JOURNAL OF VIROLOGY, Nov. 2005, p. 13630–13640 Vol. 79, No. 21
0022-538X/05/$08.00+0 doi:10.1128/JVI.79.21.13630–13640.2005

Worldwide Genomic Diversity of the High-Risk Human Papillomavirus Types 31, 35, 52, and 58, Four Close Relatives of Human Papillomavirus Type 16

Itzel E. Calleja-Macias,[1] Luisa L. Villa,[2] Jose C. Prado,[2] Mina Kalantari,[1] Bruce Allan,[3] Anna-Lise Williamson,[3,4] Lap-Ping Chung,[5] Robert J. Collins,[5] Rosemary E. Zuna,[6] S. Terence Dunn,[6] Tang-Yuan Chu,[7] Heather A. Cubie,[8] Kate Cuschieri,[8] Magnus von Knebel-Doeberitz,[9] Claudia R. Martins,[10] Gloria I. Sanchez,[11] F. Xavier Bosch,[12] Nubia Munoz,[13] and Hans-Ulrich Bernard[1]*

Department of Molecular Biology and Biochemistry, University of California—Irvine, Irvine, California 92697[1]; Ludwig Institute for Cancer Research, Sao Paulo, Brazil[2]; Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences,[3] and National Health Laboratory Service,[4] University of Cape Town, South Africa; Queen Mary Hospital and the University of Hong Kong, Hong Kong[5]; Department of Pathology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73104[6]; Department of Obstetrics and Gynecology, Budish Tzu Chi General Hospital, Hualian, Taiwan[7]; Royal Infirmary of Edinburgh, Edinburgh, Scotland[8]; Institute for Molecular Pathology, University of Heidelberg, Heidelberg, Germany[9]; Departamento de Biologia Celular, Instituto de Biologia, Universidade de Brasilia, DF, Brazil[10]; HPV Laboratory, IDIBELL, Institut Català d'Oncologia, Barcelona, Spain, and Group Infection and Cancer, Universidad de Antioquia, Medillin, Colombia[11]; Epidemiology and Cancer Registration Unit, IDIBELL, Institut Català d'Oncologia, Barcelona, Spain[12]; and International Agency for Research on Cancer, Lyon, France[13]

Received 17 June 2005/Accepted 4 August 2005

Among the more than one hundred formally described human papillomavirus (HPV) types, 18 are referred to as high-risk HPV types due to their association with anogenital cancer. Despite pathogenic similarities, these types form three remotely related taxonomic groups. One of these groups is called HPV species 9 and is formed by HPV-16, the most common and best-studied type, together with HPV-31, -33, -35, -52, -58, and -67. Previous worldwide comparisons of HPV-16 samples showed about 2% nucleotide diversity between isolates, which were subsequently termed variants. The distribution of divergent variants has been found to correlate frequently with the geographic origin and the ethnicity of the infected patients and led to the concept of unique African, European, Asian, and Native American HPV-16 variants. In the current study, we address the question of whether geography and ethnicity also correlate with sequence variations found for HPV-31, -35, -52, and -58. This was done by sequencing the long control region in samples derived from Europe, Asia, and Africa, and from immigrant populations in North and South America. We observed maximal divergence between any two variants within each of these four HPV types ranging from 1.8 to 3.6% based on nucleotide exchanges and, occasionally, on insertions and deletions. Similar to the case with HPV-16, these mutations are not random but indicate a relationship between the variants in form of phylogenetic trees. An interesting example is presented by a 16-bp insert in select variants of HPV-35, which appears to have given rise to additional variants by nucleotide exchanges within the insert. All trees showed distinct phylogenetic topologies, ranging from dichotomic branching in the case of HPV-31 to star phylogenies of the other three types. No clear similarities between these types or between these types and HPV-16 exist. While variant branches in some types were specific for Europe, Africa, or East Asia, none of the four trees reflected human evolution and spread to the extent illustrated by HPV-16. One possible explanation is that the rare HPV types that we studied spread and thereby diversified more slowly than the more abundant HPV-16 and may have established much of today's variant diversity already before the worldwide spread of humans 100,000 years ago. Most variants had prototypic amino acid sequences within the E6 oncoprotein and a segment of the L1 capsid protein. Some had one, two, or three amino acid substitutions in these regions, which might indicate biological and pathogenic diversity between the variants of each HPV type.

The phylogeny-based taxonomy of papillomaviruses (PVs) places these viruses into a separate family which is further divided into genera and species (11). On lower taxonomic levels, PVs are classified as types, subtypes, and variants. Basic and clinical research normally addresses PVs on these lower three levels of taxonomy. More than one hundred different human papillomavirus (HPV) types have been formally described (1, 9, 11). Eighteen HPV types are of particular interest, as they are frequently associated with anogenital cancer and therefore classified as high-risk HPV types (23). Paradigms for the high-risk HPVs are HPV type 16 (HPV-16) and HPV-18, which are only remotely related to one another. HPV-16 and -18 form two separate phylogenetic groups, each

* Corresponding author. Mailing address: Department of Molecular Biology and Biochemistry, University of California, Irvine, CA 92697-3900. Phone: (949) 824-5162. Fax: (949) 824-8551. E-mail: hbernard@uci.edu.

together with six other HPV types. These groups are now called HPV species 9 and 7, respectively (11).

An HPV type is defined as a separate taxon when the nucleotide sequence of its L1 gene differs from that of any other HPV type by at least 10%. The additional term "subtype" defines a HPV genome whose L1 nucleotide sequence is greater than 2% and less than 10% different from that of the closest type (11). Only a very few HPV subtypes have been described. Our recent analyses of many isolates of two HPV types (HPV-44 and HPV-68) that have given rise to subtypes led to the calculation of dichotomic phylogenetic trees, which indicated an ancient origin of the "subtype" taxa (6).

While subtypes of HPV types are rare, each HPV type comprises numerous genomic variants whose numbers are nevertheless relatively small (tens or hundreds of variants) in comparison to the quasispecies formed by rapidly evolving RNA viruses whose diversity is several orders of magnitude higher. Variants of HPV types differ by less than 2% of their L1 nucleotide sequences but slightly more in the long control region (LCR), which does not encode genes and is therefore less restricted in its ability to accumulate and tolerate mutations (15, 34). Genomic variation of HPV-16 and HPV-18 has been particularly well studied. Variants of these two types form phylogenetic trees with branches formed by variants with high prevalence in cohorts in Africa, Europe, or East Asia, with one of the East Asian phylogenetic branches of variants extending into populations of Native Americans (16, 24). Thus, these trees are reflective of the evolution and worldwide migration patterns of the human host (3) and suggested that certain variants diverged at approximately the same time when the major human ethnic groups formed.

Subsequent studies have addressed the genomic variabilities of HPV-6 and -11 in genital warts (14); HPV-2, -27, and -57 in common warts (8); HPV-5 in epidermodysplasia verruciformis (10); and rare HPV types in small numbers of cervical lesions (29). All of these studies agree that all HPV types (i) have given rise to a limited and relatively small number of genomic variants, (ii) show a small fraction of genomic diversity (less than 2% in the L1 gene and only slightly more in hypervariable regions like the LCR), and (iii) show geographic specificity, at least for some variants. However, these studies did not document a close linkage between intratype HPV evolution and the evolution and migration of humans as revealed in the studies of HPV-16 and -18 (16, 24).

This paper is the third of three recent publications that address the issue of intratype evolution of HPVs. In the first report, we studied the diversity of HPV types 44 and 68 and their subtypes and variants (6). In the second report, we addressed the diversity of HPV-53, -56, and -66 (26), which form HPV species 6. This represents the third major group in the taxonomy of HPVs in addition to the two species formed by the relatives of HPV-16 and -18. The current study reports on the diversity of four HPV types, HPV-31, -35, -52, and -58, which are closely related to HPV-16. We have been able to examine the intratype diversity of these rare HPVs in collaboration with colleagues involved in HPV studies throughout the world. The data we report confirm the notion of a limited genomic diversity for each of these HPV types and geographic clustering of some HPV variants that suggests slower rates of molecular evolution than previously proposed for HPV-16 and -18 (16, 24). In contrast to HPV-16 and -18, which diversified only by nucleotide exchanges, some of these rarer HPV types show significant amounts of deletions and insertions. This appears to be a rarely encountered mechanism of HPV evolution. Amino acid substitutions of HPV oncoproteins and capsid proteins are rare but do occur and have to be considered in studies of pathogenicity and vaccine development.

MATERIALS AND METHODS

Sample collection. In order to explore the intratype phylogeny of HPV-31, -35, -52, and -58, we examined the sequence diversity within the LCRs of isolates from seven geographically remote cohorts from Sao Paulo, Brazil (BR); Hong Kong; Monterrey, Mexico; Edinburgh, Scotland (ED); Cape Town, South Africa (SA); Oslo, Norway; Oklahoma; and Los Angeles, California. We also included fewer samples from Heidelberg, Germany; Mali; Morocco (MR); Taiwan; The Philippines; and Thailand. Most of the samples from South Africa were from patients of African ethnicity (SA samples), while some were from people of mixed ethnicity (SA* samples). All patients from Hong Kong and Taiwan were ethnic Chinese. All samples had been collected during past or ongoing diagnostic and epidemiological research of cervical smears in studies that are unrelated to the objectives of our project. No patient was sampled solely for the purpose of our research.

Phylogenetically informative amplicons of the LCR. As in our previous studies on the genomic diversity of HPV types, we aimed to amplify a segment of the LCR, which would likely contain a sufficient number (e.g., 20 to 50) of mutations to generate stable phylogenetic trees. No effort was made to amplify the whole LCR. For studying HPV-31, a 523-bp segment between the genomic positions 7527 and 137 was amplified with primers 31-8aF (5′-AGTAGTTCTGCGGTTTTTGGTTTC-3′) and 31-8aR (5′-CCGAGGTCTTTCTGCAGGATTTTT-3′). The genomic sequence was established for 503 bp of the 523-bp fragment. Of the 70 isolates, the known sequences are deposited in GenBank under the access codes AY453992 to AY454037. In order to exclude possible PCR artifacts, all samples were amplified twice and both strands were sequenced twice. The same precaution applied to the treatment of the HPV-35, -52, and -58 samples. For studying HPV-35, we amplified a 893-bp fragment between genomic positions 7146 and 187 with the primers 35LCR-F (5′-TATATTATGTGTTGTGGTGCCTGTTTG-3′) and 35LCRa-R (5′-CGTTTTCGGTCACTCCCTGTTTT-3′). The genomic sequence was established for 814 bp of the 893-bp fragment. Of the 45 isolates, known sequences are deposited in GenBank under the access codes AY454064 to AY454038. Genomic segments of HPV-52 were amplified by PCR using the primers HPV-52-LF (5′-TTGTCTGTTGGGTAATTGTCTGTG-3′) and HPV-52-LR (5′-CGTAACCGGTCGTGTAGTGC-3′), which generated a 750-bp segment between the genomic positions 7158 and 7907. The genomic sequence was established for 637 bp of the 750-bp fragment. Genomic segments of HPV-58 were amplified by PCR with the primers 58-UF (5′-TATGAGTAAGGTGCTGTCCCT-3′) and 58-UR (5′-CGGTCTGACCGAAACCGGTGC-3′), which generated a 545-bp segment between the genomic positions 7345 and 68. The genomic sequence was established for 461 bp of the 545-bp fragment.

Amplification with E6 and L1 consensus primers. The HPV-52, -58, -31, and -35 E6 genes were amplified with type-specific primers. For HPV-31, E631-F (5′-AAAAGTAGGGAGTGACCGAAAGTGG-3′) and E631-R (5′-TCGGGTAATTGCTCATAACAGTGGA-3′) were used, resulting in a 625-bp fragment; for HPV-35, E635-F (5′-CGAAAACGGTTGCCATAAAAG-3′) and E635-R (5′-TGCCTCGGGTTCCAAATCTA-3′) were used, resulting in a 578-bp fragment; for HPV-52, E652-F (5′-ACGCACGGCCATGTTTGAGGAT-3′) and E652-R (5′-TAATTGCTTGTGGCTTGTTCTGCTTGTC-3′) were used, resulting in a 622-bp fragment; and for HPV-58, E658-F (5′-AGGCTACTGCAGGACTATGTTC-3′) and E658-R (5′-AGCGTTGGGTTGTTTCCTCTCA-3′) were used, resulting in a 503-bp fragment. The genomic sequences were established for 447 bp (HPV-52) and 450 bp (HPV-58, -31, and -35). A PCR fragment (450 bp) of the L1 gene was amplified with the consensus primers MY09/MY11 (2, 4), and the genomic sequences were established for 351 bp (HPV-31 and -35) and 421 bp (HPV-52 and -58) of these 450 bp.

PCR amplification, sequence analysis, and phylogenetic evaluations. PCR mixtures contained 20 mM Tris, pH 8.0, 100 mM KCl, 200 mM of each deoxynucleoside triphosphate, 2 mM $MgCl_2$, 10 mM of each sense and antisense oligonucleotide primer, and 1 unit of *Taq* polymerase (Promega, Madison, WI). Forty amplification cycles were run in the Eppendorf Master Cycler with a 94°C denaturing step (30 s), a 60°C annealing step (30 s), and a 72°C extension step (60 s). PCR amplicons were separated electrophoretically on 2% agarose gel, puri-

fied with Exo I and alkaline phosphatase (USB, Cleveland, OH), and applied to enzymatic extension reactions for DNA sequencing using the ABI PRISM BigDye Cycle sequencing kit (AB Applied Biosystems). Both strands were sequenced with the same forward and reverse primers as those used for PCR amplification of the LCR unless stated differently. The sequencing reactions were purified by ethanol-sodium acetate precipitation and then run on an ABI Prism 3100 sequencer. The mutations were analyzed and determined by the ALIGN program at the GENESTREAM network server (25) (http://www2.igh.curs.fr/bin/align-guess.cgi). The neighbor-joining and unweighted-pair group method with arithmetic average (UPGMA) trees were constructed by using Mega version 2.1. All reference (or prototype) sequences were taken from the Los Alamos-based HPV sequence database (http://hpv-web.lanl.gov/stdgen/virus/hpv/compendium/htdocs/). In the case of HPV-35, the prototype sequence is reported at this site as HPV-35H, which corrected several mistakes published in a previous report.

Nucleotide sequence accession numbers. The new sequences of HPV-31 are published with the GenBank accession codes DQ057247 to DQ057270. The new sequences of HPV-35 are published with the GenBank access codes DQ057271 to DQ057289. The sequences of all 66 isolates of HPV-52 are published with the GenBank access codes DQ057080 to DQ057145. The sequences of all 101 isolates of HPV-58 are published with the GenBank access codes DQ057146 to DQ057246. The sequences of all isolates of E6 gene amplification are published with the following GenBank access codes: for E6-amplified HPV-31, DQ057302 to DQ057308; for E6-amplified HPV-35, DQ057309 to DQ057314; for E6-amplified HPV-52, DQ057290 to DQ057295; and for E6-amplified HPV-58, DQ057296 to DQ057301. The L1 MY09/11 GenBank codes are as follows: for HPV-52, DQ057315 to DQ057320; for HPV-58, DQ057321 to DQ057326; for HPV-31, DQ057327 to DQ057333; and for HPV-35, DQ057334 to DQ057339.

RESULTS

Genomic diversity of HPV-31 isolates. The HPV-31 reference genome was isolated from a cervical dysplasia (20). HPV-31 is considered a high-risk HPV type and the fourth most common HPV type in squamous cervical carcinomas and in asymptomatic patients worldwide (23). The genomic diversity of HPV-31 has been investigated in two regional studies (5, 12). Some of the samples described in this article are identical to those in the first of these two publications and are thus identified by the same abbreviations.

We analyzed a 503-bp fragment of the LCR of 69 HPV-31 samples and identified 28 variants (including the reference genome) with 28 point mutations and a 7-bp deletion relative to the HPV-31 genome. The deletion occurred in only four samples, namely, MX80, MX635, MX701, and MX1144. Maximal distance between any two variants was 14 mutations (2.8% of the 503-bp sequence), considering the deletion a single event. A total of 17 of the 28 variants were found in only a single sample, and 6 variants were found in two samples. The HPV-31 reference clone was found in 26 samples.

Gagnon and colleagues (12) could identify 18 variants (including the reference genome). The amplicons studied by this group and by us overlapped but extended in different directions. Based on the overlap, we observed 23 variants and Gagnon and colleagues observed 13 variants, and five of these variants were not observed in our study. Based on these numbers, our study detected more than 80% of the more common and widely distributed HPV-31 variants.

Figure 1 shows a phylogenetic tree of HPV-31 isolates based on the UPGMA algorithm (large figure) and a second tree based on the neighbor-joining method, where each variant is represented by a single isolate (small insert). These trees were calculated weighting each mutation equally except in the case of the deletion. The deletion was given the weight of two mutations, since we assumed that such a rare and specific event likely occurred only a single time. In phylogenetic trees calculated without introducing this bias, the variants MX800, MX635, MX701, and MX1144 did not cluster together.

These trees can be characterized as dichotomic, i.e., being split into two major branches rather than having several branches of similar distance (star phylogenies), such as the published trees of HPV-16 variants (16). Surprisingly, there was no deep branching based on geographic origin or on ethnicity of the patient as in HPV-16 and HPV-18 trees. A division into two branches of presumed geographic specificity, as proposed based on the previously studied small data set (5), was not supported by the addition of many new samples in this project. For example, all samples from Hong Kong and Taiwan contained the HPV-31 reference clone, which also predominated in Norway and Brazil. Two minor branches appear to be based on geographic location, namely, five of the six samples from South Africa (SA*1073 clustering with SA2157, SA*1415 with SA*1509, and SA1644). In addition, the 7-bp deletion mentioned above was observed in four variants of Mexican origin. As such a specific deletion would seem to be a rare event and unlikely to occur more than once, these four variants are likely related and may constitute a phylogenetic assembly originating from Native Americans.

Genomic diversity of HPV-35 isolates. The HPV-35 reference genome was isolated from an endocervical adenocarcinoma (21). HPV-35 is considered a high-risk HPV type and, worldwide, is the eighth most common HPV type in squamous cervical carcinomas and the seventh most common HPV type in asymptomatic control patients (23). Previously, only two genomic variants of HPV-35 were found through worldwide sampling (29), while a recent pilot study from our lab detected nine variants in a small number of cohorts (5). Some of the samples described in this article are identical to those in our previous publication and are identified by the same abbreviations. For the genomic sequence of the HPV-35 reference clone, we used the sequence HPV-35H as corrected and discussed at http://hpv-web.lanl.gov/stdgen/virus/hpv/compendium/htdocs/.

We analyzed an 814-bp fragment of the LCR of 47 HPV-35 samples and identified 12 variants with nucleotide exchanges, a single nucleotide deletion, and a 16-bp insertion relative to the reference genome. The latter was present in 33 samples. Within this 16-bp insert, we observed an additional single nucleotide exchange in two variants (10 samples) and 10 nucleotide exchanges in one sample (SA2299), suggestive of an ancient origin of this insertion. To clarify this unique observation, the raw sequencing data are shown in Fig. 2. Maximal distance between any two variants was 15 mutations (1.8%), considering the insertion a single event and counting the additional point mutations in SA2299. It is evident from the phylogenetic tree in Fig. 3 that the HPV-35 reference clone was found in only four samples (8.5%). In contrast, two variants with the insertion and specific patterns of points mutations elsewhere occurred in 16 (34%) and nine samples (19.1%), respectively, and seem to be the predominating HPV-35 genomes worldwide.

Figure 3 shows the relationship between all HPV-35 isolates based on the UPGMA algorithm (large tree) and the neighbor-joining algorithms, selecting one isolate to represent each variant (small insert). As it seems very likely that the 16-bp inser-

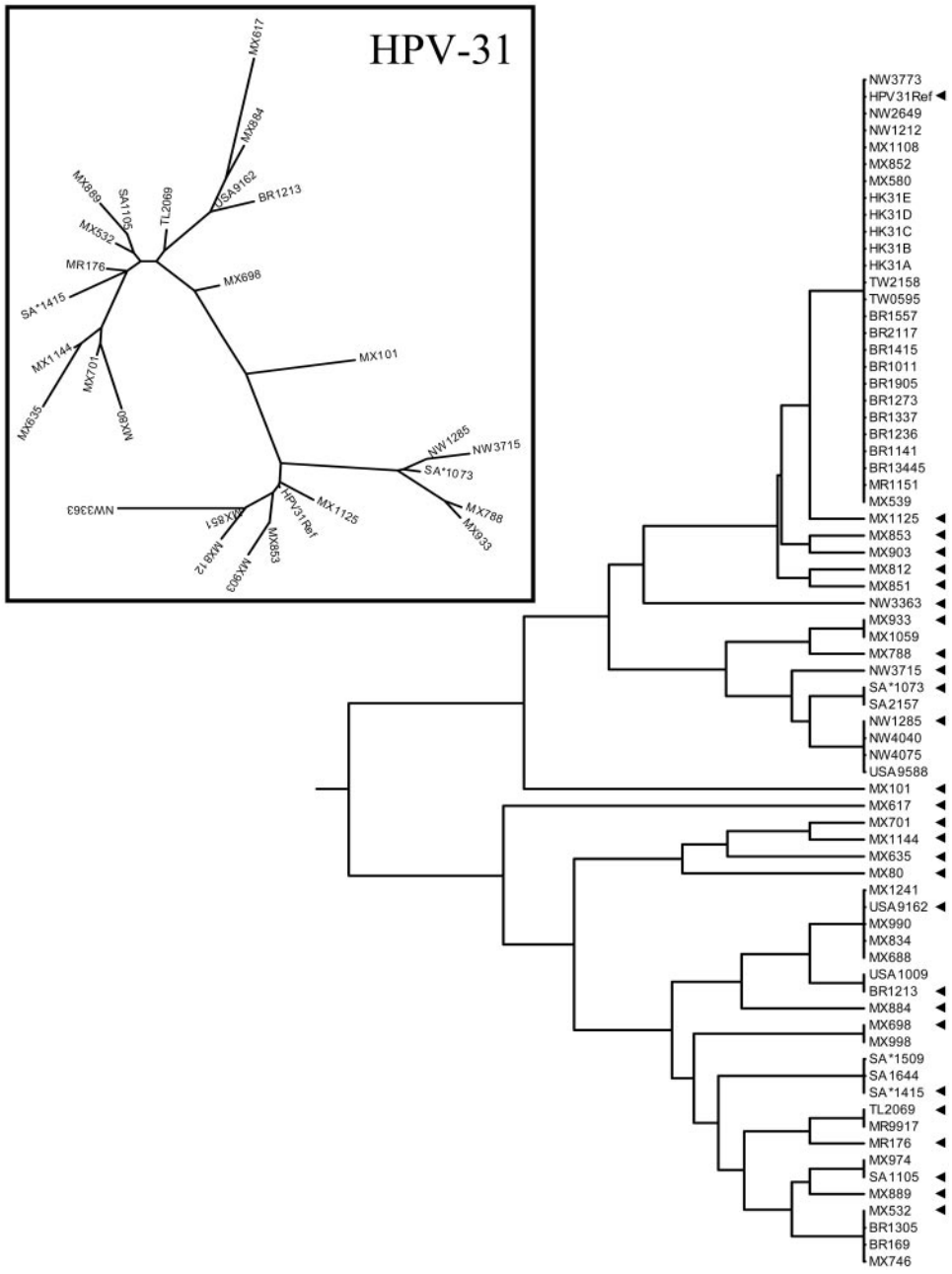

FIG. 1. The intratype diversity of HPV-31. The phylogenetic trees represent the relationship between HPV-31 variants based on a 503-bp segment of the LCR. The large tree is based on the UPGMA, and the small tree is based on the neighbor-joining algorithm. The UPGMA tree represents all isolates, while those that were chosen to represent a particular variant in the neighbor-joining tree are indicated by black triangles. BR, Sao Paulo (Brazil); ED, Edinburgh (Scotland); HE, Heidelberg (Germany); HK, Hong Kong; ML, Mali; MR, Morocco; MX, Monterrey (Mexico); OK, Oklahoma City (Oklahoma); SA, Cape Town (South Africa); PH, The Philippines; TL, Thailand; TW, Taipei (Taiwan); USA, Los Angeles (California).

tion/deletion occurred a single time, we calculated the neighbor-joining tree (Fig. 3, small insert) by overweighting the insertion by a factor of three. When the insertion was entered as a single event, the variants without or with the insertion did not form separate branches. This assumption creates a dichotomy of the tree which is mathematically artificial but likely reflects phylogenetic sequential events, considering how unlikely multiple occurrences of this highly specific mutation may be. It should be noted that African, European, and Asian isolates occurred multiple times in samples with and without this insert, which favors the hypothesis that this insertion/deletion event predated the spread of humans.

Genomic diversity of HPV-52 isolates. The HPV-52 reference genome was isolated from a cervical intraepithelial neoplasia (27). HPV-52 is considered a high-risk HPV type. Worldwide, it is the sixth most common HPV type in cervical

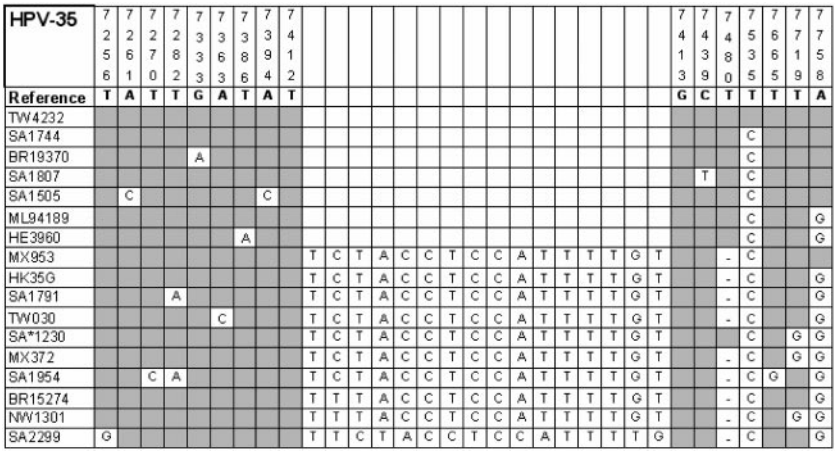

FIG. 2. Mutational patterns in HPV-35 variants. The two top rows indicate the genomic position in the HPV-35 reference clone and the corresponding nucleotides. In the following rows, nucleotide exchanges are shown by letters, deletions relative to the reference clone by a hyphen, and an insert in some variants by an open square in variants that lack this insert. The positions 7412 and 7413 are listed to indicate the position of the insert.

cancer but is not among the eight most common HPV types in cervical smears from asymptomatic patients (23). A preliminary study of the worldwide diversity of this virus based on the MY09/11 segment of L1 identified seven different variant genomes (29).

In total, we analyzed a 637-bp sequence from 66 samples, which led to the identification of 17 HPV-52 variant genomes relative to the reference HPV-52 genome (Fig. 4). All except one of our samples were missing the sequence 5′-TTATG-3′ at the genomic positions 7387 to 7391 (Fig. 4). In order to clarify this unexpected finding, Wayne D. Lancaster kindly supplied us with the original HPV-52 reference isolate. Resequencing of this isolate confirmed a lack of this sequence, indicating an error in the establishment of the original HPV-52 sequence (27). We refer to the corrected sequence, i.e., the omission of the 5-bp segment, as the HPV-52 reference sequence and have submitted this correction to GenBank. However, in order to maintain an unequivocal discussion, the numbers in the following two paragraphs refer to the genomic positions of the uncorrected reference sequence.

A single sample, BR0258, contained a 3-bp insert at exactly the same position where the five nucleotides were omitted, retaining the sequence 5′-ATG-3′ at this position. In addition to the 5′-ATG-3′ insertion, it contained a 4-bp deletion (position 7383 to 7386), two deletions of one nucleotide, a 6-bp insertion (between positions 7701 and 7702 of the original uncorrected sequence) which was not found in any other variant, and 10 nucleotide exchanges. As a result, this variant had the largest distance (14 mutations) from the reference clone of all HPV-52 variants in this study. MX706 was, with a total of 13 changes relative to the reference clone, the second most diverse variant and differed from the reference sequence by a 3-bp deletion (positions 7698 to 7700 of the original reference genomic count) and 12 nucleotide exchanges. BR0258 and MX706 differed from one another by 23 mutations, which constituted the largest distance between any two HPV-52 variants (3.6% of 637 bp). In Fig. 5, this is topologically reflected by two deep branches in the UPGMA phylogenetic tree (large figure) as well as in the neighbor-joining tree (insert). This figure shows all isolates based on the UPGMA algorithm (large figure), with a second tree based on neighbor joining, where each variant is represented by a single isolate (small insert).

Interestingly, the same 3-bp deletion observed in BR0258 was also present in all twelve samples from Hong Kong, Taiwan, and Thailand, as well as in two other Brazilian and two Mexican samples. Nine of the twelve East Asian samples carried another 2-bp deletion (position 7286 to 7287), which occurred in no other isolate. This large number of geographic-phylogenetic correlations suggests the case for an Asian-Native American phylogenetic branch (Fig. 5), similar to the Asian-American branch in HPV-16 (16) and the Asian branch in HPV-18. The only exceptions of this correlation were the presence of a variant with the same characteristics in a sample from Morocco (MR9411) and the presence of a variant that predominated in European samples in the only sample from the Philippines, which may have originated from a recent transfer. In contrast to this Asian branch, all eight samples from South Africa were identical or very similar to samples from Europe or American immigrant countries and did not establish a separate African branch as in HPV-16 and -18.

Genomic diversity of HPV-58 isolates. The HPV-58 reference genome was cloned from a squamous cervical carcinoma (22). HPV-58 is considered the seventh most common high-risk HPV type in cervical cancer and the sixth most common in asymptomatic cervical samples (23). Two previous studies identified seven (29) and eight (7) genomic variants of HPV-58 based on partial L1 sequences.

We analyzed a 461-bp segment of 101 samples, which led to the identification of 21 variants (including the reference genome) relative to the reference HPV-58 genome. Surprisingly, none of the isolates represented the sequence of the reference genome, while a variant that differed by a single nucleotide substitution was found in 61 of the 101 samples. Eighteen of the remaining variants differed only by nucleotide substitutions, while one (ED18136) had a single nucleotide deletion.

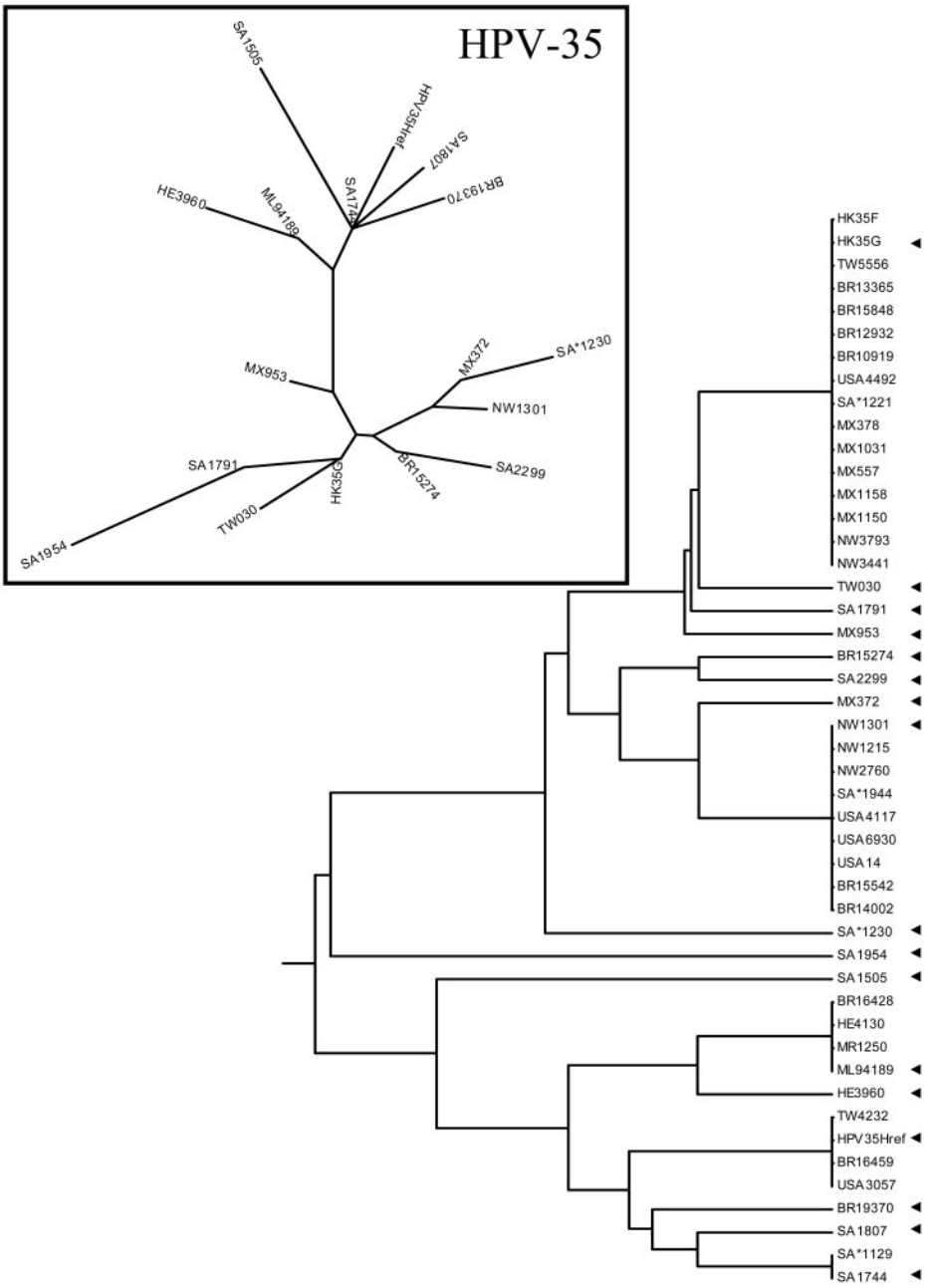

FIG. 3. The intratype diversity of HPV-35. The phylogenetic trees represent the relationship between HPV-35 variants based on an 814-bp segment of the LCR. The large tree is based on the UPGMA, and the small tree is based on the neighbor-joining algorithm. For further details, see the legend to Fig. 1.

The maximal distance between any two variants was 10 mutations (2.2%).

The most common variant, represented by BR63, occurred in South African samples (6 of 11), Scottish samples (4 of 7), and East Asian samples (2 of 7), preventing a clear geographic association of this variant. However, four samples (represented by SA013) contained a unique South African variant, and three samples contained a variant found only in Taiwan (but not in other Asian countries), showing a limited amount of diversification in unique geographic locations. Figure 6 shows a phylogenetic tree of all isolates based on the UPGMA algorithm (large figure) and a second tree based on the neighbor-joining algorithm, in which each variant is represented by a single isolate (small insert).

Diversity in the E6 oncogene and a conserved part of the L1 gene. In order to evaluate the potential for functional changes of HPV proteins, whose genes are linked to LCR variation, we compared HPV-31, -35, -53, and -58 variants selected from remote branches of the phylogenetic tree. For these 25 selected variants, we determined the sequences of the complete E6 gene and about a third of the L1 gene (i.e., that bracketed by the MY09/11 PCR primers).

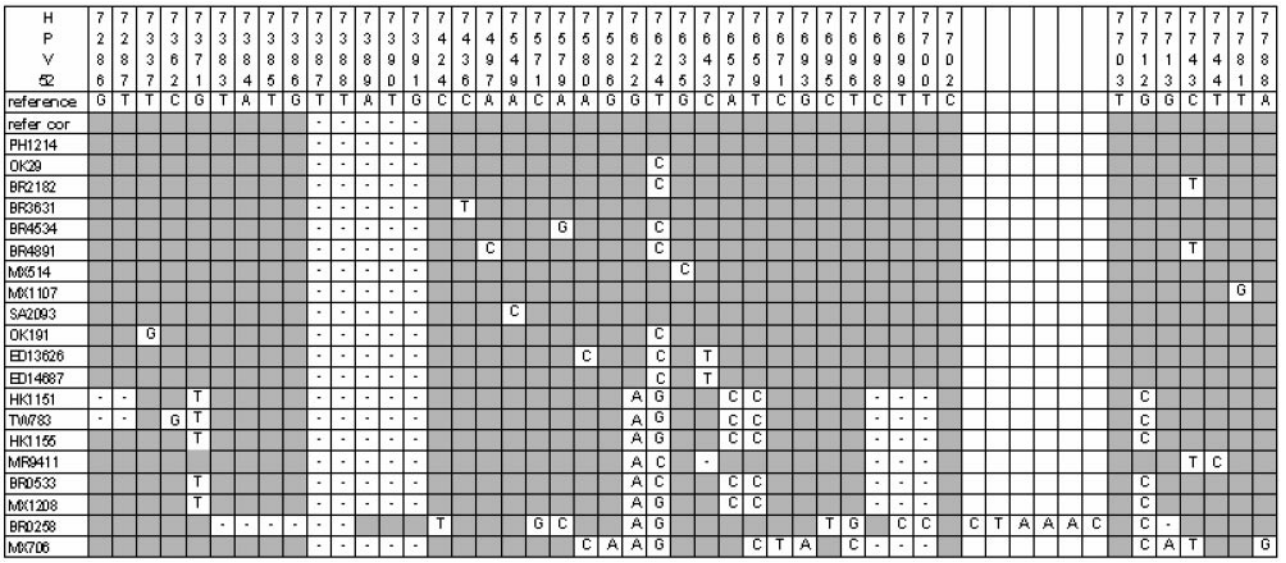

FIG. 4. Mutational patterns in HPV-52 variants. The two top rows indicate the genomic position in the published HPV-52 reference sequence and the corresponding nucleotides. Positions 7387 to 8391 could not be detected by resequencing the original reference clone or in any variants, leading to the corrected sequence of the HPV-52 reference clone in the third row (refer. cor.). In the following rows, nucleotide exchanges are shown by letters, deletions relative to the reference clone by a hyphen, and an insert in one variant (between positions 7701 and 7702) by an open square in variants that lack this insert.

The sequence data are summarized in Fig. 7 and show the nucleotide and amino acid sequence changes of six or seven variants of each of the four HPV types. Within the L1 segment of 25 variants, we found the prototype amino acid sequence in 19 variants, single amino acid sequence exchanges in 1 HPV-35 variant and 1 HPV-52 variant, and a triple amino acid exchange in an HPV-58 variant. These changes were the consequence of zero to seven nucleotide exchanges, amounting to a maximal divergence of 1.6% from the prototype and maximal intervariant diversity of 11 nucleotides (2.4%). In the E6 gene, maximal divergence from the prototype was five nucleotides (1.1%) and maximal intervariant diversity was eight nucleotides (1.8%). Altogether, the E6 protein of 14 of the 25 variants was unaltered. There were single amino acid exchanges in four variants of HPV-35, in two of HPV-52, and in one of HPV-58, and there were two amino acid exchanges in one HPV-31 variant and one HPV-52 variant.

DISCUSSION

Our research expands the knowledge about the intratype genomic diversity of HPVs, which has now been thoroughly studied for HPV-2, -5, -6, -11, -16, -18, -31, -35, -44, -45, -52, -53, -56, -58, -66, and -68 (6, 8, 10, 14, 15, 16, 24, 26, 29). An important outcome of these extensive comparisons is the fact that there is not an unlimited continuum of diverse genomes of any HPV type, i.e., divergence never exceeds 2% in L1 and 5% in the LCR. This cannot be an artifact of the sampling process, since each of the isolates in this and other studies was typed in the respective laboratory of origin by amplification with consensus primers and hybridization to type-specific probes, which would have detected very diverse genomes. The laboratories at the Ludwig Institute, Sao Paulo, Brazil, and at the University of California—Irvine also reevaluated all samples and were able to amplify almost 100% of the samples with type-specific primers. Therefore, we are confident that no hypothetical highly diverse variants were lost due to primer design constraints.

The limited amount of diversity is astonishing, considering the fact that HPV types likely originated from progenitor papillomaviruses quite unlike the extant HPVs by a process of continuous accumulation of mutations. It is therefore not trivial to reason why all the intermediary genomes that led to present day HPV types are now missing. One possibility likely involves biologic advantages acquired by genomic change. Even small improvements (e.g., slightly faster replication or higher infectivity) may, over extended periods of time, alter the composition of an HPV population. By positive selection, functionally improved genomes would eventually outnumber parent genomes over time merely by stochastic processes. Another facet of papillomavirus evolution almost certainly involves genetic drift. Genetic drift can become manifested by founder effects and bottlenecks, i.e., the limited genomic diversity of an HPV type in very small populations of infected individuals who eventually give rise to large populations, for example, at the very origin of the evolution of *Homo sapiens*. This hypothesis is supported by the fact that the chimpanzee papillomaviruses are more closely related to some HPV types (notably HPV-13) (30) than are closely related HPV types to one another, e.g., HPV-13 to HPV-6 and -11. Such observations support the idea that even closely related HPVs had already diversified into separate types in prehuman primates and the small populations of primates that evolved into new host species, such as *Homo sapiens*, may have carried only a fraction of the genomic

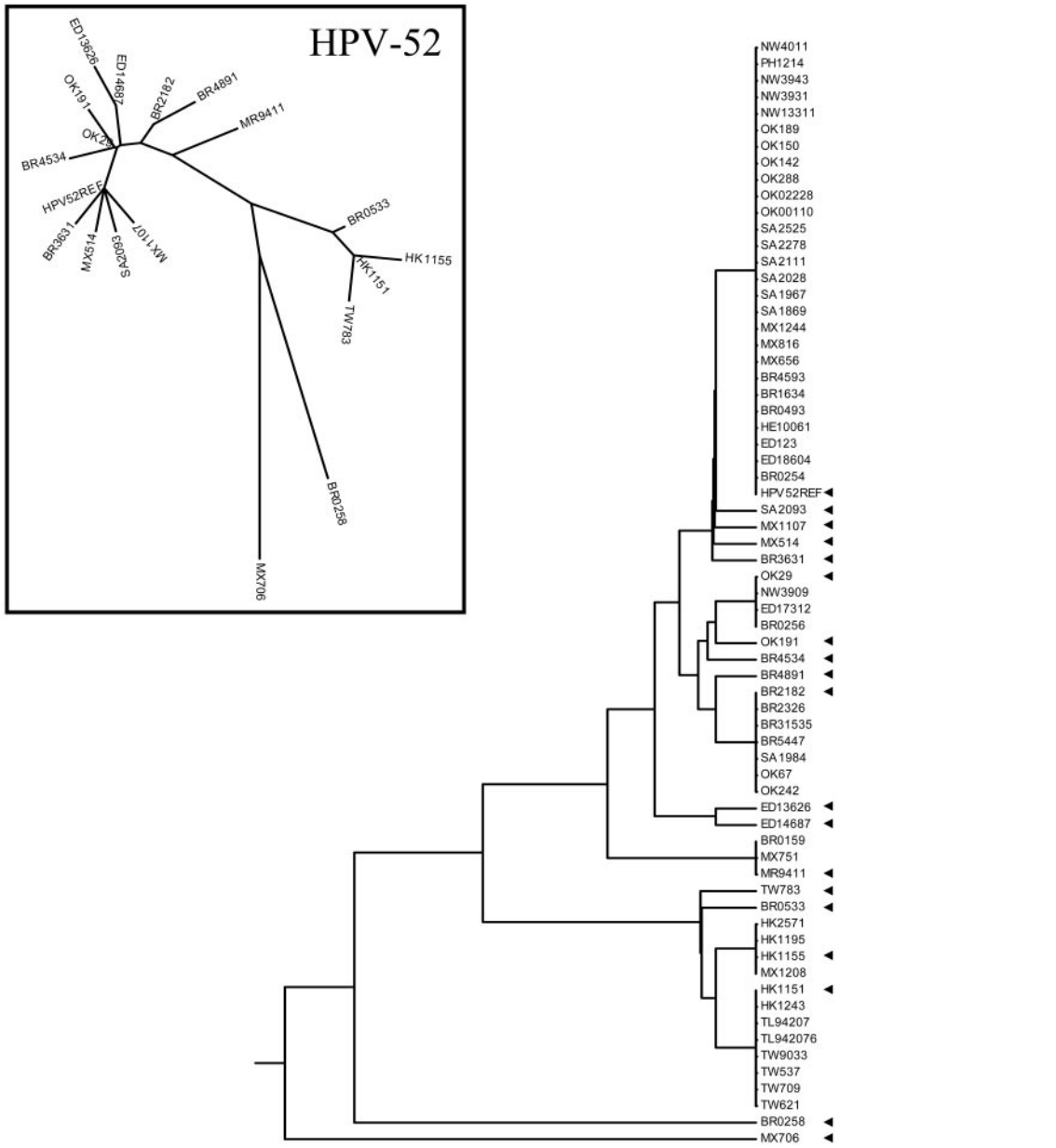

FIG. 5. The intratype diversity of HPV-52. The phylogenetic trees represent the relationship between HPV-52 variants based on a 637-bp segment of the LCR. The large tree is based on the UPGMA, and the small tree is based on the neighbor-joining algorithm. For further details, see the legend to Fig. 1.

diversity that existed in the total population of those primates. Thus, a likely mechanism for the evolution of new HPV types was the extinction of hosts close to the evolutionary roots of *Homo sapiens*. Positive selection and bottlenecks could suffice to explain the origin of closely related HPV types. Presently, available data neither support nor exclude the possibility that in addition, competition between HPV types and selection of the host against virus mutations contributed to the phylogenetic processes.

In the study of HPV-16 and -18 variation, we found a nucleotide divergence of 1 to 5% in the LCR between those variants that predominate today in Africa, Europe, and East Asia. Since the emergence of the predominant ethnic groups in these continents took 10,000 to a few tens of thousands of years, we interpret our findings to indicate that the speed of evolution of most HPV types occurred at a rate of roughly 1% per 10,000 years. This estimate may well be at the high end. For example, the ubiquitous occurrence of E variants of HPV-16 in all three continents indicates that these variants may have existed already while the other variants emerged. Our data may indicate an even lower speed of genomic change for HPV-31, -35, -52, and -58, since we found some variants of each type in

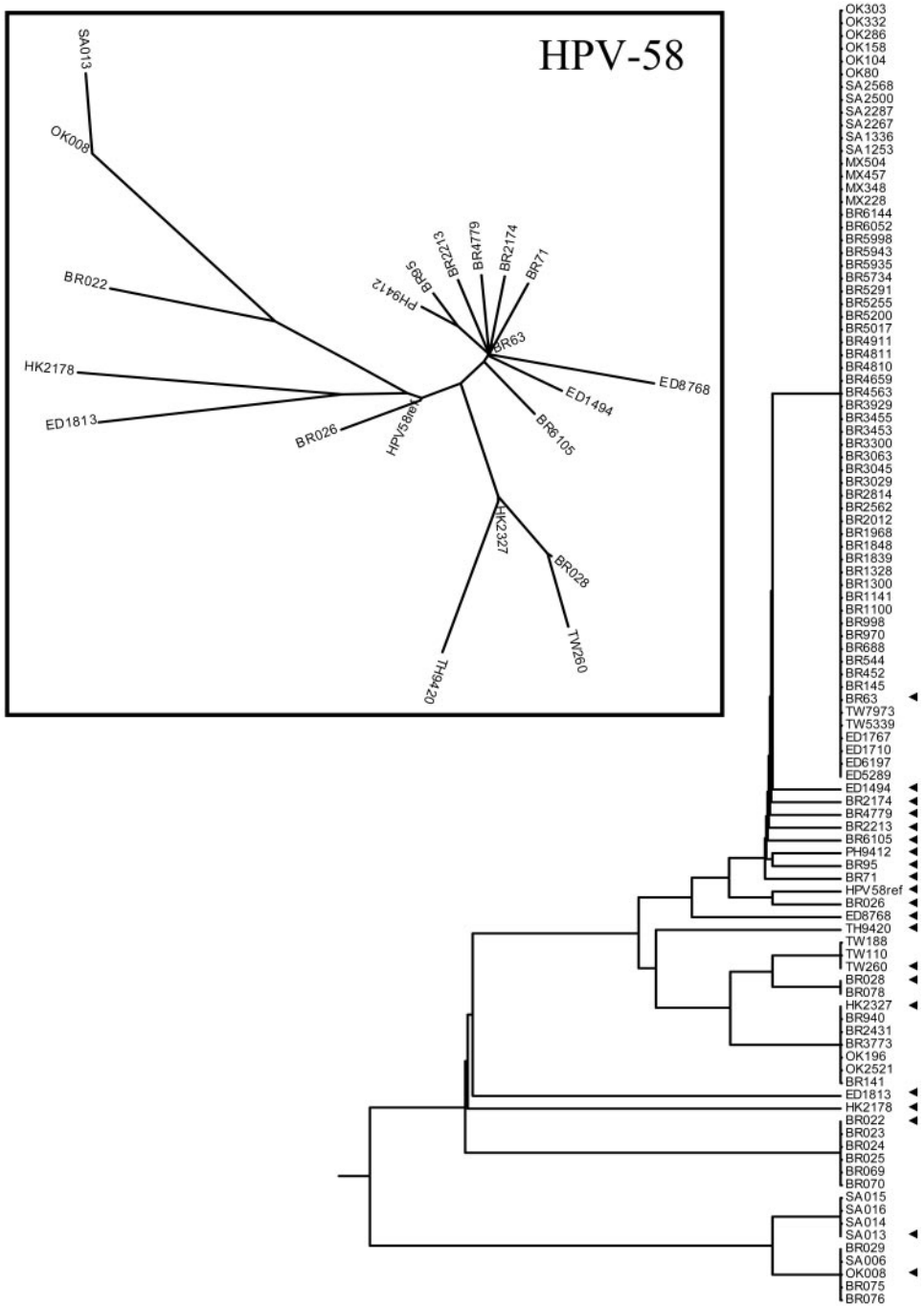

FIG. 6. The intratype diversity of HPV-58. The phylogenetic trees represent the relationship between HPV-52 variants based on a 461-bp segment of the LCR. The large tree is based on the UPGMA, and the small tree is based on the neighbor-joining algorithm. For further details, see the legend to Fig. 1.

Africa, Europe, and Asia. These widespread variants may have existed well before the spread of humans out of Africa, while clusters that are specific for ethnic groups may have evolved subsequently in specific locations.

Our data support the notion that HPV sequences encoding proteins developed a lower diversity than the noncoding LCR. However, our limited study of the E6 and L1 genes of the most distant variants of each type, aside from prototype sequences, indicate a level of 0.7 to 2.1% diversity on the level of protein sequences. This finding points to the possibility of functional differences between variants within each type, which is of relevance for epidemiological, etiological, pharmaceutical, and vaccination research. It should be noted that molecular as well as epidemiological studies have identified quite significant

	L1				
H	6	6	6	6	
P	5	6	7	8	
V	8	9	9	1	
31	6	1	6	7	
reference	**T**	**C**	**G**	**C**	Prototype
HK31A				A	Prototype
HK31B				A	Prototype
HK31C				A	Prototype
BR1213	G		A	A	Prototype
TL2069				A	Prototype
MR9917	G	T	A	A	Prototype
BR1692	G		A	A	Prototype

	E6						
H	0	0	0	0	0	0	
P	1	2	3	4	4	5	
V	7	8	2	0	2	2	
31	6	5	0	4	8	0	
reference	**C**	**C**	**A**	**G**	**A**	**C**	Prototype
HK31A							Prototype
HK31B	T						Prototype
HK31C							Prototype
BR1213		T	T	A	G	T	H60Y/A138V
TL2069							Prototype
MR9917		T	T	A		T	H60Y/A138V
BR1692		T	T	A	G	T	H60Y/A138V

	L1	
H	6	
P	6	
V	4	
35	2	
reference	**T**	Prototype
NW3793	A	S348T
NW1215	A	S348T
NW1301	A	S348T
NW2760	A	S348T
TW4232		Prototype
SA1505		Prototype

	E6					
H	0	0	0	0	0	
P	1	1	1	2	3	
V	2	3	3	9	4	
35	7	1	6	5	1	
reference	**T**	**C**	**T**	**A**	**T**	Prototype
NW3793	C	A	C	T	C	W78R
NW1215	C	A	C	T	C	W78R
NW1301	C	A	C	T	C	W78R
NW2760	C	A	C	T	C	W78R
TW4232		A		T		Prototype
SA1505		A		T		Prototype

	L1											
H	6	6	6	6	6	6	6	6	6	6	6	
P	7	7	7	7	7	7	8	9	9	9	9	
V	0	0	1	1	6	9	2	1	2	3	4	
52	1	3	1	2	4	4	4	7	0	5	1	
reference	T	A	A	G	T	A	C	C	T	T	A	Prototype
ED123												Prototype
ED18604												Prototype
HK1151	G				C	G	T	A				Prototype
HK1243	G				C	G	T	A				Prototype
HK2571	G				C	G	T	A				Prototype
BR0258		C	G	A				A	G	A	G	K380T/S383D

	E6								
H	0	0	0	0	0	0	0	0	
P	2	2	3	3	3	4	4	4	
V	0	3	5	7	7	2	6	7	
52	0	7	0	8	9	5	7	1	
reference	C	C	G	A	A	G	C	C	Prototype
ED123									Prototype
ED18604									Prototype
HK1151			T	C	G		A		K93E/N122K
HK1243			T		G				K93R
HK2571			T		G				K93R
BR0258	T	T	T			T		A	Prototype

	L1							
H	6	6	6	6	6	6	6	
P	6	7	8	8	8	8	9	
V	4	9	2	2	2	6	5	
58	1	8	2	7	8	0	0	
reference	G	A	G	C	A	A	G	Prototype
OK332	A							Prototype
ED1767	A						A	Prototype
ED1710	A						A	Prototype
ED6197	A							Prototype
ED1813		G	A	A	G			I412V/D420N/N422D
HK2178						G		Prototype

	E6			
H	0	0	0	
P	2	3	3	
V	0	0	6	
58	3	7	7	
reference	G	C	C	Prototype
OK332		T		Prototype
ED1767		T		Prototype
ED1710		T		Prototype
ED6197		T		Prototype
ED1813	C	T	A	E32Q/D86E
HK2178				Prototype

FIG. 7. Diversity of the E6 genes (four panels on the right side of the figure) and part of the L1 genes (left side of the figure) in distantly related variants of HPV-31, -35, -52, and -58. Within each panel, the first column lists the variants, whose relative phylogenetic position can be found in Fig. 1, 3, 5, and 6. The central part of the figure identifies nucleotide exchanges (letters) or maintenance of the sequence of the reference genome (gray squares). The box on the right side of each panel indicates whether the amino acid sequence of the reference clone has been maintained ("prototype") and if not, what kind of amino acid exchanges have occurred.

functional differences between variants of HPV-16, which seem to stem from diversity in the LCR as well as in protein-encoding regions of the genome (13, 17–19, 28, 29, 31, 32–34).

ACKNOWLEDGMENTS

This research was supported by funds of the Ludwig Institute for Cancer Research to L.L.V., by NIH grant ROI CA-91964, and by funds from the Chao Family Comprehensive Cancer Center of the University of California—Irvine to H.-U.B.

REFERENCES

1. **Antonsson, A., O. Forslund, H. Ekberg, G. Sterner, and B. G. Hansson.** 2000. The ubiquity and impressive genomic diversity of human skin papillomaviruses suggest a commensalic nature of these viruses. J. Virol. **74:**11636–11641.
2. **Bauer, H. M., Y. Ting, C. E. Greer, J. C. Chambers, C. J. Tashiro, J. Chimera, A. Reingold, and M. M. Manos.** 1991. Genital human papillomavirus infection in female university students as determined by a PCR-based method. JAMA **265:**472–477.
3. **Bernard, H. U.** 1994. Coevolution of papillomaviruses and human populations. Trends Microbiol. **2:**140–143.
4. **Bernard, H. U., S. Y. Chan, M. M. Manos, C. K. Ong, L. L. Villa, H. Delius, H. M. Bauer, C. Peyton, and C. M. Wheeler.** 1994. Assessment of known and novel human papillomaviruses by polymerase chain reaction, restriction digest, nucleotide sequence, and phylogenetic algorithms. J. Infect. Dis. **170:**1077–1085.
5. **Calleja-Macias, I. E., M. Kalantari, J. Huh, R. Ortiz-Lopez, A. Rojas-Martines, J. F. Gonzales-Guerrero, A. L. Williamson, B. Hagmar, D. J. Wiley, L. Villarreal, H. U. Bernard, and H. A. Barrera-Saldana.** 2004. High prevalence of specific variants of human papillomavirus-16, 18, 31, and 35 in a Mexican population. Virology **319:**315–323.
6. **Calleja-Macias, I. E., M. Kalantari, B. Allan, A. L. Williamson, L. P. Chung, R. J. Collins, R. E. Zuna, S. T. Dunn, R. Ortiz-Lopez, H. A. Barrera-Saldaña, H. A. Cubie, L. L. Villa, and H. U. Bernard.** 2005. Papillomavirus subtypes are natural and old taxa: phylogeny of the human papillomavirus (HPV) types 44/55 and 68a/b. J. Virol. **79:**6565–6569.
7. **Cerqueira, D. M., G. N. de Lima Camara, M. R. da Cruz, E. O. Silva, M. de Macedo Brigido, L. G. de Souza Carvalho, and C. R. F. Martins.** 2003. Variants of human papillomavirus types 53, 58, and 66 in central Brazil. Virus Genes **26:**83–87.
8. **Chan, S. Y., S. H. Chew, K. Egawa, E. I. Grussendorf-Conen, Y. Honda, A. Ruebben, K. C. Tan, and H. U. Bernard.** 1997. Phylogenetic analysis of the human papillomavirus type 2 (HPV-2), HPV-27, and HPV-57 group, which is associated with common warts. Virology **239:**296–302.
9. **Chan, S. Y., H. Delius, A. L. Halpern, and H. U. Bernard.** 1995. Analysis of genomic sequences of 95 papillomavirus types: Uniting typing, phylogeny, and taxonomy. J. Virol. **69:**3074–3083.
10. **Deau, M. C., M. Favre, S. Jablonska, L. A. Rueda, and G. Orth.** 1993. Genetic heterogeneity of oncogenic human papillomavirus type 5 (HPV5) and phylogeny of HPV5 variants associated with epidermodysplasia verruciformis. J. Clin. Microbiol. **31:**2918–2926.
11. **de Villiers, E. M., C. Fauquet, T. R. Broker, H. U. Bernard, and H. zur Hausen.** 2004. Classification of papillomaviruses. Virology **324:**17–27.
12. **Gagnon, S., C. Hankins, Tremblay. C., K. Pourreaux, P. Forest, F. Rouah, and F. Coutlee.** 2005. Polymorphism of human papillomavirus type 31 isolates infecting the genital tract of HIV-seropositive and HIV-seronegative women at risk for HIV infection. J. Med. Virol. **75:**213–221.
13. **Hecht, J. L., A. S. Kadish, G. Jiang, and R. D. Burk.** 1995. Genetic characterization of the human papillomavirus (HPV) 18 E2 gene in clinical specimens suggests the presence of a subtype with decreased oncogenic potential. Int. J. Cancer **60:**369–376.
14. **Heinzel, P. A., S. Y. Chan, L. Ho, M. O'Connor, P. Balaram, M. S. Campo, K. Fujinaga, N. Kiviat, J. Kuypers, H. Pfister, B. M. Steinberg, S. K. Tay, L. L. Villa, and H. U. Bernard.** 1995. Variation of human papillomavirus type 6 (HPV-6) and HPV-11 genomes sampled throughout the world. J. Clin. Microbiol. **33:**1746–1754.
15. **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.
16. **Ho, L., S. Y. Chan, R. D. Burk, B. C. Das, K. Fujinaga, J. P. Icenogle, T. Kahn, N. Kiviat, W. Lancaster, P. Mavromara, V. Labropoulou, S. Mitrani-Rosenbaum, B. Norrild, M. R. Pillai, J. Stoerker, K. Syrjaenen, S. Syrjaenen, S. K. Tay, L. L. Villa, C. M. Wheeler, A. L. Williamson, and H. U. Bernard.** 1993. The genetic drift of human papillomavirus type 16 is a means of reconstructing prehistoric viral spread and movement of ancient human populations. J. Virol. **67:**6413–6414.
17. **Hubert, W. G.** 2005. Variant upstream regulatory region sequences differentially regulate human papillomavirus type 16 DNA replication throughout the viral life cycle. J. Virol. **79:**5914–5922.
18. **Kammer, C., U. Warthorst, N. Torrez-Martinez, C. M. Wheeler, and H. Pfister.** 2000. Sequence analysis of the long control region of human papillomavirus type 16 variants and functional consequences for P97 promoter activity. J. Gen. Virol. **81:**1975–1981.
19. **Kammer, C., M. Tommasino, S. Syrjanen, H. Delius, U. Hebling, U. Warthorst, H. Pfister, and I. Zehbe.** 2002. Variants of the long control region and the E6 oncogene in European human papillomavirus type 16 isolates: implications for cervical disease. Br. J. Cancer **86:**269–273.
20. **Lorincz, A. T., W. D. Lancaster, and G. F. Temple.** 1986. Cloning and characterization of the DNA of a new human papillomavirus from a woman with dysplasia of the uterine cervix. J. Virol. **58:**225–229.
21. **Lorincz, A. T., A. P. Quinn, W. D. Lancaster, and G. F. Temple.** 1987. A new type of papillomavirus associated with cancer of the uterine cervix. Virology **159:**187–190.
22. **Matsukura, T., and M. Sugase.** 1990. Molecular cloning of a novel human papillomavirus (type 58) from an invasive cervical carcinoma. Virology **177:**833–836.
23. **Munoz, N., F. X. Bosch, S. de Sanjosé, R. Herrero, X. Castellsagué, K. V. Shah, P. J. F. Snijders, and C. J. L. M. Meijer.** 2003. Epidemiological classification of human papillomavirus types associated with cervical cancer. N. Engl. J. Med. **348:**518–527.
24. **Ong, C. K., S. Y. Chan, M. S. Campo, K. Fujinaga, P. Mavromara, V. 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.
25. **Pearson, W. R., T. Wood, Z. Zhang, and W. Miller.** 1997. Comparison of DNA sequences with protein sequences. Genomics **46:**24–36.
26. **Prado, J. C., I. E. Calleja-Macias, H. U. Bernard, M. Kalantari, S. A. Macay, B. Allan, A. L. Williamson, L. P. Chung, R. J. Collins, R. E. Zuna, S. T. Dunn, R. Ortiz-Lopez, H. A. Barrera-Saldaña, H. A. Cubie, K. Cuschieri, von M. Knebel-Doeberitz, G. I. Sanchez, F. X. Bosch, and L. L. Villa.** 2005. Worldwide genomic diversity of the human papillomaviruses-53, 56, and 66, a group of high-risk HPVs unrelated to HPV-16 and HPV-18. Virology **340:**95–104.
27. **Shimoda, K., A. T. Lorincz, G. F. Temple, and W. D. Lancaster.** 1988. Human papillomavirus type 52: a new virus associated with cervical neoplasia. J. Gen. Virol. **69:**2925–2928.
28. **Sichero, L., E. L. Franco, and L. L. Villa.** 2005. Different P105 promoter activities among natural variants of human papillomavirus type 18. J. Infect. Dis. **191:**739–742.
29. **Stewart, A. C., A. M. Eriksson, M. M. Manos, N. Munoz, F. X. Bosch, J. Peto, and C. M. Wheeler.** 1996. Intratype variation in 12 human papillomavirus types: a worldwide perspective. J. Virol. **70:**3127–3136.
30. **Van Ranst, M., A. Fuse, P. Fiten, E. Beuken, H. Pfister, R. D. Burk, and G. Opdenakker.** 1992. Human papillomavirus type 13 and pigmy chimpanzee papillomavirus type 1: comparison of the genome organization. Virology **190:**587–596.
31. **Villa, L. L., L. Sichero, P. Rahal, O. Caballero, A. Ferenczy, T. Rohan, and E. L. Franco.** 2000. Molecular variants of human papillomavirus types 16 and 18 preferentially associated with cervical neoplasia. J. Gen. Virol. **81:**2959–2968.
32. **Xi, L. F., C. W. Critchlow, C. M. Wheeler, L. A. Koutsky, D. A. Galloway, J. Kuypers, J. P. Hughes, S. E. Hawes, C. Surawicz, G. Goldbaum, K. K. Holmes, and N. B. Kiviat.** 1998. Risk of anal carcinoma in situ in relation to human papillomavirus type 16 variants. Cancer Res. **58:**3839–3844.
33. **Xi, L. F., L. A. Koutsky, D. A. Galloway, J. Kuypers, J. P. Hughes, C. M. Wheeler, K. K. Holmes, and N. B. Kiviat.** 1997. Genomic variation of human papillomavirus type 16 and risk for high grade cervical intraepithelial neoplasia. J. Natl. Cancer Inst. **89:**796–802.
34. **Yamada, T., M. M. Manos, J. Peto, C. E. Greer, N. Munoz, F. X. Bosch, and C. M. Wheeler.** 1997. Human papillomavirus type 16 sequence variation in cervical cancers: a worldwide perspective. J. Virol. **71:**2463–2472.