JOURNAL OF VIROLOGY, June 2005, p. 7014–7023 Vol. 79, No. 11
0022-538X/05/$08.00+0 doi:10.1128/JVI.79.11.7014–7023.2005

Diversifying Selection in Human Papillomavirus Type 16 Lineages Based on Complete Genome Analyses

Zigui Chen,[1] Masanori Terai,[1]† Leiping Fu,[1] Rolando Herrero,[3] Rob DeSalle,[4] and Robert D. Burk[1,2]*

Department of Microbiology & Immunology[1] and Departments of Pediatrics, Epidemiology & Population Health, and Obstetrics, Gynecology and Women's Health,[2] Albert Einstein Cancer Center, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461; Proyecto Epidemiológico Guanacaste, Costa Rican Foundation for Health Sciences, San José, Costa Rica[3]; and Division of Invertebrate Zoology, American Museum of Natural History, New York, New York 10024[4]

Received 18 October 2004/Accepted 27 January 2005

Human papillomavirus type 16 (HPV16) is the primary etiological agent of cervical cancer, the second most common cancer in women worldwide. Complete genomes of 12 isolates representing the major lineages of HPV16 were cloned and sequenced from cervicovaginal cells. The sequence variations within the open reading frames (ORFs) and noncoding regions were identified and compared with the HPV16R reference sequence (50). This whole-genome approach gives us unprecedented precision in detailing sequence-level changes that are under selection on a whole-viral-genome scale. Of 7,908 base pair nucleotide positions, 313 (4.0%) were variable. Within the 2,452 amino acids (aa) comprising 8 ORFs, 243 (9.9%) amino acid positions were variable. In order to investigate the molecular evolution of HPV16 variants, maximum likelihood models of codon substitution were used to identify lineages and amino acid sites under selective pressure. Five codon sites in the E5 (aa 48, 65) and E6 (aa 10, 14, 83) ORFs were demonstrated to be under diversifying selective pressure. The E5 ORF had the overall highest nonsynonymous/synonymous substitution rate (ω) ratio (M3 = 0.7965). The E2 gene had the next-highest ω ratio (M3 = 0.5611); however, no specific codons were under positive selection. These data indicate that the E6 and E5 ORFs are evolving under positive Darwinian selection and have done so in a relatively short time period. Whether response to selective pressure upon the E5 and E6 ORFs contributes to the biological success of HPV16, its specific biological niche, and/or its oncogenic potential remains to be established.

Papillomaviruses are a heterogeneous group of DNA viruses with closed circular double-stranded DNA genomes of about 8 kb in size that contain three general regions. An upstream regulatory region (URR) contains sequences that control transcription and replication, an early region contains genes (e.g., E6, E7, E1, E2, E4, and E5) involved primarily in enzymatic activities, and a structural region produces the L1 capsid protein and L2, which facilitates packaging of the viral DNA. Human papillomaviruses (HPVs) are classified by the sequence similarity of their genomes. A cloned HPV genome whose L1 open reading frame (ORF) displays less than 90% similarity to previously designated types is defined as a novel type. To date, more than 90 different genotypes of the HPV have been fully characterized (21). Intratypic variants and subtypes are defined as HPVs that vary by less than 10% in their L1 DNA sequences (5, 21).

HPVs are causally involved in the etiology of cervical cancer and its precursor lesions (16, 17, 43, 49). Of the high-risk HPV types associated with cervical cancer, HPV16 is the most prevalent and is found in approximately half of all cancers (7, 49). Numerous variants of HPV16 have been identified in different geographic locations and ethnic groups (35, 42, 53, 68). Although all HPV16 isolates are closely related, previous studies inferred five distinct phylogenetic branches among HPV16 variants: European (E), Asian (As), Asian-American (As-Am), African-1 (Af-1), and African-2 (Af-2), corresponding to the geographic locations from which the samples were obtained (12, 34). Subsequent studies by sequence analyses of the HPV16 variants in other genomic regions (e.g., E6, L2, and L1) expanded and complemented this phylogenetic hypothesis (69).

Although HPV16 variants are an important focus of phylogenetic studies and the molecular variants of E2, L2, L1, the URR, and especially the E6 region have been described in detail previously (28), covariation among different ORFs belonging to the same lineage or isolate have not been studied in great detail. HPV16 variants have demonstrable differences in biological properties in vitro which may be responsible, in part, for differences in pathogenicity, carcinogenic risk, and perhaps immunogenicity (28). Furthermore, HPV16 variants are associated with different cervical cancer risks (6, 32, 65, 67). Although the diversifying selection in the HPV16 E6 and E7 oncogenes has been described recently (20), the evolutionary basis of the entire genome, coevolutionary mechanisms among different HPV16 genes, and their underlying biological significance remain unknown. Obtaining whole-genome sequences representative of the major HPV16 variants allowed us to determine with certainty nucleotide and amino acid sequence changes that are of potential evolutionary importance.

Comparison of synonymous (silent; d_S) and nonsynonymous

* Corresponding author. Mailing address: Albert Einstein Comprehensive Cancer Center, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461. Phone: (718) 430-3720. Fax: (718) 430-8975. E-mail: burk@aecom.yu.edu.

† Present address: National Research Institute for Child Health & Development, 3-35-31 Taishidoh, Setagaya, Tokyo 154-8567, Japan.

TABLE 1. Clinical information on specimens

Sample	Age of patient (yr)	HPV16 class	Cervix pathology[a]	Human immunodeficiency virus infection status[b]	Study location or name	Length (bp)	GenBank accession no.
CR-1	37	As-Am	Invasive cancer	NT	Costa Rica	7,906	AF402678
CR-2	35	As-Am	ASCUS	NT	Costa Rica	7,906	AY686579
CR-3	47	As-Am	CIN grade 2	NT	Costa Rica	7,907	AY686582
CR-4	28	E	CIN grade 2	NT	Costa Rica	7,906	AY686580
CR-5	32	E	CIN grade 1	NT	Costa Rica	7,906	AY686581
CR-6	22	E	ASCUS	NT	Costa Rica	7,907	AY686583
CR-7	36	E	Normal	NT	Costa Rica	7,906	AY686584
HAPI-1	19	Af-1	CIN grade 1	NT	HAPI	7,908	AF472508
HAPI-2	27	Af-2	Normal	NT	HAPI	7,904	AF472509
WIHS-1	30	E (G131)	ASCUS	+	WIHS	7,904	AF536179
WIHS-2	24	Af-1	N/A	+	WIHS	7,906	AF536180
WIHS-3	45	E (As)	Normal	+	WIHS	7,905	AF534061

[a] ASCUS, atypical squamous cells of undetermined significance; N/A, not available.
[b] NT, not tested.

(amino acid-changing; d_N) substitution rates in protein-coding genes provides an important means for investigating the forces of molecular evolution (72). The nonsynonymous/synonymous rate ratio ($\omega = d_N/d_S$) measures selective pressure at the protein level. Nonsynonymous rates that are significantly higher than synonymous rates are accepted as evidence for molecular adaptation (71, 72). This criterion has been used to identify a number of genes under adaptive molecular evolution (45).

In this report, the complete nucleotide sequences of and codon variations within HPV16 genomes representing the five major lineages were determined. By examining the ratio of d_N to d_S substitutions per site, diversifying selection acting on all of the eight protein-coding regions was evaluated. In addition, complete genome sequences for these variants allow us to reconstruct their genealogical relationships with unprecedented precision.

MATERIALS AND METHODS

Clinical specimens and sequencing. Cervicovaginal samples were obtained from women participating in a number of different studies involving HPV genotyping in our laboratory, including a population-based study in Costa Rica (30), the Women's Interagency HIV Study (WIHS) (52), and a study of the natural history of HPV in young women (HAPI) (9, 33). Samples found to contain HPV16 DNA by MY09/MY11 PCR and dot blot analyses were further subclassified into HPV16 variant lineages by sequencing the URR and/or the E6 region from PCR products (32).

Sequence analysis of the URR revealed that HPV16 genomes were distributed into five previously defined lineages of HPV16 (69): As-Am, Af-1, Af-2, E, and As. Clinical information for each sample, together with the GenBank accession numbers of the HPV16 DNA sequences, are listed in Table 1. HPV16 whole genomes were amplified by overlapping PCR (58, 59). Two sets of primers for nested PCR were designed to amplify the entire genome in two overlapping fragments of 4,108 bp and 3,863 bp. Oligonucleotide primer sequences used in these studies are available from the authors. For overlapping PCR, an equal mixture of AmpliTaq Gold DNA polymerase (Perkin-Elmer Applied Biosystems, Foster City, CA) and Platinum *Taq* High Fidelity DNA polymerase (Invitrogen, Rockville, MD) were utilized as previously described (58, 59). The estimated rate of error using high-fidelity *Taq* polymerase is approximately 1×10^{-6} errors per bp (18, 47). PCR products were separated by electrophoresis in agarose gels, stained with ethidium bromide, and visualized under UV illumination. After confirmation of the appropriate product size, each PCR product was purified (QIAGEN gel extraction kit; QIAGEN, Valencia, CA), ligated into the pGEM-T Easy vector (Promega, Madison, WI), and sequenced on an ABI 3700 sequencer in the Einstein Sequencing Facility. Several additional primers were used to obtain supplemental sequences to clarify sequence ambiguities. Isolates within the same lineages were included to evaluate intralineage variation. HPV16 nucleotide and amino acid positions were numbered according to the

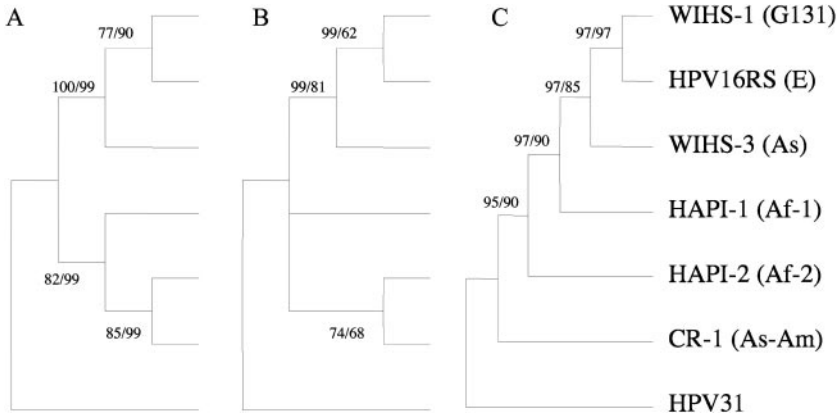


FIG. 1. Phylogenetic trees inferred from the complete genomes of representative HPV16 variants. Trees were generated using multiple algorithms. (A) Tree created using Bayesian construction and MP based on the nucleotide sequences of the eight ORFs and the URR. (B) Tree created using Bayesian construction and ML based on the nucleotide sequences of the eight ORFs and the URR. (C) Tree created using Bayesian construction and MP based on the amino acid sequences of the eight ORFs. The parsimony tree based on the concatenated amino acid and nucleotide sequences was similar to the tree shown in panel C (data not shown). The numbers on or near branches are support indices for the algorithms described for each tree.

TABLE 2. Nucleotide sequence variations between the HPV16R European variant RS and each variant genome[a]

HPV16 class (sample)	E6 ORF	E7 ORF	E1 ORF	E2 ORF
			1111111111111111111111111111222222222222222222222222222	2222222333333333333333333333333333333333333333333333
	1**1111**1122334445	667778	8990000112233444455556678888900222222222333333345566677	7889999000111111122222223333444444555555556666777788
	0**3344**77883**5**0353	143894	9260049690967127812892644568444223445689133445754903526	6263489014145668880222344678811123401123467468907880 0
	9**1235**68695**0**3372	872956	8102616370567663652794842624017077194275047345671681070	0808250803849011235469549274706851906748865544468472 5
E (RS)	TAGCGGTTACTAGTA	AATTTT	GTTATACGATTTCCCATGTTCCACAAAAACAGTCAGTCCTTAGTCTTCCTATGTA	AGCAGTTTACTACACAGCTTATTAGACTTCGGAGGACTTAGTTCTCTTGTCAT
E (CR-5)	....	......	...................C...................................	..........A.......A............C..........G.......... # #
E (CR-7)	.............C.	......	....G. #	....

E (CR-4)	..........G....
	          #
		......	.................A.....................................	............................T........................
				                            #

E (CR-6)	.....A....G....
	     # #
		......	..................A....................................
			                  #
				.............................T........................
				                            #

E (G131) (WIHS-1)	.G........G....
	 # #
		......	..........C.......A........G...................T.......
			          # # #
				...G.CC...C................C.T...............A.......
				     # # # #

E (As) (WIHS-3)	......G........ #	.G...C #	A................A......G...G...C......CC.A............ # # # # #	.A.....C....A...........A..C.T....AG..C.A....A....A.. # # # # # # ## # # #

Af-1 (WIHS-2)	..CGT..AGT.....
	  ### #
		...CG.	.......G.G..AT...C...TT.TG....T.C..G......G..T....C.A..G
			       # # # #
				G.A.A......GA.TCAT...CC.AGG..T..C.A.A....G..CAA.TGAG.
				# # # ## #### ### # # # # # # # #

Af-1 (HAPI-1)	..CGT..AGT.....
	  ### #
		G..CG.	...G..G....AT...C...TT..G....T.C.........G..T....C.A...
			      # # # #
				..A.........A.T.A.....C.AGG..T..T.A.A....G.T.AA.TGA..
				# # # ## #### ### # # # #

Af-2 (HAPI-2)	C.TGT..AGT.G...
	  ### #
		.G.CG.
		 #
			.C....GA.C.ATTGGC....T.AG....T.C.G.ACTT.....TC..TCCA...
			      ## # # # # #
				..A......T..A.T.A.......AGG..T...AA.AC.C.G...AACT.A.G
				  # # # # ### # ## ## # # # #

As-Am (CR-1)	....T..AGTG.A.G
	    # ##
		..CCG.	.CC..GGA.C.AT.G.C.AC.TGAGC...TGC.G.ACTT....CT....CCAA..
			      ## # # # ### ## # #
				..A.........AG.CA..A....AGG.CT....A.AC.C.G..CAACT.A.G.
				  # ## ## # ### ## # ## # # # #

As-Am (CR-2)	....T..AGTG...G
	    # ##
		..CCG.	.C..CGGA.C.AT.G.C.A..TGAG.G..T.C.G.ACTT....CT.C..CCAA..
			      ## # # # ## # # # # #
				..A.........A.T.A..AG..CAGG..TA...A.AC.C.G...AACT.A.G
				  # # # # ## ### ## # ## # # # #

As-Am (CR-3)	....T..AGTG....
	    # ##
		..CCG.	.C..CGGA.C.AT.G.C.A..TGAG.G..T.C.G.ACTT....CT.C..CCAA..
			      ## # # # ## # # # # #
				..A.....T...A.T.A..AG..CAGG..TA...A.AC.C.G...AACT.A.G
				  # # # # ## ### ## # ## # # # #

[a] #, nucleotide change resulting in amino acid change (nonsynonymous); ., site unchanged compared with HPV16R sequence; -, deletions/insertions of nucleotide sequences. Sequence positions in bold within HPV16 E6 and E5 are under diversifying selection (E6 nt 131/132, 143/145, and 350; E5 nt 3991 and 4042/4043) (see Table 6).

HPV16R European reference sequence (RS) described and annotated in the HPV Sequence Database (50).

Phylogenetic analyses and tree construction. The amino acid sequence of each ORF was aligned using CLUSTAL_X software (61). Codon Align (version 1.0) was implemented to align the nucleotide sequences of the coding regions corresponding to the aligned proteins. Phylogenetic trees were constructed based on the concatenated amino acid and/or nucleotide sequences from the eight complete HPV16 ORFs (i.e., E6, E7, E1, E2, E4, E5, L2, and L1) and the URR of six representative variants corresponding to the main HPV16 lineages (including the HPV16R reference genome, RS).

The computer program MRBAYES v3.0b4 (38) was used for Bayesian tree construction, with 100,000 cycles for the Markov chain Monte Carlo algorithm. The gamma model was set for among-site rate variation and allowed all substitution rates of aligned sequences to be different. Maximum parsimony (MP) and maximum likelihood (ML) trees were calculated by a heuristic search with PAUP* v4.0b10 (57). For maximum parsimony analysis, amino acid and nucleotide sequence data were reduced from 2,475 to 52 and from 7,425 to 124 phylogenetically informative sites, respectively. For maximum likelihood analysis, the computer program MODELTEST v3.06 (54) was used to identify the best evolutionary model; the phylogenetic model HKY+G was selected to provide the likelihood parameters, with a transition-transversion ratio of 1.7220 and gamma distribution shape 0.2316. Both parsimony and likelihood data were bootstrap resampled 1,000 times. HPV31 was set as the out-group taxon (29).

Positive selection estimation. The nonsynonymous/synonymous rate ratio ($\omega = d_N / d_S$) is an indicator of natural selection, with ω values of 1 representing neutral variation, ω values of <1 representing purifying selection, and ω values of >1 representing diversifying positive selection. Different amino acid sites in a protein are expected to be under different selective pressures and have different underlying ω ratios (72, 73). Six codon substitution models were used to investigate whether positive selection could be identified within the eight ORFs of HPV16 as follows: M0 (one-ratio), M1 (neutral), M2 (selection), M3 (discrete), M7 (beta), and M8 (beta and ω) (72). These models view the codon as the fundamental unit of evolutionary change and take into account genealogic history when calculating scores. Log likelihood scores evaluate the quality of the fit of the input data to the conditions of the model. In these models, ω was estimated for separate classes of codons that are assumed to have evolved independently of one another.

The six models used for the ω distribution were implemented in the CODEML program in the PAML package (70, 72). Maximum parsimony within PAUP* v4.0b10 (57) was used for tree reconstruction for each HPV16 ORF. M0 assumes the same ω ratio for all sites. M1 assumes two classes of sites in the protein characterized by the formulas $\omega_0 = 0$ (conserved sites) and $\omega_1 = 1$ (neutral sites). M2 adds a third class of sites with ω as a free parameter, thus allowing for sites with ω values of >1. M3 uses a general discrete distribution with three site classes, with the proportions ($p_0, p_1, \ldots, p_{K-1}$) and the ω ratios ($\omega_0, \omega_1, \ldots, \omega_{K-1}$) estimated from the data. M7 assumes a beta distribution $B(p, q)$, which, depending on parameters p and q, can take various shapes in the interval (0, 1). M8 adds an extra class of sites to the beta (M7) model with the proportion and the ω ratio estimated from the data, thus allowing for sites with ω values of >1 (72).

In order to assess the influence of positive selection on a particular coding region, a likelihood ratio test (LRT) was used to compare nested models (51, 72). Twice the log likelihood difference between two models follows a χ^2 distribution with 2 degrees of freedom, which is equal to the difference in the number of parameters estimated between the models. From these models, three LRTs were constructed, which compared M1 with M2, M1 with M3, and M7 with M8. When alternative models (M2, M3, and M8) all detect the same sites with ω values of >1, all three tests taken together can be considered strong evidence of positive selection.

RESULTS

The assembled sequences of the HPV16 variant genomes revealed total sizes of 7,908 bp for HAPI-1 (Af-1); 7,907 bp for CR-3 (As-Am) and CR-6 (E); 7,906 bp for CR-1 and CR-2 (As-Am), CR-4, CR-5 and CR-7 (E), and WIHS-2 (Af-1); 7,905 bp for WIHS-3 (E, As); and 7,904 bp for HAPI-2 (Af-2) and WIHS-1 (E, G131). All variant genomes displayed the same intact ORF distributions identified in the HPV16R RS (50).

To reconstruct the origin of present-day HPV16 variants, we constructed phylogenetic trees using multiple algorithms that resulted in three different topologies. The MP trees inferred from the nucleotide sequences of the eight ORFs and URR with MRBAYES and PAUP* both demonstrated two major clades for HPV16 variants (Fig. 1A), consistent with previous analysis (69). The grouping of the European taxa (E, As, and G131) was consistent in the trees generated by all algorithms. However, a different topology was noted in both ML trees based on the nucleotide sequences of the eight ORFs and URR (Fig. 1B) and the MP tree based on the amino acid sequences of the eight ORFs (Fig. 1C) using MRBAYES and PAUP* algorithms, respectively. These data indicate that the non-European taxa (Af-1, Af-2, and As-Am) are a paraphyletic group (i.e., they lack features common to all members of the European taxa). The MP trees based on the concatenated amino acid and nucleotide sequences showed the same topology as that obtained with the amino acid sequences alone (data

TABLE 2—*Continued*

E4 ORF	E5 ORF		L2 ORF
333333333333333333	333334444444444	44444444444 444444444444444 4	4444444444444444444444444444455555555555555555555555555555
333344444455555555	889990000000000	11111111111 111111111222222 2	2234444445555556666788999999000112222222222333333344444555
678811123401123467	566790123444578	13445556677 899999999000001 2	3801125681456890144258134566344341235889991366778804789016
274706851906748865	887918754239979	48591784906 413456789024694 8	4180982158570690414547184039415920669670450789696938575684
ACTTCGGAGGACTTAGTT	TGAACAGGTATTAAT	TTCACATCTTC-----ATTTTGTTTGTGTTT-T	ATGAAGTGAATTACATTTTCCAAGAAGAATGTGAAGATAACATTTCCGTGTTTCTGAC
......C..........G	....	........A..-----...G...........-C	....C..................A..........C........G..............
#			# #
....	....	..........-----.............G.-C	C......................A..........C.......................
			#
....T.............	...C.....G.....	A..........-----T..............-C	.......................A..A.......T.......................
	# #		# #
....T.............	...C.....G.....	..........-----...............TC	.......................A.......A..T.......................
	# #		#
..C.T.............	...C....G..G...	..........T-----..--G..........-C	................C......A.....C....G.....T.................
	# # #		# #
..C.T....AG..C.A..	...C....GGG..T.	..........T-----.G.-......G....-C	............C......T...A.....CA..GG..........TT.........C.
	# ###		# ## # #
GG..T..C.A.A....G.	CA.CT....T....C	.CT..GAT..T---TAG..A.--.......C-.	.CA..A.A.......C..AT..TA.....C..A..A...G..C....A.AC...CA.G
# #	# # #		# # # # # ## #
GG..T..T.A.A....G.	CA.CTG.A.T....C	..T..GAT..T---TAG....AG........-.	.CA..A.A.....T.C.CAT..TA.....C..A..A...G..C....A.AC...CA.G
# #	# # # #		# # # # # # ## #
GG..T...AA.AC.C.G.	C..CG....G..C.C	...C.GA.G.T---TAG.C..A----.....-.	.C...T...GGG...C..ATTG.A.....G..A..AG..C..C..T.A.AC..TCA.G
## # #	## #		# # # # # # ## #
GG.CT....A.AC.C.G.	C..CG.A..G....C	....AGAT.-TTTTTAG.C..A----..G..-.	.C...TCAG.....CC..AT.G.AGG.GTC..A..A.A...CA.GT.AGAC.ATCA.G
## # #	## #		# ### # ## ## ## #
GG..TA...A.AC.C.G.	C.GCG.A..G....C	....AGAT..TTTTTAG.C..A----.CG.G-.	.C.G.TCA.......C..AT.G.AGG.GTC..A..A.AC...C.GT.AGACG.TCA.G
## # #	### #		# ### # # ## ## ## #
GG..TA...A.AC.C.G.	C.GCG.A..G....C	....AGAT..TTTTTAG.C..A----.CG..-.	.C.G.TCA.......C..AT.G.AGG.GTC..A..A.AC...C.GT.AGAC..TCA.G
## # #	### #		# ### # # ## ## ## #

Continued on following page

not shown). The position of the Af-1 genome was ambiguous and the most variable; however, upon close examination of the shared and disparate sequences, it has the characteristics of a true intermediate bridging the evolution of the European and Af-2–As-Am groups.

Sequence variation within HPV16 noncoding regions. Similarities in nucleotide changes in the URR (e.g., from the stop codon of L1 to the translation initiation codon of E6) have been used to assign HPV16 variants to specific lineages. Each of the twelve HPV isolates contained nucleotide base changes previously identified in a subfragment of this region (i.e., nucleotide [nt] 7485 to 7842) (36, 68, 69). Complete URR nucleotide changes distributing to the five HPV16 lineages were classified, and an additional 26 potentially class-specific variations were identified (Table 2). Although the URR was initially thought to be the most variable HPV genomic region, full genome analyses indicated that the E4 and E5 ORFs and the noncoding region between E5 and L2 were more variable than the URR (see Table 4).

The only insertions and deletions identified were in the noncoding region between E5 and L2 (Table 2).

Sequence variation within HPV16 coding regions. Nucleotide sequence variations in the eight putative ORFs and their corresponding amino acid sequences were compared with the HPV16R reference sequence (Tables 2 and 3). Measures of variability for each region/ORF of the HPV16 genome are shown in Table 4. The frequencies of nucleotide changes in HPV16 genes varied from 2.0% to 6.3%. However, the proportion of nonsynonymous amino acid variations was highest in E5 and E2, which were followed in that regard by L2, E4, E6, E1, L1, and E7. The absolute ratios of the number of nonsynonymous to synonymous changes were over 1.00 in the E2 and E5 ORFs. This ratio is different from the d_N/d_S ratio, which is based on the relative rate ratio values.

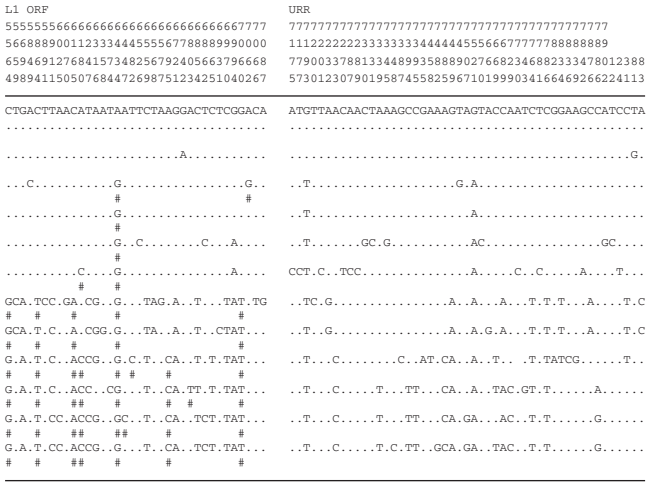
Fifty-five variable nucleotide positions (a total of 56 variations) were identified within the E1 ORF. Nucleotide and amino acid variations within the other coding regions were consistent with those of previously isolated HPV16 variants (10, 12, 24, 35, 53, 60, 62, 63, 68, 69). The Af-1 isolates (WIHS-2 and HAPI-1) have a polymorphism at nt 83 (A to C), changing an ATG codon present in all other isolates to CTG. Since the actual E6 translation initiation codon begins at nt 104 within the E6 mRNA from the p97 promoter, this variation should not affect the E6 protein.

Lineage fixation among different HPV16 regions. No evidence of genomic recombination was present in the representative HPV16 variant genomes. Since the genomes are evolving through nucleotide changes and not gross rearrangement or recombination, nucleotide changes in one region (e.g., E6) are highly correlated with and inseparable from changes in other regions (e.g., E1) within genomes from the same lineage. For instance, amino acid changes at E6 (78 [Y]), E1 (78 [E], 168 [S], and 452 [D]), E2 (35 [Q], 203 [D], 208 [A], 254 [N], 271 [V], and 341 [C]), and an additional 11 positions (shown in boldface in Table 3) all segregate together, representing ancestral changes in the non-European taxa. Similarly, amino acid changes at E6 (14 [Q] and 78 [H]), E1 (78 [Q], 168 [C], and 452 [E]), E2 (35 [H], 143 [A], 203 [N], 208 [P], 254 [T], 271 [F], and 341 [W]), and 12 other positions (Table 3, underlined in the RS) indicate ancestral changes in the European lineage. These changes have occurred as the result of lineage fixation akin to linkage disequilibrium in organisms with recombining genomes. Isolates classified by nucleotide variations in E6, the URR, and L1 would have the corresponding base changes in E1, E2, and L2. However, E4, E5, and E7 failed to supply sufficient information for variant classification, especially for the non-European isolates.

Mechanism of HPV16 variation: identification of purifying and diversifying selection. The nonsynonymous/synonymous rate ratio ($\omega = d_N/d_S$) was used to calculate whether positive selection has been a force in the evolution of HPV16 variants within each ORF. The likelihood analysis, including parameter estimates for different models, is shown in Table 5.

For each ORF, six models employing different assumptions about selection (ω) were calculated, and the model with the highest log likelihood value was used as the "best" model. In essentially all ORFs, the M3 (discrete) model was optimal. The d_N/d_S ratio is an average of all sites in an ORF. For instance, using M3 for ORF E6, the average d_N/d_S ratio is 0.41. The

TABLE 2—*Continued*



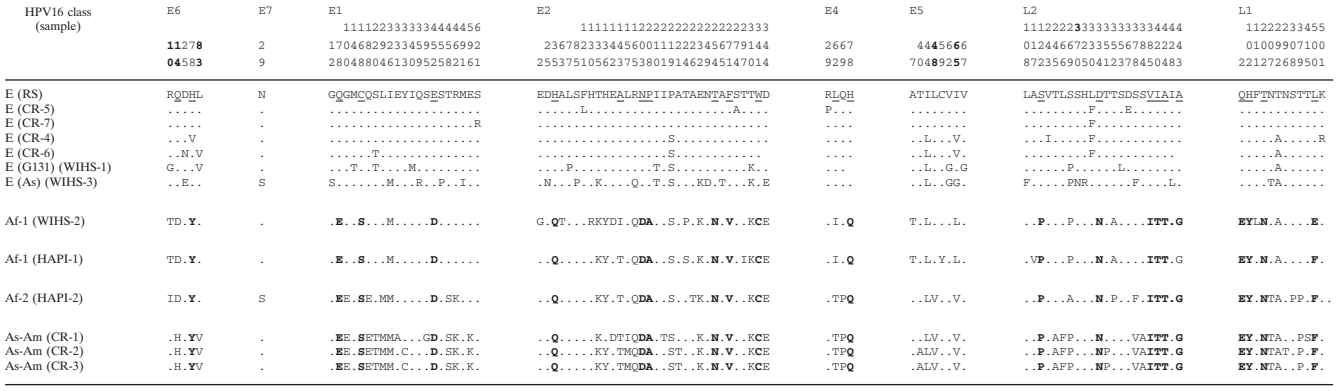
majority (94%) of sites are under purifying selection, with ω values of less than 1, but 5.6% of sites are under strong diversifying or Darwinian selection, with ω values of 7.3. These sites are HPV16 E6 aa 10, 14, and 83 (see Table 6). The E5 ORF had the highest average ω value(0.80), with about 5% of sites (E5 aa 48 and 65; Table 6) under strong diversifying selection with ω values of 10.7 by M3. To further test whether specific sites were evolving under positive selection, we used the LRT. Only the E6 and E5 ORFs consistently showed sites under positive selection (Table 5). Thus, although the E2 ORF had one of the higher average ω values (ω = 0.56), with about 24% of sites with a d_N/d_S value over 1 (ω value of 2.32 by M3), no specific sites could be confirmed to be under strong positive selection by the LRT. However, the E4 ORF overlaps with the E2 hinge region that bridges the E2 DNA binding and activation domains. This results in an increased number of nonsynonymous changes in the hinge region, elevating the overall d_N/d_S value of the E2 (data not shown). The remaining HPV16 ORFs seem to be highly conserved under purifying selection pressure, despite the presence of nucleotide and amino acid variation. Table 6 is a summary of the likelihood ratio tests, which are more sensitive for detecting selection at a single amino acid within an ORF than the other tests (72).

DISCUSSION

Phylogenetic relationships among HPV16 variants. Phylogenetic trees from maximum parsimony based on the eight ORF and URR nucleotide sequences parsed the HPV16 variants into two clades, consistent with previous studies (69), and support a distinction between the European and non-European taxa. However, the use of a variety of phylogenetic techniques revealed ambiguity in the placement of the Af-1 lineage. Sequencing of a second Af-1 isolate did not resolve the ambiguity in tree topology either as a replacement or in combination. In addition, the simplicity of the data set allows visual inspection and "back-of-the-envelope" calculations that support an intermediate position for Af-1.

Lineage fixation of genetic polymorphisms within HPV16 genomes. A number of studies have demonstrated an association between HPV16 variants and cervical high-grade disease/cancer (6, 32). Some reports have suggested that for HPV16 variants, both numerical and specific nucleotide alterations relative to the HPV16R reference sequence may affect biological behavior (65, 67, 76). Some epidemiological studies have also shown an association between high-grade lesions and cancer and an HPV16 nucleotide change within E6 at nt 350 (T to G) (1, 75), whereas another study suggested that the latter phenomenon might be population dependent (77). If causal, these associations may be explained by differences (i) in the transcriptional regulation of the virus by different variants, (ii) in the biological activities of the protein encoded by variants (e.g., enhanced transformation abilities of E6/E7), or (iii) in the abilities of the hosts to respond immunologically to specific viral epitopes encoded by variants (15, 31, 64). Alternatively, they may be markers of other linked changes in the variant genome. However, to unequivocally evaluate causal genetic effects, whole-genome analyses are needed. The present study analyzed the full genetic variability of HPV16 isolates representing the five major classes of variants. Most previously re-

TABLE 3. Amino acid sequence variations between the HPV16R European variant RS and each variant genome[a]



[a] Covariations among non-European lineage are listed in boldface, while those among European lineages are underlined in the E (RS) sequence. Sequence positions in bold within HPV16 E6 and E5 are sites under diversifying selection (E6 aa 10, 14, and 83; E5 aa 48 and 65) (see Table 6).

TABLE 4. Comparison of nucleotide and amino acid sequence variability within HPV16 genomes

ORF or region	No. of nucleotide sequences	No. of variable nucleotide positions (%)[a]	Total no. of nucleotide variations (%)[b]	No. of amino acids	No. of variable codons (%)	No. of variable aa positions (%)[c]	No. of synonymous changes (%)[d]	Nonsynonymous/ synonymous change ratio[e]
E6[f]	456	15 (3.29)	16 (3.51)	151	12 (7.95)	5 (3.31)	7 (4.64)	0.71
E7	297	6 (2.02)	6 (2.02)	98	6 (6.12)	1 (1.02)	5 (5.10)	0.20
E1	1,950	55 (2.82)	56 (2.87)	649	55 (8.47)	21 (3.24)	34 (5.24)	0.62
E2	1,098	53 (4.83)	55 (5.01)	365	47 (12.88)	32 (8.77)	15 (4.11)	2.13
E4	288	18 (6.25)	19 (6.60)	95	17 (17.89)	4 (4.21)	13 (13.68)	0.31
E5	252	15 (5.95)	17 (6.75)	83	14 (16.87)	8 (9.64)	6 (7.23)	1.33
NCR	142	29 (20.42)	31 (21.83)					
L2	1,422	58 (4.08)	63 (4.43)	473	56 (11.84)	22 (4.65)	34 (7.19)	0.65
L1	1,596	36 (2.26)	37 (2.32)	531	36 (6.78)	12 (2.26)	24 (4.52)	0.50
URR	833	48 (5.76)	49 (5.88)					
Aggregate[g]	7,908	313 (3.96)	328 (4.15)	2,452	243 (9.91)	105 (4.29)	138 (5.63)	0.76

[a] Number and percentage of positions showing nucleotide variations among 13 HPV16 complete genomes.
[b] Number of nucleotide variations, including multiple changes per position.
[c] Number and percentage of changed amino acid positions per total number of amino acids/codon in each ORF (nonsynonymous changes).
[d] Number and percentage of synonymous/silent changes per total number of amino acids/codon in each ORF.
[e] Per codon.
[f] The first ATG start codon of the HPV16 E6 ORF is predicted to be at nt 104.
[g] Each nucleotide position is counted only once.

ported variants were also detected in our sequenced isolates. This is the first report to analyze taxa within an HPV type in the context of the complete genome.

There was no evidence for recombination between HPV16 variants. A deep branch was determined between non-European and European variants, which can be inferred from linked amino acid variations throughout the HPV16 genome; 21 such positions distributing to six HPV16 ORFs represent a primordial change in the non-European taxa ancestor (in boldface in Table 3). Previous studies suggested that non-European isolates might confer a higher risk of developing high-grade cervical intraepithelial neoplasia (CIN)/cancer compared to so-called prototype-like (European) variants based on a limited set of nucleotide alterations (28). However, the array of common variants likely contains a complex set of biological changes related to a higher risk of cervical cancer. Further immunological and biochemical analyses focusing on these variable sites are needed.

Diversifying selection within HPV16 oncogenes. In this study, the nonsynonymous/synonymous rate ratio ($\omega = d_N/d_S$) was used to determine whether positive selection impacts the evolution of HPV16 variants within different ORFs. The average d_N/d_S ratio of each HPV16 ORF was less than 1, indicating that all HPV16 ORFs are under purifying selection pressure. However, in agreement with a previous report (20), the E6 oncogene contains three codon sites that are evolving under the influence of diversifying selection (HPV16 E6 aa 10, 14, 83). HPV16 E6 aa 27 was also suggested to be under selective pressure (20), but this site was not identified in the present study. In addition, we demonstrate that E5 contains two sites under diversifying selection (HPV16 E5 aa 48 and 65), which suggests that HPV16 E6 and E5 have recently evolved through positive selection pressure.

Immune selection is likely to be a force driving the evolution of HPV16 genomes. The main immune response to HPV infection is cell mediated. The presentation of viral peptides to T cells in the context of human leukocyte antigen (HLA) class I and II molecules is influenced by genetic polymorphisms of both HPV and HLA. These variations in the immune response to different viral isolates are influenced by the heterogeneity of the host immune response and influence the outcome of HPV infection. Recent suggested that the common E6 variant, L83V, is predominately associated with three distinct HLA class I alleles, namely, B*44, B*51, and B*57, for the E6 epitope stretching from aa 74 to 83 (SEYRHYCYSL/V) (1, 11, 75). The immunologic relevance of the HPV16 E6 N-terminal region and variant positions E6 aa 10 and 14 is supported by the demonstration of an endogenously processed HLA A*0201-restricted E6 peptide (KLPQLCTEL; E6 aa 11 to 19) as well as of an overlapping HLA B-7-restricted E6 peptide (RPRKLPQL; E6 aa 8 to 15) in this region (3, 23). In another study, the HPV16 E6 variant R10G was demonstrated to alter a B*07 binding epitope such that it may influence immune recognition by cytotoxic T lymphocytes (23). Certain HLA class I alleles, in concert with specific HPV variants, could be associated with a predisposition for cervical cancer development, whereas others may be protective. The variable positions predicted to be under Darwinian selection (HPV16 E6 aa 10, 14, and 83) might thus be under immune selection.

Additional functional characteristics of the sites under selection may be related to modification of E6 protein-protein interaction domains. A substantial proportion of the changes found in HPV16 E6 were located in the amino half of the protein, which represents the binding region of the cellular protein E6-AP and has a role in both cell-mediated and humoral host immune responses (26, 40, 41). Moreover, a recent report indicates that variations in codons 14 and 27 substantially altered the E6 antigenic index and may affect interactions with E6-AP (66). This provides biochemical support for selection at HPV16 E6 aa 14. In addition, the cellular tumor suppressor protein p53, usually targeted and degraded by HPV E6, is polymorphic in human populations and varies according to latitude and ethnicity (4). A p53 polymorphism at codon 72, exon 4, resulting in either a proline (Pro) or an arginine (Arg), has been identified as a possible risk factor for the development of cervical carcinoma (55). A few studies have suggested

TABLE 5. Phylogenetic analysis by ML estimation for HPV16 genes

ORF	Model	Log likelihood	d_N/d_S[a]	Parameters
E6	M0 (one-ratio)	−750.939522	0.4125	$\omega = 0.4125$
	M1 (neutral)	−744.238187	0.1477	$p_0 = 0.85229\ p_1 = 0.14771$
	M2 (selection)	−739.714214	0.4073	$p_0 = 0.94403\ p_1 = 0.00000\ p_2 = 0.05597\ \omega_2 = 7.27647$
	M3 (discrete)	−739.714190	0.4073	$p_0 = 0.12144\ p_1 = 0.82259\ p_2 = 0.05597\ \omega_0 = 0.00001\ \omega_1 = 0.00001\ \omega_2 = 7.27677$
	M7 (beta)	−744.540127	0.1000	$p = 0.00195\ q = 0.02108$
	M8 (beta and ω)	−739.714078	**0.4073**	$p = 0.00100\ q = 1.54547\ p_0 = 0.94403\ (p_1 = 0.05597)\ \omega = 7.27644$
E7	M0 (one-ratio)	−442.311583	0.1736	$\omega = 0.1736$
	M1 (neutral)	−440.244427	0.0763	$p_0 = 0.92366\ p_1 = 0.07634$
	M2 (selection)	−438.921563	**0.1577**	$p_0 = 0.98457\ p_1 = 0.0000\ p_2 = 0.01543\ \omega_2 = 10.21907$
	M3 (discrete)	−438.921757	0.1577	$p_0 = 0.01594\ p_1 = 0.96863\ p_2 = 0.01543\ \omega_0 = 0.00001\ \omega_1 = 0.00001\ \omega_2 = 10.21925$
	M7 (beta)	−440.277261	0.1000	$p = 0.00165\ q = 0.02011$
	M8 (beta and ω)	−438.921563	0.1577	$p = 0.00100\ q = 1.47425\ p_0 = 0.98457\ (p_1 = 0.01543)\ \psi = 10.21900$
E1	M0 (one-ratio)	−2937.078929	0.1538	$\omega = 0.1538$
	M1 (neutral)	−2935.147538	0.1613	$p_0 = 0.83868\ p_1 = 0.16132$
	M2 (selection)	−2935.131457	0.1577	$p_0 = 0.82138\ p_1 = 0.00000\ p_2 = 0.17862\ \omega_2 = 0.88286$
	M3 (discrete)	−2935.130917	**0.1577**	$p_0 = 0.49174\ p_1 = 0.32967\ p_2 = 0.17860\ \omega_0 = 0.00001\ \omega_1 = 0.00001\ \omega_2 = 0.88293$
	M7 (beta)	−2935.139255	0.1573	$p = 0.01252\ q = 0.06910$
	M8 (beta and ω)	−2935.131324	0.1577	$p = 0.00294\ q = 1.19859\ p_0 = 0.82138\ (p_1 = 0.17862)\ \omega = 0.88284$
E2	M0 (one-ratio)	−1876.140617	0.5218	$\omega = 0.5218$
	M1 (neutral)	−1872.769407	0.4232	$p_0 = 0.57684\ p_1 = 0.42316$
	M2 (selection)	−1871.302950	**0.5611**	$p_0 = 0.75842\ p_1 = 0.000\ p_2 = 0.24158\ \omega_2 = 2.32256$
	M3 (discrete)	−1871.303083	0.5611	$p_0 = 0.56211\ p_1 = 0.19631\ p_2 = 0.24158\ \omega_0 = 0.00001\ \omega_1 = 0.00001\ \omega_2 = 2.32256$
	M7 (beta)	−1873.007264	0.5000	$p = 0.00450\ q = 0.00441$
	M8 (beta and ω)	−1871.302950	0.5611	$p = 0.00100\ q = 1.68952\ p_0 = 0.75842\ (p_2 = 0.24158)\ \omega = 2.32255$
E4	M0 (one-ratio)	−499.426327	0.1601	$\omega = 0.1601$
	M1 (neutral)	−498.066176	0.1510	$p_0 = 0.84896\ p_1 = 0.15104$
	M2 (selection)	−498.010196	**0.1658**	$p_0 = 0.88391\ p_1 = 0.0000\ p_2 = 0.11609\ \omega_2 = 1.42792$
	M3 (discrete)	−498.010222	0.1658	$p_0 = 0.28572\ p_1 = 0.59821\ p_2 = 0.11607\ \omega_0 = 0.00001\ \omega_1 = 0.00001\ \omega_2 = 1.42799$
	M7 (beta)	−498.153076	0.1588	$p = 0.01166\ q = 0.06421$
	M8 (beta and ω)	−498.036494	0.1607	$p = 0.00213\ q = 0.01934\ p_0 = 0.95917\ (p_1 = 0.04083)\ \omega = 1.58742$
E5	M0 (one-ratio)	−435.001494	0.6705	$\omega = 0.6705$
	M1 (neutral)	−432.604998	0.3664	$p_0 = 0.63363\ p_1 = 0.36637$
	M2 (selection)	−429.272907	0.8071	$p_0 = 0.59008\ p_1 = 0.37250\ p_2 = 0.03743\ \omega_2 = 11.61198$
	M3 (discrete)	−429.254476	0.7965	$p_0 = 0.26115\ p_1 = 0.69348\ p_2 = 0.04537\ \omega_0 = 0.32659\ \omega_1 = 0.32660\ \omega_2 = 10.68596$
	M7 (beta)	−432.625891	0.4000	$p = 0.00201\ q = 0.00326$
	M8 (beta and ω)	−429.254456	**0.7966**	$p = 48.11157\ q = 99.00000\ p_0 = 0.95469\ (p_1 = 0.04531)\ \omega = 10.69102$
L2	M0 (one-ratio)	−2270.264783	0.1330	$\omega = 0.1330$
	M1 (neutral)	−2265.762403	0.1433	$p_0 = 0.85669\ p_1 = 0.14331$
	M2 (selection)	−2265.545031	0.1315	$p_0 = 0.00000\ p_1 = 0.10056\ p_2 = 0.89944\ \omega_2 = 0.03446$
	M3 (discrete)	−2265.081776	**0.1406**	$p_0 = 0.55727\ p_1 = 0.42184\ p_2 = 0.02089\ \omega_0 = 0.08231\ \omega_1 = 0.08232\ \omega_2 = 2.87503$
	M7 (beta)	−2265.639112	0.1330	$p = 0.03159\ q = 0.20124$
	M8 (beta and ω)	−2265.548829	0.1326	$p = 0.00667\ q = 0.01316\ p_0 = 0.33537\ (p_1 = 0.66463)\ \omega = 0.04454$
L1	M0 (one-ratio)	−2392.492442	0.1308	$\omega = 0.1308$
	M1 (neutral)	−2391.381730	0.1350	$p_0 = 0.86500\ p_1 = 0.13500$
	M2 (selection)	−2391.376118	**0.1330**	$p_0 = 0.85231\ p_1 = 0.00000\ p_2 = 0.14769\ \omega_2 = 0.90078$
	M3 (discrete)	−2391.376120	0.1330	$p_0 = 0.68619\ p_1 = 0.16614\ p_2 = 0.14766\ \omega_0 = 0.00001\ \omega_2 = 0.00001\ \omega_2 = 0.90089$
	M7 (beta)	−2391.392774	0.1317	$p = 0.01192\ q = 0.07147$
	M8 (beta and ω)	−2391.376158	0.1330	$p = 0.00186\ q = 2.45210\ p_0 = 0.85232\ (p_1 = 0.14768)\ \omega = 0.90083$

[a] The d_N/d_S ratios in boldface indicate the "best" models interpreted from the maximum log likelihood value.

that the combination of HPV16 E6 aa L83 and the p53 Arg/Arg genotype might represent an interactive risk factor for cervical cancer (8, 63). Therefore, selection pressure driven by protein-protein polymorphic interactions may facilitate the life cycle and pathogenicity of HPV-associated disease.

HPV16 E5 is a highly hydrophobic membrane-bound protein of 83 amino acids associated with the Golgi apparatus, endoplasmic reticulum, and nuclear membrane in infected cells. Recently, the E5 protein has been suspected of acting to alter levels of cell-cycle regulators and to play a role during the productive/early stage of the HPV16 life cycle (25, 27). Previous studies have indicated that amino acid changes within HPV16 E5 might alter the transforming activity of the protein by affecting the interactions with the epidermal growth factor receptor, the 16-kDa subunit of H^+-ATPase and, potentially, other cellular proteins (19, 56). An amino acid change, L48V, was noted between HPV16 E5 amino acids 46 and 50, a stretch of hydrophobic residues potentially representing a transmembrane domain. HPV16 E5 protein also stimulates the nuclear oncogenes c-jun, junB, and c-*fos*. E5 induction of c-jun is through an activator protein-1 binding site, and E5-activation of c-*fos* is via nuclear factor-1 (NF-1) binding sites. Therefore,

TABLE 6. Likelihood ratio tests for positive selection of amino acid sites

ORF	Model 1	Model 2	LRT statistic[a]	$d_N/d_S > 1$ detected?	Positively selected codons[b]
E6	M1	M2	**9.047946**	Yes	10 R 1.0000**
					14 Q 1.0000**
					83 L 1.0000**
	M1	M3	**9.047794**	Yes	10 R 1.0000**
					14 Q 1.0000**
					83 L 1.0000**
	M7	M8	**9.652098**	Yes	10 R 1.0000**
					14 Q 1.0000**
					83 L 1.0000**
E7	M1	M2	2.647728	Yes	None
	M1	M3	2.645340	Yes	None
	M7	M8	2.711396	Yes	None
E1	M1	M2	0.032162	No	None
	M1	M3	0.033242	No	None
	M7	M8	0.015862	No	None
E2	M1	M2	2.932914	Yes	None
	M1	M3	2.932648	Yes	None
	M7	M8	3.408628	Yes	None
E4	M1	M2	0.111960	Yes	None
	M1	M3	0.111908	Yes	None
	M7	M8	0.233164	No	None
E5	M1	M2	**6.664182**	Yes	48 L 0.6823
					65 I 0.9945**
	M1	M3	**6.701044**	Yes	48 L 0.8923
					65 I 0.9998**
	M7	M8	**6.742870**	Yes	48 L 0.8909
					65 I 0.9998**
L2	M1	M2	0.434744	No	None
	M1	M3	1.361254	Yes	None
	M7	M8	0.180566	No	None
L1	M1	M2	0.011224	No	None
	M1	M3	0.011220	No	None
	M7	M8	0.033232	No	None

[a] LRT statistics follow a χ^2 distribution with degrees of freedom equaling 2 when values were >5.99 and *P* values were <0.05 (in boldface).

[b] **, positively selected sites with posterior probability (*P*) values of >0.99 (2).

E5 may influence HPV gene expression via the activation of activator protein-1 and NF-1 (13, 14). Since several binding sites of transcription factors (e.g., NF-1) located in the HPV URR are variable, selective pressure driven by the ability of HPV16 E5 protein to active viral gene expression via these transcription factors is possible. Although mutation analyses indicated that the transforming activity of mutant E5 proteins was similar to that of wild-type E5 (37), the oncogenicity of specific HPV variants may vary geographically, possibly due to genetic differences between populations. Since the biochemical and biological functions of HPV16 E5 protein are not well characterized, efforts are still needed to decipher the complicated network existing between E6, E7, E5, and other cellular proteins.

Diversifying selection within the HPV16 E2 gene. It is worth noting that the E2 gene might be under selection pressure, since it has a relatively high ω ratio (Table 5). The HPV16 E2 protein is involved in gene-expression regulation and replication of the HPV genome and appears to be the target of both humoral and cellular immune responses (22, 44). Several humoral epitopes are known in the E2 protein, one in the transactivation domain (aa 121 to 140), three in the hinge region (aa 181 to 200, 241 to 260, and 271 to 290), and one in the DNA binding domain (aa 328 to 346) (22). In this work, all but one (i.e., aa 181 to 200) of these epitopes had changes in the E2 proteins of non-European variants, suggesting the potential for different humoral responses to HPVs of the European and non-European taxa (Table 3). Since the disruption of the E2 gene during viral integration contributes to tumor progression, the variant HPV16 E2 proteins of non-European variants (e.g., the As-Am isolate) may be better adapted for deregulating the expression of viral oncogenes (10). Hence, the high level of diversity among HPV16 E6 and E5 proteins may be a potential force driving the evolutionary selection of the E2 gene. For instance, HPV16 E2 variants frequently cosegregated with the E6 T350G (L83V) variant that has been associated with viral persistence (10). This cosegregation was proposed to act as an additional risk factor for the development of cervical cancer (46, 60). Another reason that the HPV16 E2 gene contained a relatively high level of nonsynonymous substitutions might be explained by the evolutionary constraint imposed upon it from the overlapping gene, HPV16 E4. Accordingly, synonymous substitutions of the E2 gene might be constrained by the relatively conserved E4 gene (39, 48). However, no specific sites were detected by the LRTs. Yang and Swanson (74) suggest that these tests are highly prone to sampling bias and that more variants of HPV16 E2 gene should therefore be included in future molecular evolutionary analyses.

This study provides new data demonstrating lineage fixation of variants in complete HPV16 genomes. Analysis of a representative set of HPV16 genomes identified diversifying sites under selective pressure within the HPV16 E6 and E5 ORFs. These HPV16 proteins are known to have transforming activities. Nevertheless, more effort is needed to examine the evolutionary dynamics of HPV genomes, especially those associated with high-risk HPV types, in order to understand the genetic basis of biological success and niche adaptation that probably indirectly lead to cancer causation.

ACKNOWLEDGMENTS

We thank Ziheng Yang for his useful comments and advice regarding PAML usage and David Posada for his guidance for MODELTEST operation.

This work was supported in part by Public Health Service awards CA78527 and CA85178 from the National Cancer Institute.

REFERENCES

1. **Andersson, S., M. Alemi, E. Rylander, A. Strand, B. Larsson, J. Sallstrom, and E. Wilander.** 2000. Uneven distribution of HPV 16 E6 prototype and variant (L83V) oncoprotein in cervical neoplastic lesions. Br. J. Cancer **83:** 307–310.
2. **Anisimova, M., J. P. Bielawski, and Z. Yang.** 2002. Accuracy and power of Bayes prediction of amino acid sites under positive selection. Mol. Biol. Evol. **19:**950–958.
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 A0201-restricted viral peptide from cells expressing human papillomavirus type 16 E6 oncoprotein. Eur. J. Immunol. **24:**3175–3179.
4. **Beckman, G., R. Birgander, A. Sjalander, N. Saha, P. A. Holmberg, A. Kivela, and L. Beckman.** 1994. Is p53 polymorphism maintained by natural selection? Hum. Hered. **44:**266–270.
5. **Bernard, H. U., and D. Apt.** 1994. Transcriptional control and cell type specificity of HPV gene expression. Arch. Dermatol. **130:**210–215.
6. **Berumen, J., R. M. Ordonez, E. Lazcano, J. Salmeron, S. C. Galvan, R. A. Estrada, E. Yunes, A. Garcia-Carranca, G. Gonzalez-Lira, and A. Madrigal-de la Campa.** 2001. Asian-American variants of human papillomavirus 16 and risk for cervical cancer: a case-control study. J. Natl. Cancer Inst. **93:** 1325–1330.
7. **Bosch, F. X., M. M. Manos, N. Munoz, M. Sherman, A. Jansen, J. Peto, M. Schiffman, V. Moreno, R. Kurman, K. Shah, and I. S. Group.** 1995. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. J. Natl. Cancer Inst. **87:**796–802.

8. **Brady, C. S., M. F. Duggan-Keen, J. A. Davidson, J. M. Varley, and P. L. Stern.** 1999. Human papillomavirus type 16 E6 variants in cervical carcinoma: relationship to host genetic factors and clinical parameters. J. Gen. Virol. **80:**3233–3240.
9. **Burk, R. D., G. Y. F. Ho, L. Beardsley, M. Lempa, M. Peters, and R. Bierman.** 1996. Sexual practices and partner selection are the predominant risk factors for genital HPV infection in young women. J. Infect. Dis. **174:** 679–689.
10. **Casas, L., S. C. Galvan, R. M. Ordonez, N. Lopez, M. Guido, and J. Berumen.** 1999. Asian-American variants of human papillomavirus type 16 have extensive mutations in the E2 gene and are highly amplified in cervical carcinomas. Int. J. Cancer **83:**449–455.
11. **Chakrabarti, O., K. Veeraraghavalu, V. Tergaonkar, Y. Liu, E. J. Androphy, M. A. Stanley, and S. Krishna.** 2004. Human papillomavirus type 16 E6 amino acid 83 variants enhance E6-mediated MAPK signaling and differentially regulate tumorigenesis by Notch signaling and oncogenic Ras. J. Virol. **78:**5934–5945.
12. **Chan, S.-Y., L. Ho, C.-K. Ong, V. Chow, B. Drescher, M. Durst, J. ter Meulen, L. Villa, J. Luande, H. 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.
13. **Chen, S. L., C. H. Huang, T. C. Tsai, K. Y. Lu, and Y. P. Tsao.** 1996. The regulation mechanism of c-jun and junB by human papillomavirus type 16 E5 oncoprotein. Arch. Virol. **141:**791–800.
14. **Chen, S. L., Y. K. Lin, L. Y. Li, Y. P. Tsao, H. Y. Lo, W. B. Wang, and T. C. Tsai.** 1996. E5 proteins of human papillomavirus types 11 and 16 transactivate the c-*fos* promoter through the NF1 binding element. J. Virol. **70:** 8558–8563.
15. **Cheng, G., J. P. Icenogle, R. Kirnbauer, N. L. Hubbert, M. E. St. Louis, C. Han, E. I. Svare, S. K. Kjaer, D. R. Lowy, and J. T. Schiller.** 1995. Divergent human papillomavirus type 16 variants are serologically cross-reactive. J. Infect. Dis. **172:**1584–1587.
16. **Clifford, G. M., J. S. Smith, T. Aguado, and S. Franceschi.** 2003. Comparison of HPV type distribution in high-grade cervical lesions and cervical cancer: a meta-analysis. Br. J. Cancer **89:**101–105.
17. **Clifford, G. M., J. S. Smith, M. Plummer, N. Munoz, and S. Franceschi.** 2003. Human papillomavirus types in invasive cervical cancer worldwide: a meta-analysis. Br. J. Cancer **88:**63–73.
18. **Cline, J., J. C. Braman, and H. H. Hogrefe.** 1996. PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. Nucleic Acids Res. **24:**3546–3551.
19. **Conrad, M., V. J. Bubb, and R. Schlegel.** 1993. The human papillomavirus type 6 and 16 E5 proteins are membrane-associated proteins which associate with the 16-kilodalton pore-forming protein. J. Virol. **67:**6170–6178.
20. **DeFilippis, V. R., F. J. Ayala, and L. P. Villarreal.** 2002. Evidence of diversifying selection in human papillomavirus type 16 E6 but not E7 oncogenes. J. Mol. Evol. **55:**491–499.
21. **de Villiers, E. M., C. Fauquet, T. R. Broker, H. U. Bernard, and H. zur Hausen.** 2004. Classification of papillomaviruses. Virology **324:**17–27.
22. **Dillner, J.** 1990. Mapping of linear epitopes of human papillomavirus type 16: the E1, E2, E4, E5, E6 and E7 open reading frames. Int. J. Cancer **46:** 703–711.
23. **Ellis, J. R., 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, et al.** 1995. The association of an HPV16 oncogene variant with HLA-B7 has implications for vaccine design in cervical cancer. Nat. Med. **1:**464–470.
24. **Eriksson, A., J. R. Herron, T. Yamada, and C. M. Wheeler.** 1999. Human papillomavirus type 16 variant lineages characterized by nucleotide sequence analysis of the E5 coding segment and the E2 hinge region. J. Gen. Virol. **80:** 595–600.
25. **Fehrmann, F., D. J. Klumpp, and L. A. Laimins.** 2003. Human papillomavirus type 31 E5 protein supports cell cycle progression and activates late viral functions upon epithelial differentiation. J. Virol. **77:**2819–2831.
26. **Gao, L., B. Chain, C. Sinclair, L. Crawford, J. Zhou, J. Morris, X. Zhu, and H. Stauss.** 1994. Immune response to human papillomavirus type 16 E6 gene in a live vaccinia vector. J. Gen. Virol. **75:**157–164.
27. **Genther, S. M., S. Sterling, S. Duensing, K. Munger, C. Sattler, and P. F. Lambert.** 2003. Quantitative role of the human papillomavirus type 16 E5 gene during the productive stage of the viral life cycle. J. Virol. **77:**2832–2842.
28. **Giannoudis, A., and C. S. Herrington.** 2001. Human papillomavirus variants and squamous neoplasia of the cervix. J. Pathol. **193:**295–302.
29. **Goldsborough, M. D., D. DiSilvestre, G. F. Temple, and A. T. Lorincz.** 1989. Nucleotide sequence of human papillomavirus type 31: a cervical neoplasia-associated virus. Virology **171:**306–311.
30. **Herrero, R., A. Hildesheim, C. Bratti, M. E. Sherman, M. Hutchinson, J. Morales, I. Balmaceda, M. D. Greenberg, M. Alfaro, R. D. Burk, S. Wacholder, M. Plummer, and M. Schiffman.** 2000. Population-based study of human papillomavirus infection and cervical neoplasia in rural Costa Rica. J. Natl. Cancer Inst. **92:**464–474.
31. **Hildesheim, A.** 1997. Human papillomavirus variants: implications for natural history studies and vaccine development efforts. J. Natl. Cancer Inst. **89:** 752–753.
32. **Hildesheim, A., M. Schiffman, C. Bromley, S. Wacholder, R. Herrero, A. Rodriguez, M. C. Bratti, M. E. Sherman, U. Scarpidis, Q. Q. Lin, M. Terai, R. L. Bromley, K. Buetow, R. J. Apple, and R. D. Burk.** 2001. Human papillomavirus type 16 variants and risk of cervical cancer. J. Natl. Cancer Inst. **93:**315–318.
33. **Ho, G. Y. F., R. Bierman, L. Beardsley, C. J. Chang, and R. D. Burk.** 1998. Natural history of cervicovaginal papillomavirus infection in young women. N. Engl. J. Med. **338:**423–428.
34. **Ho, L., S. Y. Chan, R. D. Burk, B. C. Das, K. Fujinaga, J. P. Icenogle, T. Kahn, N. Kiviat, W. Lancaster, N. P. Mavromara, S. Mitrani-Rosenbaum, B. 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.
35. **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.
36. **Ho, L., S. K. Tay, S. Y. Chan, and H. U. Bernard.** 1993. Sequence variants of human papillomavirus type 16 from couples suggest sexual transmission with low infectivity and polyclonality in genital neoplasia. J. Infect. Dis. **168:**803–809.
37. **Hsieh, C. H., Y. P. Tsao, C. H. Wang, C. P. Han, J. L. Chang, J. Y. Lee, and S. L. Chen.** 2000. Sequence variants and functional analysis of human papillomavirus type 16 E5 gene in clinical specimens. Arch. Virol. **145:**2273–2284.
38. **Huelsenbeck, J. P., and F. Ronquist.** 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17:**754–755.
39. **Hughes, A. L., K. Westover, J. da Silva, D. H. O'Connor, and D. I. Watkins.** 2001. Simultaneous positive and purifying selection on overlapping reading frames of the *tat* and *vpr* genes of simian immunodeficiency virus. J. Virol. **75:** 7966–7972.
40. **Huibregtse, J. M., M. Scheffner, and P. M. Howley.** 1991. A cellular protein mediates association of p53 with the E6 oncoprotein of human papillomavirus types 16 or 18. EMBO J. **10:**4129–4135.
41. **Huibregtse, J. M., M. Scheffner, and P. M. Howley.** 1994. E6-AP directs the HPV E6-dependent inactivation of p53 and is representative of a family of structurally and functionally related proteins. Cold Spring Harbor Symp. Quant. Biol. **59:**237–245.
42. **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.
43. **International Agency for Research on Cancer.** 1995. Monographs on the evaluation of carcinogenic risks to humans, vol. 64. Human papillomaviruses, International Agency for Research on Cancer, Lyon, France.
44. **Konya, J., C. Eklund, V. af Geijersstam, F. Yuan, G. Stuber, and J. Dillner.** 1997. Identification of a cytotoxic T-lymphocyte epitope in the human papillomavirus type 16 E2 protein. J. Gen. Virol. **78:**2615–2620.
45. **Li, W.-H.** 1997. Molecular evolution. Sinauer, Sunderland, Mass.
46. **Londesborough, P., L. Ho, G. Terry, J. Cuzick, C. Wheeler, and A. Singer.** 1996. Human papillomavirus genotype as a predictor of persistence and development of high-grade lesions in women with minor cervical abnormalities. Int. J. Cancer **69:**364–368.
47. **Lundberg, K. S., D. D. Shoemaker, M. W. Adams, J. M. Short, J. A. Sorge, and E. J. Mathur.** 1991. High-fidelity amplification using a thermostable DNA polymerase isolated from Pyrococcus furiosus. Gene **108:**1–6.
48. **Mizokami, M., E. Orito, K. Ohba, K. Ikeo, J. Y. Lau, and T. Gojobori.** 1997. Constrained evolution with respect to gene overlap of hepatitis B virus. J. Mol. Evol. **44**(Suppl. 1)**:**S83–S90.
49. **Munoz, N., F. X. Bosch, S. de Sanjose, R. Herrero, X. Castellsague, K. V. Shah, P. J. Snijders, and C. J. Meijer.** 2003. Epidemiologic classification of human papillomavirus types associated with cervical cancer. N. Engl. J. Med. **348:**518–527.
50. **Myers, G., H. Delius, J. Icenogle, H. U. Bernard, M. Favre, M. van Ranst, and C. M. Wheeler.** 1997. Human papillomaviruses 1997: a compilation and analysis of nucleic acid and amino acid sequences. Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, N.Mex.
51. **Nielsen, R., and Z. Yang.** 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics **148:**929–936.
52. **Palefsky, J. M., H. Minkoff, L. A. Kalish, A. Levine, H. S. Sacks, P. Garcia, M. Young, S. Melnick, P. Miotti, and R. Burk.** 1999. Cervicovaginal human papillomavirus infection in human immunodeficiency virus-1 (HIV)-positive and high-risk HIV-negative women. J. Natl. Cancer Inst. **91:**226–236.
53. **Picconi, M. A., L. V. Alonio, L. Sichero, V. Mbayed, L. L. Villa, J. Gronda, R. Campos, and A. Teyssie.** 2003. Human papillomavirus type-16 variants in Quechua aboriginals from Argentina. J. Med. Virol. **69:**546–552.

54. **Posada, D., and K. A. Crandall.** 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14:**817–818.
55. **Storey, A., M. Thomas, A. Kalita, C. Harwood, D. Gardiol, F. Mantovani, J. Breuer, I. M. Leigh, G. Matlashewski, and L. Banks.** 1998. Role of a p53 polymorphism in the development of human papillomavirus-associated cancer. Nature **393:**229–234.
56. **Straight, S. W., B. Herman, and D. J. McCance.** 1995. The E5 oncoprotein of human papillomavirus type 16 inhibits the acidification of endosomes in human keratinocytes. J. Virol. **69:**3185–3192.
57. **Swofford, D. L.** 1998. PAUP*. Phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer Associates, Sunderland, Mass.
58. **Terai, M., and R. D. Burk.** 2001. Characterization of a novel genital human papillomavirus by overlapping PCR: candHPV86 identified in cervicovaginal cells of a woman with cervical neoplasia. J. Gen. Virol. **82:**2035–2040.
59. **Terai, M., and R. D. Burk.** 2001. Complete nucleotide sequence and analysis of a novel human papillomavirus (HPV 84) genome cloned by an overlapping PCR method. Virology **279:**109–115.
60. **Terry, G., L. Ho, and J. Cuzick.** 1997. Analysis of E2 amino acid variants of human papillomavirus types 16 and 18 and their associations with lesion grade and HLA DR/DQ type. Int. J. Cancer **73:**651–655.
61. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:**4876–4882.
62. **Tornesello, M. L., F. M. Buonaguro, A. Meglio, L. Buonaguro, E. Beth-Giraldo, and G. Giraldo.** 1997. Sequence variations and viral genomic state of human papillomavirus type 16 in penile carcinomas from Ugandan patients. J. Gen. Virol. **78:**2199–2208.
63. **van Duin, M., P. J. Snijders, M. T. Vossen, E. Klaassen, F. Voorhorst, R. H. Verheijen, T. J. Helmerhorst, C. J. Meijer, and J. M. Walboomers.** 2000. Analysis of human papillomavirus type 16 E6 variants in relation to p53 codon 72 polymorphism genotypes in cervical carcinogenesis. J. Gen. Virol. **81:**317–325.
64. **Veress, G., K. Szarka, X. P. Dong, L. Gergely, and H. Pfister.** 1999. Functional significance of sequence variation in the E2 gene and the long control region of human papillomavirus type 16. J. Gen. Virol. **80:**1035–1043.
65. **Villa, L. L., L. Sichero, P. Rahal, O. Caballero, A. Ferenczy, T. Rohan, and E. L. Franco.** 2000. Molecular variants of human papillomavirus types 16 and 18 preferentially associated with cervical neoplasia. J. Gen. Virol. **81:**2959–2968.
66. **Watts, K. J., C. H. Thompson, Y. E. Cossart, and B. R. Rose.** 2002. Sequence variation and physical state of human papillomavirus type 16 cervical cancer isolates from Australia and New Caledonia. Int. J. Cancer **97:**868–874.
67. **Xi, L. F., L. A. Koutsky, D. A. Galloway, J. Kuypers, J. P. Hughes, C. M. Wheeler, K. K. Holmes, and N. B. Kiviat.** 1997. Genomic variation of human papillomavirus type 16 and risk for high grade cervical intraepithelial neoplasia. J. Natl. Cancer Inst. **89:**796–802.
68. **Yamada, T., M. M. Manos, J. Peto, C. E. Greer, N. Munoz, F. X. Bosch, and C. M. Wheeler.** 1997. Human papillomavirus type 16 sequence variation in cervical cancers: a worldwide perspective. J. Virol. **71:**2463–2472.
69. **Yamada, T., C. M. Wheeler, A. L. Halpern, A. C. Stewart, A. Hildesheim, and S. A. Jenison.** 1995. Human papillomavirus type 16 variant lineages in United States populations characterized by nucleotide sequence analysis of the E6, L2, and L1 coding segments. J. Virol. **69:**7743–7753.
70. **Yang, Z.** 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13:**555–556.
71. **Yang, Z., N. Goldman, and A. Friday.** 1994. Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. Mol. Biol. Evol. **11:**316–324.
72. **Yang, Z., R. Nielsen, N. Goldman, and A. M. Pedersen.** 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155:**431–449.
73. **Yang, Z., R. Nielsen, and M. Hasegawa.** 1998. Models of amino acid substitution and applications to mitochondrial protein evolution. Mol. Biol. Evol. **15:**1600–1611.
74. **Yang, Z., and W. J. Swanson.** 2002. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. Mol. Biol. Evol. **19:**49–57.
75. **Zehbe, I., J. Mytilineos, I. Wikstrom, R. Henriksen, L. Edler, and M. Tommasino.** 2003. Association between human papillomavirus 16 E6 variants and human leukocyte antigen class I polymorphism in cervical cancer of Swedish women. Hum. Immunol. **64:**538–542.
76. **Zehbe, I., R. Tachezy, J. Mytilineos, G. Voglino, I. Mikyskova, H. Delius, A. Marongiu, L. Gissmann, E. Wilander, and M. Tommasino.** 2001. Human papillomavirus 16 E6 polymorphisms in cervical lesions from different European populations and their correlation with human leukocyte antigen class II haplotypes. Int. J. Cancer **94:**711–716.
77. **Zehbe, I., G. Voglino, H. Delius, E. Wilander, and M. Tommasino.** 1998. Risk of cervical cancer and geographical variations of human papillomavirus 16 E6 polymorphisms. Lancet **352:**1441–1442.