JOURNAL OF VIROLOGY, Nov. 1993, p. 6413–6423
0022-538X/93/116413-11$02.00/0
Copyright © 1993, American Society for Microbiology
Vol. 67, No. 11

The Genetic Drift of Human Papillomavirus Type 16 Is a Means of Reconstructing Prehistoric Viral Spread and the Movement of Ancient Human Populations

LISA HO,[1] SHIH-YEN CHAN,[1] ROBERT D. BURK,[2] B. C. DAS,[3] KEI FUJINAGA,[4] JOSEPH P. ICENOGLE,[5] TOMAS KAHN,[6] NANCY KIVIAT,[7] WAYNE LANCASTER,[8] PENELOPE MAVROMARA-NAZOS,[9] VASSILIKI LABROPOULOU,[9] STELLA MITRANI-ROSENBAUM,[10] BODIL NORRILD,[11] M. RADHAKRISHNA PILLAI,[12] JAY STOERKER,[13] KARI SYRJAENEN,[14] STINA SYRJAENEN,[14] SUN-KUIE TAY,[15] LUISA L. VILLA,[16] COSETTE M. WHEELER,[17] ANNA-LISE WILLIAMSON,[18] AND HANS-ULRICH BERNARD[1]*

Laboratory for Papillomavirus Biology, Institute of Molecular and Cell Biology, National University of Singapore, Singapore 0511, Singapore[1]; Albert Einstein College of Medicine, Bronx, New York 10461[2]; Institute of Cytology and Preventive Oncology, Maulana Azad Medical College Campus, Delhi 110002, India[3]; Cancer Research Institute, Sapporo Medical College, Chuo-ku, Sapporo 060, Japan[4]; Centers for Disease Control and Prevention, Atlanta, Georgia 30333[5]; Deutsches Krebsforschungszentrum, Heidelberg, Germany[6]; Harborview Medical Center, Seattle, Washington 98104[7]; School of Medicine, Wayne State University, Detroit, Michigan 48230[8]; Institut Pasteur Hellenique, Athens, Greece[9]; Hadassah University Hospital, Mount Scopus, Jerusalem, Israel[10]; Panum Institute, University of Copenhagen, 2200 Kobenhavn, Denmark[11]; Regional Cancer Centre, Trivandrum 695011, India[12]; Carolina Medical Center, Charlotte, North Carolina 28232[13]; Department of Pathology, University of Kuopio, Kuopio, Finland[14]; Singapore General Hospital, Singapore 0316, Singapore[15]; Ludwig Institute for Cancer Research, São Paulo 01509, Brazil[16]; Department of Cell Biology and the New Mexico Tumor Registry, University of New Mexico Cancer Center, Albuquerque, New Mexico 87131[17]; and Department of Medical Microbiology, University of Cape Town Medical School, Observatory, 7925 Cape Town, South Africa[18]

Received 6 June 1993/Accepted 5 August 1993

We have investigated the diversity of a hypervariable segment of the human papillomavirus type 16 (HPV-16) genome among 301 virus isolates that were collected from 25 different ethnic groups and geographic locations. Altogether, we distinguished 48 different variants that had diversified from one another along five phylogenetic branches. Variants from two of these branches were nearly completely confined to Africa. Variants from a third branch were the only variants identified in Europeans but occurred at lower frequency in all other ethnic groups. A fourth branch was specific for Japanese and Chinese isolates. A small fraction of all isolates from Asia and from indigenous as well as immigrant populations in the Americas formed a fifth branch. Important patterns of HPV-16 phylogeny suggested coevolution of the virus with people of the three major human races, namely, Africans, Caucasians, and East Asians. But several minor patterns are indicative of smaller bottlenecks of viral evolution and spread, which may correlate with the migration of ethnic groups in prehistoric times. The colonization of the Americas by Europeans and Africans is reflected in the composition of their HPV-16 variants. We discuss arguments that today's HPV-16 genomes represent a degree of diversity that evolved over a large time span, probably exceeding 200,000 years, from a precursor genome that may have originated in Africa. The identification of molecular variants is a powerful epidemiological and phylogenetic tool for revealing the ancient spread of papillomaviruses, whose trace through the world has not yet been completely lost.

Papillomaviruses are involved in the etiology of various forms of epithelial neoplasia in humans and other vertebrates. Over the last decade, they have attracted particular interest because of their association with genital cancer. They are considered to be the most prevalent group of tumor viruses in humans, since carcinoma of the cervix is the most common cancer in most countries of the Third World (25). Because of the importance of these viruses, we are investigating their genomic diversity by comparing different papillomavirus types and, within each type, independent viral isolates. We have found that sequence comparison is an efficient means for reconstructing the origin and spread of these viruses and that papillomaviruses are a very good model system for the study of DNA virus evolution and for demonstration of the power of retrospective molecular epidemiology.

All mammals and probably many other vertebrates carry species-specific papillomavirus types, and even within individual hosts like humans or certain ungulates, there exist a large number of different and little related virus types that give rise to different pathologies. Consequently, evolution can be studied in correlation with various biological variables, such as host species or, because of preferential tropism of most HPV types for a subset of epithelia, cell type. The relationship of different papillomaviruses can be quantified and phylogenetically evaluated, since their genomes code for proteins with several highly conserved domains that can be unambiguously aligned (4, 23).

On a more recent level of evolution, independent isolates of human papillomavirus type 16 (HPV-16) (5, 8, 10, 14, 17, 18) and probably all other HPV types (our unpublished observa-

* Corresponding author.

tions) can be distinguished through point mutations. These are frequent in the long control region (LCR), which has a higher degree of diversity similar to that of gene-flanking regions in other organisms and unlike that of most protein-coding regions. Sequence diversity between isolates can reach 5% for this region, while it does not exceed 2% within most genes. As an exception, the sequence of the E5 gene is less conserved than that of other genes (4, 5). Mutational events in the LCR or in genes are rare, because distinct mutational patterns of HPV-16 isolates from Brazil, Germany, Singapore, and Tanzania form geographic clusters, and all isolates, even from places remote from one another, are related in the form of an evolutionary tree (5).

We concluded that these variants contain a detailed molecular record of the evolution of HPV-16 and that ancient traces of the prehistoric spread of the virus around the world have not yet been completely lost but can be revealed by means of sequence identification. It is logistically possible to obtain a large number of these sequences, since Papanicolaou smears and biopsies from cervical cancers are sampled in countless hospitals throughout the world. On the technical side, the combination of the polymerase chain reaction (PCR) and DNA sequencing procedures permits access to hundreds of sequences in reasonable time.

Given these considerations, our research aims were to compare HPV-16 variants from as many distinct ethnic groups and geographic locations as possible. Here, we report an analysis of collections from 25 geographic regions and/or ethnic groups in Africa, Eurasia, and the Americas. We have put particular emphasis on the inclusion of multiple collections of samples from the principal large, ethnically distinct populations of the world. The data suggest that HPV-16 evolved along five major branches: two predominantly present in Africa, two found mainly in Asia, and one found mainly in Europe and India. Many, but not all, details of these branches reflect the ethnicity of the group in which the infected patient lived. There are deviations from this correlation in roughly 10% of all patients. We interpret some of these deviations as traces of ancient migrations of infected human groups, but others may also represent recent transmissions of the virus across ethnic boundaries or over large geographic distances. We discuss evidence that it is apparently possible to reconstruct the pandemic spread of the virus, and indirectly the migration of infected ethnic groups, for periods of several tens of thousands of years.

MATERIALS AND METHODS

Origin of clinical specimens. This study deals with 301 partial genomic isolates of HPV-16. Eighteen of these isolates were from male patients (H samples in the Singapore cohort [15]); all others were from female patients. About 80% of the latter were from cervical carcinomas and cervical intraepithelial neoplasia (various grades), and 20% were from cervical smears, most often from patients with cytological abnormalities or indications for papillomavirus infection during the colposcopic examination. We do not report details of the pathologies, as this aspect is not relevant to this report; the study did not address the question whether molecular variation may correlate with changes in biological functions and whether certain variants may be more frequent in patients with disease than in asymptomatic patients. Initial observations (mostly from the Singaporean and Tanzanian cohorts) did not suggest such correlations. Careful analysis, however, may reveal at least minor biological differences between variants. These may be linked to the mutations identified here but not necessarily due to these particular changes of the viral LCR.

This report deals with samples from 25 different ethnic groups or geographic locations that had been gathered by 19 research groups. All samples entered into the study had originally been obtained for purposes other than the research addressed here, and DNA had been extracted independently in each of the 19 cooperating laboratories. Subsequently, these DNA samples were collected in the Singapore laboratory for centralized analysis by PCR and DNA sequencing.

Important criteria for entering a cohort into this study were that it should be (i) located as remote as possible from all other cohorts or (ii) ethnolinguistically as different as possible from all other cohorts and (iii) that it should be ethnolinguistically more or less homogeneous. With this latter goal in mind, our study did not put emphasis on a systematic coverage of immigrant societies of the New World. We made an effort, however, to trace African and European HPV-16 variants in the United States and to shed some light on certain variants, particularly common in Brazil, which may originate from Amerindians. Aside from this, we incorporated samples from six indigenous populations of the Americas and one from Greenland. Samples from Singapore (an Asian immigrant society) were included because most HPV-16 variants from this country could be interpreted by comparison with variants from other countries, in particular India and Japan. Altogether, our samples originated from 5 locations in Africa, 3 in Europe, 5 in Asia, 11 in the Americas, and 1 in Greenland. It has so far been impossible to obtain samples from New Guinea, Melanesia, Australian aborigines, and all small ethnic groups of the world.

To establish a systematic nomenclature for all samples, the country of origin or the ethnic group is identified by one or two uppercase letters followed by a hyphen (see legends to Fig. 2 to 4). The subsequent number indicates the sequence of entrance into the study. The suffixes a, b, and c identify multiple variants that were found in one sample. The letters S (for cervical smear), V (for vulvar biopsy), and L (for lymph node) have been included only for the Singaporean and Tanzanian cohorts to indicate the histological origins of the samples. This was done to retain consistency of the nomenclature with that of previous publications (5, 14). All other samples from these two cohorts and from Germany are from biopsies and are numbered as published, but the previously used suffix b (for biopsy) has been omitted. The letters H (for husband) and W (for wife) (Singapore cohort only) indicate the source of the biopsy in an independently published study of the sexual transmission of HPV-16 (15).

The phylogenetic analysis identified five principal branches of intratype evolution of HPV-16 (see below). For these branches, we created a separate nomenclature relating to the continent of principal distribution, i.e., Af1, Af2, As, AA, and E, to simplify the understanding of these trees and the discussion of the composition of individual cohorts (see below and the legend to Fig. 1).

The origins of the samples were as follows. The origins of samples from Brazil (excluding Amazon), Germany, Singapore, and Tanzania were as published previously (5). All samples from South Africa were derived from patients at Groote Schuur Hospital, Cape Town; patients SA-2, SA-3, SA-6, and SA-8 were of African race, SA-9 and SA-16 were of Caucasian race, and SA-4, SA-5, SA-10, SA-13, and SA-14 were of mixed race. DNA was prepared from biopsies from cervical carcinomas. Samples from Zaire were from patients in Kinshasa, and DNA was prepared from cervicovaginal lavages. Sierra Leone samples were derived from patients from Free-

town; DNA was prepared from cervicovaginal lavages. Senegal samples were from prostitutes from various parts of West Africa but living in Dakar; DNA was prepared from cervical smears. All Finnish patients were from Kuopio, and DNA was prepared from cervical smears. All Greek patients were from Athens, and DNA was prepared from cervical intraepithelial neoplasia lesions. For Israel, the two Arab patients in the study were from Jerusalem, and DNA was extracted from cervical carcinomas. For India, the patients designated IND were from Delhi, and DNA was extracted from carcinomas of the cervix; the patients designated INT were from Trivandrum, and DNA was extracted from cervical smears. All Japanese patients were from Sapporo, and DNA was extracted from cervical carcinomas. All Eskimo patients were from Greenland, and DNA was extracted from sections of paraffin-embedded cervical biopsies. All samples from Pueblo Indians and Navajo Indians were derived from patients in New Mexico; DNA was extracted from cervical smears or from paraffin-embedded tissue blocks. Of the 10 U.S. samples from New York, N.Y., 4 were from black, 3 were from Caucasian, and 3 were from Hispanic patients; all samples were from cervical smears. Ten samples (cervical smears) were from patients in Charlotte, N.C. (five blacks and five Caucasians). Two samples (cervical smears) were from black patients from Detroit, Mich. All Peruvian patients were from Lima, and the DNA was extracted from cervical carcinomas. Among Amazonian Indians, patients AM-1, AM-2, and AM-3 were from the Tiriyo tribe, AM-4 and AM-5 were from the Waiapi tribe, and AM-6 was from the Munduruku tribe, and all lived in the eastern Amazon of Brazil; DNA was extracted from cervical smears. All Argentinian patients were of Amerindian ethnicity and lived in the Chaco region of northern Argentina; DNA was extracted from biopsies from cervical carcinomas.

PCR and DNA sequencing. Our detailed protocol has been described previously (5, 14). In short, the segment of interest was amplified with two PCR primers that flank a 364-bp segment of the LCR of HPV-16 (genomic positions 7478 to 7841) which coincides with the viral transcriptional enhancer. Each PCR was performed in duplicate, and each PCR product was cloned into pUC18 and sequenced. Mutations that originated during the PCR amplification were identified and eliminated from further consideration as described previously (5, 14).

Phylogeny construction and evaluation. The HPV-16 genome had originally been molecularly cloned from a German patient (7). We refer to the sequence of this particular genome as the HPV-16 reference clone. This designation replaces the term prototype that we used in previous publications. This latter term had been misunderstood as referring a more ancestral virus genome, when actually it was meant to indicate isolation priority.

We refer to HPV-16 isolates that had mutational differences from the HPV-16 reference clone as HPV-16 variants. All HPV-16 variants discussed in this report differed from the reference clone by point mutations. In addition to these point mutants, we have previously described three variants that had deletions and one variant that had an insertion. We did not include these variants in this study and did not phylogenetically evaluate them, since each was found only once and each was closely related to variants without insertions or deletions and was never found to give rise to progeny.

The HPV-16 reference clone and all HPV-16 variants with point mutations were phylogenetically analyzed by various algorithms in the phylogeny inference package PHYLIP 3.4 (9). This report discusses the consensus neighbor-joining tree (100 bootstrapped replicates). The essential parts of the topology of this tree were not significantly different from those calculated by parsimony and maximum-likelihood methods. Also, there was agreement between the major branches of this tree and those of a published tree based on the unweighted pair-group method with arithmetic average cluster analysis (5). The topology of the trees was robust as examined by bootstrap and resampling algorithms.

Nucleotide sequence accession numbers. The nucleotide sequence accession numbers of 34 HPV-16 variants have been published elsewhere (5). In addition, 14 variants were newly identified. Their GenBank accession numbers are as follows: AM-6, L22662; AN-10, L22674; AN-10a, L22673; AN-12, L22665; AP-4, L22669; IND-4, L22667; IND-5, L22670; IND-7, L22666; J-5, L22668; P-1, L22663; SA-9, L22661; SH-2a, L22771; SL-6a, L22664; and SN-4, L22672.

RESULTS

Genomic variability of HPV-16. The objective of our research was to establish in a worldwide survey a comprehensive catalog of all genomic variants of HPV-16. We have previously published sequences of 118 HPV-16 isolates from four cohorts of patients in Germany, Tanzania, Singapore, and Brazil (5). Here, we report an extension of this study, which now encompasses 301 isolates from 25 different patient cohorts. New cohorts were entered by the criterion that they should either live as remote as possible from any other cohort or be ethnolinguistically as distinct as possible from any of the other cohorts. This strategy seemed to hold the greatest promise for identifying novel variants; the venereal (and possibly also vertical) mode of transmission of HPV-16 should lead to some geographic specificity of HPV-16 variants, as most infectious events take place between people living within any particular area rather than between people from remote places, and the apparently low infectivity of papillomaviruses (15) should further reduce the efficiency of their spread.

With these considerations in mind, we complemented the single previously analyzed African cohort from Tanzania with samples from western Africa (Senegal and Sierra Leone), South Africa, and central Africa (Zaire). In Europe, we added to the German samples collections from Finland and Greece. In Asia, we obtained, in addition to the large collection from Singapore, samples from two regions of India and from Japan, and we obtained two samples from Arab patients. In the Americas, we sampled from six ethnic groups ranging from New Mexico to Argentina (Materials and Methods; Fig. 1). Samples from Eskimos in Greenland can be considered as representing a seventh American ethnic group, since Greenland was colonized from the American Arctic. We also sampled in Peru and in three cities of the United States (with particular emphasis on a record of the ethnicity of each patient), and we increased the original number of Brazilian patients.

Altogether, we identified 48 different variants of the HPV-16 LCR. They showed differences in 48 genomic positions and had altogether 51 point mutational differences due to three sites with alternative substitutions. In addition to the 44 of these 51 mutations that had been found and reported previously (Fig. 1 in reference 5), we identified three variants with deletions and one with an insertion. In addition to these mutations, we observed seven point mutations, at genomic positions 7494 (T to C), 7712 (T to G [only in IND-7]), 7757 (A to C), 7764 (A to G), 7768 (C to T), 7780 (C to T), and 7790 (C to T).

It came as a surprise that after more than doubling the original sample size (301 instead of 118) and greatly increasing

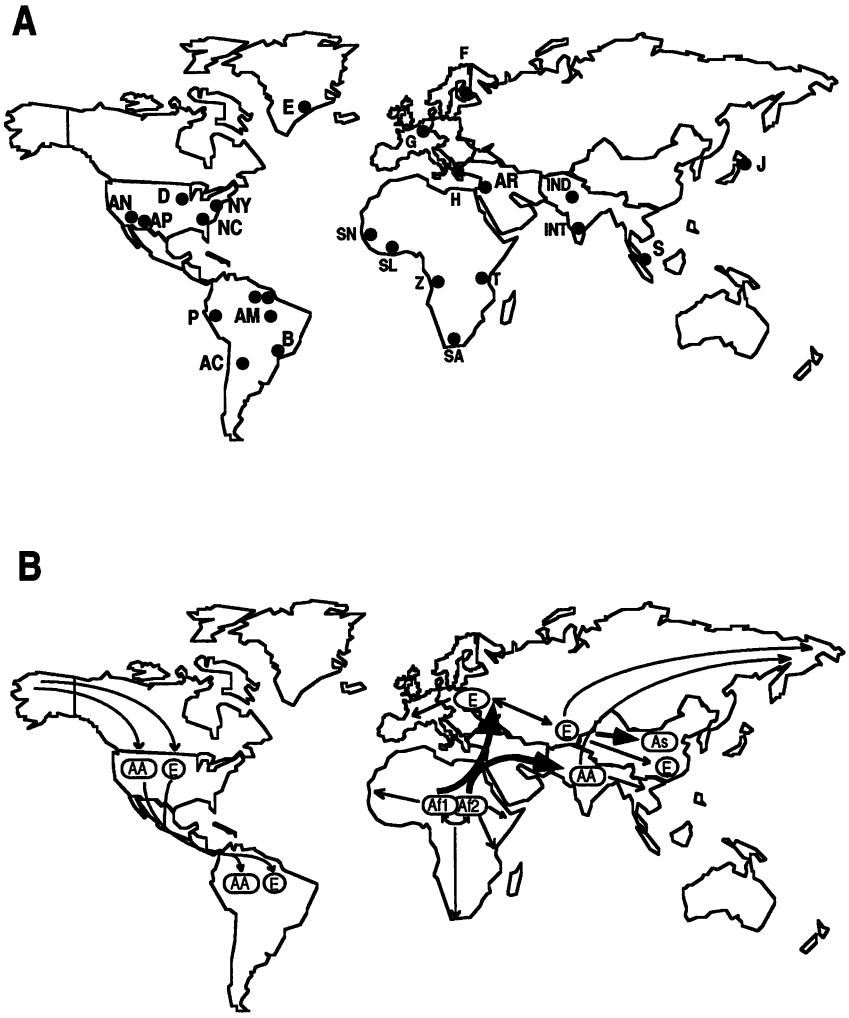


FIG. 1. Geographic origins of the samples analyzed in this study (A) and evolution and spread of HPV-16 variants in prehistoric times (B). Thin arrows represent spread without or with minor mutational changes; bold arrows represent mutational changes that gave rise to the five major branches of intratype divergence.

the number of cohorts (25 instead of 4), we did not identify many more new mutations. These mutations, or novel combinations of known mutations, led to identification of 14 new variants, but all of these new variants were closely related to the 38 that had been previously identified (see below). From these observations, we postulate that we have found all major branches of HPV-16 diversity. Of course, additional, remotely related variants may still exist, but they should be too rare to be frequently found in large populations or restricted to small isolated ethnic groups.

A phylogenetic tree of 48 HPV-16 LCR variants. Figure 2 shows a consensus neighbor-joining tree which was calculated from comparison of point mutational differences between the 48 variants. In all essential features, the tree is similar in topology to a published tree developed by the unweighted pair-group method with arithmetic average (5) as well as to trees derived from maximum-likelihood and maximum-parsimony algorithms (data not shown). A few previously described variants with deletions or insertions were not included in Fig. 2.

We have labelled the four major branches of this tree Af1, Af2, AA, and E to indicate that most variants that are assigned to either of these branches had been sampled in Africa (Af1 and Af2), Asia and the Americas (AA), or Europe (E). A fifth branch, As, emerges from a cluster of the E branch and contains variants that are frequent in East Asian populations.

All cohorts except those from Europe contain a mixture of variants that belong to several of the alternative branches, but in most cohorts, variants from only one branch predominate. Also, in none of these cohorts are there continuities between the predominant variants and the minor additions of variants from other branches. This phylogenetic discontinuity is interpreted as evidence that the HPV-16 variants of each branch evolved in some specific ethnic group or geographic location and spread subsequently to other ethnic groups or geographic locations.

Figures 3 to 5 represent all HPV-16 variants that we found in these cohorts. The first vertical column identifies each isolate by a code number, and multiple code numbers refer to the presence of these variants in several patients. A code number in parentheses refers to a variant that was identical in sequence and whose code had been entered in the phyloge-

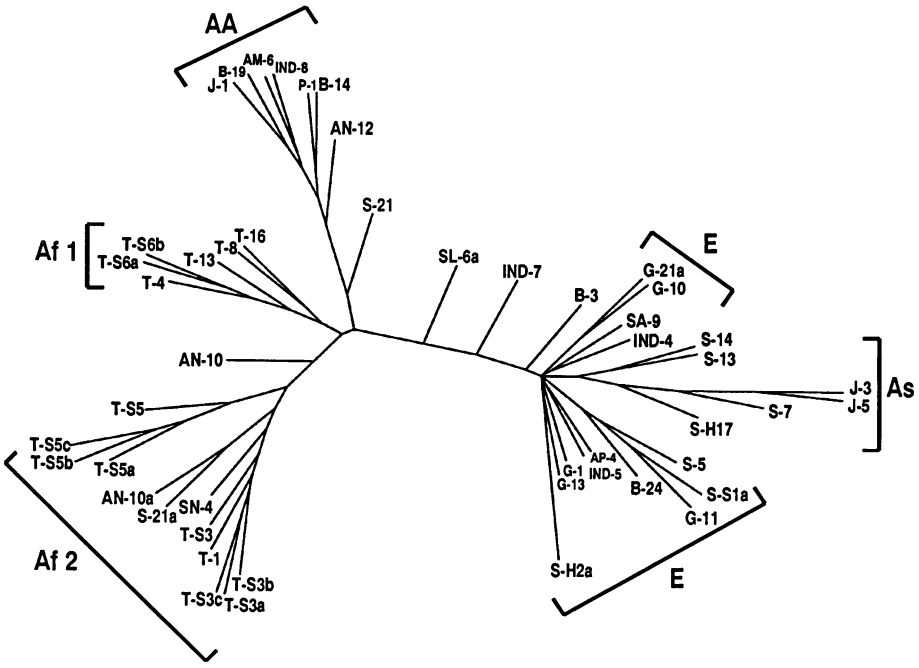


FIG. 2. Relationship between 48 genomic variants of HPV-16, all found among 301 isolates from 25 ethnic groups and geographic locations. The figure represents a consensus neighbor-joining tree (100 bootstrapped replicates) and is representative of trees computed by alternative algorithms which had the same general topology. For explanation of the codes used for the variants, see the legends to Fig. 3 to 5. Boldface letters indicate the five major branches of intratype evolution of HPV-16, i.e., the two African branches Af1 and Af2, the Asian/American branch AA, the European branch E, and the East Asian branch As. The bootstrap value for the separation of the African branches (Af1, Af2, and AA) from the Eurasian branches (E and As) is 82%, that for the AA branch at the AN-12 internode is 86%, and that for the As branch is 41%.

netic tree. The code in the last vertical column classifies the respective variant to one of the five major phylogenetic branches. The first horizontal line of each figure identifies the genomic position of the mutation; a black square indicates the presence and a white square indicates the absence of a point mutation in this particular position relative to the HPV-16 reference clone.

HPV-16 variants in Africa. Most HPV-16 variants from Tanzania differed from the HPV-16 reference clone consistently by the same five mutations (genomic positions 7487, 7519, 7762, 7784, and 7832), and many of them differed by up to seven additional mutations. The five common mutations express relationship with one another despite the deep split between branches Af1 and Af2 (Fig. 2). With 16 variants that belong to the Af1 and Af2 branches, no other ethnically homogeneous cohort was so rich in different variants. In addition to these Af1 and Af2 variants, four patients carried variants of the E branch. In consideration of the fact that these samples came from Dar-es-Salaam, a seaport and gateway of tourism, we cannot determine whether these four variants are typically present in Tanzanians or whether they were introduced recently.

We decided to compare these Tanzanian HPV-16 variants with those from four geographically remote cohorts in western, central, and southern Africa. This was done to explore the possibility that we might find a gradient of different variants across the continent. Figure 3 shows that this is obviously not the case. We found in these four cohorts 15 variants that belong to the Af1 and Af2 branches, and all except three (SL-6a, SA-9, and SN-4) were indistinguishable from Tanzanian variants. SL-6a is of interest as it is a link between the Af1 and Af2 branches and the E branch. In addition, we found 11 variants that belonged to the E branch. Two of these were isolated from the two Caucasian patients in Cape Town, the only two non-African patients examined in the African continent. Of four African patients examined in Cape Town, two (SA-2 and SA-8) had variants from the E branch. Five patients of mixed racial background carried three variants from the E branch and two from the Af branches, respectively.

Among the isolates from African patients, 39 of 52 (75%) belonged to the Af branches and 13 belonged to the E branch, and there were no isolates from the AA and As branches. It was unexpected that we found no variants of the AA branch in Africa: we had previously suspected (5) that this branch was of African origin, not only because of its linkage with the Af1 and Af2 branches but also since variants of the AA branch are common in Brazil, which was the destination of a large part of the slave trade from Africa (19, 20). Given our new findings, we assume that variants of the AA branch are not present in cohorts from five regions of Africa, some of which have been origins of the slave trade.

HPV-16 variants in Europe. The collection of samples from the European continent (Fig. 4) was unique in that this was the only continent with samples that belonged to only one major phylogenetic branch of HPV-16 diversity. Most of these variants were so closely related that they formed a dense cluster, with most variants differing by only one or two mutations. This close relationship is remarkable since we had chosen samples from three nations that are more than 2,000 km apart and have been geographically separated throughout recorded history, i.e., for several thousand years.

Because of the geographic predominance of similar variants, we termed this branch the E branch. Of the 30 samples from

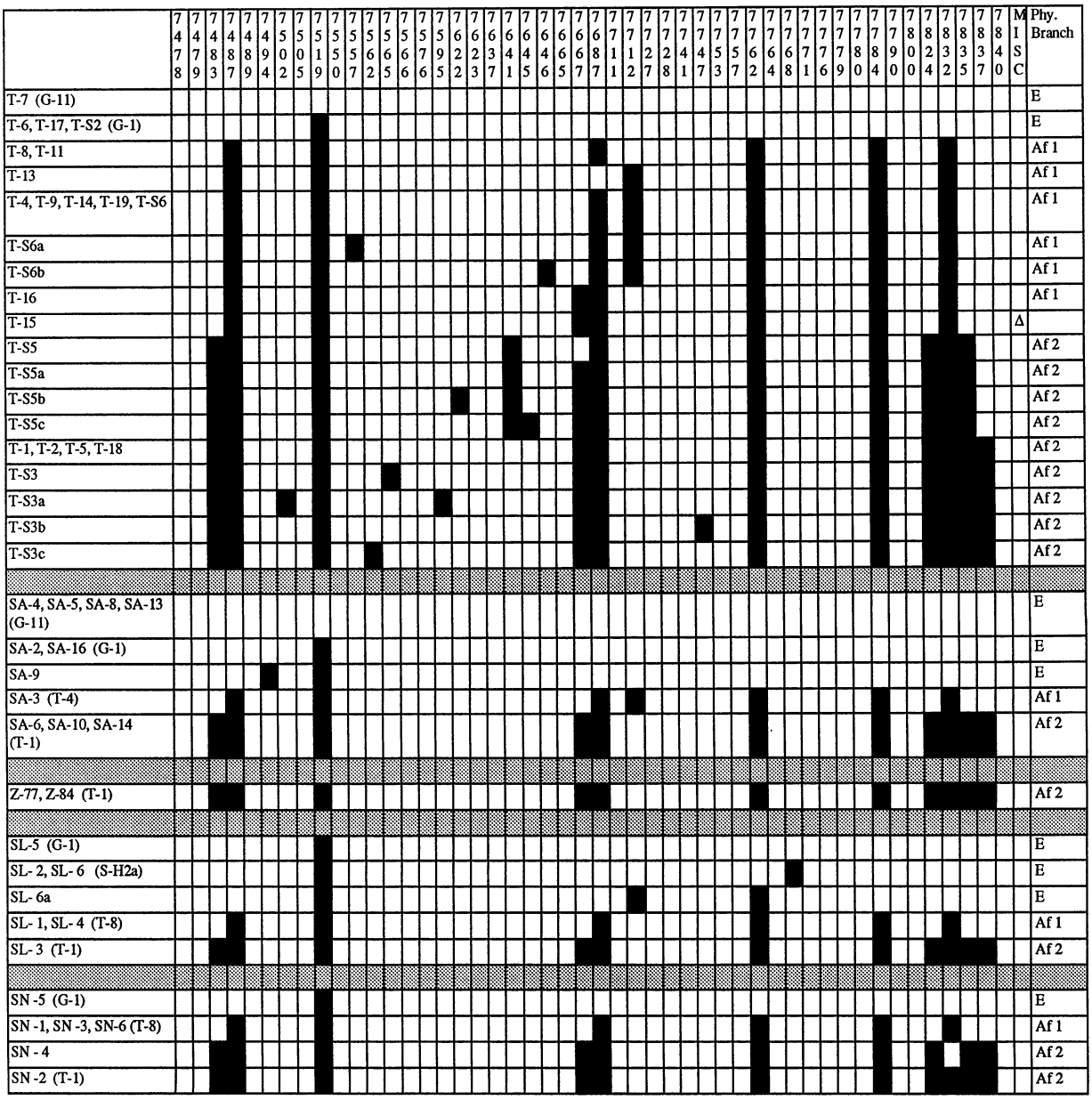


FIG. 3. Combined representation of the LCR segment of HPV-16 isolates from Africa. Each row represents one genomic variant. The first column specifies the code of the isolate. Code numbers in parentheses represent variants of identical sequence that were isolated elsewhere and had been used to represent the respective variants in the phylogenetic tree (Fig. 2). The phylogenetic assignment to either of the five major branches of intratype evolution is shown in the last vertical column. Numbers at the top identify the genomic positions where mutations have occurred. For the type of substitution, see reference 5 and Results. Black squares indicate the presence and white squares indicate the absence of the mutation relative to the HPV-16 reference clone; grey squares indicate alternative substitutions (5). Abbreviations: T, Tanzania; SA, South Africa; Z, Zaire; SL, Sierra Leone; SN, Senegal.

Europe, 25 either were identical with the HPV-16 reference clone or differed from it by a single point mutation at position 7519. This latter isolate, which we term variant 7519, is of interest because it is worldwide the most frequent HPV-16 variant (103 of 301 isolates) and is at the phylogenetic root of the E branch.

HPV-16 variants in Arabs and patients from India. We had found previously (5) that patients in Singapore, a nation in East Asia that is composed of immigrants, carry particularly diverse HPV-16 variants. We therefore wondered whether we would find a similar heterogeneity in Asian countries with an ethnically more homogeneous population or whether ethnic homogeneity would be reflected in the samples. Also, it was possible that variants from such countries may correlate with

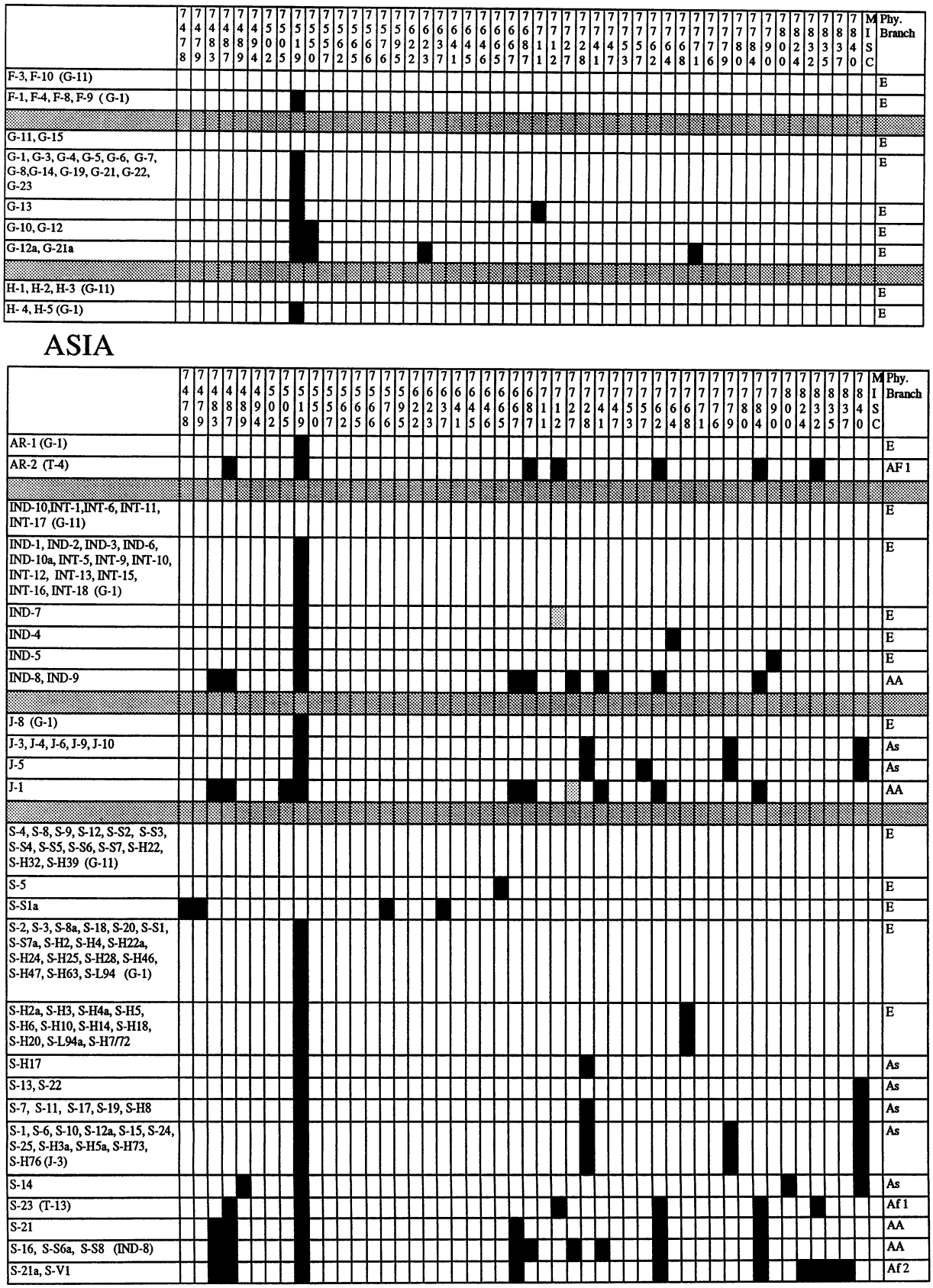


FIG. 4. Combined representation of the LCR segment of HPV-16 isolates from Europe and Asia. F, Finland; G, Germany; H, Greece; AR, Arab patient from Israel; IND, India (Delhi); INT, India (Trivandrum); J, Japan; S, Singapore. For further details, see the legend to Fig. 3.

the ethnicity of the patients, as suggested by the discrepancy between European and African variants.

Indians are of Caucasian descent, and there is evidence for ethnic components common to Europeans and Indians as recently as 10,000 years ago (2, 21). Figure 4 shows that there is a correlation of HPV-16 variants between India and Europe, since 21 of 23 variants from northern and southern India belong to the E branch of HPV-16 diversity. Interestingly, the other two samples, IND-8 and IND-9, are representatives of a novel branch of HPV-16 diversity. We termed this branch AA, as we found these variants only in Asia and the Americas. It should be noted, that the AA branch is evolutionarily closer to the African branches Af1 and Af2 than to the E branch.

Unfortunately, we obtained only two samples from Arab patients; one contained a variant that belonged to the Af1 branch, and the other, variant 7519, belonged to the E branch. Analysis of a large Arab cohort would be necessary to determine whether the African variant reflects a sporadic transfer or whether gradients of HPV-16 variants of the E and Af branches exist across the countries populated by Arabs as a result of their geographic position between Europe and Africa.

HPV-16 variants in East Asia. HPV-16 variants in the population of Singapore are very heterogenous, as can be expected in a country that began as a seaport and is composed of ethnic Chinese, Malays and Indians, and large groups of Europeans and Japanese.

To further dissect this complex picture, we analyzed samples from Japan, an ethnically fairly homogeneous East Asian country. We found in Japan one patient with variant 7519 (E branch) and one with a variant belonging to the AA branch. However, six of eight HPV-16 variants from Japan had a unique pattern characterized particularly by a mutation at position 7840 that we had previously found in 19 variants from Singapore. These 25 variants gave rise to a separate branch extending from the E branch. Since we found these variants only in East Asia, we termed this branch the As branch.

Among the 60 patients of the Singaporean cohort, 19 (32%) carried variants of the As branch, four (7%) carried variants of the AA branch, three (5%) carried variants of the Af branches, and 31 (56%) carried variants of the E branch. This latter number largely exceeds that of people in Singapore of Indian or European ethnicity (10%). Among the variants of the E branch, we found one with a mutation at position 7768 particularly frequently; this variant was otherwise found only in patients from Sierra Leone and among Eskimos from Greenland.

HPV-16 variants in American aboriginal populations. The peopling of the Americas is better understood than that of Africa, Europe, and Asia. Indigenous ethnic groups began to enter the continent from Asia via the Bering Strait probably around 12,000 years ago (13, 16), and the majority of today's population arrived during the last 500 years from Europe, Africa, and Asia. It is possible that members of each of these ethnic groups were carriers of specific HPV-16 variants. It is conceivable that these components of HPV-16 diversity can still be dissected today, although their traces will most likely have faded since none of these ethnic groups remained in reproductive and sexual isolation.

Figure 5 depicts our analysis of the HPV-16 variants found in three indigenous ethnic groups in North America and Greenland (Eskimos, Navajo Indians, and Pueblo Indians) and four in South America (three Indian tribes in the Amazon and patients of Amerindian ethnicity in the Chaco region of Argentina). Of these isolates, 34 belong to the E branch, 3 belong to the Af branches, and 4 belong to the AA branch.

HPV-16 variants in American immigrant populations. We analyzed 22 samples from the United States. As it seemed almost certain that an analysis of Caucasian cohorts would result in the homogeneous composition of HPV-16 variants with little diversity as seen in European countries, we concentrated our analysis on racially mixed cohorts. Seventeen samples contained variants of the E branch, and seven of these variants were found in black patients (two in New York, N.Y., two in Detroit, Mich., and three in Charlotte, N.C.). Four variants belonged to the Af branches and were found in a black patient from North Carolina and in two black patients and one Hispanic patient from New York. One variant isolated from a black patient in North Carolina belonged to the AA branch.

Figure 4 also includes data of two South American nations with ethnically mixed populations, Peru representing a nation with two ethnic components (European and American Indian) and Brazil representing a nation with three ethnic components (European, American Indian, and African).

Both cohorts have in common that HPV-16 variants belonging to the E branch predominate (35 of 51), while 16 samples contained variants belonging to AA branch. There were no samples with variants from the Af branches, although we have expanded the previously published isolates from the Brazilian cohort by 10, from patients of black or racially mixed ethnicity (B-27 to B-36). This was done in the assumption that the social environment and the sexual preferences of these patients may give a higher chance of finding African variants preferentially maintained in this group. Interestingly, however, these 10 patients were infected by variants of the E branch.

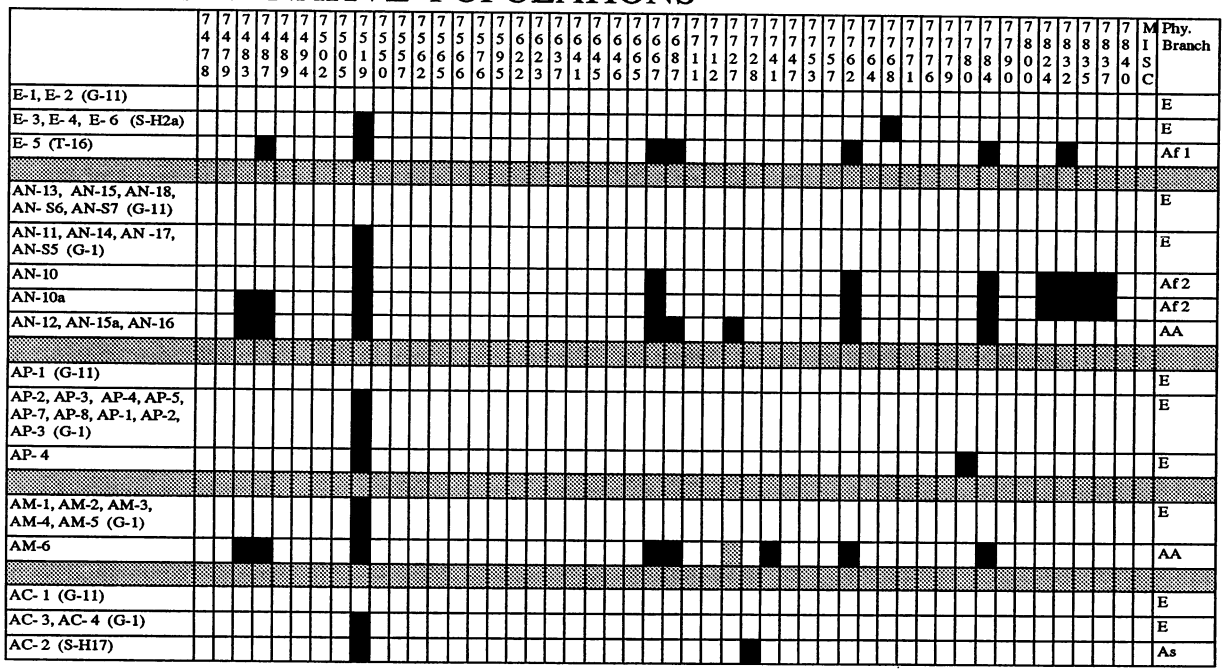
DISCUSSION

Our study aimed to be a worldwide survey of the genomic diversity of HPV-16 isolates to establish a data base for phylogenetic, taxonomic, and epidemiological research. It became clear that this virus and the particular genomic segment that we investigated were very fortunate choices. Sequence variability was large enough to be highly informative but small enough and also poor enough in convergent mutations to permit unambiguous reconstruction of the viral evolution and to suggest that we have probably identified all major lineages of intratype diversity of HPV-16. On a more technical note, it was an interesting by-product of this work to prove how rarely DNA samples become contaminated by cloned HPV-16. Despite the complex logistics involved in our research, only slightly more than 10% of all samples contained the HPV-16 reference clone, and most of them were found in cohorts where we would expect to find HPV-16 variants of the E branch. We conclude that none of the 19 laboratories involved in the study was particularly prone to generate DNA contaminations.

In this report, we restricted the phylogenetic analysis to variants of the 364-bp LCR segment. Similar mutational diversity has been recorded in the E5, E7, and L1 genes (5, 8, 10, 17). Typical patterns are linked, which is evidence for the absence (or at least rarity) of intermolecular recombination in papillomaviruses. Of course, linkage between typical patterns in different parts of the genome should not be interpreted as evidence for a causal connection, since mutations in different parts of the genome are most likely independent events. Consequently, HPV-16 variants that were defined by mutational patterns in the LCR could almost certainly be further subclassified after recording divergence through mutations in the genome elsewhere.

In the LCR, maximal evolutionary distances between some HPV-16 variants are 4.7%, i.e., 17 substitutions in 364 bp, when, for example, variant J-5 of the As branch and Tanzanian variant T-S3a of the Af2 branch are compared. It came as a

AMERICA: NATIVE POPULATIONS



AMERICA: IMMIGRANT POPULATIONS

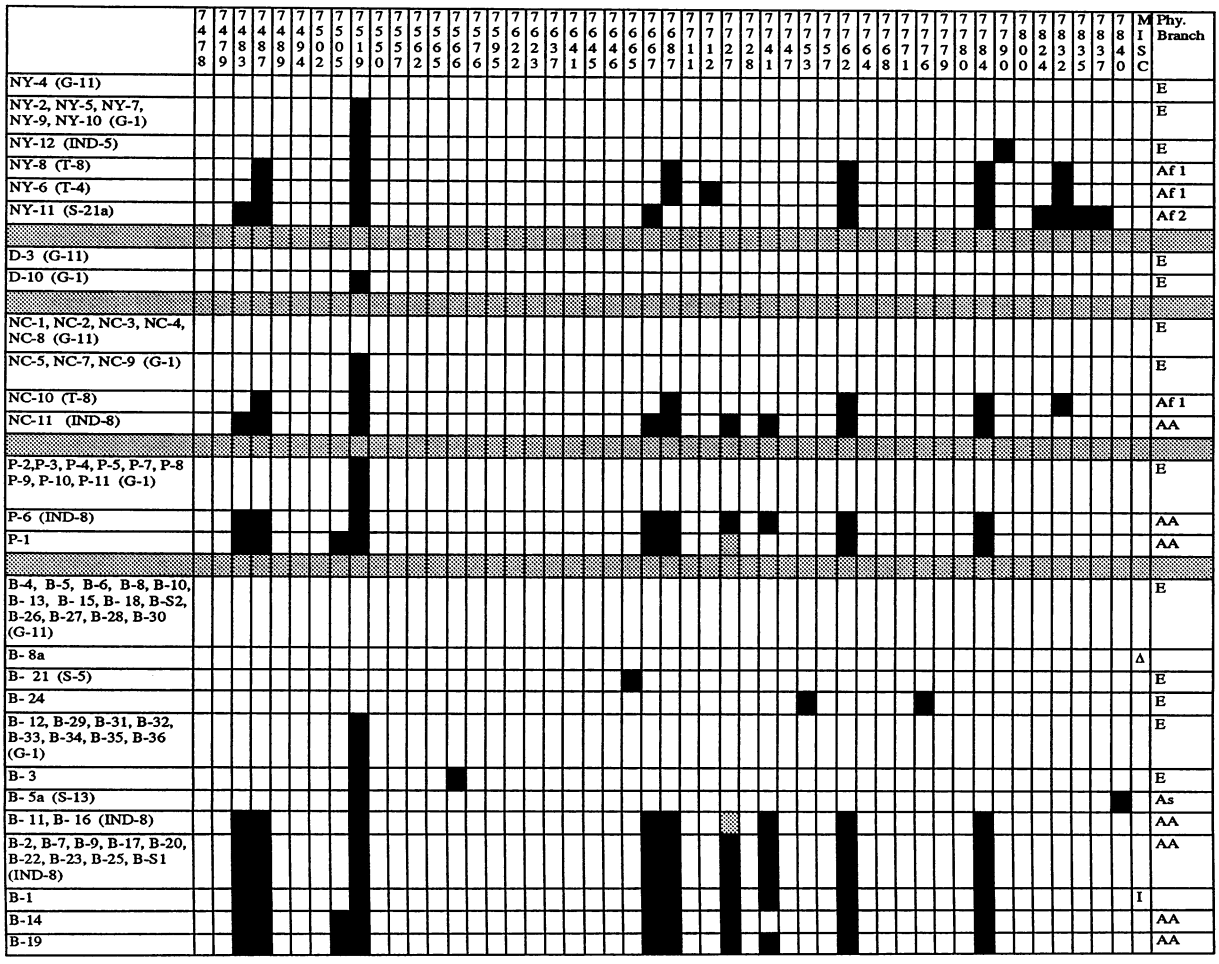


FIG. 5. Combined representation of the LCR segment of HPV-16 isolates from native and immigrant populations of the Americas and Greenland. E, Greenland (Eskimos); AN, Navajo Indian; AP, Pueblo Indian; AM, Amazonian Indian; AC, Argentina (Chaco); NY, New York City; D, Detroit; NC, North Carolina; P, Peru; B, Brazil. For further details, see the legend to Fig. 3.

surprise that this number did not increase subsequent to our previous publication, although our research now covers 25 geographic regions and ethnic groups instead of only 4. From this, we conclude that we have probably identified all major branches of HPV-16 genomic diversity that exist in the majority of all humans. It cannot be excluded, however, that remotely related HPV-16 variants exist in those ethnic groups that evolved for a long time separately from other humans, such as African Pygmies, !Kungs, or Hottentots, or in New Guinea and Melanesia and among Australian aborigines (3, 24).

The limited genomic diversity is surprising when one considers that molecular evolution is most likely a continuous process, i.e., a stepwise accumulation of mutations leading to the diversification of genomes. The HPV types that are closest related to HPV-16 are HPV-31 and HPV-35 (4, 23), and these three HPV types must have evolved from a common ancestor. Since the LCRs of these three viruses are so diverse that they cannot be aligned with confidence (our unpublished observations), HPV-16 must therefore have had precursor genomes that differed from today's isolates by much more than 4.7%. These genomes either no longer exist in humans or are too rare to be detectable. Although we do not understand the processes of extinction or outcompetition of the other genomes by today's HPV-16 genomes, it has rendered present-day HPV-16 a natural taxonomic unit. This is probably the case for all papillomavirus types, although the original definition of a type was operational rather than based on exhaustive taxonomical research (6).

In the accompanying paper (18), we report two interesting observations about the relationship of papillomaviruses and the geography and history of their origin. (i) The three ape and monkey papillomaviruses that are known today are more related to some of the genital HPV types than these human viruses are to one another. This is most likely because they had diverged before *Homo sapiens* evolved. (ii) This hypothesis predicts that some speciation events of papillomaviruses must have taken place in Africa. There is agreement that the genus *Homo* originated in Africa, because hominids evolved in this continent (2, 22, 24). This postulate led us to compare geographic variants of HPV-18 and HPV-45, two HPV types that are closer to one another than HPV-16, HPV-31, and HPV-35 are among themselves. We have found that the root of HPV-18 and HPV-45 phylogenies is in Africa and that African variants of either virus type form a phylogenetic bridge between these viruses. From a point mutation that is characteristic of HPV-18 isolates from Amazonian Indians, we could also propose a crude estimate of the speed of evolution of HPV-18. We suggest that it takes at least 12,000 years for a single point mutation in the LCR of HPV-18 to become fixed in the infected human cohort (18).

The phylogenetic tree of HPV-16 variants suggests a picture similar to the evolution of HPV-18, although we can neither root this tree against another HPV type nor find mutations that would allow us to measure the rate of HPV-16 evolution. But three other commonly accepted arguments place the root of HPV-16 evolution in Africa: (i) the African cohorts carry the largest number of different HPV-16 variants, and we find (ii) the deepest branching and (iii) the midpoint root (5) of the phylogenetic tree between the Af1, Af2, and AA branches. This observation suggests that some of the variants of these branches are closer to an unknown ancient HPV-16 precursor than to the variants of the E and As branches.

As a fourth argument for the placement of the origin, one could probably also use the rate of evolution of HPV-18 as a good approximation of that of HPV-16. These two HPV types are similar in biology, and there are similar numbers of mutational differences between the respective isolates from African, Caucasian, and East Asian peoples and a lack of mutational differences between isolates from peoples with only a few thousand years of separate histories. This time span is probably too short to permit a point mutation to become fixed in the viral population. Applying the speed of the molecular clock of HPV-18 to HPV-16, the maximal evolutionary distance of 17 mutations in the HPV-16 tree would then represent a time span of at least 200,000 years, which is similar to some estimates for the origin of modern *H. sapiens* (24).

Lastly and most convincingly, we find for HPV-16 as well as for HPV-18 (18) a phylogenetic tree topology similar to that of the three principal human races, Africans, Caucasian, and East Asians. This makes ancient coevolution of humans and of papillomaviruses much more likely than a recent spread of HPV-16 purely by geography. This hypothesis should certainly not be misunderstood to exclude ancient or recent transmission of papillomaviruses across ethnic boundaries, given sufficient mutual exposure of the members of two ethnic groups. The finding of HPV-16 variants of the E branch in some of the black patients in the United States and occasionally in Africa exemplifies these occurrences.

After the likely origin of HPV-16 in Africa in the form of variants similar to those of the Af branches, E-branch variants may have evolved while humans infected by HPV-16 spread out of Africa. Variants of the E branch must have existed outside Africa for a long time, since E-branch variant 7519 was apparently the ancestral genome of the As branch, which could have taken as much as 50,000 years to evolve.

The distribution of the AA branch in a minority of Asian Indians, East Asians, and Amerindians suggests a spread independent from the evolution of the ethnicity of today's population. It is an interesting speculation that these HPV-16 variants were widespread in ancient times in Asian people of an ethnicity that cannot be recognized today. It will be interesting to determine whether variants of the AA branch are widespread in New Guinea and Melanesia populations and in Australian aborigines, since these peoples are believed to be ethnically close to people who were widespread in parts of Asia before the arrival of the people who predominate ethnically today (1).

We have not found variants of the As branch but rather variants of the E and the AA branches in indigenous people of the Americas, although they are believed to be genetically close to East Asians. One explanation would be that HPV-16 did not exist in the New World in pre-Columbian times and that American Indians were infected subsequently. We do not believe this to be likely because it does not explain the continuity of the AA branch through Asia into the Americas. The only alternative to the spread of these variants along this route is that we have missed a major reservoir of variants of the AA branch in Africa, from where these variants could have spread with the slave trade to the Americas. This is unlikely given our intensive coverage of the African continent and the presence of African variants in many North American black patients.

It is assumed that several of the Amazonian Indian tribes investigated in this study live in isolation and probably have not yet had sexual contacts with non-Indians, and we have found that they were infected by variant AM-6 of the AA branch and variant 7519 of the E branch. Since the same people were also infected by variants of HPV-18 indistinguishable from East Asian HPV-18 variants or differing by a single point mutation (18), it is certain that the ethnic groups that colonized the Americas from East Asia carried at least this virus type. These considerations suggest that HPV-16 variants of the E and AA

branches may have been present in Amerindians in pre-Columbian times and that evolution of HPV-16 and HPV-18 is apparently so slow that no change may have occurred over 12,000 years. These findings have the implication that at the time of the peopling of the Americas, the As branch of HPV-18 diversity, but not HPV-16 diversity, existed in East Asia in nearly the same form as today. Independent bottlenecks of migration and expansion of human carrier populations could explain this phenomenon.

The population of Brazil may represent a historic example of such a scenario. People from many parts of this country carry approximately 50% Caucasian, 25% Amerindian, and 25% African genes (19, 20), and it is conceivable that infections by HPV variants reflect these ethnic components. In contrast, 4 of 6 HPV-18 variants were of Amerindian origin (18), and 14 of 50 HPV-16 variants belonged to the AA branch with its likely Amerindian source. We speculate that these asymmetries could originate from a founder effect, i.e., the establishment of a large mixed Caucasian-Amerindian population before the onset of the slave trade. Such an asymmetry could be further distorted if females represented a larger fraction of the total viral reservoir: male Caucasian settlers frequently married Amerindian women, while the opposite was rare, and fewer female than male slaves arrived from Africa.

One may ask which biological properties could make the transmission of an infectious agent mimic the Mendelian inheritance of a host gene. Certainly the fact that these viruses are sexually transmitted would contribute to this behavior. But comparison with the worldwide spread within a few decades of human immunodeficiency virus type 1 suggests that the mode of transmission alone is not sufficient to explain these observations. A factor that may further contribute to this picture is the apparently low infectivity of cutaneous papillomaviruses, which may also apply to genital HPV types. We recently found evidence for the sexual transfer of HPVs in a comparison of samples from both partners of married couples. Interestingly, however, in this study we also found several couples with partners who did not have identical HPV-16 variants. We concluded from this observation that only a large number of sexual contacts may lead to a significant probability of transmission (15). In a similar manner, the endemic nature of human T-cell leukemia (lymphotropic) virus type 1 infections could possibly also be explained by lack of efficient sexual transmission (11, 12). Lastly, the same pattern would also result if vertical transmission plays a major role in HPV infection.

ACKNOWLEDGMENTS

We thank Henning Jensen for help in obtaining clinical material from Greenland, C. M. C. Dehaeck and R. Soeters for help with samples from South Africa, S. Grinstein for help with samples from Argentina, and Alencar de Mello at the Instituto Evandro Chagas, Belem, Brazil, for support during the expeditions into the Brazilian Amazon, the Sierra Leone Ministry of Health and former chief medical officer George John Komba-Kono, and the Zaire Department of Public Health, specifically Abib Thiam Manoka.

C.M.W. is the recipient of the 1991–92 American Social Health Association/Pfizer Fellowship in Sexually Transmitted Diseases.

REFERENCES

1. **Bellwood, P.** 1991. The Austronesian dispersal and the origin of languages. Sci. Am. **265:**70–75.
2. **Cavalli-Sforza, L. L., P. Menozzi, and A. Piazza.** 1993. Demic expansions and human evolution. Science **259:**639–644.
3. **Cavalli-Sforza, L. L., E. Minch, and J. L. Mountain.** 1992. Coevolution of genes and languages revisited. Proc. Natl. Acad. Sci. USA **89:**5620–5624.
4. **Chan, S.-Y., H.-U. Bernard, C.-K. Ong, S.-P. Chan, B. Hofmann, and H. Delius.** 1992. Phylogenetic analysis of 48 papillomavirus types and 28 subtypes and variants: a showcase for the molecular evolution of DNA viruses. J. Virol. **66:**5714–5725.
5. **Chan, S.-Y., L. Ho, C.-K. Ong, V. Chow, B. Drescher, M. Duerst, J. ter Meulen, L. Villa, J. Luande, H. N. Mgaya, and H.-U. Bernard.** 1992. Molecular variants of human papillomavirus-16 from four continents suggest ancient pandemic spread of the virus and its coevolution with humankind. J. Virol. **66:**2057–2066.
6. **Coggin, J. R., Jr., and H. zur Hausen.** 1979. Workshop on papillomaviruses and cancer. Cancer Res. **39:**545–546.
7. **Duerst, M., L. Gissmann, H. Ikenberg, and H. zur Hausen.** 1983. A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions. Proc. Natl. Acad. Sci. USA **80:**3812–3815.
8. **Eschle, D., M. Duerst, J. ter Meulen, J. Luande, H. C. Eberhardt, M. Pawlita, and L. Gissmann.** 1992. Geographical dependence of sequence variation in the E7 gene of human papillomavirus type 16. J. Gen. Virol. **73:**1829–1832.
9. **Felsenstein, J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:**789–791.
10. **Fujinaga, Y., K. Okazawa, Y. Ohashi, Y. Yamakawa, M. Fukushima, I. Kato, and K. Fujinaga.** 1990. Human papillomavirus type 16 E7 gene sequences in human cervical carcinoma analyzed by polymerase chain reaction and direct sequencing. Tumor Res. **25:**85–91.
11. **Gessain, A., E. Boeri, R. Yanagihara, R. C. Gallo, and G. Franchini.** 1993. Complete nucleotide sequence of a highly divergent human T-cell leukemia (lymphotropic) virus type I (HTLV-I) variant from Melanesia: genetic and phylogenetic relationship to HTLV-I strains from other geographical regions. J. Virol. **67:**1015–1023.
12. **Gessain, A., R. C. Gallo, and G. Franchini.** 1992. Low degree of human T-cell leukemia/lymphoma virus type I genetic drift in vivo as a means of monitoring viral transmission and movement of ancient human populations. J. Virol. **66:**2288–2295.
13. **Gibbons, A.** 1993. Geneticists trace the DNA trail of the first Americans. Science **259:**312–313.
14. **Ho, L., S.-Y. Chan, V. Chow, T. Chong, S.-K. Tay, L. L. Villa, and H.-U. Bernard.** 1991. Sequence variants of human papillomavirus type 16 in clinical samples permit verification and extension of epidemiological studies and construction of a phylogenetic tree. J. Clin. Microbiol. **29:**1765–1772.
15. **Ho, L., S.-Y. Chan, S.-K. Tay, and H.-U. Bernard.** Sequence variants of human papillomavirus type 16 from couples suggest sexual transmission with low infectivity and polyclonality in genital neoplasia. J. Infect. Dis., in press.
16. **Hoffecker, J. F., W. R. Powers, and T. Goebel.** 1993. The colonization of Beringia and the peopling of the New World. Science **259:**46–50.
17. **Icenogle, J. P., P. Sathya, D. L. Miller, R. A. Tucker, and W. E. Rawls.** 1991. Nucleotide and amino acid sequence variation in the L1 and E7 open reading frames of human papillomavirus type 6 and type 16. Virology **184:**101–107.
18. **Ong, C.-K., S.-Y. Chan, M. S. Campo, K. Fujinaga, P. Mavromara-Nazos, Vassiliki Labropoulou, H. Pfister, S.-K. Tay, J. ter Meulen, L. L. Villa, and H.-U. Bernard.** 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.
19. **Saldanha, P. H.** 1989. Mistura de raças, mistura de genes. Cien. Hoje **9:**48–54.
20. **Salzano, F. M.** 1986. Em busca das raizes. Cien. Hoje **5:**48–53.
21. **Sokal, R. R., N. L. Oden, and B. A. Thomson.** 1992. Origins of the Indo-Europeans: genetic evidence. Proc. Natl. Acad. Sci. USA **89:**7669–7673.
22. **Thorne, A. G., and W. H. Wolpoff.** 1992. The multiregional evolution of human. Sci. Am. **266:**28–32.
23. **van Ranst, M., J. B. Kaplan, and R. D. Burk.** 1992. Phylogenetic classification of human papillomaviruses: correlation with clinical manifestations. J. Gen. Virol. **73:**2653–2660.
24. **Vigilant, L., M. Stoneking, H. Harpending, K. Hawkes, and A. C. Wilson.** 1991. African populations and the evolution of human mitochondrial DNA. Science **253:**1503–1507.
25. **zur Hausen, H.** 1991. Viruses in human cancers. Science **254:**1167–1173.