VIROLOGY 221, 189–198 (1996)
ARTICLE NO. 0365

Variability and Phylogeny of the L1 Capsid Protein Gene of Human Papillomavirus Type 5:

MASAAKI KAWASE,* GERARD ORTH,* STEFANIA JABLONSKA,† CLAUDINE BLANCHET-BARDON,‡ LUIS-ALFREDO RUEDA,§ and MICHEL FAVRE*,[1]

*Unité Mixte Institut Pasteur–INSERM (Unité des Papillomavirus, Unité INSERM 190), Institut Pasteur, 75724 Paris Cedex 15, France; †Department of Dermatology, Warsaw School of Medicine, Warsaw, Poland; ‡Département de Dermatologie, Hôpital Saint Louis, Paris, France; and §Centro Dermatologico F. Lleras A., Bogota, Colombia

Received March 15, 1996; accepted May 7, 1996

We analyzed the variability and established the phylogeny of the L1 capsid protein gene of 33 isolates of human papillomavirus type 5 (HPV5) obtained from epidermodysplasia verruciformis patients from different continents. By comparing the sequences of a 419-bp fragment with those published for two Japanese isolates, we found 12.9% variable nucleotide positions, defining 25 variants with mutation rates ranging from 0.2 to 8.8%. Such a high intratypic diversity is unusual among HPVs. Nine of the 139 encoded amino acids were variable and 12 protein variants were identified. Fifteen of the 16 substitutions observed were clustered in two short regions. A 9-amino-acid insert, already reported for the Japanese HPV5b isolate, was found within one of the regions in five isolates. Our data support that the insert arose from the duplication of a 30-nucleotide sequence. Phylogenetic trees distributed the DNA variants into three subtypes (a to c) with a divergence higher than 4.5% and allowed the recognition of European and African lineages. By contrast with the trees based on the HPV5 E6 gene, HPV5a DNA variants and the HPV5b variants lacking the insert constituted a single group in the L1 amino acid tree, probably reflecting different levels of structural constraints for the HPV5 L1 and E6 proteins. In that respect, the short variable L1 sequences should represent less constrained regions.

INTRODUCTION

Human papillomaviruses (HPVs) constitute a large group of DNA viruses causing a variety of epithelial proliferations of the squamous epithelia of the skin and mucous membranes (Orth, 1994; zur Hausen and de Villiers, 1994). Among the 77 HPV genotypes characterized so far (zur Hausen, 1995), some are etiologically associated with the carcinomas of the uterine cervix (mainly HPV16 and 18) and others with the cutaneous carcinomas observed in patients suffering from epidermodysplasia verruciformis (mainly HPV5) (Orth, 1987; zur Hausen and de Villiers, 1994). HPV genotypes were first distinguished on the basis of a cross-hybridization less than 50% under the most stringent conditions. A more recent definition, grossly equivalent, is that genotypes share less than 90% identical nucleotides in the open reading frames (ORFs) encoding the viral oncoproteins (E6 and E7 ORFs) and the major capsid protein (L1 ORF). HPV isolates sharing more than 98% identical nucleotides are considered variants, and a DNA sequence identity in the range of 90 to 98% defines subtypes (Bernard *et al.,* 1994; de Villiers, 1994; Deau *et al.,* 1993; Longuet *et al.,* 1996; Orth, 1994). Phylogenetic trees constructed from nucleotide or amino acid sequences disclosed a grouping of HPV genotypes reflecting their cutaneous or mucous tropism, their pathogenicity, and their oncogenic potential (Chan *et al.,* 1995; Van Ranst *et al.,* 1993).

Evidence for an intratypic diversity of oncogenic HPV types has recently been disclosed. Comparison of nucleotide sequences of different genomic regions among a large number of HPV16 and HPV18 isolates demonstrated the existence of numerous stable DNA variants (Chan *et al.,* 1992; Ho *et al.,* 1993; Icenogle *et al.,* 1991, 1995; Ong *et al.,* 1993; Pushko *et al.,* 1994; Smits *et al.,* 1994; Yamada *et al.,* 1995). The amount of variable nucleotide positions among HPV16 isolates was found to vary between 2.2% for the L1 ORF and 4.9% for the long control region (LCR), and the maximum pairwise divergence in the coding regions did not exceed 1.2% for the E6 ORF, 1.3% for the L1 ORF, and 2.4% for the L2 ORF (Icenogle *et al.,* 1995; Yamada *et al.,* 1995). A higher degree of intratypic diversity is displayed by HPV types 5 and 8 associated with epidermodysplasia verruciformis (EV) (Deau *et al.,* 1991, 1993). This rare disease is a genetically conditioned abnormal susceptibility both to cutaneous infection by a subgroup of specific, phylogenetically related HPV types that are innocuous for the general population and to the oncogenic potential of some of them, mostly HPV5 (Majewski and Jablonska, 1995; Orth, 1987; Orth *et al.,* 1979). Comparison of 13 HPV5 isolates

[1] To whom correspondence and reprint requests should be addressed at Unité Mixte Institut Pasteur–INSERM U190, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France. Fax: 33.1.45 68.89.66.

0042-6822/96 $18.00

disclosed a great variability in the 3′ end of the LCR and in the E6 gene, affecting 13.3 and 11% of the nucleotide positions, respectively, with a maximal divergence of 8.8% between variants (Deau *et al.,* 1993). Phylogenetic analysis of HPV5 variants distributed them into three subtypes, designated a, b, and c (Deau *et al.,* 1993).

The functional significance of the intratypic variability of HPV genotypes remains poorly understood. No evidence for distinct HPV16 serotypes has been obtained so far (Cheng *et al.,* 1995; Yaegashi *et al.,* 1993). There is some evidence, however, that the variability of the E6 protein constitutes a mechanism by which tumor cells escape immune surveillance (Ellis *et al.,* 1995). Due to its great genetic heterogeneity compared to HPV16, HPV5 constitutes a valuable model to analyze the genetic diversity and the evolution of HPV genotypes and to investigate to which extent this diversity affects the biological properties of the variants.

As a prerequisite to the study of the antigenic variability of the major capsid protein, we analyzed the genetic diversity of the HPV5 L1 ORF of 33 isolates originating from Europe, Africa, and America. We sequenced a 419-bp L1 region reported to contain a 27-nucleotide insert in a Japanese isolate (Yabe *et al.,* 1991). Our data disclosed a high degree of nucleotide substitutions, in the range of that observed previously in the 3′ end of the LCR and the E6 gene (Deau *et al.,* 1993). We also demonstrate the existence of a group of phylogenetically related variants containing the 27-nucleotide insert. Amino acid changes or insertion occurred in two short sequences that probably correspond to less constrained regions of the L1 protein.

MATERIALS AND METHODS

Tissue specimens

Scrapings or biopsies were taken from skin flat warts and macules of 31 EV patients originating from different continents (Table 1). A biopsy of a squamous cell carcinoma was also obtained from one of these patients (isolate E11). Patients 1, 2, 3, 5, 10, 16, 17, 29, and 30 correspond to patients 1, 7, 8, 10, 6, 2, 3, 4, and 5 in Deau *et al.* (1993), patients 4 and 6 to patients 1 and 5 in Orth *et al.* (1979), patients 9, 15, and 16 to patients 8, 4, and 5 in Kremsdorf *et al.* (1984), patient 12 to patient 2 in Guilhou *et al.* (1980), patients 18 and 19 to patients 5 and 6 in Lutzner *et al.* (1984), and patients 26, 27, 28, and 31 to patients 9, 10, 8, and 4 in Rueda (1993).

Polymerase chain reaction

The molecular cloning of the DNA of HPV5 E1, NAf1, SAm4, E10, and E11 isolates has been reported elsewhere (Deau *et al.,* 1991; Kremsdorf *et al.,* 1982). The prototypical HPV5 DNA (J1) was kindly provided by R. Ostrow (University of Minnesota, Minneapolis). The L1 419-nucleotide fragments encompassing nucleotide positions 6124 to 6582 were obtained by the polymerase chain reaction (PCR) technique with primers corresponding to the nucleotide positions 6124 to 6143 (5′-CAACACAGAGTATTTCGCCT) and 6563 to 6582 (5′-TAAACTGACATCTGATCTAC). PCR reaction mixtures contained 100 to 500 ng of total cellular DNA, 50 m*M* KCl, 10 m*M* Tris–HCl (pH 8.3), 100 μM each deoxynucleoside triphosphate, 1.5 m*M* $MgCl_2$, 25 pmol of each primer, and 2 U of *Thermus aquaticus* DNA polymerase (Perkin–Elmer Cetus Instruments, Emeryville, CA). Mixtures were incubated at 100° for 5 min for DNA denaturation. Subsequently, 30 cycles of amplification were performed with a PCR processor (Hybaid, Ltd.) as previously described (Deau *et al.,* 1993). After Klenow fragment of DNA polymerase I was used for end-filling, the fragments were inserted by blunt ligation at the unique *Sma*I site of a pBluescript plasmid (Stratagene GmbH, Heidelberg).

DNA sequencing

For each HPV5 isolate, sequencing was performed on both strands of recombinant plasmids by the dideoxy chain termination method (Sanger *et al.,* 1977) using the modified T7 DNA polymerase (Sequenase; United States Biochemical Corp., Cleveland) under the conditions described by the manufacturer. The nucleotide sequence of the L1 fragment of HPV5 E1, NAf1, SAm4, E10, and E11 isolates was determined on full-length cloned DNA with primers used for PCR amplification and an internal primer encompassing nucleotides 6340 to 6359 in both orientations. The nucleotide sequence of the other isolates was determined with M13 oligonucleotide primers complementary to sequences flanking the insert and the internal L1 primer in both orientations. Several independent clones (2 to 5) were sequenced for each PCR product in order to take into account the possibility of PCR errors. Sequences have been submitted to the Genbank nucleotide sequence database and Accession Nos. are listed in Table 1.

Phylogenetic trees

Pairwise alignments of the DNA and deduced amino acid sequences of HPV5 variants with those of the prototype HPV5 (J1) (Zachow *et al.,* 1987) and the subtype HPV5b (J2) (Yabe *et al.,* 1991) were performed without taking into account the 27-nucleotide insert, using the fast method of Wilbur and Lipman (1983) executed with the CLUSTAL V option (Higgins *et al.,* 1992), as previously described (Deau *et al.,* 1993). Multiple alignments were generated with the same program, and the similarity scores were used to construct phylogenetic trees by unweighted pair-group maximum-average (UPGMA) analysis (Sneath and Sokal, 1973). Results of multiple sequence alignments were used to construct HPV5 trees by the maximum parsimony (DNA PARS) or the maximum likelihood methods with the PHYLIP 3.4 programs (Phylogeny Inference Package) (Felsenstein, 1988). To root

TABLE 1

Origin and Nomenclature of HPV5 Isolates

Patient	Geographic origin	Isolate designation[a]	Subtype[b]	Accession No.[c]
1	Poland	E1 (5a2)	a	U49465
2	Poland	E2 (5a4)	a	U49463
3[d]	Poland	E3 (5a5)	a	U49462
4[d]	Poland	E4	a	U49462
5[d]	Poland	E5 (5b1)	b(+)	U49477
6[d]	Poland	E6	a	U49462
7	Poland	E7	a	U49466
8	Poland	E8	a	U49460
9	Poland	E9	a	U49467
10	France	E10 (5a3)	a	U49461
		E11 (5c1)	c	U49476
11	France	E12	a	U49462
12	France	E13	b(+)	U49479
13	France	E14	b(+)	U49481
14	Spain	E15	a	U49459
15	Italy	E16	b(+)	U49478
16	The Netherlands	E17 (5a2)	a	U49464
17	Algeria	NAf1 (5b3)	b	U49469
18	Algeria	NAf2	b(+)	U49480
		NAf3	b	U49474
19	Algeria	NAf4	b	U49468
20	Algeria	NAf5	b	U49470
21	Senegal	WAf1	b	U49470
22	Cape Verde Islands	WAf2	b	U49471
23	French Caribbean	FC1	b	U49472
24	French Caribbean	FC2	b	U49473
25	United States	NAm1	c	U49475
26[d]	Colombia	SAm1	a	U49466
27[d]	Colombia	SAm2	a	U49466
28[d]	Colombia	SAm3	a	U49466
29[d]	Colombia	SAm4 (5b2)	b	U49468
30[d]	Colombia	SAm5 (5b2)	b	U49468
31[d]	Colombia	SAm6	b	U49468
J1	Japan	J1 (5a1)	a	M17463
J2	Japan	J2 (5b)	b(+)	D90252

[a] Isolates were designated by letters according to the geographic origin of the patients (E, Europe; NAf, North Africa; WAf, West Africa; FC, French Caribbean; NAm, North America; SAm, South America; J, Japan) and by numbers, taking into account the chronological order of their characterization and the familial occurrence of the disease. J1 and J2 correspond to Japanese isolates sequenced by Zachow *et al.* (1987) and Yabe *et al.,* (1991), respectively. Previous phylogenetic classification (Deau *et al.,* 1993) is given in parentheses.

[b] Deduced from the phylogenetic analysis shown in Fig. 4A. (+) indicates the presence of the 27-nucleotide insert.

[c] Assigned to the sequences of the L1 419-bp fragment upon submission to the Genbank database.

[d] Patients 3 to 5, 26 to 28, and 29 to 31 are siblings from three different families. Patient 6 is the daughter of patient 5.

the trees, HPV47 sequences (Kiyono *et al.,* 1990) were used as an outgroup. Bootstrap resampling allowed estimation of the confidence limits of tree grouping which were significant at the 95% level (Felsenstein, 1988).

RESULTS

A high level of nucleotide substitutions in the HPV5 L1 ORF

The sequence of a 419-bp segment localized in the 5′ half of the HPV5 L1 ORF (nt 6144 to 6562; Zachow *et al.,* 1987) was established for 33 HPV5 isolates obtained from 31 EV patients (Table 1). Patients originated from Europe (16 cases, mainly from Poland and France), North Africa (4 cases), West Africa (2 cases), French Caribbean (2 cases), North America (1 case), and South America (6 cases). Isolates were designated by letters and numbers, taking into account the geographic origin of patients, the chronological order of their identification, and the familial occurrence of the disease (Table 1). Previous sequence analysis of the 3′ end of the LCR and of the entire E6 ORF of 10 of these isolates had distributed them into three subtypes, HPV5 a, b, and c (Deau *et al.,* 1993). We compared the L1 sequences with those published for two Japanese isolates corresponding to the prototypical HPV5 (Zachow *et al.,* 1987) and to the HPV5b subtype,

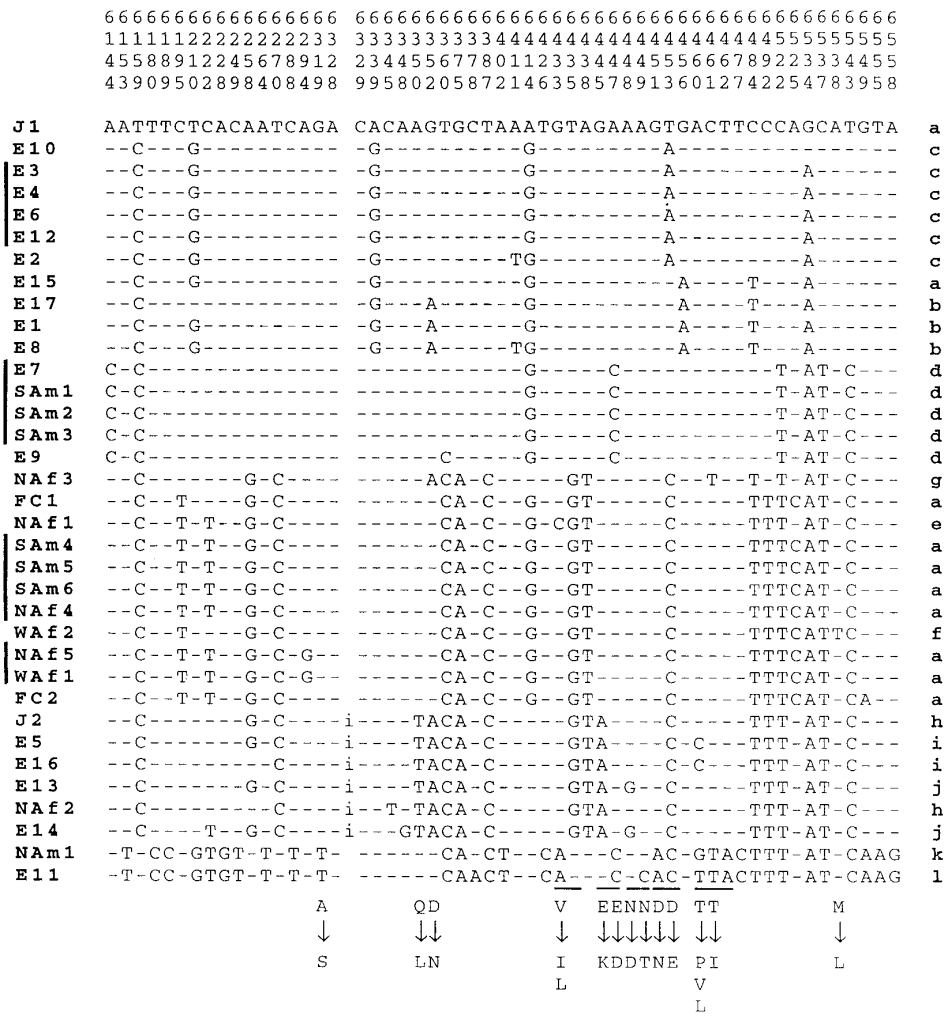

FIG. 1. Nucleotide sequence variations in the L1 ORF among HPV5 variants. Nucleotide positions at which variations were observed are written vertically on the top, as well as the nucleotides 6328 and 6329 flanking the 27-nucleotide insert (i) found in six isolates. The numbering of nucleotides refers to the nucleotide sequence of the prototypical HPV5 isolate (Zachow *et al.,* 1987), as corrected in this study (J1). Isolates (left column) are described in Table 1. Sequence data for the J2 isolate are from Yabe *et al.* (1991). Only substitutions observed in the variants are indicated. Bars on the left indicate identical isolates. The protein variant (a to l) corresponding to each isolate (Fig. 4C) is given on the right. Variable nucleotide positions within the same codon are underlined and amino acid changes resulting from nonsynonymous mutations are indicated at the bottom.

an isolate reported to contain an additional 27-nt sequence in this region (Yabe *et al.,* 1991). These two Japanese isolates will be referred to as J1 and J2, respectively.

All isolates, including J2, were found to differ from the prototypical HPV5 J1 by three C → G nucleotide substitutions at positions 6175, 6265, and 6502, corresponding to three amino acid changes (Pro87 → Ala; Arg117 → Gly; Gln196 → Glu). This difference was no longer found upon resequencing this L1 region of the original J1 DNA clone (a generous gift from R. Ostrow). We shall refer hereafter to the corrected sequence as the prototypical HPV5. Nucleotide sequence comparison revealed a high level of nucleotide substitutions among isolates (Fig. 1). Fifty-four (12.9%) of the 419 nucleotide positions were found variable and 58 nucleotide substitutions were observed. A hot spot of mutations was located between nucleotides 6433 and 6462, with 13 (43%) variable nucleotide positions and 16 substitutions. Moreover, five isolates were found to harbor a 27-nucleotide insert (Fig. 2). Without taking into account this insert, 23 HPV5 variants were identified, in addition to J1 and J2 variants (Fig. 1). Mutations involving C → T or T → C transitions accounted for 39% of the nucleotide changes, and CC → TT double mutations characteristic of UV-induced mutagenesis (Brash *et al.,* 1991) were not observed.

Pairwise comparison revealed nucleotide substitution rates ranging from 0.2% (a single mutation) to 8.8% (37 mutations). Two patients (cases 10 and 18) were found infected with two variants, and two variants were detected in a Polish family with four affected members (cases 3 to 6) (Table 1). Some variants were found in patients from different countries of the same continent (e.g., patients infected with E3, 4, 6, and 12 isolates) or from different continents (patients infected with E7, SAm1, 2, and 3 isolates)

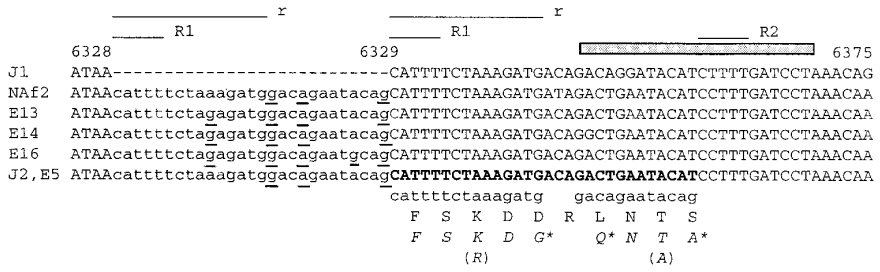

FIG. 2. Sequence alignment of the HPV5 variants in the L1 region containing a 27-nucleotide insert. The J2, E5, E13, E14, E16, and NAf2 sequences were aligned with that of the prototypical J1. The position of the insert reported by Yabe *et al.* (1991) in the J2 isolate is shown by a shadowed box above the J1 sequence. According to the alignment presented, the additional sequences (small letters) are likely to have resulted from a duplication of the 30-nucleotide sequence downstream of position 6328, by a slipped mispairing mechanism (Roth *et al.,* 1985) involving two imperfect repeats R1 and R2. The direct repeats (R1, r) thus created, and theoretically allowing the deletion of the insert, are also shown. Mutations in the duplicated sequences are underlined. The alignments of the 27-nucleotide insert and of the deduced 9-amino-acid sequence (italics) of the J2 and E5 isolates with the downstream sequences (bold) are given at the bottom. The three amino acid changes (stars) resulting from the 3-nucleotide deletion and from the two mutations shared by all variants are indicated, as well as the amino acid changes (parentheses) observed among variants.

(Table 1; Fig. 1). This points to the stability and the worldwide distribution of HPV5 variants.

Two short variable amino acid sequences in the HPV5 L1 protein

Sixteen of the 58 nucleotide changes were nonsynonymous mutations and resulted in 16 amino acid substitutions involving 9 (6.4%) of the 139 amino acid positions (amino acids 77 to 215) (Figs. 1 and 3). Pairwise comparisons of the deduced L1 amino acid sequences of the 25 DNA variants disclosed 12 protein variants, designated a to l. Variants differed by 1 to 6 amino acid changes (Fig. 3). Amino acid changes affecting 4 positions were nonconservative, namely, Gln145 → Leu, Asp146 → Asn, Glu177 → Lys, and Thr182 → Pro, Leu, Val, or Ile (Risler *et al.,* 1988). All variable residues but 1 were located in two short sequences of 10 and 12 amino acids. The 10-amino-acid sequence (amino acids 173 to 182) showed 12 substitutions affecting 5 residues. The 12-amino-acid sequence (amino acids 135 to 146) contained 3 variable residues and, for variants h to j, an insertion of 9 amino acids (Fig. 3). When the HPV5 L1 amino acid sequences were aligned with that of HPV47 (Kiyono *et al.,* 1990), the type most related to HPV5 in this region, only 8 (5.7%) differences were found. Five of them (3 amino acid changes and a deletion of 2 residues) were located in regions homologous to the two HPV5 variable sequences. Both regions are the least conserved in the aligned HPV16 L1 sequence (Seedorf *et al.,* 1985) (Fig. 3).

A subgroup of HPV5 L1 variants with a 27-nucleotide/9-amino-acid insert

A 27-nucleotide insert was found in variants obtained from a North African (NAf2) and four European (E5, E13, E14, and E16) patients (Fig. 2). A similar insert had been reported by Yabe *et al.* (1991) to be present in the HPV5b (J2) genome at position 6349 or 6350 (corresponding to nucleotide 6343 or 6344 in the protypical J1 HPV5) (Fig. 2). From the alignment of the nucleotide sequences of all six variants with that of J1, it is likely that the insert maps at nucleotide position 6328 and results from a duplication of the immediately downstream 30 nucleotides (nucleotides 6329 to 6358). In the variants containing the insert, the 5 nucleotides following position 6328 (CATTT, R1) are perfectly repeated 27 nucleotides downstream, and an imperfect repeat (CCTTT, R2) is found 30 nucleotides farther. This would fit with the generation of an overlapping duplication of 30 nucleotides during viral DNA replication by a slipped mispairing mechanism, according to the model described by Roth *et al.* (1985). Compared with the 30-nucleotide sequence, all inserts share a deletion of 3 nucleotides affecting codons 143 and 144 and two mutations, leading to the loss of the residue Arg144 and to three amino acid changes (Fig. 2). Two variants (E5 and NAf2) share the same insert as J2 while the others harbor mutations at positions 6328 +10 (E13, E14, and E16), +17 (E14), and +24 (E16), two of them leading to amino acid changes (Figs. 2 and 3). The first 15 nucleotides (r) in the duplication are directly repeated in the downstream sequences (Fig. 2) and this repeat could allow the deletion of the insert by a similar slipped mispairing mechanism.

The phylogenetic analysis of the HPV5 L1 variants

Multiple alignments of the nucleotide and deduced amino acid sequences of the L1 fragment were performed without taking into account the insert, and phylogenetic trees of the HPV5 variants were constructed by a distance matrix approach (UPGMA) (Sneath and Sokal, 1973), the maximum parsimony analysis (DNA PARS), and the maximum-likelihood method (Felsenstein, 1988). The L1 nucleotide trees distributed the isolates into three groups (subtypes a, b, and c) differing by more than 4.5% of nucleotides, as illustrated in Fig. 4A. The HPV5a L1 variants include 12 of the 17 European isolates, a variant found in three siblings from Colombia, and the J1 isolate.

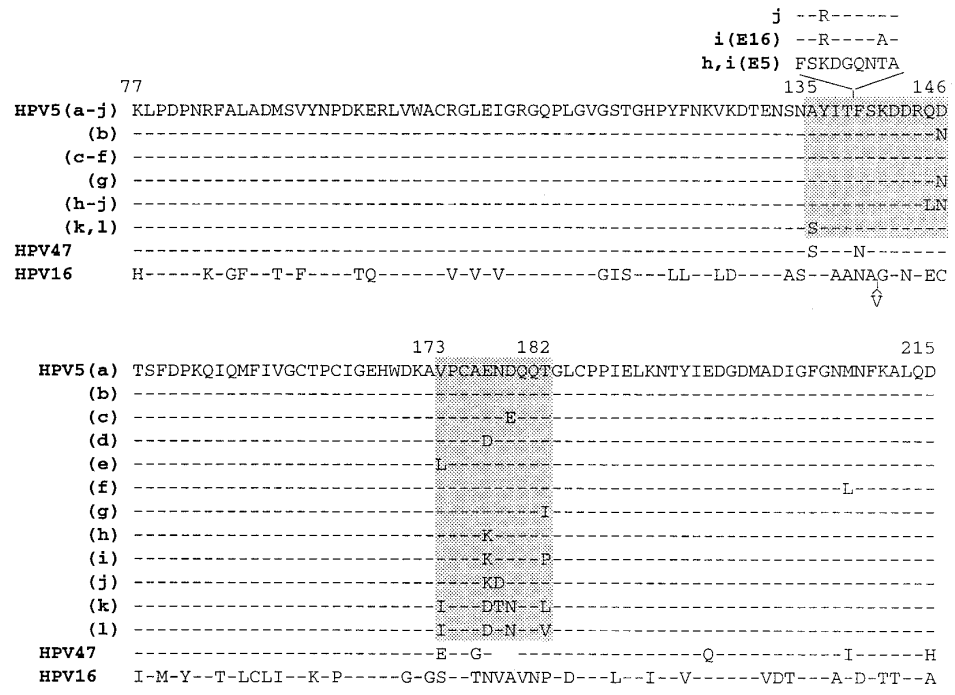

FIG. 3. Amino acid variations in the L1 protein among HPV5 variants. The L1 amino acid sequence encoded by the 419-bp fragment (amino acid positions 77 to 215) of the prototypical HPV5 (protein variant a) was deduced from the nucleotide sequence reported by Zachow *et al.* (1987), as corrected in this study. Substitutions found in the protein variants b to l are indicated. The sequence of the 9 amino acids inserted at position 138 in the J2 isolate (protein variant h) and amino acid changes observed in variants i and j are presented above the sequences. The two variable regions are shadowed. The alignment of L1 amino acid sequences of HPV 47 (Kiyono *et al.,* 1990) and HPV16 (Seedorf *et al.,* 1985) are shown below the HPV5 sequences.

The HPV5b L1 variants are split into two subgroups differing by less than 2% of nucleotides. One corresponds to isolates from 5 of the 6 African patients, 2 French Caribbean patients of African origin, and 3 Colombian siblings. The other subgroup contains all variants with the 27-nucleotide insert and 1 (NAf3) lacking this insert. All the variants with the insert share two specific mutations at nucleotide positions 6350 and 6445 (Fig. 1). The HPV5c L1 variants include a French and a North American isolate very closely related but greatly divergent from the others (Fig. 4A).

The phylogenetic trees based on the deduced amino acid sequences also distributed the 12 L1 protein variants into three groups. However, striking differences were observed for HPV5a and b isolates when nucleotide and amino acid L1 trees were compared (Figs. 4A and 4B). The 10 HPV5a DNA variants and the 7 HPV5b DNA variants lacking the insert encode closely related L1 protein variants that are clustered into a single group and differ from each of the others by only one (protein variants a to f) or two (variant g) amino acid changes. The prototypical J1 and the E15 HPV5a isolates encode the same L1 protein variant as most non-European HPV5b isolates (Figs. 1, 4A, and 4B). The HPV5 variants with the L1 insert constitute a distinct group and encode 3 related L1 protein variants (h to j) that share two specific, nonconservative amino acid changes (Gln145 → Leu and Glu177 → Lys) (Fig. 3). In addition, the E5 and E16 isolates encoding the variant protein i differ by two amino acids within the insert. The most divergent HPV5c DNA variants encode two closely related protein variants (k, l) characterized by four specific amino acid changes (Fig. 3).

The nucleotide sequence of the entire E6 ORF (471 nucleotides) was established previously for 10 of the 33 isolates of this study (Deau *et al.,* 1993). The trees constructed from these sequences distributed the variants into three groups corresponding to those found for the L1 DNA variants, with more than 4.5% nonidentical nucleotides. (Figs. 4A and 4C). However, by contrast with the L1 variants, the HPV5 E6 DNA and protein variants shared the same phylogenetic grouping (Figs. 4C and 4D).

DISCUSSION

The analysis of the variability of a 419-bp sequence within the 5′ half of HPV5 L1 ORF has disclosed an intratypic diversity as high as that of the E6 ORF and the 3′ region of the LCR, which had previously allowed us to distinguish three subtypes, HPV5 a, b, and c (Deau *et al.,* 1993). Fifty-four (12.9%) nucleotide positions were found variable and 23 HPV5 L1 DNA variants were identified among the 33 isolates studied, in addition to the Japanese variants J1 and J2 (Zachow *et al.,* 1987; Yabe *et al.,* 1991). The most divergent isolates (variants E10

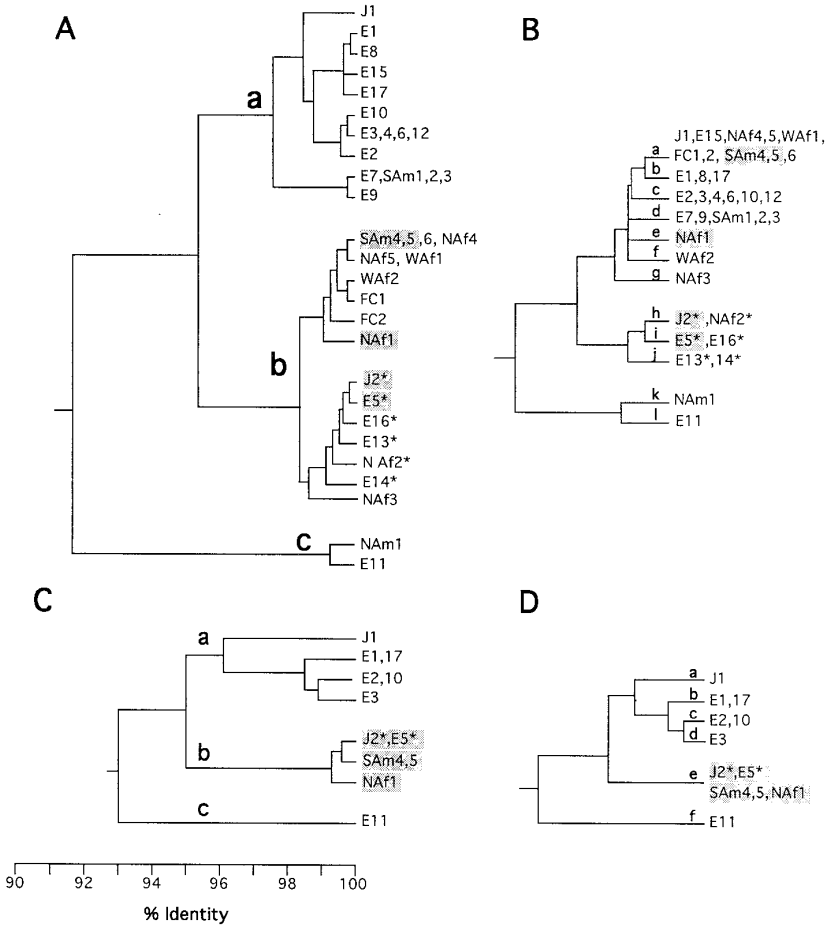

FIG. 4. Phylogenetic trees of HPV5 variants. Trees were generated by use of the UPGMA distance matrix method from the alignment of the sequences of a 419-nucleotide fragment in the L1 ORF of 35 HPV5 isolates (nucleotide positions 6144 to 6562), without taking into account the 27-nucleotide insertion (A), and the deduced 139-amino acid sequences (B), and from the nucleotide sequences of the E6 ORF (471 nucleotides) of 12 of the 35 HPV5 isolates (C), and the deduced 157-amino-acid sequences (D) (Deau *et al.,* 1993; Yabe *et al.,* 1991; Zachow *et al.,* 1987). The HPV47 nucleotide sequence (Kiyono *et al.,* 1990) was used as an outgroup. Isolates are designated as described in Table 1 and HPV5b isolates with the L1 insert are indicated by a star. HPV5b isolates included in both the L1 and the E6 trees are shadowed. The 12 L1 and the 6 E6 protein alleles are designated a to l and a to f, respectively.

and E11 found in the same French patient) show the same percentage (8.8%) of nonidentical nucleotides in the E6 and L1 ORFs. The HPV5 L1 DNA variants are stable and have a worldwide distribution since some of them were found in patients originating from different countries of the same continent or from different continents. The intratypic diversity of HPV5 is greater than that of HPV16. Only 42 variable nucleotide positions have been identified in the HPV16 L1 ORF (Yamada *et al.,* 1995), and variants were found to display a maximum pairwise difference of 1.2 and 1.3% for the L1 and E6 ORFs, respectively (Icenogle *et al.,* 1995; Yamada *et al.,* 1995). If one admits that all HPV genotypes share the same slow evolution rate, it would mean that HPV5 emerged before HPV16. Another hypothesis would be that HPV5 evolved more rapidly, under the influence of the mutagenic effect of ultraviolet B radiations. HPV5 replicates in EV patients in very flat, abnormally keratinized, skin lesions located in sun-exposed areas (Orth, 1987). The relatively low frequency (39%) of C → T or T → C transitions and the absence of CC → TT double mutations, pathognomonic of UVB-induced mutations (Brash *et al.,* 1991), do not support a major role of ultraviolet radiations in the evolution of HPV5.

The phylogenetic grouping of HPV5 L1 DNA variants fits with their geographic origin, resulting in the identification of three lineages, a European lineage including the majority of the variants of HPV5a subtype, an African lineage comprising the majority of HPV5b isolates, and a third lineage, probably of European origin, corresponding to the HPV5c subtype. The phylogenetic analysis of genital oncogenic HPV types has also disclosed the existence of African, European, Asian, and Asian-American lineages of this virus (Bernard *et al.,* 1994; Ho *et al.,* 1993; Ong *et al.,* 1993). In view of the greater genetic heterogeneity of African HPV16 variants, it has been postulated that these viruses emerged in Africa, at least 200,000 years ago, and coevolved with the major human

races (Bernard *et al.,* 1994). In contrast to HPV16, the most divergent HPV5 variants (E10 and E11) are of European origin, which could suggest that HPV5 emerged outside Africa, in an ancestor of *Homo sapiens* who migrated to Europe and Asia (Wanpo *et al.,* 1995).

The percentage (72.5%) of silent nucleotide substitutions observed within the L1 ORF is higher than that (59.3%) previously found for the HPV5 E6 ORF (Deau *et al.,* 1993). It is likely that this reflects the existence of negative selection pressures to comply with the structural and functional constraints imposed by capsomer formation and capsid assembly. This could explain why HPV5 variants belonging to two distinct groups in the L1 nucleotide phylogenetic tree (subtype a and subtype b without the insert) encode a family of closely related L1 proteins, whereas the corresponding E6 DNA and protein variants remain distributed into two subtypes (Fig. 4). An example of a mutation most probably counterselected is the His202 → Asp change found in a cervical cancer specimen (Seedorf *et al.,* 1985), responsible for the inefficient self-assembly of the HPV16 L1 protein into a capsid (Kirnbauer *et al.,* 1993).

The analysis of the intratypic diversity of the HPV5 L1 ORF has further substantiated a feature characteristic of the evolution of the major L1 capsid protein of papillomaviruses, i.e., the existence of short regions allowing the accumulation of amino acid substitutions and insertions or deletions. Fifteen of the 16 amino acid substitutions observed are localized into two short sequences (amino acids 135 to 146 and 173 to 182). In certain HPV5b variants originating from North Africa and Europe, one of these sequences contains a 9-amino-acid insertion similar to that previously described in a Japanese variant (Yabe *et al.,* 1991). We postulate that this insertion arose from the duplication of a 30-nucleotide sequence generated by a slipped mispairing mechanism (Roth *et al.,* 1985; Streisinger *et al.,* 1966). It is tempting to assume that the 3-nucleotide deletion and the 3 amino acid changes within the insert shared by all variants, as well as 2 specific amino acid changes downstream of the insert, were required to stabilize the duplication. Three of these 5 amino acid changes affected a charged residue, namely the deletion of an Arg residue and an Asp → Gly change within the insert and a Glu → Lys change located in the second variable region. That the duplication generated a specific selective pressure on the evolution of the capsid protein is shown by the fact that 2 HPV5b L1 DNA variants with the insert and 2 without the insert, encoding the same E6 protein, belong to two groups in the L1 amino acid tree constructed without taking into account the insert (Fig. 4). The alignment of the protein L1 amino acid sequence of 19 of the phylogenetically related HPV types associated with EV has revealed the existence of at least seven hypervariable regions, localized in the HPV5 L1 protein at amino acid positions 51 to 62, 133 to 141, 173 to 183, 273 to 302, 358 to 375, 457 to 461, and 482 to 510 (Myers *et al.,* 1995). These regions are also the most variable when multiple papillomavirus phylogenetic groups are compared (Myers *et al.,* 1995). Insertions or deletions of 1 to 4 amino acids responsible for the different sizes of the L1 proteins of EV HPVs (ranging from 506 to 518 amino acids) are found within these sequences. The two regions of intratypic variability found in this study correspond to two of these regions, the plasticity of which seems compatible with the preservation of the functional integrity of the L1 protein.

The observed variability could also reflect an immunological selective pressure. Antibodies reacting with conformational or linear epitopes of the L1 protein are raised during natural HPV infections (Cheng *et al.,* 1995; Christensen *et al.,* 1992; Dillner *et al.,* 1995; Kirnbauer *et al.,* 1994). Immunization with L1 fusion proteins or with virus-like particles obtained by self-assembly of the L1 protein induces neutralizing antibodies (Breitburd *et al.,* 1995; Lin *et al.,* 1992; Roden *et al.,* 1994). So far, the existence of distinct serotypes has never been demonstrated for any of the known HPV genotypes (Galloway, 1994). By using virus-like particles in an ELISA test, it has been shown recently that the most divergent HPV16 L1 DNA variants, with L1 proteins differing by 7 amino acids, are serologically cross-reactive (Cheng *et al.,* 1995). Compared with that of HPV16, the intratypic variability of the HPV5 L1 protein involves a greater number of variable amino acids, a higher maximum pairwise divergence between variants (21 amino acids for E10 and E11 variants; M. Kawase *et al.,* unpublished results) and, even, the insertion of 9 amino acids in some of the variants. By using virus-like particles obtained by self-assembly of L1 proteins of different HPV5 variants, we are currently investigating whether such a variability interferes with capsid morphogenesis and stability and results in variations of the antigenic properties of viral particles.

ACKNOWLEDGMENTS

We thank F. Breitburd for stimulating discussions and critical reading of the manuscript and D. Senlecques for expert assistance in the preparation of the manuscript.

REFERENCES

Bernard, H. U., Chan, S.-Y., and Delius, H. (1994). Evolution of papillomaviruses. *Curr. Top. Microbiol. Immunol.* 186, 33–54.

Brash, D. E., Rudolph, J. A., Simon, J. A., Lin, A., McKenna, G. J., Baden, H. P., Halperin, A. J., and Pontén, J. (1991). A role for sunlight in skin cancer: UV-induced p53 mutations in squamous cell carcinoma. *Proc. Natl. Acad. Sci. USA* 88, 10124–10128.

Breitburd, F., Kirnbauer, R., Hubbert, N. L., Nonnenmacher, B., Trin-Dinh-Desmarquet, C., Orth, G., Schiller, J. T., and Lowy, D. R. (1995). Immunization with virus-like particles from cottontail rabbit papillomavirus (CRPV) can protect against experimental CRPV infection. *J. Virol.* 69, 3959–3963.

Chan, S.-Y., Ho, L., Ong, C.-K., Chow, V., Drescher, B., Dürst, M., ter Meulen, J., Villa, L., Luande, J., Mgaya, H. N., and Bernard, H.-U. (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.

Chan, S.-Y., Delius, H., Halpern, A. L., and Bernard, H.-U. (1995). Analysis of genomic sequences of 95 papillomavirus types: Uniting typing, phylogeny, and taxonomy. *J. Virol.* 69, 3074–3083.

Cheng, G., Icenogle, J. P., Kirnbauer, R., Hubbert, N. L., StLouis, M. E., Han, C. L., Svare, E. I., Kjaer, S. K., Lowy, D. R., and Schiller, J. T. (1995). Divergent human papillomavirus type 16 variants are serologically cross-reactive. *J. Infect. Dis.* 172, 1584–1587.

Christensen, N. D., Kreider, J. W., Shah, K. V., and Rando, R. F. (1992). Detection of human serum antibodies that neutralize infectious human papillomavirus type 11 virions. *J. Gen. Virol.* 73, 1261–1267.

Deau, M.-C., Favre, M., and Orth, G. (1991). Genetic heterogeneity among human papillomaviruses (HPVs) associated with epidermodysplasia verruciformis: Evidence for multiple allelic forms of HPV5 and HPV8 E6 genes. *Virology* 184, 492–503.

Deau, M.-C., Favre, M., Jablonska, S., Rueda, L.-A., and Orth, G. (1993). Genetic heterogeneity of oncogenic human papillomavirus type 5 (HPV5) and phylogeny of HPV5 variants associated with epidermodysplasia verruciformis. *J. Clin. Microbiol.* 31, 2918–2926.

de Villiers, E.-M. (1994). Human pathogenic papillomavirus types: An update. *Curr. Top. Microbiol. Immunol.* 186, 1–12.

Dillner, J., Wiklund, F., Lenner, P., Eklund, C., Fredriksson-Shanazarian, V., Schiller, J. T., Hibma, M., Hallmans, G., and Stendahl, U. (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.

Ellis, J. R. M., Keating, P. J., Baird, J., Hounsell, E. F., Renouf, D. V., Rowe, M., Hopkins, D., Duggan-Keen, M. F., Bartholomew, J. S., Young, L. S., and Stern, P. L. (1995). The association of an HPV16 oncogene variant with HLA-B7 has implications for vaccine design in cervical cancer. *Nature Med.* 1, 464–470.

Felsenstein, J. (1988). Phylogenies from molecular sequences: Inference and reliability. *Annu. Rev. Genet.* 22, 521–565.

Galloway, D. A. (1994). Human papillomavirus vaccines: A warty problem. *Infect. Agent Dis.* 3, 187–193.

Guilhou, J.-J., Malbos, S., Barnéon, S., Habib, A., Baldet, P., and Meynadier, J. (1980). Epidermodysplasie verruciforme (2 observations): Etude immunologique. *Ann. Dermatol. Venereol.* 107, 611–619.

Higgins, D. G., Bleasby, A. J., and Fuchs, R. (1992). CLUSTAL V: Improved software for multiple sequence alignment. *CABIOS* 8, 189–191.

Ho, L., Chan, S. Y., Burk, R. D., Das, B. C., Fujinaga, K., Icenogle, J. P., Kahn, T., Kiviat, N., Lancaster, W., Mavromara-Nazos, P., Labropoulou, V., Mitrani-Rosenbaum, S., Norrild, B., Pillai, M. R., Stoerker, J., Syrjänen, K., Syrjänen, S., Tay, S. K., Villa, L. L., Wheeler, C. M., Williamson, A. L., and Bernard, H. U. (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.

Icenogle, J. P., Sathya, P., Miller, D. L., Tucker, R. A., and Rawls, W. E. (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.

Icenogle, J. P., Clancy, K. A., and Lin, S. Y. (1995). Sequence variation in the capsid protein genes of human papillomavirus type 16 and type 31. *Virology* 214, 664–669.

Kirnbauer, R., Taub, J., Greenstone, H., Roden, R., Dürst, M., Gissmann, L., Lowy, D. R., and Schiller, J. T. (1993). Efficient self-assembly of human papillomavirus type 16 L1 and L1-L2 into virus-like particles. *J. Virol.* 67, 6929–6936.

Kirnbauer, R., Hubbert, N. L., Wheeler, C. M., Becker, T. M., Lowy, D. R., and Schiller, J. T. (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, 494–499.

Kiyono, T., Adachi, A., and Ishibashi, M. (1990). Genome organization and taxonomic position of human papillomavirus type 47 inferred from its DNA sequence. *Virology* 177, 401–405.

Kremsdorf, D., Favre, M., Jablonska, S., Obalek, S., Rueda, A. L., Lutzner, M. A., Blanchet-Bardon, C., Van Voorst Vader, P. C., and Orth, G. (1984). Molecular cloning and biochemical characterization of the genomes of nine newly recognized human papillomavirus types associated with epidermodysplasia verruciformis. *J. Virol.* 52, 1013–1018.

Kremsdorf, D., Jablonska, S., Favre, M., and Orth, G. (1982). Biochemical characterization of two types of human papillomaviruses associated with epidermodysplasia verruciformis. *J. Virol.* 43, 436–447.

Lin, Y.-L., Borenstein, L. A., Selvakumar, R., Ahmed, R., and Wettstein, F. O. (1992). Effective vaccination against papilloma development by immunization with L1 or L2 structural protein of cottontail rabbit papillomavirus. *Virology* 187, 612–619.

Longuet, M., Beaudenon, S., and Orth, G. (1996). Two novel genital human papillomavirus (HPV) types, HPV68 and HPV70, related to the potentially oncogenic HPV39. *J. Clin. Microbiol.* 34, 738–744.

Lutzner, M., Blanchet-Bardon, C., and Orth, G. (1984). Clinical observations, virologic studies, and treatment trials in patients with epidermodysplasia verruciformis, a disease induced by specific human papillomaviruses. *J. Invest. Dermatol.* 83, 18s–25s.

Majewski, S., and Jablonska, S. (1995). Epidermodysplasia verruciformis as a model of human papillomavirus-induced genetic cancer of the skin. *Arch. Dermatol.* 131, 1312–1318.

Myers, G., Delius, H., Icenogle, J., Bernard, H.-U., Baker, C., Halpern, A., and Wheeler, C. (1995). L1 protein alignments. *In* "Human Papillomaviruses 1995: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences," Part II, pp. II.L1.2–II.L1.20. Los Alamos Natl. Lab., Los Alamos, NM.

Ong, C. K., Chan, S. Y., Campo, M. S., Fujinaga, K., Mavromara-Nazos, P., Labropoulou, V., Pfister, H., Tay, S.-K., ter Meulen, J., Villa, L. L., and Bernard, H.-U. (1993). Evolution of human papillomavirus type-18: An ancient phylogenetic root in Africa and intratype diversity reflect coevolution with human ethnic groups. *J. Virol.* 67, 6424–6431.

Orth, G. (1987). Epidermodysplasia verruciformis. *In* "The Papillomaviruses" (P. M. Howley and N. P. Salzman, Eds.), Vol. 2, pp. 199–243. Plenum, New York.

Orth, G. (1994). Human papillomaviruses: general features. *In* "Encyclopedia of Virology" (R. G. Webster and A. Granhoff, Eds.), Vol. 2, pp. 1013–1021. Academic Press, London.

Orth, G., Jablonska, S., Jarzabek-Chorzelska, M., Rzesa, G., Obalek, S., Favre, M., and Croissant, O. (1979). Characteristics of the lesions and risk of malignant conversion as related to the type of the human papillomavirus involved in epidermodysplasia verruciformis. *Cancer Res.* 39, 1074–1082.

Pushko, P., Sasagawa, T., Cuzick, J., and Crawford, L. (1994). Sequence variation in the capsid protein genes of human papillomavirus type 16. *J. Gen. Virol.* 75, 911–916.

Risler, J. L., Delorme, M. O., Delacroix, H., and Henaut, A. (1988). Amino acid substitutions in structurally related proteins: A pattern recognition approach. Determination of a new and efficient scoring matrix. *J. Mol. Biol.* 204, 1019–1029.

Roden, R. B. S., Weissinger, E. M., Henderson, D. W., Booy, F., Kirnbauer, R., Mushinski, J. F., Lowy, D. R., and Schiller, J. T. (1994). Neutralization of bovine papillomavirus by antibodies to L1 and L2 capsid proteins. *J. Virol.* 68, 7570–7574.

Roth, D. B., Porter, T. N., and Wilson, J. H. (1985). Mechanisms of nonhomologous recombination in mammalian cells. *Mol. Cell. Biol.* 5, 2599–2607.

Rueda, L. A. (1993). Specific cytopathic effects of human papillomavirus in epidermodysplasia verruciformis. *In* "Dermatology. Progress and Perspectives: The Proceedings of the 18th World Congress of Dermatology" (W. H. C. Burgdorf and S. I. Katz, Eds.), pp. 226–229. Parthenon Pub., New York.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with

chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.

Seedorf, K., Krämmer, G., Dürst, M., Subai, S., and Röwekamp, W. G. (1985). Human papillomavirus type 16 DNA sequence. *Virology* 145, 181–185.

Smits, H. L., Traanberg, K. F., Krul, M. R. L., Prussia, P. R., Kuiken, C. L., Jebbink, M. F., Kleyne, J. A. F. W., van den Berg, R. H., Capone, B., Debruyn, A., and ter Schegget, J. (1994). Identification of a unique group of human papillomavirus type 16 sequence variants among clinical isolates from Barbados. *J. Gen. Virol.* 75, 2457–2462.

Sneath, P. H. A., and Sokal, R. R. (1973). "Numerical Taxonomy" (W. H. Freeman, Ed.). Freeman, San Francisco.

Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E., and Inouye, M. (1966). Frameshift mutations and the genetic code. *Cold Spring Harbor Symp. Quant. Biol.* 33, 77–84.

Van Ranst, M. A., Tachezy, R., Delius, H., and Burk, R. D. (1993). Taxonomy of the human papillomaviruses. *Papillomavirus Rep.* 4, 61–65.

Wanpo, H., Ciochon, R., Yumin, G., Larick, R., Qiren, F., Schwarcz, H., Yonge, C., de Vos, J., and Rink, W. (1995). Early *Homo* and associated artefacts from Asia. *Nature* 378, 275–278.

Wilbur, W. J., and Lipman, D. J. (1983). Rapid similarity searches of nucleic acid and protein data banks. *Proc. Natl. Acad. Sci. USA* 80, 726–730.

Yabe, Y., Sakai, A., Hitsumoto, T., Kato, H., and Ogura, H. (1991). A subtype of human papillomavirus 5 (HPV-5b) and its subgenomic segment amplified in a carcinoma: Nucleotide sequences and genomic organizations. *Virology* 183, 793–798.

Yaegashi, N., Xi, L., Batra, M., and Galloway, D. A. (1993). Sequence and antigenic diversity in two immunodominant regions of the L2 protein of human papillomavirus type 6 and type 16. *J. Infect. Dis.* 168, 743–747.

Yamada, T., Wheeler, C. M., Halpern, A. L., Stewart, A. C. M., Hildesheim, A., and Jenison, S. A. (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.

Zachow, K. R., Ostrow, R. S., and Faras, A. J. (1987). Nucleotide sequence and genome organization of human papillomavirus type 5. *Virology* 158, 251–254.

zur Hausen, H. (1995). Are human papillomavirus infections not necessary or sufficient causal factors for invasive cancer of the cervix? *Int. J. Cancer* 63, 315–316.

zur Hausen, H., and de Villiers, E. M. (1994). Human papillomaviruses. *Annu. Rev. Microbiol.* 48, 427–447.