with the RT and 5' E6 and E7 primers. The product was extracted with phenol and chloroform-isoamyl alcohol and precipitated with ethanol. One-fifth of the purified product was then electrophoresed on a 3% GTG agarose gel that consisted of 3:1 NuSieve to SeaPlaque in Tris-borate buffer. The resolved products were isolated from the gel as previously described (64) and analyzed for the presence of sequences homologous to E6 or E7 by slot blot hybridization with 32 P-labeled probes that were specific for one or the other of these genes.

Nucleotide sequence accession number. The HPV sequence in this paper was assigned GenBank accession number M62877.

RESULTS AND DISCUSSION

Nucleotide sequence of the HPV-51 genome. To further characterize the HPV-51 genome, a detailed restriction

1 AACAAATTATC TTGTAAAAAC TAGGGTGTAA CCGAAAAGGG TTATGACCGA AAACGGTGCA TATAAAGTGC CAGTGGTAAA AGTATAGAAG AACACCATGT
101 TCGAAGACAA GAGGGGAAAG CCACGAACGC TGCATGAATT ATGTGAAGCT TTGAACGTTT CTATGCACAA TATACAGGTA GTGTGTGTGT ATTGTAAAAA
201 GGAATTATGT AGAGCAGATG TATATAATGT AGCATTTTAC TGTATATAGG GGATAAATAAT CCATATGCCAG TATGCAAAAC ATGTTTACTG
301 TTTTATTCAA AAATTAGAGA GTATAGACGT TATAGCAGGT CTGTGTATGG TACTACATTA GAGGCAATTA CTAAAAAAG CTTATATGAT TTATCGATAA
401 GGTGTATAG ATGTCAAAGA CCACCTGGGC CTGAAGAAAA GCAAAAATTG GTGGACGAAA AAAAAAGGTT CCATGAAATA GCGGGACGTT GGACGGGGCA
501 ATGCGCTAAT TGCTGGCAAC GTACACGACA ACGTAACGAA ACCCAAGTGT AATAAAGCCA TCGGTGGTAA TGTACCACAA TTAAGAATG TAGTATTGCA
601 TTTAACACCA CAGACTGAAA TTGACTTGCA ATGCTACGAG CAATTTGACA GCTCAGAGGA GGAGGATGAA GTAGATAATA TGCCTGACCA GCTACCAGAA
701 AGACGGGCTG GCAGGGCTAC GTGTTACAGA ATTGAAGCTC GTGTTTCAGT GTGTTTCAGT GTAGTACAA TGGCAGTGGG AAGCAGTGGG GACACCTTC
801 GCGTTGTACA GACAGGTGTA ATGGGCGAAC TAAGCTGGGT TTGCCCCGTG TGTGCGAACA ACTAGCAACG GCGATGGACT GTGAAGGTAC AGAGGATGAG
901 GGGGGGGGGT GTAATGGGTG GTTTTTTGTG GAAGCAATAG TAGAAAAAAA AACAGGAGAT AATGTTTCGG ATGATGAGGA TGAATAATGCA GATGATACAG
1001 GATCTGATTT AATAAATCTT ATAGATAGTG AAACCTAGTAT TTGCAGTCAG CCGGAAACAGG AGACAGCACG GCGCTGTGTT CAGGCCCAAG AATTACAGGC
1101 AAACAAAGAG GCTGTGCATC AGTTAAACAG AAAGTTTCTA GTCAGCCCGC GAAGCAGCCC ATTAGGAGAC ATTACAAATC AAAACAACAC ACACAGCCAT
1201 AGTCAGGCAA ACGAGTCACA AGTTAAAGAG AGATTACTGG ACAGTTATCC GGACAGCGGA TATGGCAATA CACAAGTGGG AACTGTGGAA GCAACGTTGC
1301 AGGTAGATGG GCAACATGGC GGTTCACAGA ACGTGTGTG TAAGCTGGGT TTGCGAGGTT TATGGATGTT GGAAACAACA GAAAGCTGTG CAAATGTAGA
1401 ACTAAACAGT ATATGTGAAG TATTAATAAG CAGTAATGCA AAAGCAACGT TAATGGCAAA ATTTAAAGAG TTGTATGGTA TTAGTTATAA TGATGTGGTA
1501 CCGGTGTGTTA AAGTGTGATA AACATGTTGT ATAGATTTGG TTTGTGCTAT GTTTGCGGTT TCCCAATGTT TAGCAGAAAT TTAATAACA CTAATTAAGC
1601 CATTTTGCAT GTACTACCAT ATACAATGTT TATCATGTGA TTGGGGCACC ATTGTATTAA TGCTAATTAG GTTTTCATGT GCAAAAAACA GAACAACAAT
1701 TGCTAAGTGT TTAAGTACAT TAGTAAATAT CCCACAATCA CAAATGTTTA TAGAACACCC AAAAATTACGT AGTACACCTG TGCCATTATA TTTTATTAGA
1801 ACAGGCATAT CAAACATTAG CAATACATAT CTGAATGGAT TACACGACAA ACGCAACTAC AGAGCATGTT TGAGGATAGT ACCTTTGAAT GATTTAGAGC
1901 TATCACAAT GGTGCAATGG GCATTTGACC ATGAAGTATT AGATGATAGT GAAATAGCAT TTCATTATGC ACAATTAGCA GATATAGATA GTAATGCTGC
2001 AGCGTTTTTA AAGAGTAATT GCCAAGCAAA ATATGTAAAG GATTGTGGGA CCATGGCAGC GCATTACAAA CGAGCACAAA GAAAAATCATT ATCTATGCTA
2101 GCGTGGATAA GGTATAGATG TGATAGAGCA AAGGATGGAG CCAACTGAGG AGAATTTGCT GATATCAAGG GATATCAAGG AAGAGCTGTG CAAATGTAGA
2201 TTTCAATGTT TAAACAGTTT TTAAGAGGAA CACCAAAACA CAATTGCATA GTCATATATG CCCCACCAAA CACAGGCAAG TCATTATTTG CAATGAGGCT
2301 AATGAAGTTT ATGCAAGGTT CCATTATTTT ATATGTAAAC TCTGTGATGC ATTTTGTGTT TACGAGCTAT ACAGCCACTA GAGGATGCTA GTTAGATGAT
2401 GCTACGTATG GGTGTGGAC ATATATTGAT CAGTATTTAA GAACTTTTTT AGATGGTAAT CCATGTAGTA TAGATAGAAA ACATAGGAGT TTAATACAAAT
2501 TAGTATGCTC ACCATTACTA ATACAGTCAA ACATAATCC AAGAAGGAGT GCAACCTTAA TGTATTACA TACAAGGGTA ACAGATTAA AGTTTTTAA
2601 TACATTTTCC TTTGATACAA ATGGGAATGC TGTGTATATG TGAATGATG TCAATGTGAA AATTTTTTTT TCCACCAATG GGTCCAGATT AGTTTGGAG
2701 GAGGAAGAGG ACAAAGAAAA TGGAGACCT ATGCCACCGT TTAATGTGTG GCCAGGAGAA AATACTAGAC TGTATTGAAC TGGACAGTGA TAAATTAGTA
2801 GATCAAAATA ACTATTGGAC ATTGTTACGA TATGAAGCTG CTATGTGTTT TGCAGCAGCG GAAAGAAACT TACGAACAAT CAATACCAGG TAGTACCAG
2901 CAACCAACAGT ATCAAAACAA AAGGCTGTGC AAGCAATTGA ATGTGCATG GCCTTACAA CGCTTACAA ATCAGACTAT AACATGGAAC CATGGACAAT
3001 GCGGGAGACA TGTATGAAC TATGGTGTGT GCGTCCCAAG CAATGTTTCA AAAAGGGGGG CATAACTGTA ACAGTTATAT TTGATGGAAT TAAGGACAAT
3101 GCAATGGACT ATACAAAGCTG GAAATTTATA TATATATATG ATAATGATA GTGGTATGAA ACAAATGGAA ATGTGAGCTA TACGGGTATA TATTACACTG
3201 TAAATTCAAA AAAAGAATAT TATGTACAGT TTAAGATGA AGCCAAAATA TATGGGGCAC AACAGTGGGA GGTCTATATG TATGGTACTG TAATAACATG
3301 TCCTGAATAT GATCTAGTA CCTGCAGCGA CCGGTTATCC ACTACTACAA CTGTGAAACA ACTATCAAA ACCCCAACGA CCAATCCCCT TACCACCTGC
3401 GTGGGCGCCA ATGAAGCCCA GACACAACAG CCGAAACGAC AGCACTTAC TGAGCCCGAC TCCTCCACAA TCTCCCACTG TCCGCTGGAC AATACAAACA
3501 ACCAAATACA CTGTGGAGAT GGAAGCACTA ACACCTGGAG GACCAAACTG GCAACTCAGA CTGCGTTTAT AGTGCAATTA AAGGTGGATA CAAATGTGTT
3601 AAAATGTTTT AGATACAGAT TTACAAAACA CAAAGGGTTA TATCCTAAC CTGCGATTGG ACCAGTAAAT TAAAGCAAGG CATTGTTACC
3701 ATTGTGTTTG ACAGTGCACA TCAACGGGAA ACATTTATAA AAACCATTAA AGTACCCCCA AGTGTAAAC TGTCATTGGG AATTATGACA CTGTAACATG
3801 TGTAATATAT GTATTGTACA TATATATCTG TCAAGGCTAA TATGTGCTGC TAAATGTATA GACATATTGT AACCATTTGA GTGTATTATA TTTTGCTATT
3901 TGTGCTTTGC TTGTGTGTGT GTCTTTGTGT GTGTTGTTTG TTGCGCTTAC ATACGTGTTT GAGCTGCTCT TATTATTATT TTTATGTTTT
4001 TGGTTTGTG TTGCAACATC CCAATTAAC ACATTTTGTG TATATTGAT TTTTTTTAC TTACTTGTGT TACTTTTACA TCTATATACA TTTTACTTT
4101 TGCAATAAAC TTGTATATT TTTGTGATTA AATATGGTGC CTACAGCTGC ACGGCTCGG AAGCGAGCAT CTGTAAACA ATTATTTCT ACATGCAAG
4201 CTGCTGGTAC ATGTTCTCCT GATGTTGTGA ATAAGGTTGA AGGTACTACA TTGGCCGATA TGTGCTGCTA TGTGCTGCTA TGTGCTGCTA TTTTGGTGG
4301 CCTAGGTATT GGTACTGGGT CTGGATCTGG GGGGCGTACT GGATATATCC CTTTAGGTGG TGGGGGTCGC CCAGGCTGGG TGGATATTG TCCCTCAAGG
4401 CCACCTTATA TAAATGACCT ATGGCACCAT CAGCAATACCT CTATAGTAA TTTTGTGAG GACTCTAGTA TTAATCAGT TGGGTCTCCT ATACCTACCT
4501 TTACTGGTAC CGATGGCTTT GAAATTACTT CATCTTCCAC AACAAACCCCT GCTGTGTTGG ACATCACCCC ATCTGCTGCT ACCTGATACAT TTTCTAGTAC
4601 TAACATTGAA AATCCTTTAT ATATTGAAC TCCATCCATT GAGGCTCCAC AATCTGGAGA AGTGTACAGT ATATATTAC TAGTACACTA CTCTGGTACT
4701 CATGGGTATG AAGAAATACC TATGGAAGTG TTTGATCCCA TTGCGATGAC TGACAGCAAG GTGTATTGCG CACTGTGCTA TTTGCTGCTA CAGAGAATA
4801 CTGCTCCCCG CTTGTATAGT AAGTCTCTACA CACAGGTAAA AGTATCAAA CTCTGATTTA TTAGTAAGCC ATCCACATTT GTTACATTTA ATAACTCTGC
4901 TTTTGAGCCT ATTCACATAT CCATAACTTT TGAGGAACCT GATGCTGTTG ACCTGTATCC CACTGTATCC GATATTATTA GATATTATTA CACTGACCC
5001 ACATCTCGTA GAGGCACAGT ACGCTTTAGT AGGTTAGGTC AAAAGGCCAC CATGCGCACT CGTAGTGCCA AACAAATTGG TGCTCGTGTA CATTATTATC
5101 ATGATATTAG TAGAATTGCA CAGCTGTATG AACTTGAATG CGAGCCTTTA CTTTCACTTT CTAATAATTA TAGTATTAGC ATTTATGCTG ATTTAGATGA
5201 AGCTGAAACA GGTTTTATAC AGCCACACACA CACCAACCT ATGTACACT CTCTTTGTG TAGGCAGTTG CCCCCTTTAT CTTTCTCTAT GTCTTACTCT
5301 TATGCAATG TTACTATTCC ATTTTCAACT ACATATTCTG TTCTATTCA TACAGGCGCT GATGTGGTAT TGCCCAATC TCCTACAGTA TGGCCTTATG
5401 TTCCCAACAC TTCCATTGAC ACCAAGCATT CTATTGTTAT ACTAGGTGGG GATTACTATT TGTGGCCCTA TACACATTTA CTACGCAAC GCCGTAAACG
5501 TATACCCTAT TTTTGTACAG ATGGCATTGT GGGCGACTAA TTGACAGCAAG GTGTATTGCG CACCTGTACC TGTGTCTCGA ATTTGTAATA CAGAAGAATA
5601 TATCACACGC ACCGGCATAT ATACTATGC AGGCAGTTCC AGACTAATA CATTAGGACA TCCCTATTTT CCAATACCTA AAACCTCAAC CGGTGCTGCT
5701 ATTCCTAAG TATCTGCATT TCAATACAGG GTATTTAGG TACAGTTACC AGATCCTAAC CAGTTTGGAC TCCCGGATCC AAATTTATAT AATCCAGACA
5801 CAGATAGGTT GGTGTGGGTG TGTGTGGCG TTAGGTTGGG CAGAGGACAG CCCCTTGGTG TTGGCCTTAG TGGTCTATCC TTTATTAATA AATATGATGA
5901 CACGAAAAAT TCAGGCATAG CAAATGGCAA TGCACAACAA GATGTTAGAG ATAAACATCT TGTGTGACAA AACACAGACT AGTTATGTAT AATAGGCTGT
6001 GCTCCACCTA TTGGGGAACA CTGGGGTATT CCGCAATACCT ACCTGTACTC CCAGGAGACT GCGCCCCCTT GGAACCTGTA TCCCTGTCTA
6101 TTCAGGATGG CGATATGAT GATACAGGGT TTGGAGCTAT GGATTTGCTT GGCCTACAGC CCACCAAAATC AGACGTCCTT TTGATATTTT CACAGTCTGT
6201 TTGTAATAT CCTGATTATT TAAAAATGTC TGCAGACACA TATGTTAATT CCATGTTTTT TCATTTACGC AGGAGCAAAA TCTTTGCTAG GCATATTAT
6301 AATAAACTTG TAGGTGTTGG GGAAGACATT CTAACGATT ATTTATTTAA GGGTAGTGGT AATGGCCGTG ACCCTATAGA AAGTTATATA TACTCTGCTA
6401 CTCCAGTGG GTCTATGATA ACATCTGATT CTCAAATTTT TAATAAGCCT TATTGGCTCC ACCGTGCGCA GGGTCACAA TATGGCATTT GCTGGAACAA
6501 TCAGCTTTTT ATTACCTGTG TTGATACTAC CAGAAGTACA AATTTAATCA TTAGCATGCG CACTGCTGCG GTTTCACCAA CATTACTCTT CAGTAACCTT
6601 AAGCAATATA TTAGCATGAG GGAAGAGTAT GAATTGCAT TTTATTTTCA ATTATGAAA ATTACTTTAA CTACAGAGGT AATGGCTTAT TTACACACAA
6701 TGGATCCTAC CATTCTTGAA CAGTGGAAAT TTGGATTAACT ATTACCCTCC TCTGCTAGTT TGGAGGATGC ATATAGGTTT GTTAAAAATG CAGCTACTAG
6801 CTGTCAAAAG GACACCCCTC CACAGGCTAA GCGAGATCCT TTGGCCCAAT ATAAATTTTG GGATGTTGAT TTAAGGAAAC ATTTTCTTTT AGATTATTAGC
6901 CAATTTGCAAT TGGGTGCGAA GTTTTTGTG CAGGTGGGCG TACAAGCAAA GCCCAGACCA GGCCTTAAAC GCGCGGCTC ATCGGCATCC TCTTCTCTT
7001 CCTTTTCAGC CAAAGCTAAA CGTGTAAAAA CAGTAATGAT TTTAGTTTTT TATGCTTTGT TATGCTTTGT GATGCTGCT GTGTGTATGT TGTGTATGT
7101 ACTGTATGTG TTTTGTGTG TGTGTGTGT GTTGTCTCTG TATGTATGAG TTTATGTATG TTTATTTATA TAAACTATGT GGTGTGTGTG TGTGTGTGTT
7201 TGCACTGACTG CATTGTGATG ACATGTACGG GTGTATGTGG GTATTACATT ATCCCGTAG GTCAAGGGTG GTGTGTTGGT GGGCTCCCTA TTGCGCTACC
7301 CATTTTGTGC AGCACAACAG TTTATATTGG TGCTATTATG TTTATCTTTG TTTATCTTTG TTTGTTACAG CTGACAGACT TTTGAGTGA ACAGATTTCG
7401 GTTCGTGTAC TTTTAGTATA TTTGCCAAGT TTTAAACCAT ACCTGCCAGT TGTGTTTGGC ATAAACCATC ATTTTTTAT GACATAGTGC ATACATCCCG
7501 CGGCCACGC CTGTACTTGT GCGCGCTTAA CCGCGCTTAC CACTACAACC TTTATGTCAT TTTGCTTTA ACAATTTGTT GCACACTGTT TCCGCGCTA
7601 TAAATAATTG ATGCTTTATA GGCATGTATT TTTTGGCATA TTTTATCTTA CTAATTGTCAT AGTTGGCAGG TCAAACTATA TGTTTTTAGT GCCAAGTTTC
7701 TATCCTACT ATAAACCATC TTACTCATAT GCAGGTGTGC TACACAAATG TGTACCTTAA CCGATTGTTG TCTGCTCTAT GCTTGCACAA TTTTTCCTTA
7801 TAACATTT

FIG. 2. Complete 7,808-bp sequence of the HPV-51 genome. The position of nucleotide 1 was determined by alignment with HPV-16 DNA.

TABLE 1. Homologies of HPV-51 and other HPV ORFs^a

HPV-51 ORF	% Homology											
	HPV-6		HPV-11		HPV-16		HPV-18		HPV-31		HPV-33	
	NA	AA	NA	AA	NA	AA	NA	AA	NA	AA	NA	AA
E6	59	57	57	56	70	72	58	61	67	73	69	77
E7	58	63	55	63	60	67	62	62	58	64	54	67
E1	66	67	63	67	66	65	63	67	67	74	67	73
E2	61	62	62	62	60	66	62	61	61	65	51	64
E4	58	54	57	54	58	62	65	60	50	60	58	58
E5	58	55	59	56	59	63	62	65	65	66	57	61
L2	58	70	59	70	63	68	66	77	63	70	54	67
L1	66	82	66	80	66	78	66	84	68	83	50	83

^a Sequences were aligned and compared by using the Needleman-Wunsch algorithm (46) as implemented by the DNASIS analysis package running on an MS-DOS-based computer. NA, nucleic acid homology; AA, predicted amino acid homology within the ORF. Complete ORFs were used for comparison of the nucleic acids, and ORFs beginning with the first in-frame ATG codon were used for the amino acids, except for the E4 of HPV-16, -31, and -33 and the E5 of HPV-16 and -51, which have no in-frame ATG; in these instances the complete ORFs were used. Only the E4A ORF of HPV-6 and the 3' E1 ORF of HPV-16 were used for comparison.

endonuclease map of the original *Hind*III fragment was prepared and each of the *Pst*I-*Pst*I and *Pst*I-*Hind*III fragments was subcloned in both orientations in M13 vectors. These subclones were subjected to sequence analysis as described in Materials and Methods. Where the ability to read the sequence was exceeded by the length of the clone, that piece of DNA was subcloned or pared with exonuclease III, or specific oligonucleotide primers were synthesized and used to "walk" through the sequence. The sequencing strategy and clones analyzed are displayed in Fig. 1; more than 95% of the sequence was determined in both orientations. The sequences derived from the original *Hind*III clone were assembled and aligned with the genomes of other HPVs. This analysis revealed that the original clone lacked the potential to encode an E6 protein. To determine if the missing sequence resulted from isolation of a defective virus or if a small *Hind*III fragment had gone undetected in our original analysis, primers which flanked the single *Hind*III site were synthesized and used to amplify DNA from several HPV-51-containing lesions, including DNA from which the original clone was isolated. Each sample tested was shown to contain DNA that served as a template for the synthesis of a 340-bp DNA which released a 232-bp fragment after digestion with *Hind*III. The entire PCR-amplified fragment was cloned and sequenced. Insertion of this information into the original compiled sequence resulted in generation of a viral genome of 7,808 bp.

The complete nucleotide sequence of the HPV-51 genome is shown in Fig. 2. The first nucleotide was assigned on the basis of homology with HPV-16 DNA. The G+C content of this DNA is 39%. Alignment of the nucleotide sequence of HPV-51 with that of the other HPVs for which complete sequence information was available revealed that HPV-51 most closely resembled HPV-31 (Table 1).

ORFs. The pattern of ORFs was determined and shown to be conserved with other sequenced HPVs. All of the ORFs, with the exception of that encoding the E5 homolog, contain an initiating methionine (Tables 1 and 2). A schematic representation of the major ORFs is shown in Fig. 1. Note that all of the major ORFs are derived as a result of translation of mRNAs that would be transcribed from only one strand of the DNA. Comparisons of the amino acids

TABLE 2. ORFs of HPV-51

ORF	Nucleotide position			No. of bases	No. of amino acids ^a
	Start ORF	First ATG	Stop codon		
E6	85	97	550	465	151
E7	479	560	863	666	101
E1	733	874	2776	2,043	634
E2	2690	2720	3794	1,104	358
E4	3294	3309	3570	276	87
E5	3854		4106	255	84
L2	4125	4134	5538	1,413	468
L1	5443	5521	7033	1,590	504

^a Number of amino acid which would constitute the predicted proteins if translation initiates at the first ATG of each ORF, except in the case of E5, which does not have an initiating ATG and therefore begins at the first in-frame amino acid.

encoded by these ORFs allowed us to localize them with respect to their homologs in other HPVs. The first nucleotide of each of the major ORFs along with the positions of the first ATG and stop codon is shown in Table 2. The number of amino acids composing each of the presumptive ORFs is also presented in this table.

The E6 and E7 ORFs are adjacent but in different reading frames, and the E6 stop codon is separated from the first ATG in E7 by 10 nucleotides. E6 from the high-risk HPVs associates with p53 and targets it for the ubiquitin-dependent degradation pathway (54, 66). The E6 ORF in HPV-51 is 151 amino acids and contains two repeats of a Cys-X₂-Cys-X₂₉-Cys-X₂-Cys motif that are separated by 36 amino acids. The sequence and arrangement of this motif, which is involved in coordinating Zn²⁺ (2, 29), is conserved in all sequenced HPVs (14). Cole and Danos (14) proposed that the periodicity of these repeats and the presence of other highly conserved amino acids at regularly spaced intervals suggest that E6 has arisen as a result of a duplication event. In addition to the motifs they describe we note that, among the high-risk types (HPV-16, -18, -31, -33 and -51), there are additional highly conserved blocks of amino acids at invariant distances from the metal-binding domains (Table 3). The role of these amino acids is unknown; however, their conservation in the E6 ORF may potentiate their oncogenicity by targeting p53 for degradation. Furthermore, HPV-51, like HPVs associated with anogenital cancer and in contrast to those associated with benign condylomata, possesses potential splice donors at positions 177 and 219 and a consensus acceptor at 402 within E6 (57, 60, 61). If the donor at 177 is used and if translation initiates at the same AUG codon as for E6, then the resulting product would be a 43-amino-acid E6*-like protein (58). Use of the donor at 219 would require internal initiation to maintain the reading frame and therefore would not conserve the amino terminus of the E6* protein.

The E7 proteins from high-risk viruses associate with p105-RB with high affinity, whereas those from low-risk groups do not (43). This protein family shares significant amino acid homologies, which include the conserved Cys-X₂-Cys-X₂₉-Cys-X₂-Cys motif (14) and a string of seven amino acids (DLXC_HEQ), which in HPV-18 are involved in binding pRB-105. Mutagenic analysis of this region of the HPV-16 E7 ORF demonstrated that the amino acids which compose this string are important in binding pRB-105 (43). Analysis of the E7 ORF from HPV-51 reveals that it encodes a 101-amino-acid protein with both the Cys motif and the binding domain for the product of the retinoblastoma gene.

TABLE 3. Organization of E6 ORFs from high-risk HPVs

HPV	Sequence ^a
HPV-51	CVYCKLELCRADVYNVAETEIKIVYRDNNPYAVCKQCLLFYSLIREYRRY
HPV-33	CVECKKPLQRSEYVDFAFADLTVVYREGNPFQICKLCLRELKISEYRHY
HPV-31	CVYCKGQLTETEYLDFAETDLTIVYRDDTPHGVCTKCLREYSKVSEFRWY
HPV-18	CVYCKTVLELTEYFEFAFKDLFVVYRDSIPHAACHKCIDEYSRIRELRHY
HPV-16	CVYCKQQLLRREYVDFAERDLCIVYRDGNPYAVCDKCLKFYSKISEYRHY
HPV-51	SRSVYGTTLLEAITKKSLYDLSIRCHRCQRPLGPPEEKQKLVAEKKRHFQIA
HPV-33	NYSVYGNLTLEQTVKKPLNEILIRCIICQRPLCPQEKRRHVDLNKRFEHNIS
HPV-31	RYSVYGTTLLEKLTNGICDLLIRCITCQRPLCPEEKQRHLDKKKRFHNIG
HPV-18	SDSVYGDITLEKLTNTGLYNLLIRCLRCQRPLNPAEKLRLHLNEKKRFEHNIA
HPV-16	CYSLYGTTLLEQQYNKPLCDLLIRCINCQRPLCPEEKQRHLDKKKRFHNIR
HPV-51	GRYTGQCANCW
HPV-33	GRWAGRCAACW
HPV-31	GRWTGRCIACW
HPV-18	GHYRGQCHSCC
HPV-16	GRWTGRCMSCC

^a The amino acid sequences of the E6 ORFs from these high risk HPVs were aligned as described in Cole and Danos (14). Highly conserved amino acids previously noted by these authors are in boldface. Those conserved amino acids which were detected in this more extensive analysis, which included HPV-31 and -51, are underlined.

The E1 ORF is the largest found in papillomaviruses, and it encodes a protein which has been shown to dock to the virus-specified E2 DNA-binding protein when it interacts with sequences in the upstream regulatory region (URR) (6, 40). Phylogenetic analysis of this ORF in human viruses suggests that it may be composed of two domains (14). The E1 ORF in HPV-51 has the potential to encode a protein of 634 amino acids. There is considerable conservation of sequence homology at both the nucleotide and amino acid levels between the HPV-51 E1 ORF and other members of the HPV family (Table 1).

E2 is a dimeric DNA-binding protein which recognizes the sequence ACCN₆GGT (1, 41) and has both activator and repressor functions on the URR from bovine papillomavirus (BPV). These functions can be attributed to the presence of two functional domains which are separated by a spacer whose length is associated with the virus subgroup from which it is derived. Thus, viruses associated with epidermodysplasia verruciformis contain longer and more hydrophilic spacers than those associated with infections of the genital tract (36). The N- and C-terminal segments encode the transactivation and repressor-dimerization domains, respectively. Repressor may be independently derived by transcription from an internal promoter in cells infected with BPV (62, 68) or by alternative splicing as detected in transcriptional analyses of cells and biopsy samples containing HPV-6, -11, or -16 (11, 22, 52). In the genital papillomaviruses E2 binds to two sites that are proximal to transcriptional elements of the E6 promoter and represses transcription (5, 51, 63). The E2 ORF of HPV-51 can encode a 358-amino-acid protein that has the characteristic three-domain structure of other papillomaviruses (28). The 185-amino-acid N-terminal domain is rich in arginine and lysine residues and terminates at a conserved V-185. The 70-amino-acid C-terminal domain which begins at G-289 is preceded by a 104-amino-acid spacer. The relatively short size of the middle segment and the paucity of arginine and lysine residues distinguish this virus from those associated with epidermodysplasia verruciformis (36).

In HPVs the mRNA encoding the E4 ORF is most likely derived by splicing from a donor in E1 to an acceptor in E4 (11, 12, 45), because the bulk of E1 RNA, as detected by in

situ hybridization analysis, is found in the nucleus in cervical intraepithelial neoplasms (19). The spliced mRNA can encode a functional E4 gene product, and immunohistochemical analysis of warts containing HPV-1 and early neoplasias containing HPV-16 reveals that E4 is expressed principally as a cytoplasmic protein in cells that stain positive for L1 (16, 20, 21). The E4 ORF in HPV-51 is completely contained within E2, as it is in other HPVs. Just upstream of the E4 ORF, at position 3270, is a potential splice acceptor which could be used to generate this species from any of several potential donor sites within the E1 ORF. Demonstration of this spliced species will require the analysis of mRNAs expressed in lesions containing HPV-51 or cell lines expressing the virus genome.

The E5 ORF from BPV has been studied in great detail. This 44-amino-acid protein is present in transformed cells as a membrane-associated homodimer (55). The protein contains a hydrophobic core and a hydrophilic carboxyl-terminal tail which defines a single functional domain that is responsible for its ability to induce focus formation and stimulate cellular DNA synthesis (33, 59). Recently, this protein was shown to activate the endogenous β -type receptor for the platelet-derived growth factor in transformed cells (48). Although no clear function has been ascribed to the E5 ORFs in HPVs, the E5a ORF of HPV-6c was shown to transform 3T3 cells (10). The E5 ORF in HPV-51 lacks an initiator methionine and can encode a protein of 84 amino acids with a predicted molecular weight of 9,764. As in the other HPV E5 ORFs there is a conserved Cys-X-Cys in the amino-terminal third of the protein, and there are three hydrophobic domains which have been suggested to constitute a membrane spanning region (8).

The L2 and L1 ORFs which encode the minor and major capsid proteins, respectively, are the products of late gene expression and are highly conserved between HPVs (Table 1). They are typically only expressed in well-differentiated epithelium in the apical portion of a lesion (17, 44). This sequence conservation has formed the basis for the development of a sensitive PCR assay which uses primers that specifically amplify the L1 ORF for the detection and typing of HPV genomes (37). Amplification of viral DNA or DNA from lesions containing HPV-51 with the MY09 and MY11

FIG. 3. Amplification of HPV DNAs. DNA was isolated from biopsies containing HPV sequences, and the virus type was determined by the restriction endonuclease digestion pattern of the DNA after Southern blot hybridization. (A) The products of DNA from these samples were amplified with the MY09 and MY11 primers described in Manos et al. (37). (B) The two potential hybridization sites within the L1 region of HPV-51 DNA for the MY09 primer; annealing at these sites results in generation of the two PCR products seen in panel A.

primers described by Manos et al. (37) generates two products (Fig. 3A): one of 452 bp is similar in size to what is amplified in other genital HPVs, and another that is 272 bp which results from mispriming of the MY09 primer at position 6718 to 6737 (Fig. 3B). The shorter amplification product can be eliminated if the annealing temperature is raised to 55°C. The L1 ORF in HPV-51 can encode a 504-amino-acid protein which is approximately 80% homologous at the amino acid level to L1s from other HPVs. The L2 ORF consists of 468 amino acids and contains the motif T-T-P-A-V-L at position 148 along with clusters of basic residues at the amino and carboxy termini (56); the carboxy terminus overlaps the amino terminus of L1 by 6 amino acids. The T-T-P-A-V-L motif is highly conserved among the mucosal HPVs, and its presence further cements the classification of this virus among these members of the HPV family (32).

Regulatory motifs in the HPV-51 genome. The URR of HPVs is operationally defined as the sequence between the

FIG. 4. Transformation of CREF cells. Freshly plated CREF cells were transfected with HPV-51 (a), salmon sperm (b), or HPV-18 DNA (c) as described in Materials and Methods. The cells were split 1:3 after 24 h, and the medium was changed every third day. Foci were photographed after 21 days by using a Nikon inverted microscope and T-MAX 400 film (Kodak).

last stop codon of L1 and the initiator ATG in E6. Computer-based searches of HPV URRs and DNA-binding studies of the URR from HPV-16 have revealed that the genital HPVs contain a glucocorticoid response element (GRE) and a keratinocyte-dependent enhancer (15). In addition, there are other overlapping *cis*-acting sequences which respond to and interact with general, tissue-specific, and virus-specified *trans*-acting proteins. These include sites for interaction with AP1, NF1, TFIID, Sp1, and the E2 activator-repressor protein. The URR of HPV-51 is 869 bp and contains three copies of the sequence ACCN₆GGT, to which E2 is known

TABLE 4. Location of sequence motifs in the HPV-51 genome

Sequence motif ^a	Consensus sequence	Sequence in HPV-51	Location ^b
Keratinocyte-dependent enhancer	TTTGGCTT	TTTcGgTT TTTGGCaT TTTaTCTT TTTGgcTT TTTcGgTT	7396–7403 (+) 7455–7462 (+) 7402–7409 (+) 7765–7772 (+) 29–37 (–)
E2 binding site	ACCN ₆ GGT	ACCGATTTCGGT ACCGAAAAGGGT	7391–7402 (+) 30–41 (+) 46–57 (+)
NFI	TTGGCA	ACCTTTTACTGG TTGGCA	4497–4508 (+) 7423–7428 (–) 7456–7461 (+) 7578–7583 (+) 7633–7638 (+) 7563–7568 (+) 7690–7695 (–) 7497–7502 (–) 7501–7506 (–) 7593–7598 (–)
Sp1	GGGCGG	GGGCGG	60–64 (+) ^c 53–59 (+) 7653–7660 (+) 7356–7372 (+) 5090–5105 (+) 4030–4044 (+) 3818–3833 (+) 4103–4109 (+) 7169–7173 (+) 181–190 (+) 3913–3922 (+) 7116–7125 (+) 7186–7195 (+)
TFIID	ATATA	ATATA	
AP1	TGA ₆ TCA	TtAGTCA	
Oct-1	ATTTCAT	AaTGCAT	
GRE	GGTACANNNTGTTCTC ^d TGTACATTGTGCAT ^e TCTACATTTTATACT ^f AGAACATACTGTCCC ^f	GcTtcATTTtTTgT TGTACATTaTtATCAT aCTACATTTTtTg_T AcA ₁ tATACTGTCCC	
Poly(A) addition	AATAAA	AATAAA	
GT repeat	GTGTGTGTGT	GTGTGTGTGT	

^a Sequence motifs for interaction with known transcription factors were as described in Mitchell and Tjian (39).

^b +, motifs that are in the sense orientation; –, motifs in the antisense orientation.

^c There are additional copies of this sequence throughout the genome; only the site in URR is shown here.

^d The GRE consensus used was from Chan et al. (9).

^e HPV-16 GRE.

^f HPV-39 GRE (65).

to bind (1, 41). These sequences are located at nucleotides 7391, 30, and 46 within URR, and a fourth copy of this sequence motif is found at position 4497 within L2. A mobility shift gel electrophoresis assay with end-labeled fragments from the HPV-51 URR (that contained the canonical E2 binding site sequence) was used to demonstrate that recombinant E2 from BPV would bind and alter the electrophoretic mobility of those fragments which contained these sites (data not shown).

Nucleotide sequence analysis of the HPV-51 URR reveals the presence of the many tightly clustered and overlapping regulatory elements; there is a homolog of the GRE, and there are seven copies of the NFI binding motif. In addition, there are single sites for AP1, Oct-1, Sp1, and TFIID. The orientation and location of these sites is summarized in Table 4. In addition to the single GRE in URR there are four other homologs of this sequence, three of which are clustered near the E5 ORF and one which is in L2.

Poly(A) signals for mRNA termination are bipartite, consisting of a signal AATAAA followed at a distance of 20 to 30 nucleotides by a GT-rich sequence (50). Sequences homologous to this termination signal are found in all HPVs. In HPV-51 there are six copies of the AATAAA motif, and two of these, located at 4104 and 7169, are downstream of the early and late genes, respectively, and contain both components of the polyadenylation signal.

Finally, we note the occurrence of a decameric sequence

that is composed of alternating G · T residues which is repeated four times in the virus genome.

Transforming activity of HPV-51 DNA. Cloned DNAs from those HPVs which are associated with lesions that have a high risk for malignant progression have the capacity to transform rodent cells in culture (69) and to immortalize human keratinocytes (23, 35, 49). Immortalization of keratinocytes results from cooperation between the E6 and E7 gene products (31). Sequence analysis revealed that HPV-51 was more closely related to the oncogenic HPVs than to those associated with benign lesions. To extend this association further, we next asked if DNA from HPV-51 would transform a continuous line of rat embryo fibroblasts. Because the original clone of virus DNA was isolated as a *Hind*III fragment and lacked the E6 ORF, the intact genome was reconstructed after cleavage with *Hind*III and self-ligation to generate concatamers. This DNA was digested at the unique *Hinc*II site in the original clone (47), downstream of E7, and cloned into a *Hind*III[–] vector at a unique *Hinc*II site. The missing 232-bp *Hind*III fragment (which encodes most of E6) was then inserted at the unique *Hind*III site in the virus genome. This DNA was released from the resulting plasmid, and 10 µg (each) of HPV-51 and HPV-18 DNA was compared for their abilities to transform CREF cells. HPV-51 DNA induced transformed cells with a frequency similar to that of HPV-18 DNA (30 CFU/µg versus 20 CFU/µg). Moreover, control transfections that included

FIG. 5. Analysis of RNA and DNA in a cell line transformed with HPV-51 DNA. (A) The products of PCR amplification of DNA derived from purified plasmid (lanes 2 and 4) or transformed cells (lanes 3 and 5) with primers that would amplify E6 and E7 (lanes 4 and 5) or E6 (lanes 2 and 3) along with the PCR amplification products of the cDNAs synthesized with the RT primer on RNA isolated from the same transformed cell line (lane 6) were analyzed by electrophoresis in 3% GTG agarose gels. Lane 1 contains an *Msp*I digest of pBR322 DNA as a size marker. (B) A schematic of the E6 and E7 regions of the HPV-51 genome is shown along with the positions of the primers used for amplification of E6 (5' E6 [nt 79 to 99] and 3' E6 [nt 399 to 419]) and E6 and E7 DNA (5' E6 and E7 [nt 91 to 111] and RT [nt 866 to 887]) and for synthesis of cDNA (RT [nt 866 to 887]) and its amplification. The theoretical spliced products with their potential donor and acceptor sites are indicated along with the predicted size of the amplification products. The probes used to analyze the products of amplification of the cDNAs (panel A, lane 6), i.e., F1, 780 bp; F2, 580 bp; F3, 220 bp; F4, 162 bp; F5, 135 bp; F6, 120 bp; F7, 88 bp; and F8, 80 bp, for the presence of E6 (Probe 1) or E7 (Probe 2) sequences are also shown. (C) Results of hybridization of each of the amplified products of the cDNA reaction to ³²P-labeled probes specific for E6 (Probe 1) and E7 (Probe 2).

HPV-6, HPV-11, or placental DNA gave rise to no transformed colonies, whereas transfection with similar amounts of *src* DNA resulted in >500 CFU/μg. Comparison of the morphologic transformation frequency of these same DNAs on BALB 3T3 cells resulted in too few transformants with either HPV-18 or HPV-51 DNAs to quantitate. A typical morphologically transformed CREF colony is shown in Fig. 4. Several transformants were isolated and recloned twice. The cloned colonies were then expanded and examined for the presence of sequences homologous to HPV-51 DNA by PCR amplification of the E6, E7, and L1 regions of this virus. The transformed cells contained amplification products of the appropriate size (Fig. 5A and B). Analysis of these products by digestion with restriction endonucleases demonstrated that they were derived from HPV-51 DNA (data not shown).

The state of virus DNA in a single PCR-positive clone was examined by Southern blot hybridization. Hybridization of transformed-cell DNA cleaved with *Hinc*II (which should not cleave integrated virus DNA because the site is destroyed after release of the insert from the vector) revealed

the presence of a single 14-kb species; after digestion with *Eco*RV (a single cutter), two bands were detected (data not shown). This result suggests that only a single copy of virus DNA was integrated into the genome of the transformed cell. Because products of both E6 and E7 are required for complete morphologic transformation (3, 30, 31, 34, 42), we next examined the RNAs present in a line of transformed cells that were homologous to virus sequences from these regions of the HPV-51 chromosome. RNA isolated from the single-copy transformant was reverse transcribed with a primer (RT primer) that is downstream of the E7 termination codon (Fig. 5B). The resulting cDNAs were then amplified with primers that flank these ORFs and examined by gel electrophoresis. The electrophoretic mobilities of the products are shown in Fig. 5A. The slowest migrating band resulted from contamination of the RNA with virus DNA, as judged by its appearance in the absence of any reverse transcription and the fact that it was not present if RNA was preselected for poly(A)-containing molecules (data not shown). Bands corresponding to products that might be derived as a result of utilization of the various splice donors

and acceptors in this region of the genome (Fig. 5B) were excised from the gel and examined for the presence of sequences homologous to either E6 or E7 by molecular hybridization by using probes that were specific for each ORF. The results of this analysis are shown in Fig. 5C. The three slowest migrating bands contained amplimers that hybridized with both the E6 and E7 probes, while the other products appeared to contain sequences that were predominantly derived from E7. Because the E6 probe only contained 30 bp of homology with the region upstream of the splice donor at position 177, its ability to hybridize with E6 sequences is somewhat limited. We conclude from these data that there are spliced RNAs present in this transformed cell that can encode both the E6 and E7 proteins. Further characterization of the products of these reactions will be required to determine the utilization of donors and acceptors and the exact coding potential of each of the RNAs identified in this preliminary analysis.

In summary, we have determined the nucleotide sequence of HPV-51. Analysis of the genome sequence, in conjunction with assay of its transformation potential, places this isolate among the high-risk group of HPVs.

ACKNOWLEDGMENTS

Studies in the authors' laboratories were supported by grants CA23767 (S.S.) and CA47676 (C.P.C.) from the National Institutes of Health and SIG13 (S.S.) from the American Cancer Society. We gratefully acknowledge the Perkin-Elmer Foundation for the donation of a thermal cycler.

We thank Rey Sia for performing the E2 binding assays, Elliott Androphy for a clone expressing BPV E2, and Li Hui for some of the clones used in determining the sequence of HPV-51.

REFERENCES

- Androphy, E. J., D. R. Lowy, and J. T. Schiller. 1987. Bovine papillomavirus E2 trans-activating gene product binds to specific sites in papillomavirus DNA. *Nature (London)* **325**:70-73.
- Barbosa, M. S., D. R. Lowy, and J. T. Schiller. 1989. Papillomavirus polypeptides E6 and E7 are zinc-binding proteins. *J. Virol.* **63**:1404-1407.
- Barbosa, M. S., W. C. Vass, D. R. Lowy, and J. T. Schiller. 1991. In vitro biological activities of the E6 and E7 genes vary among human papillomaviruses of different oncogenic potential. *J. Virol.* **65**:292-298.
- Bauer, H. M., Y. Ting, C. E. Greer, J. C. Chambers, C. J. Tashiro, J. Chimera, A. Reingold, and M. M. Manos. 1991. Genital human papillomavirus infection in female university students as determined by a PCR-based method. *JAMA* **265**:472-477.
- Bernard, B. A., C. Bailly, M.-C. Kebiur, M. Darmon, F. Thierry, and M. Yaniv. 1989. The human papillomavirus type 18 (HPV18) E2 gene product is a repressor of the HPV18 regulatory region in human keratinocytes. *J. Virol.* **63**:4317-4324.
- Blitz, I. L., and L. A. Laimins. 1991. The 68-kilodalton E1 protein of bovine papillomavirus is a DNA binding phosphoprotein which associates with the E2 transcriptional activator in vitro. *J. Virol.* **65**:649-656.
- Boshart, M., L. Gissman, H. Ikenberg, A. Kleinheinz, W. Scheurlen, and H. zur Hausen. 1984. A new type of papillomavirus DNA, its presence in genital cancer biopsies and in cell lines derived from cervical cancer. *EMBO J.* **3**:1151-1157.
- Bubb, V., D. J. McCance, and R. Schlegel. 1988. DNA sequence of the HPV-16 E5 ORF and the structural conservation of its encoded protein. *Virology* **163**:243-246.
- Chan, W.-K., T. Chong, H.-U. Bernard, and G. Klock. 1990. Transcription of the transforming genes of the oncogenic papillomavirus-16 is stimulated by tumor promoters through AP1 binding sites. *Nucleic Acids Res.* **18**:763-769.
- Chen, S.-L., and P. Mounts. 1990. Transforming activity of E5a protein of human papillomavirus type 6 in NIH 3T3 and C127 cells. *J. Virol.* **64**:3326-3333.
- Chow, L. T., M. Nasser, S. M. Wolinsky, and T. R. Broker. 1987. Human papillomavirus 6 and 11 mRNAs from genital condylomata acuminata. *J. Virol.* **61**:2581-2588.
- Chow, L. T., S. S. Reilly, T. R. Broker, and L. B. Taichman. 1987. Identification and mapping of human papillomavirus type 1 RNA transcripts recovered from plantar warts and infected epithelial cell cultures. *J. Virol.* **61**:1913-1918.
- Coggin, J. R., and H. zur Hausen. 1979. Workshop on papillomaviruses and cancer. *Cancer Res.* **39**:545-546.
- Cole, S. T., and O. Danos. 1987. Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome. *J. Mol. Biol.* **193**:599-608.
- Cripe, T. P., T. H. Haugen, J. P. Turk, F. Tabatashi, P. G. Schmid, M. Durst, L. Gissman, A. Roman, and L. P. Turek. 1987. Transcriptional regulation of the human papillomavirus-16 E6-E7 promoter by a keratinocyte-dependent enhancer, and by viral E2 trans-activator and repressor gene products: implications for cervical carcinogenesis. *EMBO J.* **6**:3745-3753.
- Crum, C. P., S. Barber, M. Symbula, K. Snyder, A. M. Saleh, and J. K. Roches. 1990. Coexpression of the human papillomavirus type 16 E4 and L1 open reading frames in early cervical neoplasia. *Virology* **178**:238-246.
- Crum, C. P., G. Friedman, G. Nuovo, and S. J. Silverstein. 1987. Morphological correlates of genital HPV infection: viral replication, transcription and gene expression, p. 355-369. *In* R. Gallo, W. Haseltine, G. Klein, and H. zur Hausen (ed.), *Viruses and human cancer*. Alan R. Liss, Inc., New York.
- Crum, C. P., M. Mitao, R. U. Levine, and S. Silverstein. 1985. Cervical papillomaviruses segregate within morphologically distinct precancerous lesions. *J. Virol.* **54**:675-681.
- Crum, C. P., G. Nuovo, D. Friedman, and S. J. Silverstein. 1988. Accumulation of RNA homologous to human papillomavirus type 16 open reading frames in genital precancers. *J. Virol.* **62**:84-90.
- Doorbar, J., D. Campbell, R. J. A. Grand, and P. H. Gallimore. 1986. Identification of the human papilloma virus-1a E4 gene products. *EMBO J.* **5**:355-362.
- Doorbar, J., I. Coneron, and P. H. Gallimore. 1989. Sequence divergence yet conserved physical characteristics among the E4 proteins of cutaneous human papillomaviruses. *Virology* **172**:51-62.
- Doorbar, J., A. Parton, K. Hartley, L. Banks, T. Crook, M. Stanley, and L. Crawford. 1990. Detection of novel splicing patterns in a HPV-16 containing keratinocyte cell line. *Virology* **178**:254-262.
- Durst, M., R. T. Dzarlieva-Petrusevska, P. Boukamp, N. E. Fusenig, and L. Gissman. 1987. Molecular and cytogenetic analysis of immortalized human primary keratinocytes obtained after transfection with human papillomavirus type 16 DNA. *Oncogene* **1**:251-256.
- Durst, M. L., L. Gissman, H. Ikenberg, and H. zur Hausen. 1985. A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions. *Proc. Natl. Acad. Sci. USA* **82**:3812-3815.
- Dyson, N., P. M. Howley, K. Munger, and E. Harlow. 1989. The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. *Science* **243**:934-936.
- Fisher, P. B., L. E. Babiss, I. B. Weinstein, and H. S. Ginsberg. 1982. Analysis of type adenovirus transformation with a cloned rat embryo cell line (CREF). *Proc. Natl. Acad. Sci. USA* **79**:3527-3531.
- Giri, I., and O. Danos. 1986. Papillomavirus genomes: from sequence data to biological properties. *Trends Genet.* **2**:227-232.
- Giri, I., O. Danos, and M. Yaniv. 1985. Genomic structure of the cottontail rabbit (Shope) papillomavirus. *Proc. Natl. Acad. Sci. USA* **82**:1580-1584.
- Grossman, S. R., and L. A. Laimins. 1989. E6 protein of human papillomavirus type 18 binds zinc. *Oncogene* **9**:1089-1093.
- Halbert, C. L., G. W. Demers, and D. A. Galloway. 1991. The E7

- gene of human papillomavirus type 16 is sufficient for immortalization of human epithelial cells. *J. Virol.* **65**:473-478.
31. Hawley-Nelson, P., K. H. Vousden, N. L. Hubbert, D. R. Lowy, and J. T. Schiller. 1989. HPV16 E6 and E7 proteins cooperate to immortalize human foreskin keratinocytes. *EMBO J.* **8**:3905-3910.
 32. Hirsch-Behnam, A., H. Delius, and E.-M. de Villiers. 1990. A comparative sequence analysis of two human papillomavirus (HPV) types 2a and 57. *Virus Res.* **18**:81-98.
 33. Horwitz, B. H., A. C. Burkhardt, R. Schlegel, and D. DiMaio. 1988. 44-Amino acid E5 transforming protein of bovine papillomavirus type 1 requires a hydrophobic core and specific carboxyl-terminal amino acids. *Mol. Cell. Biol.* **8**:4071-4078.
 34. Hudson, J. B., M. A. Bedell, D. J. McCance, and L. S. Laimins. 1990. Immortalization and altered differentiation of human keratinocytes in vitro by the E6 and E7 open reading frames of human papillomavirus type 18. *J. Virol.* **64**:519-526.
 35. Kaur, P., and J. K. McDougall. 1988. Characterization of primary human keratinocytes transformed by human papillomavirus type 18. *J. Virol.* **62**:1917-1924.
 36. Kiyono, T., A. Adachi, and M. Ishibashi. 1990. Genome organization and taxonomic position of human papillomavirus type 47 inferred from its DNA sequence. *Virology* **177**:401-405.
 37. Manos, M. M., Y. Ting, D. K. Wright, A. J. Lewis, T. R. Broker, and S. M. Wolinsky. 1989. Use of polymerase chain reaction amplification for the detection of genital human papillomaviruses. *Cancer Cells* **7**:209-214.
 38. Meisels, A., and C. Morin. 1981. Human papillomavirus and cancer of the uterine cervix. *Gynecol. Oncol.* **12**:111-123.
 39. Mitchell, P. J., and R. Tjian. 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**:371-378.
 40. Mohr, I. J., R. Clark, S. Sun, E. J. Androphy, P. MacPherson, and M. R. Botchan. 1990. Targeting the E1 replication protein to the papillomavirus origin of replication by complex formation with the E2 transactivator. *Science* **250**:1694-1699.
 41. Moskaluk, C., and D. Bastia. 1987. The E2 "gene" of bovine papillomavirus encodes an enhancer-binding protein. *Proc. Natl. Acad. Sci. USA* **84**:1215-1218.
 42. Munger, K., W. C. Phelps, V. Bubb, P. H. Howley, and R. Schlegel. 1989. The E6 and E7 genes of the human papillomavirus type 16 together are necessary and sufficient for transformation of primary human keratinocytes. *J. Virol.* **63**:4417-4421.
 43. Munger, K., B. A. Werness, N. Dyson, W. C. Phelps, E. Harlow, and P. M. Howley. 1989. Complex formation of human papillomavirus E7 proteins with the retinoblastoma tumor suppressor gene product. *EMBO J.* **8**:4099-4105.
 44. Nagai, N., G. Nuovo, D. Friedman, and C. Crum. 1987. Detection of papillomavirus nucleic acids in genital precancers with the in situ hybridization technique: a review. *Int. J. Gynecol. Pathol.* **6**:366-379.
 45. Nasser, M., R. Hirochika, T. R. Broker, and L. T. Chow. 1987. A human papillomavirus type 11 transcript encoding an E1^{*}E4 protein. *Virology* **159**:433-439.
 46. Needleman, S. B., and C. D. Wunsch. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**:443-453.
 47. Nuovo, G., E. DeVilliers, R. Levine, S. Silverstein, and C. Crum. 1988. The isolation of a novel human papillomavirus (HPV 51) from a cervical condyloma. *J. Virol.* **62**:1452-1455.
 48. Petti, L., L. Nilson, and D. DiMaio. 1991. Activation of the platelet-derived growth factor receptor by the bovine papillomavirus E5 transforming protein. *EMBO J.* **10**:845-856.
 49. Pirisi, L., S. Yasumoto, M. Feller, J. Doniger, and J. DiPaolo. 1987. Transformation of human fibroblasts and keratinocytes with human papillomavirus type 16 DNA. *J. Virol.* **61**:1061-1066.
 50. Proudfoot, N. 1991. Poly(A) signals. *Cell* **64**:671-674.
 51. Romanczuk, H., F. Thierry, and P. M. Howley. 1990. Mutational analysis of *cis* elements involved in E2 modulation of human papillomavirus type 16 P₉₇ and type 18 P₁₀₅ promoters. *J. Virol.* **64**:2849-2859.
 52. Rotenberg, M. O., L. T. Chow, and T. R. Broker. 1989. Characterization of rare human papillomavirus type 11mRNAs coding for regulatory and structural proteins, using the polymerase chain reaction. *Virology* **172**:489-497.
 53. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.
 54. Scheffner, M., B. A. Werness, J. M. Huibregtse, A. J. Levine, and P. M. Howley. 1990. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell* **63**:1129-1136.
 55. Schlegel, R., M. Wade-Glass, M. S. Rabson, and Y. C. Yang. 1986. The E5 transforming gene of bovine papillomavirus encodes a small hydrophobic polypeptide. *Science* **233**:464-467.
 56. Schwarz, E., M. Dürst, C. Demankowski, O. Lattermann, R. Zech, E. Wolfspurger, S. Suhai, and H. zur Hausen. 1983. DNA sequence and genome organization of genital human papillomavirus type 6b. *EMBO J.* **2**:2341-2348.
 57. Schwarz, E., U. K. Freese, L. Gissman, W. Mayer, B. Roggenbuck, A. Stremlau, and H. zur Hausen. 1985. Structure and transcription of human papillomavirus sequences in cervical carcinoma cells. *Nature (London)* **314**:111-114.
 58. Schwarz, E., A. Schneider-Gädick, and H. zur Hausen. 1987. Human papillomavirus type-18 transcription in cervical carcinoma cell lines and in human cell hybrids. *Cancer Cells* **5**:47-53.
 59. Settleman, J., A. Fazeli, J. Malicki, B. H. Horwitz, and D. DiMaio. 1989. Genetic evidence that acute morphologic transformation, induction of cellular DNA synthesis, and focus formation are mediated by a single activity of the bovine papillomavirus E5 protein. *Mol. Cell. Biol.* **9**:5563-5572.
 60. Smotkin, D., H. Prokoph, and F. O. Wettstein. 1989. Oncogenic and nononcogenic human genital papillomaviruses generate the E7 mRNA by different mechanisms. *J. Virol.* **63**:1441-1447.
 61. Smotkin, D., and F. O. Wettstein. 1986. Transcription of human papillomavirus type 16 early genes in a cervical cancer and a cancer-derived cell line and identification of the E7 protein. *Proc. Natl. Acad. Sci. USA* **83**:4680-4684.
 62. Stenlund, A., J. Zabielski, H. Ahola, J. Moreno-Lopez, and U. Pettersson. 1985. Messenger RNAs from the transforming region of bovine papilloma virus type 1. *J. Mol. Biol.* **182**:541-544.
 63. Thierry, F., and P. M. Howley. 1991. Functional analysis of E2-mediated repression of the HPV18 p105 promoter. *New Biol.* **3**:90-100.
 64. Vogelstein, B., and D. Gillespie. 1979. Preparative and analytical purification of DNA from agarose. *Proc. Natl. Acad. Sci. USA* **76**:615-619.
 65. Volpers, C., and R. E. Streeck. 1991. Genome organization and nucleotide sequence of human papillomavirus type 39. *Virology* **181**:419-423.
 66. Werness, B., A. J. Levine, and P. M. Howley. 1990. Association of human papillomavirus types 16 and 18 E6 proteins with p53. *Science* **248**:76-79.
 67. Wigler, M., A. Pellicer, S. Silverstein, R. Axel, G. Urlaub, and L. Chasin. 1979. Transformation of the APRT locus in mammalian cells. *Proc. Natl. Acad. Sci. USA* **76**:1373-1376.
 68. Yang, Y.-C., H. Okayama, and P. M. Howley. 1985. Bovine papillomavirus contains multiple transforming genes. *Proc. Natl. Acad. Sci. USA* **82**:1030-1034.
 69. Yasumoto, S., A. L. Burkhardt, J. Doniger, and J. A. DiPaolo. 1986. Human papillomavirus type 16 DNA-induced malignant transformation of NIH 3T3 cells. *J. Virol.* **57**:572-577.
 70. zur Hausen, H. 1977. Human papillomaviruses and their possible role in squamous cell carcinomas. *Curr. Top. Microbiol. Immunol.* **78**:1-30.
 71. zur Hausen, H. 1987. Papillomaviruses in human cancer. *Cancer* **59**:1692-1696.