JOURNAL OF VIROLOGY, Dec. 1995, p. 7743–7753 Vol. 69, No. 12
0022-538X/95/$04.00+0

Human Papillomavirus Type 16 Variant Lineages in United States Populations Characterized by Nucleotide Sequence Analysis of the E6, L2, and L1 Coding Segments

TAKASHI YAMADA,[1,2] COSETTE M. WHEELER,[2,3]* AARON L. HALPERN,[4] ANN-CHARLOTTE M. STEWART,[2,3] ALLAN HILDESHEIM,[5] AND STEVEN A. JENISON[1,2,6]

Departments of Medicine,[1] Microbiology,[6] and Cell Biology[3] and the New Mexico Cancer Research and Treatment Center,[2] University of New Mexico School of Medicine, Albuquerque, New Mexico 87131; Theoretical Biology and Biophysics, Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545[4]; and Epidemiology and Biostatistics Program, Division of Cancer Etiology, National Cancer Institute, Bethesda, Maryland 20892[5]

Received 14 June 1995/Accepted 24 August 1995

Human papillomavirus type 16 (HPV16) nucleotide sequence variations in the E6 (nucleotide positions [nt] 104 to 559), L2 (nt 4272 to 5657), and L1 (nt 5665 to 7148) open reading frames (ORFs), and the long control region (nt 7479 to 7842), were examined in 29 selected United States isolates. Of 3,690 nucleotide positions, 129 (3.5%) varied. The maximum pairwise distance was 66 nucleotide differences, or 1.8%. Nucleotide variations within different genome segments were phylogenetically compatible, and nucleotide changes within E6, L2, and L1 contained phylogenetic information beyond that provided in the long control region. Most isolates were classified as members of HPV16 lineages that have been described previously. However, two novel phylogenetic branches were identified. The L2 ORF was the most variable coding segment. L2 synonymous and nonsynonymous nucleotide changes were distributed asymmetrically. The amino-terminal half of the L2 protein was remarkably conserved among all isolates, suggesting that the region is under evolutionary constraint. The amino-terminal region of the E6 ORF was relatively varied, especially at E6 amino acid positions 10 and 14. Several amino acid differences in the L1 ORF were observed between lineages. Forty-nine amino acid variations across all sequenced coding regions were observed. These amino acid differences may be relevant to differences in the generation of humoral or cell-mediated immune responses to HPV16 variants. Our data form a basis for considering HPV16 sequence variation in the rational design of vaccine strategies and as an epidemiologic correlate of cervical cancer risk.

Human papillomavirus type 16 (HPV16) infections are associated with the majority of invasive squamous cell carcinomas of the cervix worldwide (8, 46, 78). Proteins encoded by the HPV16 E6 and E7 open reading frames (ORFs) immortalize human keratinocytes in vitro (35, 48, 49, 55) and are likely to be involved in the malignant transformation of cervical epithelium in vivo. Effective vaccines might prevent HPV16 infections and thereby reduce the incidence of HPV-associated cervical carcinomas. HPV virus-like particles that result from the self-assembly of recombinant capsid proteins are being evaluated as vaccine candidates (29, 40, 42, 50, 59, 73, 77).

HPV isolates historically were classified into types by comparing their DNA with a set of HPV reference genomes in a standardized liquid hybridization assay (17, 19). Isolates that belong to the same HPV type are closely related to one another on the basis of nucleotide sequence and represent a discrete phylogenetic group (7, 10). Although all HPV16 isolates are closely related, previous investigations have identified five distinct phylogenetic branches among 301 HPV16 isolates that were collected from Europe, Asia, Africa, and North and South America (33). The branches were designated E (European), As (Asian), AA (Asian-American), Af1 (African-1), and Af2 (African-2). The HPV16 reference genome is a German isolate and is a member of the E lineage (11, 33, 34, 64).

The phylogeny inferred by Ho and coworkers (33, 34) was based on nucleotide sequence analysis of the HPV16 long control region (LCR), a noncoding segment that regulates transcription and contains the origin of replication. In the LCR-based analysis, E and As branches were distinct but very closely related, such that As appeared to be a subclass of the E lineage. AA, Af1, and Af2 branches were well separated from each other and from the E and As branches.

Limited information has been available on sequence variations within the HPV16 L1 and L2 ORFs. All HPV16 clinical isolates for which sequence is available contain a C-to-G nucleotide change, relative to the HPV16 reference genome, at L1 nucleotide coordinate (nt) 6241 (42, 56). This results in a histidine-to-aspartic acid change at L1 amino acid coordinate (aa) 202. HPV16 L1 recombinant proteins that contain aspartic acid at aa 202, but which otherwise are identical to the HPV16 prototype, assemble efficiently into virus-like particles (Fig. 1 and 2, isolate 114K) (42). In contrast, L1 recombinant polypeptides encoded by the HPV16 reference genome do not form virus-like particles efficiently. In L2, many HPV16 isolates contain a G-to-A change at nt 4937 (42, 56), but this change does not alter the predicted amino acid sequence.

Pushko and coworkers sequenced HPV16 L1 and L2 ORFs from cervical cell samples collected from seven women in the United Kingdom and from four women in Trinidad (Fig. 1 and 2) (56). The L1 and L2 segments were amplified independently

* Corresponding author. Mailing address: University of New Mexico Cancer Research and Treatment Center, Dept. of Cell Biology, 900 Camino de Salud NE, Albuquerque, NM 87131. Phone: (505) 277-9151. Fax: (505) 277-9494. Electronic mail address: cwheeler@medusa.unm.edu.

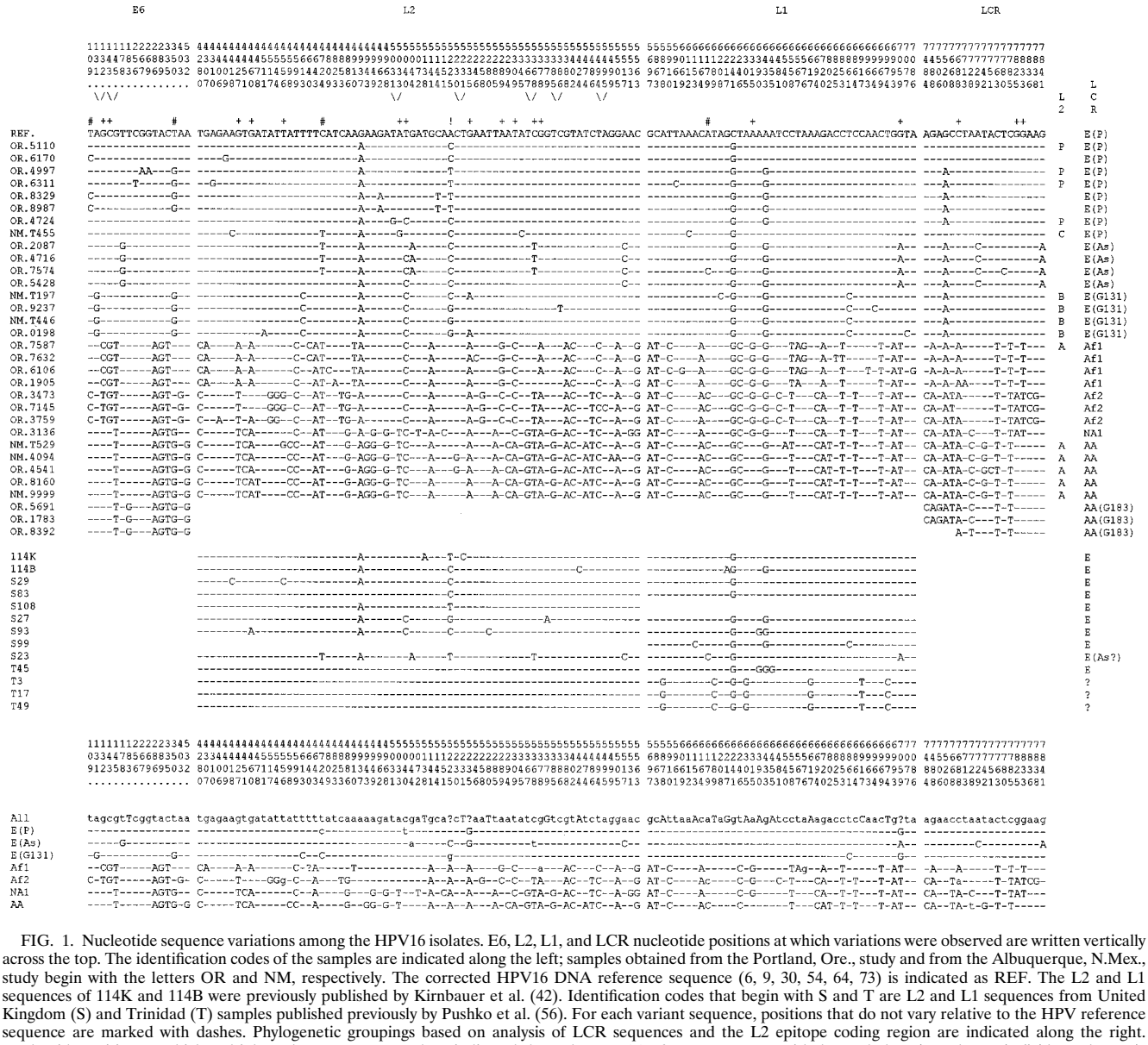

FIG. 1. Nucleotide sequence variations among the HPV16 isolates. E6, L2, L1, and LCR nucleotide positions at which variations were observed are written vertically across the top. The identification codes of the samples are indicated along the left; samples obtained from the Portland, Ore., study and from the Albuquerque, N.Mex., study begin with the letters OR and NM, respectively. The corrected HPV16 DNA reference sequence (6, 9, 30, 54, 64, 73) is indicated as REF. The L2 and L1 sequences of 114K and 114B were previously published by Kirnbauer et al. (42). Identification codes that begin with S and T are L2 and L1 sequences from United Kingdom (S) and Trinidad (T) samples published previously by Pushko et al. (56). For each variant sequence, positions that do not vary relative to the HPV reference sequence are marked with dashes. Phylogenetic groupings based on analysis of LCR sequences and the L2 epitope coding region are indicated along the right. Nucleotide positions at which multiple replacements occurred are indicated above the HPV16 reference sequence with the symbols + (two changes), # (three changes), and ! (four or more changes). Varied nucleotide positions contained within the same codon are marked with the symbol \/. Consensus sequences for each of the phylogenetic groups are shown at the bottom; distinctive changes present in all sequences of a class are indicated by a capital letter, while those present in 50% or more (but not all) members of a class are indicated in lowercase. Question marks indicate positions at which no majority (50%) consensus exists. The LCR sequence obtained for isolate OR.8392 begins at nt 7672.

from each cell sample by using PCR. The United Kingdom samples closely resembled the HPV16 reference genome. In three of the Trinidad samples, the L2 sequences were identical to the HPV16 reference genome but the L1 sequences contained seven nucleotide sequence changes, of which six were nonsynonymous (Fig. 1 and 2, isolates T3, T17, and T49). Smits and coworkers observed a similar lack of linkage between genome segments among HPV16 DNA-containing samples from Barbados (65). This apparent lack of linkage suggests the existence of HPV16 genetic recombinants. However, recombination between HPV types has not been reported, and other investigators (11, 38) have demonstrated linkage of nucleotide changes between HPV16 genomic segments of the same isolate. Possible explanations for the appearance of recombination include a high prevalence of recombinant genomes in the Caribbean region, preferential amplification of genomic segments when different HPV16 variants are present in a single specimen, and simple PCR contamination events.

To date, variations within the HPV16 E6 ORF have not been described in detail. Ellis et al. recently reported an HPV16 E6 variant that contained an A-to-G change at HPV16 E6 nt 131 (21). We report here the characterization of HPV16 variants that contain an A-to-G change at HPV16 nt 131, and we find that these variants constitute a distinct subclass of the HPV16 E lineage.

In this study, HPV16 isolates obtained in the United States were initially examined for nucleotide sequence variations within an L2 segment (nt 4551 to 4871) that encodes human

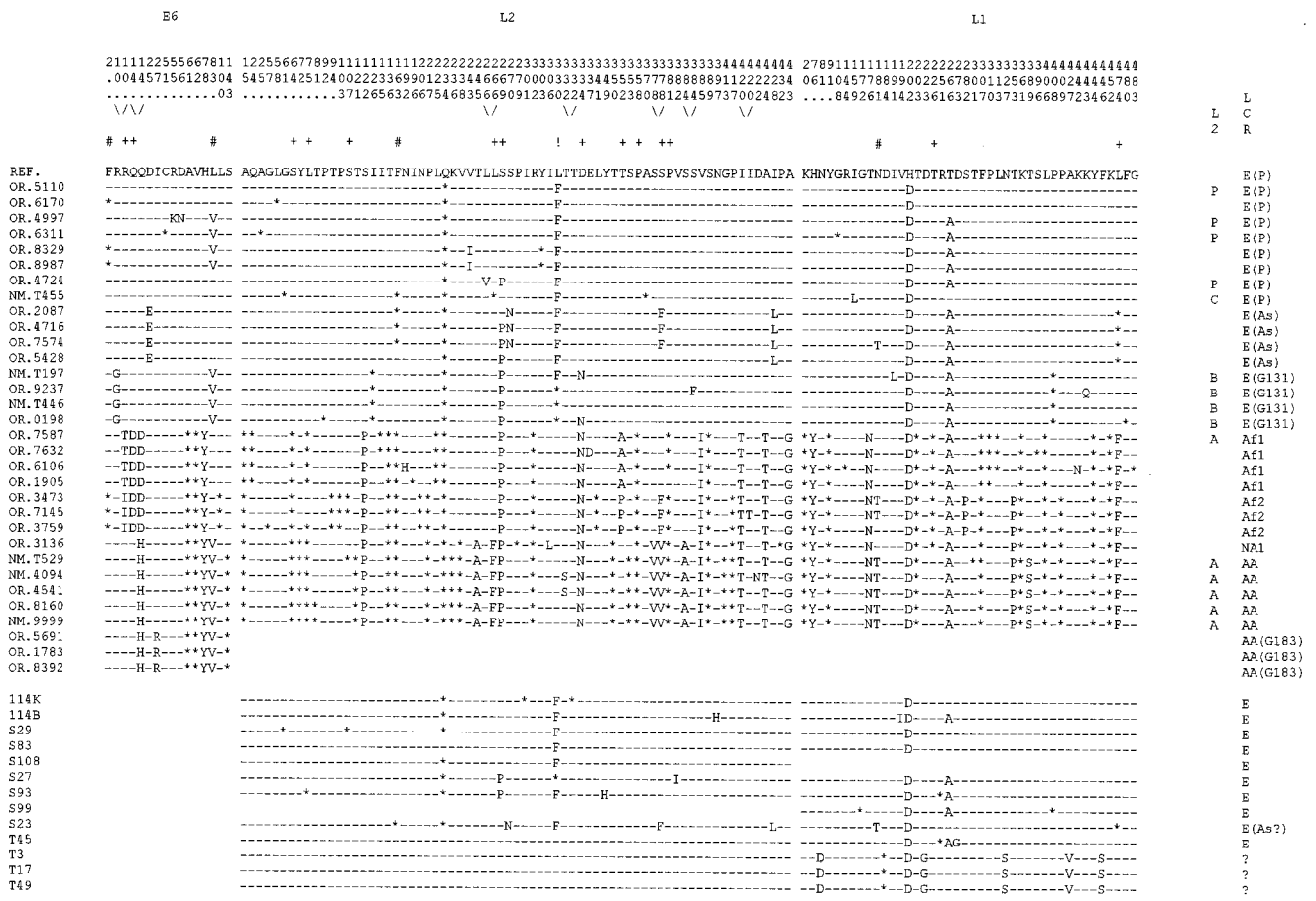

FIG. 2. Amino acid sequence variations among the HPV16 isolates. E6, L2, and L1 amino acid positions at which variations were observed are written vertically across the top. For amino acid positions affected by variations at two different nucleotide positions within the codon, the amino acid position is listed twice and marked with \ /. Conventions used in this figure are otherwise similar to those used in Fig. 1. Amino acid positions whose codons contain more than one nucleotide replacement are marked with +, #, and ! as in Fig. 1. Codons that contain synonymous nucleotide changes are marked with asterisks.

antibody-reactive epitopes (39) and within the E6 coding region (nt 104 to 559). Nucleotide and amino acid sequence alignments were used to identify variant clusters. Expanded E6 (nt 104 to 559), L2 (nt 4272 to 5657), L1 (nt 5665 to 7148), and LCR (nt 7479 to 7842) sequences were obtained from selected isolates of each variant cluster. L1 and L2 were examined in order to determine the prevalence, the extent, and the distribution of amino acid changes that might affect immune responses to HPV16 capsid proteins. E6 was studied to detect sequence variations that may be associated with differences in cervical cancer risk or that may be relevant to differences in cellular immune responses to HPV16-infected cells. The LCR was analyzed in order to determine the relationship between our variant isolates and those described previously (11, 33, 34). These data permit a robust analysis of HPV16 phylogeny, mutation patterns, and linkage of genomic regions.

MATERIALS AND METHODS

Clinical specimens. HPV16 DNA-containing cervical cell samples were obtained from subjects enrolled in epidemiologic investigations conducted in Albuquerque, N.Mex. (n = 30) (5, 71), and Portland, Ore. (n = 137) (62). Specimens included a range of cytologic diagnoses and were predominantly normal.

HPV16 nucleotide position numbering. HPV16 DNA nucleotide positions are numbered according to the published sequence of the reference clone (64), revised to include corrections reported by Bubb et al. (9), Halbert and Galloway (30), and Parton (54). We have included an additional correction, replacing the published GC at position 7432 with CGG, communicated to us by Bernard (6) and observed in the LCR sequences of our HPV16 isolates. The numbering system used here corresponds to that of original HPV16 sequence (64) as follows: (i) same as the original sequence from their nt 1 to 3905; (ii) one greater than the original sequence from their nt 3906 to 6900; (iii) four greater than the original sequence from their nt 6901 to 6950; (iv) one greater than the original sequence from their nt 6954 to 7431 (their nt 6951 to 6953 are not present in the corrected sequence); and (v) two greater than the original sequence from their nt 7434 to the end of the LCR segment discussed here.

PCR. For the *Thermus aquaticus* (*Taq*)-based PCRs, 100-μl amplification reaction mixtures contained 10 mM Tris (pH 8.3), 50 mM KCl, 200 μM each deoxynucleoside triphosphate (dNTP) (dATP, dCTP, dGTP, and dTTP), 2.5 or 4.0 mM $MgCl_2$, 0.125 to 0.5 μM each sense-strand and antisense-strand oligonucleotide primers, and 2.5 U of *Taq* DNA polymerase (Perkin-Elmer, Foster City, Calif.). In some cases, sequential nested PCRs were performed in order to increase the yields of the desired amplicons. In each case, a 2-μl aliquot of the PCR products from the outer reaction was used as the template for the inner reaction. PCR conditions and oligonucleotide primer sequences are shown in Table 1.

For the long-PCR technique (2, 14, 52), we used an XL PCR kit (Perkin-Elmer). Amplification reaction mixtures contained 1.1× DNA polymerase XL buffer, 200 μM each dNTP, 1.1 mM magnesium acetate, 0.25 μM each oligonucleotide primer, and 2 U of recombinant *Thermus thermophilus* DNA polymerase XL. PCR conditions and oligonucleotide primer sequences are shown in Table 1. The long PCRs were preceded by a manual hot start (15) at 80°C for 5 min.

HPV16 L2 and L1 amplicons were isolated from low-melting-temperature agarose gel (SeaPlaque; FMC Bioproducts, Rockland, Maine) following electrophoresis and ligated to pCRII vector DNA (Invitrogen, San Diego, Calif.).

Nucleotide sequence analysis. Single-stranded DNA templates were prepared from pCRII recombinant plasmids by alkaline denaturation (31, 32) and from biotinylated PCR amplicons by using streptavidin-coated magnetic beads (Dynabeads; Dynal, Lake Success, N.Y.). Nucleotide sequence was determined by the

TABLE 1. PCR conditions and primer sequences

ORF or LCR	nt range		PCR type	Primers[a]	PCR conditions[b]					
	5′	3′			Pol	No. of cycles	Temp (°C), time(s)			
							D	A	E	TE
E6	31	683	Nested (outer rxn[c])	CGTAACCGAAATCGGTTGAAC GGACCATCTATTTCATCCTCCT	*Taq*	35	95, 60	55, 60	72, 60	72, 300
	31	640	Nested (inner rxn)	CGTAACCGAAATCGGTTGAAC GCTCATAACAGTAGAGATC	*Taq*	25	95, 60	55, 60	72, 60	72, 300
	31	640	Unnested	CGTAACCGAAATCGGTTGAAC GCTCATAACAGTAGAGATC	*Taq*	40	95, 60	55, 60	72, 60	72, 300
L2	4377	5092	Nested (outer rxn)	TATGGIAGIATGGGIGTITTTTTTGG ATRTCIAKRAARTCIGGRTCIGGIGC	*Taq*	35	95, 60	55, 60	72, 60	72, 300
	4392	4954	Nested (inner rxn)	GTITTYTTYGGIYTIGGIATIGG GGRTCIAYIACYTTIACYTGYTGIGT	*Taq*	25	95, 60	55, 60	72, 60	72, 300
	4236	5690	Long	GCGGCCGCAAATATGCGACACAAACGTTCTGCAAAACGCACAAAACGT CCGCGGACTGGGACAGGAGGCAAGTAGACAGTGGCCTCA	r*Tth*[d]	40	94, 15	62, 60	68, 240	72, 600
	4001	5776	Nested (outer rxn)	AACAGCAGCCTCTGCGTTTAGGTG GTCCAACTGCAAGTAGTCTGGATG	*Taq*	35	95, 60	55, 60	72, 150	72, 300
	4239	5693	Nested (inner rxn)	CGACACAAACGTTCTGCAAAACGC GTCCAACTGCAAGTAGTCTGGATG	*Taq*	25	95, 60	55, 60	72, 150	72, 300
L1	5638	7181	Long	AGATCTAATCATGTCTCTTTGGCTGCCTAGTGAGGCC AAGCTTACACTAATTCAACATACATACAATACTTACAGC	r*Tth*	40	94, 15	62, 60	68, 240	72, 600
	5594	7313	Nested (outer rxn)	CATGTTACGAAAACGACGTAAACG GCATGACACAATAGTTACACAAGC	*Taq*	35	95, 60	55, 60	72, 150	72, 300
	5633	7222	Nested (inner rxn)	TTCAGATGTCTCTTTGGCTGCCTA GCATGACACAATAGTTACACAAGC	*Taq*	25	95, 60	55, 60	72, 150	72, 300
	6038	6370	Nested (inner rxn)	CTTATGCAGCAAATGCAGGTG TGTCGCCATATGGTTCTGAC	*Taq*	25	95, 60	55, 60	72, 60	72, 300
LCR	7336	7866	Unnested	CAACACCTACTAATTGTGTTGTGG AAATCGGTTTGCACACACCCATGT	*Taq*	40	95, 60	55, 60	72, 90	72, 300
	7291	7890	Nested (outer rxn)	CTTGTGTAACTATTGTGTCATGC AGTTGCTTGTAAATGTGTAACCC	*Taq*	35	95, 60	55, 60	72, 90	72, 300
	7336	7866	Nested (inner rxn)	CAACACCTACTAATTGTGTTGTGG AAATCGGTTTGCACACACCCATGT	*Taq*	25	95, 60	55, 60	72, 90	72, 300

[a] Sequences (top, sense; bottom, antisense) are written in the 5′-to-3′ direction. Positions at which more than one nucleotide was incorporated during the synthesis are indicated as follows: K = G + C, R = A + G, Y = T + C. I = deoxyinosine. Underlined segments are restriction enzyme recognition sequences that were added to the 5′ ends of some primers.

[b] PCR conditions include the polymerase (Pol), number of cycles, and temperature and time of the denaturation (D), annealing (A), extension (E), and terminal extension (TE) steps.

[c] rxn, reaction.

[d] r*Tth*, recombinant *T. thermophilus*.

dideoxy termination method (60), using a Sequenase version 2.0 kit (United States Biochemical, Cleveland, Ohio) with HPV16-specific or pCRII-specific oligonucleotide primers.

To reduce the likelihood that nucleotide variations were artifacts resulting from *Taq* polymerase error, changes were accepted only when they were reproducible from a second PCR. This strategy has been used previously for the identification of HPV16 molecular variants (11, 56). Because cloned DNA molecules from the long PCR may also contain polymerase artifacts (68), complete nucleotide sequences were determined for three or more clones from each PCR amplification; nucleotide changes were accepted as valid only when they were contained in at least three clones.

The nucleotide sequences were determined directly from the affinity-purified PCR amplicons without molecular cloning for all E6 amplicons, all LCR amplicons, and the L2 and L1 amplicons from 14 isolates (Portland, Ore., samples OR.6170, OR.8329, OR.8987, OR.2087, OR.4716, OR.7574, OR.5428, OR.7632, OR.6106, OR.1905, OR.3473, OR.7145, OR.3759, and OR.3136).

HPV16 L1 RFLP assay. HPV16 L1 nt 6038 to 6370 PCR amplicons were generated in a nested PCR by using either the L1 nt 5594 to 7313 amplicons or the L1 nt 5638 to 7181 amplicons as templates. The 333-bp L1 nt 6038 to 6370 PCR amplicons were digested with *Rsa*I at 37°C for 2 h. The DNA fragments were subjected to electrophoresis through a 15% polyacrylamide gel and stained with ethidium bromide. PCR products that showed *Rsa*I restriction fragment length polymorphisms (RFLPs) compared with the HPV16 reference sequence were digested with *Mae*III at 55°C and were analyzed similarly for the presence of RFLPs. The nt 6164 C-to-A change present in all HPV16 L2-A cluster variants (described in Results) results in an *Mae*III site (GTAAC), whereas an *Rsa*I site (GTACC) is present in non-L2-A variants (nt 6164 is indicated in boldface type; the restriction enzyme recognition sequence is underlined). HPV16 L2-A variants with an additional A-to-C change at nt 6179 contained a second *Mae*III site.

Phylogenetic analysis. Phylogenetic analyses were performed over the E6, L2, L1, and LCR of 30 HPV16 DNA samples, including the corrected HPV16 reference sequence (6, 9, 30, 54, 64, 73). All sequences from a given sample were combined, and the combined sequences were used to construct an alignment. Parsimony analyses were performed by using the Phylogenetic Analysis Using Parsimony (PAUP) package version 3.1.1 (69) on a Macintosh Quadra 800, running system software 7.1. MacClade version 3.03 (47) was used on the same system for analyzing the results of the PAUP analyses. A branch-and-bound search was conducted to find the most parsimonious tree or trees. Bootstrapping (100 bootstrap replicates, 10 random-addition and subtree pruning-regrafting search iterations per bootstrap) was used to confirm the robustness of the major phylogenetic groups. Maximum-likelihood analyses were performed with fastDNAml (53) on a Sun Sparc 10 running Sun OS 4.1.3. Ten-random addition replicates (jumbles) were performed, using the search option "global 1 1" and a transition/transversion ratio of 1.4; this transition/transversion ratio gave trees with the highest likelihood, indicating that it was approximately the optimal value (26). Neighbor-joining trees were computed on the same Sparc 10 system, using programs from PHYLIP 3.5 (23–25), specifically DNADIST version 3.53c, with the Kimura two-parameter model and a transition/transversion ratio of 1.4, and NEIGHBOR version 3.52c; for neighbor-joining bootstrap analysis, SEQBOOT 3.52c was used in conjunction with DNADIST and NEIGHBOR.

Analysis of synonymous and nonsynonymous mutations. Determination of degrees of synonymous and nonsynonymous nucleotide divergence for coding regions was performed by the method of Nei and Gojobori (44, 51).

Nucleotide sequence accession numbers. Nucleotide sequences have been submitted to GenBank and will be released upon publication. Accession numbers have been assigned as follows: U34078 to U34106 to the LCR sequences, U34107 to U34135 to the E6 sequences, U34136 to U34164 to the L2 sequences, and U34165 to U34193 to the L1 sequences.

RESULTS

Nucleotide sequences of an L2 coding segment (nt 4551 to 4871) and the complete E6 coding region (nt 104 to 559) were determined for HPV16 DNAs amplified from human cervicovaginal cell samples. HPV16 DNA-containing specimens were obtained from subjects enrolled in epidemiologic investigations conducted in Albuquerque, N.Mex. (n = 30) (5, 71), and Portland, Ore. (n = 137) (62). After analysis of the L2 and E6 nucleotide changes, isolates were grouped into variant classes. Nucleotide sequence variations in the E6 (nt 104 to 559), L2 (nt 4272 to 5657), and L1 (nt 5665 to 7148) ORFs, and the LCR (nt 7479 to 7842), were examined in 29 selected isolates that were representative of the L2 and E6 variant clusters. Previously reported corrections to the original HPV16 reference sequence (54, 73) were in agreement with sequences obtained from all isolates. The nucleotide and amino acid sequence data are summarized in Fig. 1 and 2. Three isolates

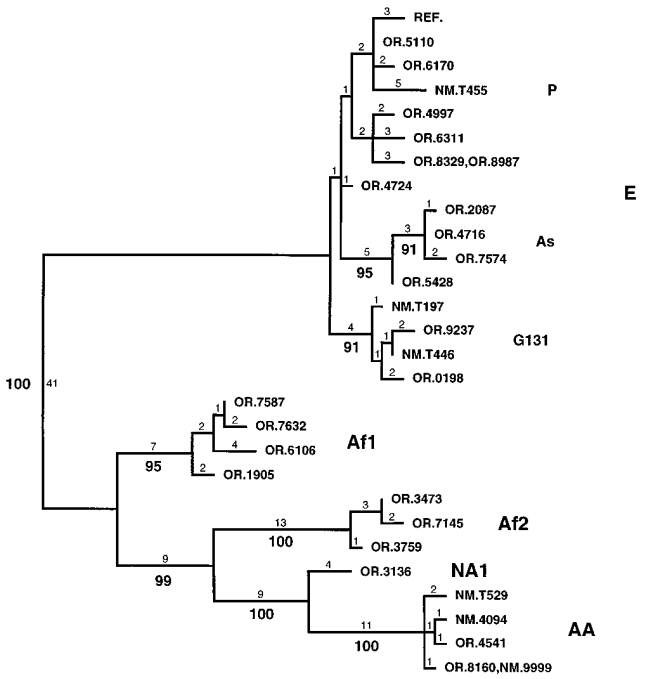

FIG. 3. Phylogenetic tree of the HPV16 variants based on parsimony analysis. Combined sequences from E6, L2, L1, and the LCR for 30 isolates (including the HPV16 reference [REF.] genome) yielded an alignment of 129 variable positions. The single most parsimonious tree (157 steps) is shown. Small numbers above branches indicate the number of steps (reconstructed point mutations) along the corresponding branch; the horizontal length of each branch is proportional to the number of steps, while vertical branch length is for layout only. Large numbers below branches indicate bootstrap values that are greater than 90%. The portions of the tree corresponding to variant classes E, Af1, Af2, NA1, and AA are indicated by the larger labels, and the E lineage subclasses P, As, and G131 are indicated with smaller labels.

representing a distinct AA subclass were identified during the E6 and LCR sequence analyses (OR.5691, OR.1783, and OR.8392 in Fig. 1 and 2), but they were not included in the phylogenetic studies discussed below; L2 and L1 sequences were not obtained for these isolates. L2 and L1 sequences for our isolates are compared with sequences of previously described variants: 114K and 114B and HPV16 isolates described by Kirnbauer et al. (42), and Sx and Tx are United Kingdom and Trinidad isolates, respectively, described by Pushko et al. (56).

The phylogenetic studies are presented first to facilitate a detailed discussion of variation in the HPV16 classes and subclasses.

Phylogenetic studies. Parsimony analysis was performed on a combined E6-L2-L1-LCR nucleotide sequence alignment of 3,690 nucleotides from each isolate, including the corrected HPV16 reference sequence (n = 30). The absence of insertions or deletions resulted in an unambiguous optimal alignment. The alignment contained 129 varied positions, of which 92 gave segregations which were potentially informative for parsimony analyses. The analysis yielded a single most parsimonious tree shown in Fig. 3. Bootstrapping confirmed the robustness of the four major lineages (E, AA, Af1, and Af2) with bootstrap values of 95% or more in each case; we will refer to these groups as classes. An isolated sequence (OR.3136) was identified. OR.3136 formed a novel branch that was substantially removed from the class AA isolates; however, OR.3136 and AA isolates are more closely related to each other than either is to Af2 isolates. We refer to this variant as North American

1 (NA1). Within the E class, two divergent subclasses, As and G131, can be recognized (bootstrap values of 95 and 91%, respectively). The As subclass has been identified previously (33), while the G131 subclass has not. The G131 variants would not have been identified through sequence analysis of the LCR alone. Moreover, the E6 coding region permitted discrimination of the maximum number of HPV16 variants across the shortest sequence segment located between nt 109 and 183 (Fig. 1). Although the parsimony analysis provided a single most parsimonious reconstruction of the complete tree, little confidence should be placed in the details of the arrangement of individual isolates within classes or subclasses, since these were based on only a few mutations.

Maximum-likelihood analysis also gave a single optimal tree (6 of 10 jumbles gave identical trees; the other 4 gave slightly less likely trees which maintained the same major branchings). The best tree from the maximum-likelihood analysis differed from that of the parsimony analysis only in the branching within the E class variants. Within the E class, the G131 and As subclasses were maintained. Bootstrapping of neighbor-joining trees gave 95% or greater bootstrap values for the branches defining the major classes and 90% or greater for the As and G131 subclasses (not shown). Neighbor-joining trees for each of the regions (E6, L2, L1, and LCR) gave the same major classes (not shown).

Sequence variations within the HPV16 L2 human antibody-reactive region. Jenison and coworkers previously identified an HPV16 L2 polypeptide segment between aa 149 and 204 that contains three distinct human antibody-reactive epitopes (39). The L2 genome segment (nt 4551 to 4871) that encodes the epitope-containing region was used to screen HPV16 DNA-containing samples for nucleotide sequence variations.

A subset of HPV16 DNA-containing human cervicovaginal cell samples (n = 51) was analyzed from study subjects in Albuquerque, N.Mex. (n = 30), and Portland, Ore. (n = 21). The HPV16 L2 nt 4392 to 4954 segment was amplified by PCR using degenerate HPV L2 consensus primers (Table 1). Nested L2 consensus primer sequences were designed from conserved amino acid segments of HPV16, HPV31, HPV33, HPV35, and HPV58. The primers efficiently amplified HPV16, HPV31, HPV33, HPV35, HPV52, HPV58, and HPV67 DNAs from cervicovaginal cell samples obtained from the same study populations (data not shown). The HPV16 L2 nt 4551 to 4871 segment of the PCR amplicons was sequenced for all samples.

Of the 51 samples, 39 (76.5%) contained L2 nt 4551 to 4871 sequences that were identical to the HPV16 reference genome (64); we refer to this cluster of isolates as L2-P, or prototype-like. Twelve samples (23.5%) contained sequences with one or more nucleotide differences and fell into three distinct clusters, which we have designated L2-A (n = 6), L2-B (n = 5), and L2-C (n = 1) (Fig. 1 and 2). As stated previously, extended E6, L2, L1, and LCR sequences were obtained for selected L2-P (n = 4), L2-A (n = 6), and L2-B (n = 4) variants and the L2-C variant; results are included in Fig. 1 and 2. Subsequent LCR analysis indicated that the L2-A variants were contained within HPV16 classes AA (n = 5) and Af1 (n = 1) (11, 33, 34). The L2-P (n = 4) and L2-B (n = 4) isolates and the L2-C isolate were contained within the HPV16 E class.

Additional L2-A variants were identified among Portland, Ore., samples by using an HPV16 L1 RFLP assay that differentiated HPV16 L2-A from non-L2-A isolates. Of 137 HPV16 DNA-containing samples examined, 19 (13.9%) contained RFLPs consistent with cluster L2-A isolates. By LCR analysis, all L2-A variants identified by the RFLP assay were found to be members of the lineage composed of the HPV16 AA, Af1, Af2, and NA1 classes.

Of the seven nucleotide differences detected within the HPV16 L2 epitope region, only the T-to-C change at nt 4599 results in a change in the predicted L2 amino acid sequence (serine to proline at aa 122). This L2 aa 122 change was present in all AA, Af1, Af2, and NA1 isolates. Thirty-nine HPV16 E class isolates were similarly analyzed. None contained amino acid changes within the HPV16 L2 epitope region relative to the HPV16 reference genome.

HPV16 E6 sequence variations. The complete E6 coding region was amplified and sequenced from the same specimens (n = 137) characterized in the HPV16 L1 RFLP assay described above. Representative HPV16 E6 sequences are included in Fig. 1 and 2. HPV16 E6 variant isolates segregated into two main clusters, those that contained zero to three nucleotide changes and those that contained five to eight nucleotide changes relative to the HPV16 reference genome. The former group included all E class isolates, and the latter group included all AA, Af1, Af2, and NA1 isolates.

Within the HPV16 E class, the G131 (n = 4) and As (n = 4) subclasses contained distinctive E6 nucleotide changes relative to the reference genome. Four of four G131 isolates contained E6 changes at nt 131 (A to G) and at nt 350 (A to G), and four of four As variants contained an E6 change at nt 178 (T to G). Of the HPV16 isolates examined in this study, only the isolates from the G131 subclass contained the A-to-G change at nt 131, which results in an arginine-to-glycine change at E6 aa 10. On the basis of E6 sequence analysis, all L2-B isolates described above were subsequently found to be contained within the G131 subclass. The single L2-C isolate was found to be a minor variant of the prototypic HPV16 E class and was identical to the HPV16 reference genome within the E6 coding segment. E6 changes observed in more than one of the eight prototype-like E isolates (P subclass isolates) examined included (i) an E6 T-to-G change nt 350, resulting in a leucine-to-valine change at E6 aa 83 (n = 4), and (ii) a synonymous E6 T-to-C change at nt 109 (n = 3). Among the P subclass variants, the E6 T-to-G change at nt 350 was linked to a distinctive A-to-T change at L2 nt 5225.

All HPV16 AA, Af1, Af2, and NA1 isolates (n = 16) contained E6 changes at nt 145 (G to T), nt 286 (T to A), nt 289 (A to G), and nt 335 (C to T). In addition to these consensus changes, class-specific changes were present. AA variants (n = 8) contained changes at nt 532 (A to G) and at nt 350 (T to G); a unique AA subclass containing an additional T-to-G change at nt 183 was identified in three isolates. We refer to this AA subclass as G183. Further support for the distinction of this variant subclass was provided by subsequent LCR sequence determinations. Af1 variants (n = 4) contained a change at nt 132 (G to C), and Af2 variants (n = 3) contained three specific changes at nt 109 (T to C), at nt 132 (G to T), and at nt 403 (A to G). An E6 nt 143 C-to-G change was common to the Af1 and Af2 isolates. The NA1 isolate contained the four AA-Af1-Af2 E6 consensus changes and the E6 nt 350 T-to-G change seen among AA isolates and some E isolates but lacked the E6 nt 532 change present in all AA isolates.

At the amino acid level, the only HPV16 E6 change shared by all AA, Af1, Af2, and NA1 isolates was a histidine-to-tyrosine change at E6 aa 78. AA variants contained additional amino acid changes at E6 aa 14 (glutamine to histidine) and at E6 aa 83 (leucine to valine); the AA subclass G183 isolates also contained a change at aa 27 (isoleucine to arginine). Af1 variants showed a specific change at E6 aa 10 (arginine to threonine), and Af2 variants showed a different change at E6 aa 10 (arginine to isoleucine). Af1 and Af2 isolates shared a change at E6 aa 14 (glutamine to aspartic acid). The distribution of E6

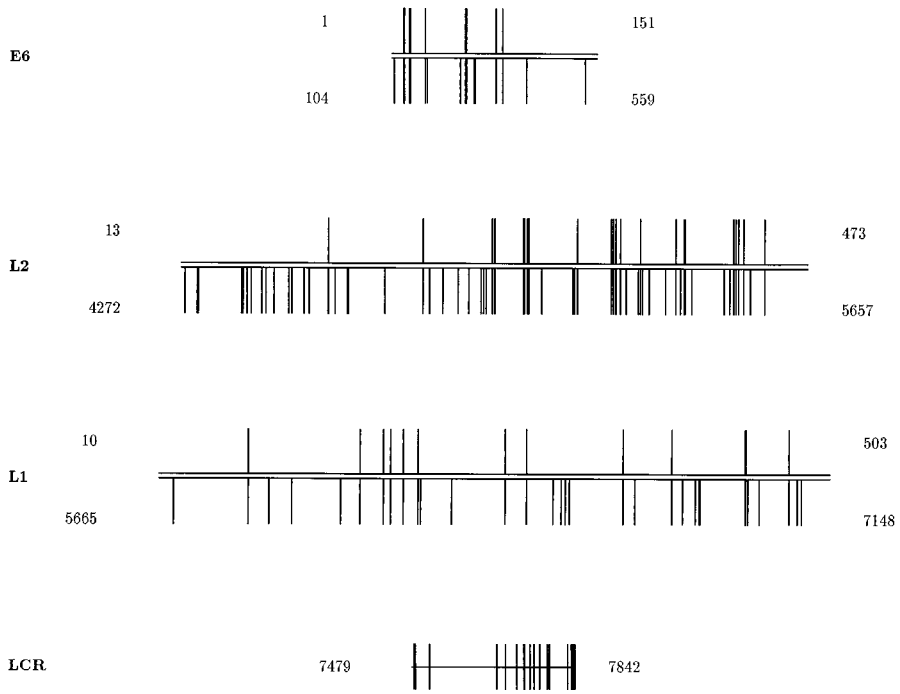

FIG. 4. Distribution of nucleotide and amino acid changes among HPV16 variants within E6, L2, and L1 coding sequences and a segment of the LCR. The horizontal lines represent the sequenced portions of each region. For the coding regions, the beginning and ending nucleotide positions are indicated below the lines to the left and right, respectively; the beginning and ending residues of the predicted amino acid sequences are indicated above the lines. Vertical bars below the lines represent the positions of nucleotide substitutions, while vertical bars above the lines represent predicted amino acid changes. For the LCR, the beginning and ending nucleotide positions are indicated to the left and the right of the line, while bars through the line represent the positions of nucleotide substitutions.

nucleotide and amino acid variations is shown in Fig. 4. The amino-terminal region of E6 is relatively varied.

HPV16 L2 and L1 sequence variations. The HPV16 AA-Af1-Af2-NA1 isolates (n = 13) all contained nine nucleotide changes in L2 and 11 nucleotide changes in L1 that were restricted to this group. In addition to these changes, the AA variants (n = 5) contained five class-specific changes at nt 4598 (A to C), nt 4943 (A to G), nt 5294 (A to C), nt 5474 (T to A), and nt 6802 (A to T). The Af1 variants (n = 4) contained five class-specific changes at nt 4307 (G to A), nt 4427 (G to A), nt 4910 (A to T), nt 5289 (A to G), and nt 6567 (T to A). The Af2 variants (n = 3) contained six class-specific changes at nt 4517 (A to G), nt 4544 (T to G), nt 4853 (C to T), nt 5258 (A to G), nt 5289 (A to C), and nt 6481 (T to C). Among the HPV16 E class isolates (n = 16), the As subclass (n = 4) showed two subclass-specific changes at nt 5517 (A to C) and nt 7059 (G to A). The G131 isolates (n = 4) contained two subclass-specific changes at nt 4613 (T to C) and nt 6861 (T to C). A total of 15 and 5 class- and subclass-specific nucleotide changes are found in the L2 and L1 coding regions, respectively.

At the amino acid level, 13 of 13 HPV16 AA-Af1-Af2-NA1 isolates contained five changes in L2 and three changes in L1 that were restricted to this group. Specific amino acid changes in the L2 coding region were observed at aa 122 (serine to proline), aa 385 (valine to isoleucine), aa 420 (isoleucine to threonine), aa 424 (aspartic acid to threonine), and aa 443 (aspartic acid to glycine). Specific amino acid changes in the L1 coding region were observed at aa 76 (histidine to tyrosine), aa 176 (threonine to asparagine), and aa 474 (leucine to phenylalanine). The As isolates contained one subclass-specific change within the L2 coding region at aa 428 (isoleucine to leucine). The G131 isolates did not contain subclass-specific amino acid changes.

The distribution of nucleotide and amino acid variations observed in the L2 and L1 coding regions is shown in Fig. 4. Amino acid variations in the L1 protein are distributed relatively evenly over the L1 coding region. In contrast, amino acid variations in the HPV16 L2 protein are not distributed evenly over the length of the polypeptide. Although the two halves of L2 have similar overall nucleotide mutation rates, the ratios of synonymous to nonsynonymous changes are quite different. Nearly all of the substitutions in the 5′ half of L2 are synonymous changes, while the majority of substitutions in the 3′ half are nonsynonymous. The effect is highly significant according to a chi-square test of dependence (chi squared = 14.75, 1 df, $P < 0.01$). This asymmetry in the distribution of HPV16 L2 amino acid changes has been also noted in an HPV16 Af2 variant from Zaire (36).

HPV16 LCR sequence variations. Nucleotide changes observed in the LCR (n = 32, including the three AA subclass G183 isolates) were the same as those previously reported by Ho and coworkers (33) except for the NA1 isolate. Although we did not observe all of the unique HPV16 LCR sequence variants reported elsewhere, all previously documented major classes are represented in our data. The greater variation observed by Ho and coworkers is most likely related to the number of isolates studied and the worldwide sampling strategy. As noted above, we distinguished a greater number of clusters than are defined in the LCR by characterizing nucleotide changes present in E6, L2, and L1 coding regions; a short segment of E6 (nt 109 to 183) is capable of distinguishing all classes and subclasses.

Comparison of sequenced regions. Several measures of variability for each of the sequenced regions are shown in Table 2. We considered the following: (i) the number and percentage of varied positions; (ii) the number of mutations and the number

TABLE 2. Comparison of nucleotide sequence variability between E6, L2, L1, and the LCR among HPV16 isolates, for several measures of variability

Region	No. of bases sequenced	Variable positions[a]		Mutations[b]		Maximum pairwise distance[c]		Maximum pairwise pS[d] (%)	Maximum pairwise pN[e] (%)
		n	%	*n*	%	*n*	%		
E6	456	15	3.3	21	4.6	6	1.3	3.3	0.82
L2	1,386	64	4.6	79	5.7	33	2.4	5.8	1.2
L1	1,484	32	2.2	36	2.4	18	1.2	3.0	0.69
LCR	364	18	4.9	21	5.8	9	2.5		
Aggregate[f]	3,690	129	3.2	157	4.2	66	1.8	4.3[g]	0.94[g]

[a] Number and percentage of positions showing variation among 30 isolates (29 clinical isolates plus the HPV16 reference sequence).
[b] Number of mutations for the most parsimonious phylogenetic reconstruction, including multiple changes per position, and percentage of changes over positions.
[c] Number and percentage of positions with different bases for NM.T455 with NM.T529, two of the most diverged isolates.
[d] Percent divergence for synonymous changes between NM.T455 and NM.T529, calculated according to method 1 of Nei and Gojobori (44, 51).
[e] Percent divergence for nonsynonymous changes between NM.T455 and NM.T529, calculated according to method 1 of Nei and Gojobori (44, 51).
[f] Aggregate based on combined data from E6, L2, L1, and the LCR.
[g] Aggregate value based on combined data from E6, L2, and L1.

of mutations per site as reconstructed from the most parsimonious tree (47); (iii) the simple (Hamming) distance, or dissimilarity, between two of the most divergent sequences; and (iv) the rate or degree of divergence for synonymous and nonsynonymous substitutions (51) between the same two sequences. Only differences between L2 and L1 achieve significance ($P < 0.05$) by Student's *t* test.

DISCUSSION

In this study, we describe HPV16 intratypic diversity over a span of 3,690 nucleotide bases among 29 selected United States isolates. We also present E6 and LCR sequences from a select group of HPV16 AA variants that would most likely represent an AA subclass G183. The data set complements and expands previous descriptions of worldwide HPV16 variants based on nucleotide sequences of E7, E5, L2, L1, and LCR segments (4, 11, 22, 27, 28, 33, 34, 37, 38, 56, 65, 76).

We identified a unique subclass within the HPV16 class E that we have designated G131. G131 isolates did not contain distinctive nucleotide changes within the LCR segment (nt 7479 to 7842) compared with other class E isolates and would not have been identified by LCR analysis alone. Likewise, the HPV16 isolate OR.3136 did not fall within any of the previously defined classes. NA1 appears to be a descendant of an evolutionary intermediate that diverged from the AA lineage well after the AA lineage diverged from the Af2 lineage. As such, the G131 and NA1 nucleotide sequences provide novel contributions to HPV16 phylogeny.

With the exception of the subclass G131 isolates and the NA1 isolate, the United States HPV16 samples were members of the major HPV16 classes and subclasses described previously (11, 33, 34). The major groupings deduced from L2 and E6 sequence analysis corresponded to the LCR-based HPV16 E, AA, Af1, and Af2 classes (33). The E6 sequence provided more phylogenetic information than the LCR sequence, because it identified all LCR-based groups and also identified the E subclass G131 (Fig. 1). Similar to LCR studies, E6 information can easily be obtained by targeting a short continuous segment of the HPV16 genome. Of the regions examined here, these features make E6 a particularly good choice for distinguishing HPV16 variants.

We identified a set of signature patterns for each HPV16 lineage in each of the sequenced regions that allows tentative classification of HPV16 isolates (Fig. 1). This catalog of signature positions is currently being used by our group to develop oligonucleotide probes for specific HPV16 lineages. These assays should assist future epidemiological studies of HPV16 infections.

Nucleotide changes present in the E6, L2, L1, and LCR segments were phylogenetically compatible. None of the isolates that we examined appeared to be the products of genetic recombination between viruses of different HPV16 lineages. This observation is in agreement with those of Icenogle et al. (38) and Chan et al. (11) and fails to confirm previous data suggesting a lack of linkage between HPV16 genome segments observed in Caribbean HPV16 isolates (56, 65). The indication in our data that recombination between variants of HPV16 is rare or nonexistent implies that nucleotide changes in one region can be used as markers of changes found in other regions within the same lineage.

The data presented here permit good resolution of certain details of the HPV16 phylogenetic tree shown in Fig. 3. As with previous phylogenetic analyses based on LCR variations (33), we find that the bulk of the mutations separate the E class from the AA, Af1, Af2, and NA1 classes. The midpoint of the tree, which corresponds approximately to the probable HPV16 ancestral root, lies along the branch separating the E class from the AA, Af1, Af2, and NA1 classes. AA and Af2 lie closer to one another than to Af1; this is a robust finding, with a bootstrap value for parsimony analysis of 99%.

When rates of observed mutation for different HPV16 genome regions are compared (Table 2), only differences between L2 and L1 achieve significance ($P < 0.05$) by Student's *t* test. Nonetheless, it is interesting that in all measures of variability, L2 is more varied than the other coding regions and is approximately as varied as the LCR. L1 is the most conserved region. With regard to the LCR, the overall rate of mutation in the sequenced region falls between the synonymous and nonsynonymous rates of substitution for the coding regions. This finding suggests that the LCR segment, while under less selectional pressure than the coding regions, is far from unconstrained. Overall, the nonsynonymous-to-synonymous ratios are typical of human nuclear genes, in contrast to the higher ratios seen in some retroviral genes (44).

Nucleotide sequence variations at specific positions were examined. By direct inspection of nucleotide sequence alignments, sites at which there are three or four different bases are seen (nt 5225, for example). For the current data set, the hypothetical ancestry of each nucleotide position could be re-

constructed according to the principle of maximum parsimony, and additional positions which had undergone parallel or back mutations could be identified. Sites that had undergone at least two mutations in different lineages are indicated in Fig. 1 and 2. Some such sites may represent mutational or selectional hot spots, although many of these sites involve only silent (synonymous) changes. Further correlation with functional information is needed to evaluate the significance of variation at these positions. The large number of changes in the 5′ region of E6 is an intriguing example.

The codons for E6 residues 10 and 14 show amino acid-replacing changes at two nucleotide positions each. A possible explanation for this concentration of changes could be selective immune pressures on the region. The HPV16 E6 protein appears to be the target of both humoral and cellular immune responses (3, 21, 67, 70). Immunologic relevance of the E6 N-terminal region is supported by the demonstration of an endogenously processed HLA A*0201-restricted peptide (KLPQLCTEL; E6 aa 11 to 19) as well as an overlapping HLA B-7-restricted peptide (RPRKLPQL; E6 aa 8 to 15) in this region (3, 21). Cellular immune responses may be modified by host factors such as presenting HLA molecules, viral factors such as E6 and E7 sequence variability, or a combination of these factors.

In addition to the potential immunologic relevance of E6 and E7, these proteins are expressed in HPV-associated genital cancers and immortalize human keratinocytes in vitro, suggesting that they may act in the generation and maintenance of the transformed phenotype (1, 63, 66). HPV16 E6 binds to the cellular protein p53 and promotes its degradation by the ubiquitin pathway, one mechanism by which E6 mediates its transforming activities (18, 49, 61). Recent studies have identified an interaction between HPV16 E6 and a putative calcium-binding protein that may be important to p53-independent transforming activities (13).

The amino-terminal half of the HPV16 L2 ORF was extremely conserved, and most amino acid changes were observed in the carboxy-terminal half of the protein (Fig. 4). Several lines of evidence suggest that the amino-terminal segments of the papillomavirus L2 polypeptides are exposed on the viral surface, are immunoreactive, and can generate neutralizing antibody responses (16, 39, 43, 45, 57, 58, 72, 74, 75). Mouse antibodies generated by immunization with purified HPV1 virions recognize a linear epitope between HPV1 L2 aa 102 and 108 (75). A bovine papillomavirus type 1 (BPV-1) L2 neutralizing epitope has been mapped to the segment between BPV L2 aa 45 and 173 (58). BPV-1 L2 antisera neutralize BPV infectivity in vitro, perhaps by interfering with viral entry into cells (57), and a virus neutralizing epitope has been localized to a segment between aa 101 and 120 in BPV-4 (12). Immunization with cottontail rabbit papillomavirus L2 recombinant proteins protects rabbits from experimental infections with this virus (16, 45). Within the HPV16 L2 polypeptide, multiple human antibody-reactive epitopes have been mapped to a region between L2 aa 149 and 204 (39). Homologous L2 segments of HPV6, HPV11, and HPV18 are highly divergent and react with human antibodies in an HPV type-specific manner (39, 74).

Given these observations, we expected that immune pressures might lead to more amino acid variations within the amino-terminal half of L2. However, our findings suggest that this region of the HPV16 L2 polypeptide is subject to strong negative selectional pressures. The amino-terminal region may be vital to the structure and the functions of the L2 protein such that mutations are poorly tolerated. If the HPV16 L2 protein generates protective antibodies in response to vaccination, it is encouraging that very little amino acid sequence variation exists within this region among different HPV16 isolates.

The HPV major capsid protein, encoded by the L1 ORF, is the target of neutralizing antibody responses (20, 40, 45, 57, 58, 59), and naturally occurring L1 antibodies react almost exclusively with conformation-dependent epitopes (41). Therefore, little is known about the locations or structures of the epitopes with which L1 antibodies react. It is clear from the findings of Kirnbauer and coworkers that a single amino acid change can markedly affect the efficiency with which HPV16 L1 recombinant proteins self-assemble into virus-like particles (42). It is possible that one or more of the L1 amino acid changes present among the HPV16 variants described here could similarly affect L1 conformation-dependent epitopes that are relevant to virus neutralization. It will be difficult to examine cross-protection between HPV16 variants with the lack of an HPV16 neutralization model system. In this regard, caution should be taken when interpreting the results of serologic assays that propose to evaluate cross-reactivity of antibodies to HPV16 variants. Polyclonal serological responses measured in these assays are not necessarily measures of neutralizing antibody, and neutralization will be critical for HPV16 prophylactic vaccine efficacy.

Our data set provides information demonstrating the genetic linkage of various HPV16 genomic segments and the relative degrees of selectional pressure between genomic regions. Additionally, these data form a basis for considering HPV16 sequence variations in the development of HPV16 L1-, L2-, or E6-based vaccine strategies, as correlates of cervical disease risk, and as tools for studying epidemiological issues such as viral persistence and transmission.

ACKNOWLEDGMENTS

We thank C. Parmenter, J. Herron, C. Morris, T. Mulcahy, and A. Pastusyn for technical assistance; J. Icenogle, J. Neidhart, J. Oliver, C. Greer, and T. Becker for helpful advice and support; and G. Myers for critical review of the manuscript.

This work was supported by Public Health Service grants CA57975 to S.A.J. and AI32917 to C.M.W. from the National Institutes of Health and by the University of New Mexico Cancer Research and Treatment Center (UNM-CRTC) Molecular Analysis Facility. T.Y. is a postdoctoral fellow sponsored by the UNM-CRTC and the UNM Department of Medicine. A.L.H. is a postdoctoral research associate sponsored by the Division of Microbiology and Infectious Diseases of the National Institute of Allergy and Infectious Diseases through an interagency agreement with the Los Alamos National Laboratory. A.-C.M.S. is a postdoctoral fellow sponsored by the International Agency for Research on Cancer.

REFERENCES

1. **Androphy, E. J., N. L. Hubbert, J. T. Schiller, and D. R. Lowy.** 1987. Identification of the HPV-16 E6 protein from transformed mouse cells and human cervical carcinoma cell lines. EMBO J. **6:**989–992.
2. **Barnes, W. M.** 1994. PCR amplification of up to 35-kb DNA with high fidelity and high yield from bacteriophage templates. Proc. Natl. Acad. Sci. USA **91:**2216–2220.
3. **Bartholomew, J. S., S. N. Stacey, B. Coles, D. J. Burt, J. R. Arrand, and P. L. Stern.** 1994. Identification of a naturally processed HLA A02011-restricted viral peptide from cells expressing human papillomavirus type 16 E6 oncoprotein. Eur. J. Immunol. **24:**3175–3179.
4. **Bavin, P. J., P. G. Walker, and V. C. Emery.** 1993. Sequence microheterogeneity in the long control region of clinical isolates of human papillomavirus type 16. J. Med. Virol. **39:**267–272.
5. **Becker, T. M., C. M. Wheeler, N. S. McGough, C. A. Parmenter, S. W. Jordan, C. A. Stidley, R. S. McPherson, and M. H. Dorin.** 1994. Sexually transmitted diseases and other risk factors for cervical dysplasia among southwestern Hispanic and non-Hispanic white women. JAMA **271:**1181–1188.
6. **Bernard, H.-U.** Personal communication.
7. **Bernard, H.-U., S.-Y. Chan, M. M. Manos, C.-K. Ong, L. L. Villa, H. Delius,**

C. L. Peyton, H. M. Bauer, and C. M. Wheeler. 1994. Identification and assessment of known and novel human papillomaviruses by polymerase chain reaction amplification, restriction fragment length polymorphisms, nucleotide sequence, and phylogenetic algorithms. J. Infect. Dis. **170:**1077–1085.

8. **Bosch, F. X., M. M. Manos, N. Muñoz, M. Sherman, A. M. Jansen, J. Peto, M. H. Schiffman, V. Moreno, R. Kurman, K. V. Shah, and the I.B.S.C.C. Study Group.** 1995. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. J. Natl. Cancer Inst. **87:**796–802.
9. **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.
10. **Chan, S.-Y., H. Delius, A. L. Halpern, and H.-U. Bernard.** 1995. Analysis of genomic sequences of 95 papillomavirus types: uniting typing, phylogeny, and taxonomy. J. Virol. **69:**3074–3083.
11. **Chan, S.-Y., L. Ho, C.-K. Ong, V. Chow, B. Drescher, M. Dürst, J. ter Meulen, L. Villa, F. Luande, H. N. Mgaya, and H.-U. Bernard.** 1992. Molecular variants of human papillomavirus type 16 from four continents suggest ancient pandemic spread of the virus and its coevolution with humankind. J. Virol. **66:**2057–2066.
12. **Chandrachud, L. M., G. J. Grindlay, G. M. McGarvie, B. W. O'Neil, E. R. Wagner, W. F. H. Jarrett, and M. S. Campo.** 1995. Vaccination of cattle with the N-terminus of L2 is necessary and sufficient for preventing infection by bovine papillomavirus-4. Virology **211:**204–208.
13. **Chen, J. J., C. E. Reid, V. Band, and E. J. Androphy.** 1995. Interaction of papillomavirus E6 oncoproteins with a putative calcium-binding protein. Science **269:**529–531.
14. **Cheng, S., C. Fockler, W. M. Barnes, and R. Higuchi.** 1994. Effective amplification of long targets from cloned inserts and human genomic DNA. Proc. Natl. Acad. Sci. USA **91:**5695–5699.
15. **Chou, Q., M. Russell, D. E. Birch, J. Raymond, and W. Bloch.** 1992. Prevention of pre-PCR mis-priming and primer dimerization improve low-copy-number amplification. Nucleic Acids Res. **20:**1717–1723.
16. **Christensen, N. D., J. W. Kreider, N. C. Kan, and S. L. DiAngelo.** 1991. The open reading frame L2 of cottontail rabbit papillomavirus contains antibody-inducing neutralizing epitopes. Virology **181:**572–579.
17. **Coggins, J. R., and H. zur Hausen.** 1979. Workshop on papillomaviruses and cancer. Cancer Res. **39:**545–546.
18. **Crook, T., J. A. Tidy, and K. H. Vousden.** 1991. Degradation of p53 can be targeted by HPV E6 sequences distinct from those required for p53 binding and trans-activation. Cell **67:**547–556.
19. **de Villiers, E.-M.** 1989. Heterogeneity of the human papillomavirus group. J. Virol. **63:**4898–4903.
20. **Dillner, J., F. Wiklund, P. Lenner, C. Eklund, V. Fredriksson-Shanazarian, J. T. Schiller, M. Hibma, G. Hallmans, and U. Stendahl.** 1995. Antibodies against linear and conformational epitopes of human papillomavirus type 16 that independently associate with incident cervical cancer. Int. J. Cancer **60:**377–382.
21. **Ellis, J. R. M., P. J. Keating, J. Baird, E. F. Hounsell, D. V. Renouf, M. Rowe, D. Hopkins, M. F. Duggan-Keen, J. S. Bartholomew, L. S. Young, and P. L. Stern.** 1995. The association of an HPV16 oncogene variant with HLA-B7 has implications for vaccine design in cervical cancer. Nature (Medicine) **1:**464–470.
22. **Eschle, D., M. Dürst, J. ter Meulen, J. Luande, H. C. Eberhardt, M. Pawlita, and L. Gissmann.** 1992. Geographical dependence of sequence variation in the E7 gene of human papillomavirus type 16. J. Gen. Virol. **73:**1829–1832.
23. **Felsenstein, J.** 1982. Numerical methods for inferring evolutionary trees. Q. Rev. Biol. **57:**379–404.
24. **Felsenstein, J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:**783–791.
25. **Felsenstein, J.** 1989. PHYLIP—phylogeny inference package (version 3.2). Cladistics **5:**164–166.
26. **Felsenstein, J.** Personal communication.
27. **Fujinaga, Y., K. Okazawa, A. Nishikawa, Y. Yamakawa, M. Fukushima, I. Kato, and K. Fujinaga.** 1994. Sequence variation of human papillomavirus type 16 E7 in preinvasive and invasive cervical neoplasias. Virus Genes **9:**85–92.
28. **Fujinaga, Y., K. Okazawa, Y. Ohashi, Y. Yamakawa, M. Fukushima, I. Kato, and K. Fujinaga.** 1990. Human papillomavirus type 16 E7 gene sequence in human cervical carcinoma analysed by polymerase chain reaction and direct sequencing. Tumor Res. **25:**85–91.
29. **Hagensee, M. E., N. Yaegashi, and D. A. Galloway.** 1993. Self-assembly of human papillomavirus type 1 capsids by expression of the L1 protein alone or by coexpression of the L1 and L2 capsid proteins. J. Virol. **67:**315–322.
30. **Halbert, C. L., and D. A. Galloway.** 1988. Identification of the E5 open reading frame of human papillomavirus type 16. J. Virol. **62:**1071–1075.
31. **Haltiner, M., T. Kempe, and R. Tjian.** 1985. A novel strategy for constructing clustered point mutations. Nucleic Acids Res. **13:**1015–1025.
32. **Hattori, M., and Y. Sakaki.** 1986. Dideoxy sequencing method using denatured plasmid templates. Anal. Biochem. **152:**232–238.
33. **Ho, L., S.-Y. Chan, R. D. Burk, B. C. Das, K. Fujinaga, J. P. Icenogle, T. Kahn, N. Kiviat, W. Lancaster, P. Mavromara-Nazos, V. Labropoulou, S. Mitrani-Rosenbaum, M. Norrild, M. R. Pillai, J. Stoerker, K. Syrjaenen, S. Syrjaenen, S.-K. Tay, L. L. Villa, C. M. Wheeler, A.-L. Williamson, and H.-U. Bernard.** 1993. The genetic drift of human papillomavirus type 16 is a means of reconstructing prehistoric viral spread and the movement of ancient human populations. J. Virol. **67:**6413–6423.
34. **Ho, L., S.-Y. Chan, V. Chow, T. Chong, S.-K. Tay, L. L. Villa, and H.-U. Bernard.** 1991. Sequence variants of human papillomavirus type 16 in clinical samples permit verification and extension of epidemiological studies and construction of a phylogenetic tree. J. Clin. Microbiol. **29:**1765–1772.
35. **Howley, P. M.** 1991. Role of the human papillomaviruses in human cancer. Cancer Res. **51:**S5019–S5022.
36. **Icenogle, J. P.** Personal communication.
37. **Icenogle, J. P., M. Laga, D. Miller, R. A. Tucker, and W. C. Reeves.** 1992. Genotypes and sequence variants of human papillomavirus DNAs from human immunodeficiency virus type 1-infected women with cervical intraepithelial neoplasia. J. Infect. Dis. **166:**1210–1216.
38. **Icenogle, J. P., P. Sathya, D. L. Miller, R. A. Tucker, and W. E. Rawls.** 1991. Nucleotide and amino acid sequence variation in the L1 and E7 open reading frames of human papillomavirus type 6 and type 16. Virology **184:**101–107.
39. **Jenison, S. A., X.-P. Yu, J. M. Valentine, and D. A. Galloway.** 1991. Characterization of human antibody-reactive epitopes encoded by human papillomavirus types 16 and 18. J. Virol. **65:**1208–1218.
40. **Kirnbauer, R., F. Booy, N. Cheng, D. R. Lowy, and J. T. Schiller.** 1992. Papillomavirus L1 major capsid protein self-assembles into virus-like particles that are highly immunogenic. Proc. Natl. Acad. Sci. USA **89:**12180–12184.
41. **Kirnbauer, R., N. L. Hubbert, C. M. Wheeler, T. M. Becker, D. R. Lowy, and J. T. Schiller.** 1994. A virus-like particle enzyme-linked immunosorbent assay detects serum antibodies in a majority of women infected with human papillomavirus type 16. J. Natl. Cancer Inst. **86:**494–498.
42. **Kirnbauer, R., J. Taub, H. Greenstone, R. Roden, M. Dürst, L. Gissmann, D. R. Lowy, and J. T. Schiller.** 1993. Efficient self-assembly of human papillomavirus type 16 L1 and L1-L2 into virus-like particles. J. Virol. **67:**6929–6936.
43. **Köchel, H. G., M. Monazahian, K. Sievert, M. Höhne, C. Thomssen, A. Teichmann, P. Arendt, and R. Thomssen.** 1991. Occurrence of antibodies to L1, L2, E4 and E7 gene products of human papillomavirus types 6b, 16 and 18 among cervical cancer patients and controls. Int. J. Cancer **48:**682–688.
44. **Korber, B. M., B. MacInnes, R. F. Smith, and G. Myers.** 1994. Mutational trends in V3 loop protein sequences observed in different genetic lineages of human immunodeficiency virus type 1. J. Virol. **68:**6730–6744.
45. **Lin, Y.-L., L. A. Borenstein, R. Selvakumar, R. Ahmed, and F. O. Wettstein.** 1992. Effective vaccination against papilloma development by immunization with L1 or L2 structural protein of cottontail rabbit papillomavirus. Virology **187:**612–619.
46. **Lowy, D. R., R. Kirnbauer, and J. T. Schiller.** 1994. Genital human papillomavirus infection. Proc. Natl. Acad. Sci. USA **91:**2436–2440.
47. **Maddison, W. P., and D. R. Maddison.** 1992. MacClade: analysis of phylogeny and character evolution. Sinauer, Sunderland, Mass.
48. **Münger, K., W. C. Phelps, V. Bubb, P. M. 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.
49. **Münger, K., M. Scheffner, J. M. Huibregtse, and P. M. Howley.** 1992. Interactions of HPV E6 and E7 oncoproteins with tumour suppressor gene products. Cancer Surv. **12:**197–217.
50. **Muñoz, N., L. Crawford, and P. Coursaget.** 1995. HPV vaccines and their potential use in the prevention and treatment of cervical neoplasia. Papillomavirus Rep. **6:**54–55.
51. **Nei, M., and T. Gojobori.** 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:**418–426.
52. **Ohler, L. D., and E. A. Rose.** 1992. Optimization of long-distance PCR using a transposon-based model system. PCR Methods Appl. **2:**51–59.
53. **Olsen, G. J., H. Matsuda, R. Hagstrom, and R. Overbeek.** 1994. fastDNAml: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. Comput. Appl. Biosci. **10:**41–48.
54. **Parton, A.** 1990. Nucleotide sequence of the HPV16 L1 open reading frame. Nucleic Acids Res. **18:**3631.
55. **Pirisi, L., S. Yasumoto, M. Feller, J. Doniger, and J. A. DiPaolo.** 1987. Transformation of human fibroblasts and keratinocytes with human papillomavirus 16 DNA. J. Virol. **61:**1061–1066.
56. **Pushko, P., T. Sasagawa, J. Cuzick, and L. Crawford.** 1994. Sequence variation in the capsid protein genes of human papillomavirus type 16. J. Gen. Virol. **75:**911–916.
57. **Roden, R. B. S., R. Kirnbauer, A. B. Jenson, D. R. Lowy, and J. T. Schiller.** 1994. Interaction of papillomaviruses with the cell surface. J. Virol. **68:**7260–7266.
58. **Roden, R. B. S., E. M. Weissinger, D. W. Henderson, F. Booy, R. Kirnbauer, J. F. Mushinski, D. R. Lowy, and J. T. Schiller.** 1994. Neutralization of

bovine papillomavirus by antibodies to L1 and L2 capsid proteins. J. Virol. **68:**7570–7574.
59. **Rose, R. C., R. C. Reichman, and W. Bonnez.** 1994. Human papillomavirus (HPV) type 11 recombinant virus-like particles induce the formation of neutralizing antibodies and detect HPV-specific antibodies in human sera. J. Gen. Virol. **75:**2075–2079.
60. **Sanger, F., S. Nicklen, and A. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
61. **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.
62. **Schiffman, M. H., H. M. Bauer, R. N. Hoover, A. G. Glass, D. M. Cadell, B. B. Rush, D. R. Scott, M. E. Sherman, R. J. Kurman, S. Wacholer, C. K. Stanton, and M. M. Manos.** 1993. Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. J. Natl. Cancer Inst. **85:**958–964.
63. **Schneider-Gadicke, A., and E. Schwarz.** 1986. Different human cervical carcinoma cell lines show similar transcription patterns of human papillomavirus type 18 early genes. EMBO J. **5:**2289–2292.
64. **Seedorf, K., G. Krämmer, M. Dürst, S. Suhai, and W. G. Röwenkamp.** 1985. Human papillomavirus type 16 DNA sequence. Virology **145:**181–185.
65. **Smits, H. L., K. F. Traanberg, M. R. L. Krul, P. R. Prussia, C. L. Kuiken, M. F. Jebbink, J. A. F. W. Kleyne, R. H. van den Berg, B. Capone, A. de Bruyn, and J. ter Schegget.** 1994. Identification of a unique group of human papillomavirus type 16 sequence variants among clinical isolates from Barbados. J. Gen. Virol. **75:**2457–2462.
66. **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.
67. **Stacey, S. N., C. Eklund, D. Jordan, N. K. Smith, P. L. Stern, J. Dillner, and J. R. Arrand.** 1994. Scanning the structure and antigenicity of HPV 16 E6 and E7 oncoproteins using antipeptide antibodies. Oncogene **9:**635–645.
68. **Stewart, A.-C. M., P. E. Gravitt, S. Cheng, and C. M. Wheeler.** 1995. Generation of entire human papillomavirus genomes by long PCR and frequency of errors produced during amplification. Genome Res. **5:**79–88.
69. **Swofford, D. L.** 1993. Phylogenetic analysis using parsimony: PAUP, version 3.1. Illinois Natural History Survey, Champaign, Ill.
70. **Viscidi, R. P., Y. Sun, B. Tsuzaki, F. X. Bosch, N. Muñoz, and K. V. Shah.** 1993. Serologic response in human papillomavirus-associated invasive cervical cancer. Int. J. Cancer **55:**780–784.
71. **Wheeler, C. M., C. A. Parmenter, W. C. Hunt, T. M. Becker, C. E. Greer, A. Hildesheim, and M. M. Manos.** 1993. Determinants of genital human papillomavirus infection among cytologically normal women attending the University of New Mexico Student Health Center. Sex. Transm. Dis. **20:**286–289.
72. **Wikström, A., C. Eklund, G. von Krogh, P. Lidbrink, and J. Dillner.** 1992. Levels of immunoglobulin G antibodies against defined epitopes of the L1 and L2 capsid proteins of human papillomavirus type 6 are elevated in men with a history of condylomata acuminata. J. Clin. Microbiol. **30:**1795–1800.
73. **Xi, S.-Z., and L. M. Banks.** 1991. Baculovirus expression of the human papillomavirus type 16 capsid proteins: detection of L1-L2 protein complexes. J. Gen. Virol. **72:**2981–2988.
74. **Yaegashi, N., S. A. Jenison, M. Batra, and D. A. Galloway.** 1992. Human antibodies recognize multiple distinct type-specific and cross-reactive regions of the minor capsid proteins of human papillomavirus types 6 and 11. J. Virol. **66:**2008–2019.
75. **Yaegashi, N., S. A. Jenison, J. M. Valentine, M. Dunn, L. B. Taichman, D. A. Baker, and D. A. Galloway.** 1991. Characterization of murine polyclonal antisera and monoclonal antibodies generated against intact and denatured human papillomavirus type 1 virions. J. Virol. **65:**1578–1583.
76. **Yaegashi, N., L. Xi, M. Batra, and D. A. Galloway.** 1993. Sequence and antigenic diversity in two immunodominant regions of the L2 protein of human papillomavirus types 6 and 16. J. Infect. Dis. **168:**743–747.
77. **Zhou, J., X. Y. Sun, D. J. Stenzel, and I. H. Frazer.** 1991. Expression of vaccinia recombinant HPV 16 L1 and L2 ORF proteins in epithelial cells is sufficient for assembly of HPV virion-like particles. Virology **185:**251–257.
78. **zur Hausen, H.** 1991. Viruses in human cancers. Science **254:**1167–1173.