Identification of a unique group of human papillomavirus type 16 sequence variants among clinical isolates from Barbados

Henk L. Smits,1*K Kristine F. Traanberg,1 Mark R. L. Krul,2 Patsy R. Prussia,3 Carla L. Kuiken,1 Maarten F. Jebbink,1 John A. F. W. Kleyne,2 Rocco H. van den Berg,1 Bruno Capone,1 Anna de Bruyn1 and Jan ter Schegget1

1Human DNA Virus Laboratory, Department of Virology, Academic Medical Centre, University of Amsterdam, Amsterdam, 2Laboratory of Virology, National Institute of Public Health and Environmental Protection, Bilthoven, The Netherlands and 3Department of Pathology and Surgery, University of West Indies, Cave Hill Campus, Barbados

The naturally occurring sequence variation of human papillomavirus type 16 (HPV-16) was analysed by direct sequence analysis of the PCR products of the long control region (LCR), the E5 and E7 open reading frames (ORFs), a segment of the L2 ORF overlapping the early viral poly(A) signal and a small segment of the L1 ORF or clinical isolates from Barbados and The Netherlands. Despite the widely different geographical and ethnic origin of the two groups of specimens, sequence analysis revealed relatively few mutational differences. Analysis of the LCR and the E5 ORF appeared to be the minimum requirement for the correct positioning of these variants in the HPV-16 phylogenetic tree. Most of the Barbadian variants appeared to be located at a unique position in the HPV-16 phylogenetic tree, at the internal branch close to the point where the European and Asian branches diverge. In contrast, most of the Dutch samples were located on the European branch.

Epidemiological and biological evidence indicates a causative role for certain human papillomaviruses (HPVs) in cervical carcinogenesis (Muñoz et al., 1988; zur Hausen, 1991). These HPVs include types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56 and 58, of which type 16 is found in more than 40% of all cervical carcinomas (de Villiers, 1989; Koutsoky et al., 1988; Resnick et al., 1990; Vousden, 1989; Wilczynski et al., 1988; Lórrincz et al., 1992). Sequence variation among different HPV isolates provides an interesting tool to study the evolution of these viruses and may also be used for molecular epidemiological studies. The sequence variation occurring in different HPV-16 isolates has been reported previously (Ho et al., 1991, 1993; Icenogle et al., 1991; Chan et al., 1992; Eschle et al., 1992). These studies have shown that the mutation rate of this virus is low and have indicated that the virus coevolved with human races. Similar phylogenetic trees could be constructed for sequence information obtained for the long control region (LCR; nucleotides 7478 to 7841) and for the E5 open reading frame (ORF; nucleotides 3847 to 4258) of isolates derived from Brazil, Germany, Singapore and Tanzania (Chan et al., 1992). Recently a phylogenetic tree was constructed for sequence variation observed within the 364 bp segment of the LCR of HPV-16 isolates of a large number of geographical regions and ethnic groups (Ho et al., 1993). Despite the large number of isolates analysed, remarkably little variation was observed. The phylogenetic tree revealed an internal branch that diverges at one end into two African and one Asian–American branch and at the other end into European and Eurasian branches. HPV-16 sequence variation has also been reported for parts of the early E7 ORF (nucleotides 562 to 855) and the late L1 ORF (nucleotides 6244 to 6656) of genomes present in patients' samples collected from North America (Icenogel et al., 1991) and of the E7 ORF (nucleotides 500 to 964) of samples from Germany and Tanzania (Eschle et al., 1992).

We have analysed the variation of HPV-16 isolates derived from The Netherlands and Barbados. Surprisingly, despite the two widely different geographical locations, very little sequence variation was observed when we analysed the region of the LCR (7478 to 7841) also studied by Ho et al. (1991, 1993) and Chan et al. (1992). Between nucleotides 7478 and 7841, five of the 19
isolates were identical to the prototype (Seedorf et al., 1985) and the remaining 14 isolates differed at only one position (nucleotide 7519). These two sequence variants of the HPV-16 LCR were previously detected in almost all of 25 different ethnic groups and geographical locations (Ho et al., 1993) and these two variants were found to be, worldwide, the most frequent HPV-16 sequence variants. However, sequence analysis of the complete LCR (7171 to 7851), the E5 (3836 to 4082) and E7 (543 to 863) ORFs, a segment (4023 to 4421) of the L2 ORF overlapping the early poly(A) signal and part of the HPV-16 genome will enable the detection of additional unique sequence variants with distinct geographical distribution and/or ethnic origin.

From the Dutch population, five cervical lesions [Dc141, cervical intraepithelial neoplasia (CIN) phase I; Dc255, CIN III; Dc212, CIN II; Dc207, CIN I; Dc269, CIN III] and three cervical carcinomas (Dt4, Dt24 and Dt42) were selected, based on their HPV-16-positive status, from the study groups presented by Cornelissen et al. (1992) and Resnick et al. (1990). Another 11 HPV-16-positive cervical carcinomas were selected from samples from Barbados (Prusia et al., 1994). For sequence analysis, the different parts of the HPV-16 genomes present in paraflin-embedded material from patients was amplified by PCR for 40 cycles. To increase yield, amplification of the LCR was carried out in two overlapping segments of 603 bp (nucleotides 7152 to 7754) and 579 bp (7457 to 132) using primer combinations 1 and 7, and 2 and 4, respectively (Fig. 1). Also the region containing the E5 ORF and the early poly(A) signal was amplified in two overlapping segments of 282 bp (nucleotides 3818 to 4099) and 437 bp (4004 to 4440) using primer combinations 13 and 17 and 14 and 16, respectively. The E7 ORF was amplified by the primers 11 and 12 and a segment of the L1 ORF (6374-6595) was amplified using primers 18 and 19. For PCR analysis, DNA was extracted by Proteinase K treatment of SDS lysates of the cell lines and from deparaffinized thin tissue sections by the method described by Shibata et al. (1988). PCR products were purified by agarose gel electrophoresis and the appropriate bands were extracted from the gel by the silica–guanidinium isothiocyanate method (Boom et al., 1990), or by using a Costar Spin X centrifuge filter unit. Sequence analysis was done directly on the purified PCR product. Locations and sequences of the primers used for PCR and sequencing are given in Fig. 1. To monitor sequence artefacts produced by the Taq polymerase used in the PCR, most PCRs were performed in duplicate. To allow direct comparison with other studies reporting HPV-16 sequence variation we have followed the nucleotide numbering used by Chan et al. (1992), Ho et al. (1991), Icenogle et al. (1991) and Eschle et al. (1992).

Fig. 1. Location and nucleotide sequence of primers. Localization and nucleotide sequence of primers used in PCR and sequence analysis are indicated below a linear representation of the HPV-16 genome. Numbers of the sequences are according to Seedorf et al. (1985). Primers 11 and 12 include a 5' (18 bases) M13 sequence used for sequencing (not depicted in the sequence). Primer pairs 1 and 7, 2 and 4, 11 and 12, 13 and 17, 14 and 16, and 18 and 19 were used for PCR amplification. All but primer 4 were used for sequencing. The nucleotide sequences of the primers used were: primer 1, 5' TAAAGTTTGAGATCGTATCCGATGAGTCTTAACCTG 3' (7171); primer 2, 5' TCGGTTGCATGGTTTTGTGGC 3' (7457 to 7476); primer 3, 5' AGGTGTTAATTAAAAAGCG 3' (7749 to 7768); primer 4, 5' CTGGGTTCGTCTGTTGGTC 3' (132 to 113); primer 5, 5' TAACCCAAAATCCGTTGTGCAC 3' (7871 to 7851); primer 6, 5' CAATGTATGACTAACCTTTAC 3' (7823 to 7803); primer 7, 5' AAACTTTATGGAATATGTC 3' (7754 to 7735); primer 8, 5' AGTGATATGAATGTTGCACA 3' (7638 to 7619); primer 9, 5' TAAGTATTGTATGTATGTTG 3' (nucleotide positions 7152 to 7163); primer 10, 5' AAACCTTATGCCAAATATGC 3' (7519 to 7539); primer 11, 5' AGATCTACAGACAAAGCGTAC 3' (524 to 543); primer 12, 5' ACCTCGAGCTCAAGCAATGCG 3' (882 to 863); primer 13, 5' CTATACAGTGGCTTACTACG 3' (3818 to 3836); primer 14, 5' GCAGGCCTCTGGTTTTAGGTC 3' (4004 to 4023); primer 15, 5' GGTTGAGGAACTAAGTAC 3' (4135 to 4155); primer 16, 5' CTCGCCCTGTTACCGAC 3' (4440 to 4421); primer 17, 5' TTATGTATTTAAAAAGCG 3' (4099 to 4082); primer 18, 5' GTGTCTCAGAACCATAATGCGACG 3' (6347 to 6368); primer 19, 5' TGTGCGCCTGTGCTGCTTGGT 3' (6595 to 6576).

The results of the individual sequence variation of the 19 HPV-16-positive patient samples are presented in Fig. 2. Among these variants three major groups (I to III) could be distinguished that differed from each other in at least three positions. Variants belonging to group I showed
the closest homology to the HPV-16 prototype sequence (Seedorf et al., 1985). The Dutch and Barbadian variants in groups I and II most probably belong to the European branches of the HPV-16 phylogeny because, firstly, the German isolates analysed by Eschle et al. (1992) showed the same mutations present at positions 789 (T) and 795 (T) in the E7 ORF. Secondly, the nucleotide variation at certain positions, such as 7728 and 7840 in the LCR and 4077 in the E5 ORF, which are characteristic of some isolates belonging to the Eurasian lineage (variants similar to Sb7 and Sb1; Chan et al., 1992), was not observed. Finally, the only German sample (Gbl0) sequenced by Chan et al. (1992) for the LCR as well as the E5 ORF had a nucleotide pattern that was intermediate between groups I and II. In contrast, some of the Barbadian variants showed several mutations in common with the Tanzanian variants described by Chan et al. (1992) and Eschle et al. (1992). These are the mutations at positions 3868 (A), 3991 (T) and 4042 (T) in the E5 ORF of the variants of group IIIB and at positions 789 (C) and 795 (G) in the E7 ORF of the variants of groups IIIA and B. However, in the LCR at positions 7483, 7487, 7667, 7687, 7762, 7784 and 7832 these Barbadian variants clearly differed from many of the Tanzanian variants and showed the European type of variation (Chan et al., 1992).

To depict the relationship of the variants, phylogenies were constructed by transforming the sequence data into a distance matrix by the Kimura two parameter model (Kimura, 1980), followed by applying the neighbour-joining method (Saitou & Nei, 1987), which was executed by the NEIGHBOR program in the PHYLIP version 3.4 phylogeny inference package of Felsenstein (1985, 1992). The DRAWTREE program in this package was

---

**Table:**

<table>
<thead>
<tr>
<th></th>
<th>LCR</th>
<th>E7</th>
<th>E5</th>
<th>L2</th>
<th>L1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>9</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

**Fig. 2.** Grouping of HPV-16 sequence variants. Sequence variant groups I, II and IIIA and IIIB were categorized, based on their common sequence variation. Nucleotide changes relative to the prototype are boxed and in bold. None of the nucleotide changes in the LCR affect known functional domains. An A at position 616 results in a codon for serine and a T in a codon for threonine. C at position 3872 results in a codon for serine and an A in tyrosine. An A at 3979 results in a codon for isoleucine and a C for leucine. C at 3883 results in a codon for leucine and an A for methionine. C at 3979 results in a codon for leucine and a G for valine. An A at 4042 results in a codon for isoleucine and a G for valine. An A at 4364 results in a codon for glutamic acid and a T for aspartic acid. An A at 6433 results in a codon for threonine and a G for alanine. The other nucleotide changes within the ORFs were silent. N, not determined.
Fig. 3. Phylogeny of HPV-16 sequence variants. (a) Phylogenetic tree of HPV-16 sequence variants detected in The Netherlands (prefixed Dt and Dc) and Barbados (Bt) and of the prototype (Proto). The phylogeny was based on the sequence variation observed in the LCR and the E7, E5, L2 and L1 ORFs at the nucleotide positions indicated in Fig. 2. (b) Phylogenetic tree of HPV-16 sequence variants detected in Tanzania (Tb and Ts), Brazil (Bt), Singapore (Sb and Sv), Germany (Gb; Chan et al., 1992), The Netherlands and in Barbados and of the prototype. The phylogeny was based on sequence variation observed in part of the LCR (7478 to 7841) and the E5 ORF (3847 to 4258) only.
CONSENSE programs of the PHYLIP package gave identical clusters in 83 and 88% of the variants, respectively, indicating high reproducibility for these clusters. The distribution, over two clusters, of the samples collected in Barbados suggests that the introduction of HPV-16 variants into the Barbadian population occurred from geographically different areas. This is consistent with the multinational composition of the Barbadian population. Similarly, Dutch society has, over the centuries, assimilated immigrants from different parts of Europe as well as from the rest of the world. Detailed sequence analysis of HPV-16 variants from other European locations may show whether the Dutch variants belonging to the two distinct clusters have different European origins.

In order to determine more precisely the relationship of the HPV-16 variants detected in Barbados and in The Netherlands with variants detected in such widely different geographic areas as Germany, Tanzania, Brazil and Singapore, we attempted to place our samples in the UPGMA phylogenetic trees constructed for the E5 ORF and for the LCR of 22 HPV-16 isolates (Chan et al., 1992), but with conflicting results. For instance, the variants from group III could be placed in the LCR tree near the Singaporian variant Sb2 in the Eurasian branch but in the E5 tree these variants appeared to be distributed over the African branch containing Tb4, Sb16 and Svl at the opposite side of the tree (see Chan et al., 1992). This could indicate that the Barbadian samples belonging to group III are located at or near the centre of the tree, in between Sb2 and the branches containing Tb4, Sb16 and Svl, instead of at the external branches as would appear by considering either of the segments of the genome alone. We therefore constructed a phylogenetic tree using the sequence information obtained for a segment of both the LCR (7478 to 7841) and the E5 ORF (3847 to 4258) of the Dutch and Barbadian variants as well as of the Tanzanian variants (Tb1, Tb4, Tb13, Tb16 and Ts3), the Singaporian variants (Sb2, Sb4, Sb5, Sb7, Sb10, Sb13, Sb16, Sb17, Sb19, Sb21a and Svl), the Brazilian variants (Bb2, Bb4, Bb11) and the German variant Gb10, published by Chan et al. (1992). This tree is presented in Fig. 3(b). In this tree, which resembles the phylogenetic tree reported by Ho et al. (1993), a European branch accommodating all Dutch, the German, three of the Barbadian (Bt6, Bt11 and Bt23), two Singaporian (Sb4 and Sb5) and one Brazilian (Bb4) variant can be distinguished. A second distinct Asian branch (bootstrap value 92%) contains variants (Sb7, Sb10, Sb13, Sb17 and Sb19) from Singapore. Most of the Barbadian samples (Bt7, Bt8, Bt9, Bt10, Bt12, Bt15, Bt20 and Bt22) are located on the internal branch of the tree (bootstrap value 83%), which extends into the African branch (bootstrap value 86%) containing the African variants (Tb1, Tb4, Tb13, Tb16 and Ts3) collected in Tanzania, two variants (Bb2 and Bb211) collected in Brazil and two of the variants (Sb21a and Svl) collected in Singapore. The location of the Barbadian variants on the internal branch clearly demonstrates that sequence information from at least the LCR and the E5 ORF is required to construct a reliable phylogenetic tree. An alternative explanation could be that the samples collected in Barbados contain mixtures of European and African sequence variants that are preferentially amplified by the different primer pairs used to amplify the different regions of the genome; i.e. the African variants by the E5 primers and the European variants by the PCR primers used to amplify the LCR. However, we used six different primer pairs for the amplification of the different segments of the genome and in most of the segments, including the LCR, mutations that were specific for the Barbadian variants were detected. These are the nucleotide changes at positions 7231 and 7232 in the LCR, at 789 and 795 in E7 and at 4228 in E5. These nucleotide changes in the LCR are outside the segment sequenced by Chan et al. (1992). Secondly, given the very low number of mutations observed, we consider it unlikely that each of the primer pairs used for PCR amplification would preferentially amplify one of these highly homologous sequence variants.

As stated above, within the segment of the LCR used by Ho et al. (1993) to construct the phylogenetic tree, most of the variants detected in our samples were either the same as the prototype or mutated at position 7519 only. As a consequence these variants would be positioned together with the European variant G1 or G11 on the same branch (Ho et al., 1993). This result again indicates that the LCR does not contain sufficient sequence variation to construct a detailed and reliable phylogenetic tree for many isolates. In fact Ho et al. (1993) observed that variants G1 and G11 are the most frequent variants worldwide and are present in almost all ethnic groups and geographical locations. Our study supports their assumption that such variants could be further classified by mutations elsewhere in the genome. For some variants this could be E5, for others this could be a different segment of the genome. We expect that further study of the variation of different parts of the genome of a larger number of isolates may reveal additional types of sequence variants with unique geographical distribution and/or ethnic origin and may show the consecutive steps by which HPV-16 variants have evolved during its coevolution with mankind.

At present we can only speculate about the ancestral, geographical and ethnic origin of the variants detected in Barbados. In view of the African origin of a substantial part of the Barbadian population, an African origin for
these variants could be hypothesized, but as can be seen from the tree, the Barbadian samples are located at a substantial distance from the nearest African variant.

The existence of sequence variants could be used to perform particular types of molecular epidemiological studies such as transmission studies. Given the very low number of mutations found, such studies would however require sequence analysis of large portions of the genome.

Finally, we did not find a striking difference in the number of nucleotide variations in the LCR, nor other parts of the genome of samples obtained from patients with either a CIN I lesion, a CIN III lesion or a cervical carcinoma. This contrasts with the findings of Bavin et al. (1993), who found significantly more nucleotide variation in the LCR derived from a patient with CIN III compared to the LCR from a patient with no evidence of cervical disease.

The Dutch patient samples were kindly provided by Professor Dr F. B. Lammes of the Department of Gynaecology and Obstetrics and by Professor Dr F. ten Kate at the Department of Pathology of the Academic Medical Centre, Amsterdam. Wim van Est is acknowledged for excellent photography, Theo H.G. Dirksen for assistance in sectioning the paraffin blocks and Dr S. Valli for critical reading of the manuscript. This work was supported by grants from the Dutch Cancer Society (grants IKA 90-01 and IKA 91-10).

References


(Received 26 January 1994; Accepted 8 April 1994)