

Sequence variations and viral genomic state of human papillomavirus type 16 in penile carcinomas from Ugandan patients

Maria Lina Tornesello, Franco M. Buonaguro, Antonella Meglio, Luigi Buonaguro, Elke Beth-Giraldo and Gaetano Giraldo

Division of Viral Oncology, National Cancer Institute, 'Fondazione Pascale', I-80131 Naples, Italy

Sequence variations in the E6/E7 (nt 34–880) and the L1 (nt 6584–7035) ORFs, and in the long control region (LCR) (nt 7289–93) of human papillomavirus type 16 (HPV-16) were analysed in five penile carcinoma biopsies obtained from Ugandan patients. Uganda is a country with a high incidence of genital cancers. All five isolates were classified as members of African-1 lineage (Af1) by phylogenetic analysis based on LCR sequences. The E6 gene phylogenetic analysis, however, showed that four isolates fell into a new subclass designated Af1-u. This subclass, characterized by three point mutations located at the 5' end of the E6 gene with resulting changes in amino acids at positions 10 and 14, is distinguishable from the Af1 class by the absence of synonymous mutations at nt 286 and 289. The nonsynonymous substitution at nt 335 was present in three out of five samples. The E6 Af1

mutation pattern was present in only a single Ugandan HPV-16 isolate. Nucleotide sequence analysis of the E7 and L1 regions did not allow any Af1 subclass identification. The physical state of the viral DNA in these samples was characterized by PCR and Southern blot analysis. Oligonucleotides which enable amplification of the full length E2 region (nt 2734–3872) failed to amplify the target sequence in four out of five samples, suggesting disruption of the E2 ORF and integration of the HPV genome into the human DNA. Southern blot analysis confirmed the virus integration status. Our results contribute to the characterization of the HPV-16 'African lineages' with the identification of the Af1-u subclass; furthermore, this is also the first report showing that in male genital cancers HPV-16 is integrated into the human genome with disruption of the E2 ORF.

Introduction

In the late 1970s human papillomaviruses (HPVs) were proposed as the sexually transmitted aetiological agents of human lower-genital-tract malignancies (reviewed in IARC, 1995). Of the more than 70 different HPV types which have been hitherto identified, only a small subset has been associated with anogenital cancers (de Villiers, 1994; Van Ranst *et al.*, 1994). Several studies, performed worldwide, reported that 'high risk' HPV types 16 and 18 are present in more than 90%

of cervical cancer and in more than 48% of penile carcinoma biopsies (Dürst *et al.*, 1983; McCance *et al.*, 1986; zur Hausen, 1991; Tornesello *et al.*, 1992; Wiener *et al.*, 1992).

High-risk-HPV transforming activity has been attributed mainly to E6 and E7 nucleotide sequences and/or to their expression levels, as these sequences are consistently detectable in cervical cancer biopsies/cell lines (Schwarz *et al.*, 1985) and are sufficient to immortalize primary human cells. The intrinsic binding affinity of HPV-16 E6 and E7 to the cellular tumour suppressors p53 and pRb, and the modulation of their expression by the LCR as well as by the P₉₇ promoter, derepressed by the disruption of E2 during HPV integration into the cellular genome, have been correlated with their oncogenic properties in the pathogenesis of neoplasia of the female lower genital tract (Dyson *et al.*, 1989; Werness *et al.*, 1990; Scheffner *et al.*, 1993, 1995; Lambert & Howley, 1988; Lamberti *et al.*, 1990; Schiller *et al.*, 1989).

Author for correspondence: Gaetano Giraldo.
Fax +39 81 545 1276. e-mail IRCCS@UNINA.IT

The GenBank accession numbers for the nucleotide sequences reported in this paper are AF003013–AF003019 (E6 sequences), AF003020–AF003026 (E7 sequences), AF003027–AF003031 (L1 sequences) and AF003032–AF003036 (LCR sequences).

The relevance of specific amino acid substitutions within E6 and E7 proteins for the binding to and the E6-dependent degradation rate of oncosuppressor products has been shown by *in vitro* studies and their role in the HPV-transforming activity has been proposed (Crook *et al.*, 1991; Sang & Barbosa, 1992). There have been only limited studies, however, on the presence and the type of E6 variants in nature and the prevalence of naturally occurring intragenotypic high-risk HPV variants as well as their geographical clustering and oncogenicity (Yamada *et al.*, 1995).

Studies and phylogenetic characterization of the 364 bp long control region (LCR) fragment (nt 7478–7841) from HPV-16 variants from four continents allowed the identification of four major clusters: E (European), AA (Asian/American), Af1 (African-1) and Af2 (African-2) (Chan *et al.*, 1992; Ho *et al.*, 1993). The AA, Af1 and Af2 classes form together a coherent phylogenetic group and are referred to as Ax classes (Myers *et al.*, 1994, 1996). Eschle *et al.* (1992), analysing the nucleotide sequence of the HPV-16 E7 gene in 32 genital tumours collected from Tanzania and from Germany identified two major clusters: the first included all ten German and four Tanzanian HPV-16 isolates, without any nucleotide substitution compared to the prototype; and the second the remaining 18 Tanzanian HPV-16 isolates, which are characterized by two silent mutations at nt 789 (T to C) and 795 (T to G). Eight samples of the latter cluster show an additional nonsynonymous mutation at nt 647 (A to G). More recently, Yamada *et al.* (1995) reported a phylogenetic study based on sequence analysis of the E6, L1, L2 and LCR regions. The 29 selected HPV-16 isolates were obtained in the United States from cervical cell samples. Nucleotide variations within different genomic segments were phylogenetically compatible; furthermore phylogenetic analysis of the E6 gene allowed the classification of the HPV-16 variant G131, with the single transition A to G at nt 131 (Ellis *et al.*, 1995), as a distinct subclass of the HPV-16 E lineage. This subclass is neither detectable by LCR nor by L2/L1 phylogenetic analysis.

The purpose of the study reported here was to verify the variability of HPV-16 isolates in a further sub-Saharan central African country, namely Uganda, which has a high incidence of genital cancers (Dodge *et al.*, 1973). Sample biopsies were collected from patients with penile cancers and cell lines have been established (Gentile *et al.*, 1987) in order to identify HPV-16 variants from male genital cancers and to determine the molecular status of HPV-16 in these neoplasms.

Methods

■ **Source of HPV-16 Ugandan isolates.** Biopsies of histologically confirmed penile squamous-cell carcinomas were obtained from 13 Ugandan patients attending the New Mulago Hospital-Makerere University in Kampala (Uganda) from 1972 to 1980 and cryo-preserved in liquid nitrogen at -192°C until further use for molecular hybridization analysis or establishment of long-term tumour-cell lines such as those (PCA-5 and PCA-23) described by Gentile *et al.* (1987).

CaSki and SiHa human carcinoma cell lines were obtained from the ATCC (Rockville, Md., USA).

■ **DNA isolation and Southern blot analysis.** DNA from frozen biopsy samples and cell lines was digested by proteinase K treatment and extracted once with phenol and once with phenol-chloroform-isoamyl alcohol (1:1:40), followed by ethanol precipitation. HPV-16 molecular status was determined by Southern blot experiments. Aliquots (10 μg) of genomic DNA were cleaved with the appropriate restriction enzyme, DNA fragments were separated on a 0.7% agarose gel and transferred onto Hybond-N nylon membranes (Amersham). Following pre-hybridization at 65°C for 1 h, filters were hybridized with a ^{32}P -labelled HPV-16 probe (Church & Gilbert, 1984).

■ **PCR amplification.** DNA samples, HPV-16 positive by Southern blot analysis (Tornesello *et al.*, 1992), were subjected to single-round PCR amplification of the E6/E7 region (nt 34–880), the E2 gene (nt 2733–3873) and the L1 (nt 6584–6631) and LCR (nt 7289–114) fragments. PCR oligonucleotide primer sequences are summarized in Table 1. HPV-16 DNA nucleotide positions are numbered according to the published sequence of the reference clone (Seedorf *et al.*, 1985), revised as described by Icenogle *et al.* (1991), Chan *et al.* (1992), Eschle *et al.* (1992) and Ho *et al.* (1993). The reaction mixtures (50 μl) for all sets of primers contained 200 ng of target DNA, 20 pmol of each primer, 50 mM KCl, 2.5 mM MgCl_2 , 100 mM Tris-HCl pH 8.3, 0.1% Triton X-100, 50 μM of each dNTP and 1.8 units of thermostable *Taq* DNA polymerase (Perkin-Elmer Cetus). DNA was amplified in a Perkin-Elmer GenAmp PCR System 9600 thermal cycler with the following steps: an initial 5 min denaturation at 94°C , followed by 30 cycles of 55°C for 45 s, 72°C for 60 s, 94°C for 15 s, and a final annealing at 55°C for 45 s with 5 min elongation at 72°C . PCR amplification products, extracted with phenol and chloroform-isoamyl alcohol and purified by precipitation with 10% polyethylene glycol (PEG 6000) in 1.25 M NaCl, were subjected to direct nucleotide sequencing and sequence analysis after cloning into *Sma*I-dephosphorylated pBluescript (Stratagene).

■ **DNA sequencing.** Recombinant plasmids containing HPV inserts were prepared from minicultures (5 ml Luria broth) by the Wizard Minipreps DNA purification system (Promega). Sequencing reactions were performed by the Sanger method using a Sequenase 2.0 kit (United States Biochemical) according to the manufacturer's instructions. Several primers were used: (a) the 17-mer universal (–20), (b) the 16-mer reverse M13 sequencing primers and (c) primers specific for HPV sequences, as described in Table 1. The primer set designated 16E6-1 (5' ATGCATAGTATATAGAGATGGGAAT 3') and 16E6-2 (5' ATGCATGAT-TACAGCTGGGTTTCTC 3') was previously described by Shibata *et al.* (1989), the E2 oligo primer set by Krajnović & Savić (1991) and the degenerated L1 primer pair by Manos *et al.* (1989). The remaining HPV-specific primers were chosen with the PrimerSelect program of the Lasergene software (DNASTAR).

Direct sequencing of the amplified and PEG-purified DNA was performed using a rapid method modified from Winship (1989). Briefly, DNA samples were denatured at 95°C in the presence of 10% DMSO, immediately cooled in liquid nitrogen and subsequently sequenced with the Sequenase protocol modified in the labelling step (3 min on ice). Amplification primers were also used in direct sequencing reactions. All samples were amplified and analysed in duplicate to identify point mutations possibly resulting from the PCR reaction. Sequences were analysed on a 6% polyacrylamide wedge sequencing gel.

■ **HPV phylogenetic analysis and distance measures.** Multiple sequence alignments and phylogenetic analysis were performed with the MegAlign program of the Lasergene software (DNASTAR) and with the TREECON software package (Van de Peer *et al.*, 1993), respectively.

Table 1. Primers used for PCR amplification and sequence analysis

ORF	Nucleotide position	Primer name*	Primer sequences	Application	Product length (nt)
E6	34	16-E613N	5' AACCGAAATCGGTTGAACCG	PCR and sequencing	847
	880	16-E713C	5' TGCAGGATCAGCCATGGTAGAT		
	253	16-E61N	5' ATGCATAGTATATAGAGATGGGAAT	PCR and sequencing	
	567	16-E62C	5' ATGCATGATTACAGCTGGGTTTCTC		
	289	16-E63N	5' TGTATGTGATAAATGTTTAAAGTTTATTC	Sequencing	
328	16E6-3C	5' GAATAAAACTTTAAACATTTATCACATACA	Sequencing		
E2	2734	16E2-1N	5' AGGACGAGGACAAGGAAAA	PCR	1139
	3872	16E2-2C	5' GGATGCAGTATCAAGATTTG		
L1†	6584	MY11-N	5' GCMCAGGGWCATAAAYAATGG	PCR and sequencing probe	452
	7035	MY09-C	5' CGTCCMARRGGAWACTGATC		
	6631	GP-N	5' CTGTGGTGWATACYACWCGCAGTAC		
LCR	7289	16LCR-1N	5' GCTTGTGTAACACTATTGTGTCA	PCR and sequencing	730
	114	16LCR-2C	5' GTCCAGAAACATTGCAGTTCT		
	7717	16LCR-3C	5' CAAGCCAAAAATATGTGCCTAAC		

*N, normal strand; C, complementary strand.

† Degenerated primers: M = A/C, R = A/G, W = A/T, Y = C/T.

Synonymous and nonsynonymous substitution rates were calculated on the basis of the method of Nei & Gojobori (1986) and Gojobori *et al.* (1990).

Results and Discussion

Data from nucleotide sequencing analysis of the E6, E7 and L1 genes and the LCR region of seven HPV isolates obtained from all five penile carcinoma (PC) samples and two HPV-16-positive cervical cell lines are summarized in Table 2. The analysed genomic regions of Ugandan HPV-16 isolates are distinct from the HPV-16 reference sequence (E class).

HPV-16 E6 sequence mutations

All five Ugandan isolates contained changes in the E6 gene at nt 132 (G to C) with substitution of Arg with Thr at position 10, and two mutations at nt 143 (C to G) and nt 145 (G to T) resulting in a change from Gln to Asp at position 14. Only sample PC-8, also containing mutations at nt 286 (T to A), 289 (A to G) and 335 [C to T (His to Tyr)], showed all six nucleotide changes consistently detected by Yamada *et al.* (1995) in the E6 gene of USA Af1 isolates. Two further samples contained the C to T transition at nt 335. Additional changes were observed at nt 442 (A to C, Gln to Asp), present also in SiHa cells, 196 (G to A), 350 [T to G (Leu to Val)] and 419 [T to G (Cys to

Table 2. Nucleotide sequence variations in HPV-16 E6, E7 and L1 ORFs and LCR of Ugandan penile carcinoma (PC) HPV-16 isolates and HPV-16-positive cervical cell lines

	E6	E7	L1	LCR
	11111122334445	66777	556666666666666667	7777777777777777
	03344988350143	44899	891122344555678888990	44445666777778888
	91235669503922	57905	616744138567920566695	26881689124682333238
			304936631876402314939	00489080382355368513
E (Ref.)	TAGCGGTACTATAA	AATCT	CTCACTAATCTAAGACTCCGG	TCAGGCCATATCCGGAATCA
Af1 (cons.)	--CGT-AGT-----	--C-G	TCA-GCGG-TAG-A-T--TAT	..-AA-A-a--TT-T--..
Af2 (cons.)	C-TGT-AGT-G---	-GC-G	TCACGCGGCT--CA-T-TTAT	..CAATa----TTATCG..
AA (cons.)	----T-AGTG---G	TCACGC-G-T--CATT-TTAT	..CAATA--CGTT-----
PC-4	--CGTA---G--C-	--CTG-A-T--TA.	-G-AA-A-A--TT-T---TC
PC-7	--CGT-----GC-	--C-G-A-T--TA.	-G-AA-A-A--TT-T---TC
PC-8	--CGT-AGT-----	--C-G	TCA-GCGG-----A-T--TAT	---AA--TA--TT---CTC
PC-15	--CGT---T-----	--C-G-A-T--TA.	-G-AA-A-A--TT-T---TC
PC-17	--CGT---T-----	--C-G-A-T--TA.	CG-AA-A-A--TT-T---TC
Af1-u (cons)	--CGT---t---c-	--C-G-A-T--T..	-G-AA-A-A--TT-T---TC
CaSki	-G-----G----	-------C--	----A-----
SiHa	-G-----G--C-	C----	----A-----

Gly)). The transversion T to G at nt 350 observed in PC-4, was also detected in the HPV-16 isolated from CaSki and SiHa cell lines, members of the E-lineage subclass designated G131 for the A to G transition at nt 131.

A point mutation at nt 419 present in sample PC-7 changes the cysteine to glycine at aa position 106, with loss of the corresponding Cys-x-x-Cys zinc-binding motif. Crook *et al.* (1991) have shown that deletion of amino acids 106–110 reduces to 12% and 10% E6 binding to and degradation of p53, respectively. The relevance of this mutation, which has never been described in HPV-16 natural variants, for E6 transforming activity remains to be investigated; only substitutions of Cys-66 and Cys-136 by glycine have been associated with a severe reduction in transforming activity (Kanda *et al.*, 1991).

It is peculiar that the only two nonsynonymous point mutations consistently present in all HPV-16 variants of the Af lineages, of both male and female genital lesions, are localized within the N-terminal coding region at aa positions 10 and 14, associated with changes of arginine to threonine and glutamine to aspartic acid, respectively. The arginine at position 10 is adjacent to a proline (aa 9) conserved throughout oncogenically associated HPV types; the glutamine at position 14 is conserved among all genital HPVs. The highly conserved residue changes at aa 10 and 14, possibly due to selective humoral as well as cellular immune pressure (Ellis *et al.*, 1995), could result in a different binding affinity to p53 and altered degradation rate of this oncogene. Such a possibility is suggested by a 180% higher binding affinity to p53 and a consequent reduction to 20% of the oncosuppressor degradation rate shown with E6 protein deleted of amino acids 9–13 (Crook *et al.*, 1991).

Detection of HPV-16 natural variants by PCR and nucleotide sequence analysis allowed the identification of subclasses and their geographical distribution pattern. However, the biological diversity of such variants, in particular their oncogenic properties, will possibly be elucidated only by epidemiological studies on the association between virus variants and disease progression rate and by *in vitro* transformation experiments.

HPV-16 E7 and L1 sequence mutations

Nucleotide sequencing analysis of the E7 gene consistently showed in all five PC samples two synonymous point mutations, at nt positions 789 (T to C) and 795 (T to G). Both mutations have been previously reported in HPV-16 variants, without any specific class association, from cervical carcinoma biopsies of non-European geographical regions: 5 out of 11 samples from Panama, Georgia and Alabama (Icenogle *et al.*, 1991); 18 out of 22 samples from Tanzania (Eschle *et al.*, 1992); and 8 out of 11 samples from Barbados (Smits *et al.*, 1994). The linkage of both mutations to specific mutation patterns in LCR, L1 and/or E5, and to the E6 pattern in the Ugandan samples,

suggests that the position 789 and 795 changes are distinctive of the Ax group.

An additional mutation at nt 790 (C to T), which changes arginine to cysteine at aa position 77, was present only in sample PC-4. Ugandan samples do not show the mutation at nt 647 (A to G) identified by Eschle *et al.* (1992) in 8 out of 18 HPV isolates from Tanzanian cervical carcinoma. These eight samples, on the basis of LCR sequence, have been subsequently classified as Af2 variants (Ho *et al.*, 1993). Therefore, the transition A to G at nt 647, which changes asparagine to serine, absent in all analysed Ugandan isolates and in the African samples classified as Af1 variants on the basis of the LCR sequence, could represent a distinctive feature of the Af2 lineage. SiHa cells show a single point mutation at position 645 which creates a codon for leucine at aa position 28 instead of a phenylalanine. These findings further show that the E7 gene is highly conserved in different HPV-16 isolates, with an intragenotypic divergence of < 1%, and that nucleotide changes within this ORF do not greatly affect the amino acid sequence.

The L1 region, strongly conserved among all genital HPVs, was amplified with the degenerated primers MY09/MY11 (Manos *et al.*, 1989). All five samples contained three changes at nt positions 6720 (G to A), 6853 (C to T) and 6969 (C to T), which identify Ax variants. The L1 gene of the PC-8 sample was completely sequenced and additional mutations were observed: nucleotide changes common to Ax isolates at nt 5863 (C to T), 5910 (T to C), 6164 (C to A) and 6246 (T to C); and changes that identify the Af classes, previously reported in samples from the USA and UK (Hsu *et al.*, 1993), at nt 6315 (A to C) and 6433 (A to G).

The transition T to C at nt 6246 is of particular relevance, resulting in substitution of histidine with asparagine at aa position 202. This amino acid change has in fact been shown to increase *in vitro* the assembly efficiency of L1 and L2 viral proteins during the production of virus-like particles in the baculovirus system (Kirnbauer *et al.*, 1993). This observation suggests that single nonsynonymous point mutations could alter the *in vivo* production of viral proteins, their conformational epitopes and the binding efficiency to neutralizing antibodies.

HPV-16 LCR sequence mutations

Sequencing analysis of the Ugandan HPV-16 LCR (nt 7289–93) showed seven nucleotide changes typical of the Af1 lineage identified by Ho *et al.* (1993) in HPV isolates from malignant and benign genital-tract lesions of female subjects recruited in South Africa, Tanzania, Sierra Leone and Senegal, and by Yamada *et al.* (1995) in the USA. All five penile carcinoma samples showed the common mutation at nt 7519 (G to A) reported in > 80% of the isolates sampled throughout the world, and mutations at nt 7488 (G to A), 7763 (C to T) and 7785 (C to T) which are distinctive of the Ax group. Four

Table 3. HPV-16 nucleotide sequence variability

(a) Comparison of nucleotide sequence variability of E6, E7 and L1 ORFs and LCR among Ugandan HPV-16 isolates

Region	No. of nucleotides sequenced	Variable positions ^{a1} (%)	Max. pairwise distance ^{b1} (%)	Synonymous mutations	Max. pairwise pS ^{c1} (%)	Max. pairwise pN ^{d1} (%)	ds/dn
E6	456	7 (1.53)	6 (1.31)	3	3.31	0.83	4.05
E7	303	1 (0.33)	1 (0.33)	0	0.00	0.43	
L1	450	0	0	0	0.00	0.00	
LCR	716	6 (0.83)	5 (0.69)				

(b) Comparison of nucleotide sequence variability of E6, E7 and L1 ORFs and LCR between Af1-u variants and E reference sequence

Region	No. of nucleotides sequenced	Variable positions ^{a2} (%)	Max. pairwise distance ^{b2} (%)	Synonymous mutations	Max. pairwise pS ^{c2} (%)	Max. pairwise pN ^{d2} (%)	ds/dn
E6	456	10 (2.19)	6 (1.31)	1	1.10	1.37	0.80
E7	303	3 (0.99)	3 (0.99)	2	3.23	0.43	7.65
L1	450	4 (0.88)	4 (0.88)	4	3.90	0.00	
LCR	716	10 (1.39)	10 (1.39)				

(c) Comparison of nucleotide sequence variability of E6, E7 and L1 ORFs and LCR between Af1-u variants and Af2 consensus sequence

Region	No. of nucleotides sequenced	Variable positions ^{a3} (%)	Max. pairwise distance ^{b3} (%)	Synonymous mutations	Max. pairwise pS ^{c3} (%)	Max. pairwise pN ^{d3} (%)	ds/dn
E6	456	9 (1.97)	9 (1.97)	5	5.52	1.10	5.17
E7	303	2 (0.66)	2 (0.66)	0	0.00	0.86	
L1	450	3 (0.66)	3 (0.66)	2	1.96	0.28	7.07
LCR	530	9 (1.69)	5 (0.94)				

a1–3 Number and percentage of positions showing variations among five Ugandan isolates (a1), among five Ugandan isolates and the E reference sequence (a2) and among Ugandan isolates and Af2 consensus sequence (a3).

b1–3 Number and percentage of positions with different bases between the most diverged isolate PC-4 and PC-8 (b1), between PC-4 and the E reference sequence (b2) and between PC-4 and Af2 consensus sequence (b3).

c1–3 Percentage divergence for synonymous changes between PC-4 and PC-8 (c1), between PC-4 and the E reference sequence (c2) and between PC-4 and Af2 consensus sequence (c3), calculated according to method 1 of Nei & Gojobori (1986) and Gojobori *et al.* (1990).

d1–3 Percentage divergence for nonsynonymous changes between PC-4 and PC-8 (d1), between PC-4 and the E reference sequence (d2) and between PC-4 and Af2 consensus sequence (d3), calculated according to method 1 of Nei & Gojobori (1986).

Table 4. Medians for pairwise comparisons of sequences showing inter- and intrasubtype similarity relationships between E6 sequences of Ugandan penile carcinoma (PC) HPV-16 isolates and E reference sequence

Matrices show the ratios of the synonymous substitution rates to nonsynonymous substitution rates (ds/dn values); the number of synonymous substitutions divided by potential synonymous substitutions (pS); and the number of nonsynonymous substitutions divided by potential nonsynonymous substitutions (pN).

Compared value	Samples	Intra-class						
		avg	E	PC-4	PC-7	PC-8	PC-15	PC-17
ds/dn	E	0.71		0.80	0.21	2.03	0.27	0.27
	PC-4	2.18			2.00	4.06	1.33	1.33
	PC-7	1.13				2.68	0.36	0.36
	PC-8	8.13					8.13	8.13
	PC-15							0
	PC-17							
pS (%)	E	0.66		1.11	< 0.003	2.22	< 0.003	< 0.003
	PC-4	1.65			1.10	3.30	1.10	1.10
	PC-7	0.73				2.20	< 0.003	< 0.003
	PC-8	2.14					2.14	2.14
	PC-15							0
	PC-17							
pN (%)	E	1.21		1.38	1.38	1.10	1.10	1.10
	PC-4	0.76			0.55	0.83	0.83	0.83
	PC-7	0.83				0.83	0.83	0.83
	PC-8	0.27					0.27	0.27
	PC-15							0
	PC-17							

out of five isolates contain mutations at nt 7833 (G to T) that are specific to Af1 and Af2 classes, and at nt 7713 (T to A) observed only in Af1. In addition, the C to G transversion was observed in four out of five samples at nt 7460, which falls within an E2-binding site. Furthermore, in all five samples two unreported point mutations were observed at nt 31 (C to T) and 83 (A to C), a nucleotide region not described in previous studies of African isolates. The presence of such mutations should be verified in HPV-16 variants, particularly Af1 and Af2 Tanzanian isolates as well as the Ax lineage, in order to possibly identify mutations within LCR sequences specific to virus subclasses.

Sequence comparison and phylogenetic analysis of Ugandan HPV-16

Several measures of variability for each of the sequenced regions are shown in Table 3. Synonymous and nonsynonymous substitution rates were calculated on the basis of the method of Nei & Gojobori (1986) and Gojobori *et al.* (1990). pS is the number of observed synonymous mutations divided by the number of all possible synonymous substitutions; pN is the number of observed divided by the number of possible nonsynonymous nucleotide changes. The synonymous substitution rate, ds , is the Jukes-Cantor trans-

formation of pS , $ds = -\frac{3}{4} \ln(1 - \frac{4}{3} pS)$; dn is calculated in an analogous manner from pN .

When rates of observed mutations between different HPV-16 genome regions are compared (Table 3), although no comparison achieves significance by Fisher's exact test, E6 is more variable than the other coding regions and is approximately as varied as the LCR. E7 and L1 are the most conserved regions. In order to verify the variability rate and to identify the co-factors involved the ratio of synonymous divided by nonsynonymous substitutions (ds/dn ratio) is evaluated. This value, which is directly related to the stability of the protein sequence, being the result of selective pressures against any amino acid change (for virus tropism) and towards epitope variations (for virus escape mutants), is lower between Ugandan isolates and the European (E) reference sequence than among Ugandan samples (Table 4). This result suggests that a greater rate of changes at the protein level are present between Ugandan and the E reference E6 variants than within Ugandan isolates (Table 4).

Pairwise comparison analysis was done by alignment of nucleotide sequences of the Ugandan HPV-16 variants with the HPV-16 E prototype, isolated from cervical carcinoma of a German woman (Seedorf *et al.*, 1985), and with HPV-16 isolates of the four classes reported by Yamada *et al.* (1995). Phylogenies were constructed by transforming the sequence

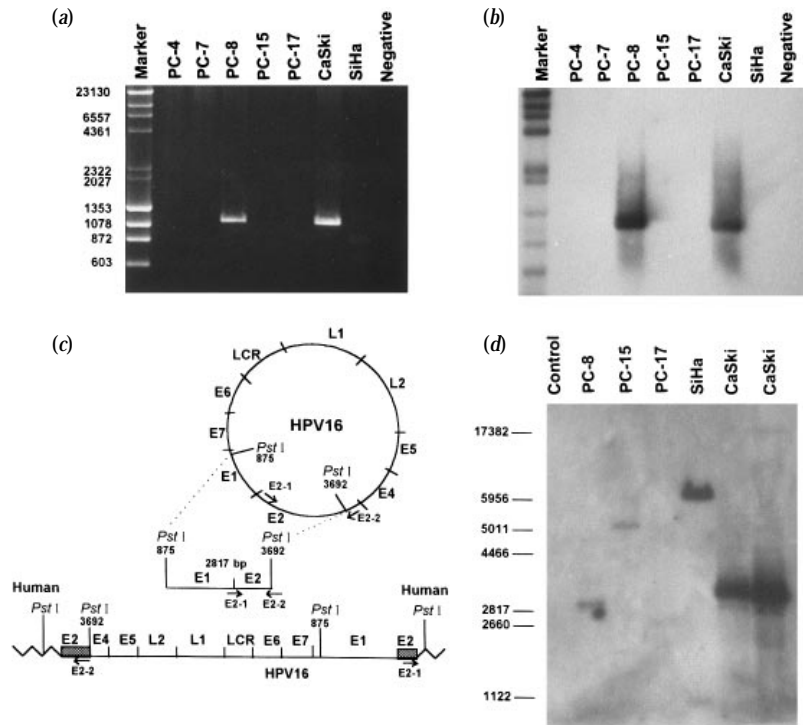


Fig. 2. (a) Fluorescence pattern of PCR-amplified E2 products on 1% agarose gel. The lane labelled 'Marker' contains molecular-size marker *HindIII*-digested lambda DNA and *HaeIII*-digested phi DNA. (b) Southern blot analysis of the agarose gel shown in (a) probed with a ^{32}P -labelled HPV-16 fragment corresponding to the E2 gene. (c) Schematic representation of episomal and integrated HPV-16 DNA structure; arrows indicate positions and orientations of primer pairs used to amplify the E2 gene. (d) Southern blot analysis of DNA fragments isolated from penile carcinoma samples and cervical cell lines digested with *PstI* and hybridized with a HPV-16 E2-specific probe. The pointers indicate the molecular mass of bands homologous to the HPV-16 E2 gene. The lane labelled 'Control' contains HPV-negative human placental DNA.

junction fragment is in fact rarely detectable because it shares < 160 bp with the labelled E2 probe and the virus is present at low copy number in the tumour cells. In contrast, samples PC-4, PC-7, PC-15 and PC-17, following digestion with *PstI*, released fragments of 4500, 3800, 5011 and 2660 bp, respectively, showing that all E2 sequences are integrated and rearranged within human DNA. Southern blot analysis of SiHa cell DNA, where the integration of HPV-16 with interruption of the E2 gene has been previously demonstrated, showed two fragments of 5956 bp and 1122 bp. The presence of a unit-length band (2800 bp) as well as smaller and bigger fragments in CaSki cell DNA indicates that the majority of HPV copies have a full-length, intact E2 gene and are flanked by a partial copy of HPV with E2 genes disrupted by integration events (Fig. 2).

This result shows that episomal HPV is rarely detected in neoplastic tissue and also suggests that in penile carcinoma the integration stage could play a relevant role in malignant progression, as recently reported in cervical intra-epithelial neoplasia (Daniel *et al.*, 1995). In this respect, *in vitro* experiments have shown that inactivation of the HPV-16 E2 gene results in higher levels of E6 and E7 expression and cellular growth advantage (Romanczuk & Howley, 1992; Jeon *et al.*, 1995).

Conclusions

This report describes the identification in Ugandan penile carcinomas of the E7 and LCR genetic pattern specific to the

HPV-16 Af1 class identified in cervical carcinoma from Africa (Tanzania) and America (Georgia, Alabama, Panama and Barbados). Furthermore, the genetic analysis of E6 gene allowed the identification in these lesions of an HPV-16 Af1-u subclass. Analysis of E6 (nt 34–560) thus seems to be more sensitive and accurate than LCR analysis for the identification and characterization of a greater number of subclasses. Characterization of this genomic region, moreover, could allow the identification of natural variants relevant for oncogenic studies, as well as for serological screening and design of vaccine strategies.

Furthermore, we have demonstrated that integration of HPV-16 into the human genome with interruption of the E2 gene is a frequent event in the pathogenesis of penile cancer, as described for cervical cancer.

This work was supported by grants from Ministero della Sanità (Ricerca Corrente 1996), IX Progetto AIDS-Istituto Superiore di Sanità (no. 9404-14) and the ICSC-World Laboratory (Project MCD-2/7). We thank Miss Chris Hattie McBride for polishing the English form of the manuscript.

References

- Chan, S., Ho, L., Ong, C., Chow, V., Drescher, B., Dürst, M., ter Meulen, J., Villa, L., Luande, J., Mgaya, H. N. & Bernard, H. (1992). Molecular variants of human papillomavirus type 16 from four continents suggest ancient pandemic spread of the virus and its coevolution with humankind. *Journal of Virology* **66**, 2057–2066.
- Church, G. H. & Gilbert, W. (1984). Genomic sequencing. *Proceedings of the National Academy of Sciences, USA* **81**, 1991–1995.

- Crook, T., Tidy, J. A. & Vousden, K. (1991).** Degradation of p53 can be targeted by HPV E6 sequences distinct from those required for p53 binding and transactivation. *Cell* **67**, 547–556.
- Daniel, B., Mukherjee, G., Seshadri, L., Vallikad, E. & Krishna, S. (1995).** Changes in the physical state and expression of human papillomavirus type 16 in the progression of cervical intraepithelial neoplasia lesions analysed by PCR. *Journal of General Virology* **76**, 2589–2593.
- de Villiers, E. (1994).** Human pathogenic papillomavirus types: an update. *Current Topics in Microbiology and Immunology* **186**, 1–12.
- Dodge, O. G., Owor, O. & Templeton, A. C. (1973).** Tumors of the male genitalia. In *Tumours in a Tropical Country*, pp. 132–144. Edited by A. C. Templeton. Berlin: Springer.
- Dürst, M., Gissman, L., Ikenberg, H. & zur Hausen, H. (1983).** A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions. *Proceedings of the National Academy of Sciences, USA* **80**, 3812–3815.
- Dyson, N., Howley, P. M., Munger, K. & Harlow, E. (1989).** The human papillomavirus-16 E7 oncoprotein is able to bind the retinoblastoma gene product. *Science* **243**, 934–937.
- 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. & Stern, P. L. (1995).** The association of an HPV16 oncogene with HBLA-B7 has implications for vaccine design in cervical cancer. *Nature Medicine* **1**, 464–469.
- Eschle, D., Dürst, M., ter Meulen, J., Luande, J., Eberhardt, H. C., Pawlita, M. & Gissmann, L. (1992).** Geographical dependence of sequence variation in the E7 gene of human papillomavirus type 16. *Journal of General Virology* **73**, 1829–1832.
- Gentile, G., Giraldo, G., Stabile, M., Beth-Giraldo, E., Lonardo, F., Kyalwazi, S. K., Perone, L. & Ventruto, V. (1987).** Cytogenetic study of a cell line of human penile cancer. *Annales de Genetique* **30**, 164–169.
- Gojobori, T., Moriyama, E. N. & Kimura, M. (1990).** Statistical methods for estimating sequence divergence. *Methods in Enzymology* **183**, 531–549.
- Ho, L., Chan, S.-Y., Burk, R. D. and 19 other authors (1993).** The genetic drift of human papillomavirus type 16 is a means of reconstructing prehistoric viral spread and movement of ancient human populations. *Journal of Virology* **67**, 6413–6423.
- Hsu, E. M., McNicol, P. J., Guijon, F. B. & Paraskevas, M. (1993).** Quantification of HPV-16 E6-E7 transcription in cervical intraepithelial neoplasia by reverse transcriptase polymerase chain reaction. *International Journal of Cancer* **55**, 397–401.
- IARC (1995).** *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans*, vol. 64, *Human Papillomaviruses*, pp. 47–52. Lyon: International Agency for Research on Cancer, World Health Organization.
- Igenogle, J. P., Sathya, P., Miller, D. L., Tucker, R. & 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.
- Jeon, S., Allen-Hoffmann, L. & Lambert, P. F. (1995).** Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. *Journal of Virology* **69**, 2989–2997.
- Kanda, T., Watanaba, S., Zanna, S., Sato, H., Furuno, A. & Yoshiike, K. (1991).** Human papillomavirus type 16 E6 proteins with glycine substitution for cysteine in the metal-binding motif. *Virology* **185**, 536–543.
- Kimura, M. (1980).** A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**, 111–120.
- Kirnbauer, R., Taub, J., Greenstone, H., Roden, R., Dürst, M., Gissmann, L., Lowy, D. R. & Schiller, J. T. (1993).** Efficient self-assembly of human papillomavirus type 16 L1 and L1-L2 into virus-like particles. *Journal of Virology* **67**, 6929–6936.
- Krajnović, M. & Savić, A. (1991).** Sequencing data on the long control region of human papillomavirus type 16. *Journal of General Virology* **72**, 2573–2576.
- Lambert, P. F. & Howley, P. M. (1988).** Bovine papillomavirus type 1 E1 replication-defective mutants are altered in their transcriptional regulation. *Journal of Virology* **62**, 4009–4015.
- Lamberti, C., Morrissey, L. C., Grossmann, S. R. & Androphy, E. J. (1990).** Transcriptional activation by the papillomavirus E6 zinc finger protein. *EMBO Journal* **9**, 1907–1913.
- McCance, D. J., Kalache, A., Ashdown, K., Andrate, L., Menzes, F., Smith, P. & Doll, R. (1986).** Human papillomavirus type 16 and 18 in carcinomas of the penis from Brazil. *International Journal of Cancer* **37**, 55–59.
- Manos, M. M., Ting, Y., Wright, D. K., Lewis, A. J., Broker, T. R. & Wolinsky, S. M. (1989).** The use of polymerase chain reaction amplification for the detection of genital human papillomaviruses. *Cancer Cells* **7**, 209–214.
- Myers, G., Bernard, H.-U., Delius, H., Favre, M., Icenogle, J., Van Rast, M. & Wheeler, C. (editors) (1994).** *Human Papillomaviruses 1994: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences*. Los Alamos, N.Mex., USA: Los Alamos National Laboratory.
- Myers, G., Bernard, H.-U., Delius, H., Favre, M., Icenogle, J., Van Rast, M. & Wheeler, C. (editors) (1996).** *Human Papillomaviruses 1996: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences*. Los Alamos, N.Mex., USA: Los Alamos National Laboratory.
- Nei, M. & Gojobori, T. (1986).** Simple method for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* **3**, 418–426.
- Romanczuk, H. & Howley, P. M. (1992).** Disruption of either the E1 or the E2 regulatory gene of human papillomavirus type 16 increases viral immortalization capacity. *Proceedings of the National Academy of Sciences, USA* **89**, 3159–3163.
- Saitou, N. & Nei, M. (1987).** The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406–425.
- Sang, B. & Barbosa, M. S. (1992).** Single amino acid substitutions in “low risk” human papillomavirus (HPV) type 6 E7 protein enhances features characteristic of the “high-risk” HPV E7 oncoproteins. *Proceedings of the National Academy of Sciences, USA* **89**, 8063–8067.
- Scheffner, M., Huibregtse, J. M., Vierstra, R. D. & Howley, P. M. (1993).** The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. *Cell* **75**, 495–505.
- Scheffner, M., Nuber, U. & Huibregtse, J. M. (1995).** Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. *Nature* **373**, 81–83.
- Schiller, J. T., Kleiner, E., Androphy, E. J., Lowy, D. R. & Pfister, H. (1989).** Identification of bovine papillomavirus E1 mutants with increased transforming and transcriptional activity. *Journal of Virology* **63**, 1775–1782.
- Schwarz, E., Freese, U. K., Gissmann, L., Mayer, W., Roggenbuck, B., Stremlau, A. & zur Hausen, H. (1985).** Structure and transcription of human papillomavirus sequences in cervical carcinoma cells. *Nature* **314**, 111–114.

- Seedorf, K., Krammer, G., Dürst, M., Suhai, S. & Rowekamp, W. G. (1985). Human papillomavirus type 16 DNA sequences. *Virology* **145**, 181–185.
- Shibata, D. K., Arnheim, N. & Martin, W. J. (1988). Detection of human papillomavirus in paraffin-embedded tissue using the polymerase chain reaction. *Journal of Experimental Medicine* **167**, 225–230.
- 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., de Bruyn, A. & ter Schegget, J. (1994). Identification of a unique group of human papillomavirus type 16 sequence variants among clinical isolates from Barbados. *Journal of General Virology* **75**, 2457–2462.
- Tornesello, M. L., Buonaguro, F. M., Beth-Giraldo, E., Kyalwazi, S. K. & Giraldo, G. (1992). Human papillomavirus (HPV) DNA in penile carcinomas and in two cell lines from high-incidence areas for genital cancers in Africa. *International Journal of Cancer* **51**, 587–592.
- Van de Peer, Y. & de Wachter, R. (1993). TREECON: a software package for the construction and drawing of evolutionary trees. *Computer Applications in the Biosciences* **9**, 177–182.
- Van Ranst, M., Tachezy, R. & Burk, R. D. (1994). Human papillomavirus nucleotide sequences: what's in stock? *Papillomavirus Report* **5**, 65–75.
- Werness, B. A., Levine, A. J. & Howley, P. M. (1990). Association of human papillomavirus type 16 and 18 E6 proteins with p53. *Science* **248**, 76–79.
- Wiener, J. S., Effert, P. J., Humphrey, P. A., Yu, L., Liu, E. T. & Walther, P. J. (1992). Prevalence of human papillomavirus types 16 and 18 in squamous-cell carcinoma of the penis: a retrospective analysis of primary and metastatic lesions by differential polymerase chain reaction. *International Journal of Cancer* **50**, 694–701.
- Winship, P. R. (1989). An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. *Nucleic Acids Research* **17**, 1266.
- Yamada, T., Wheeler, C. M., Halpern, A. L., Stewart, A.-C. M., Hildesheim, A. & 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. *Journal of Virology* **69**, 7743–7753.
- zur Hausen, H. (1991). Viruses in human cancers. *Science* **254**, 1167–1173.

Received 10 March 1997; Accepted 8 April 1997