Journal of Virology, Apr. 1992, p. 2057–2066 Vol. 66, No. 4
0022-538X/92/042057-10$02.00/0

Molecular Variants of Human Papillomavirus Type 16 from Four Continents Suggest Ancient Pandemic Spread of the Virus and Its Coevolution with Humankind

SHIH-YEN CHAN,[1] LISA HO,[1] CHI-KEONG ONG,[1] VINCENT CHOW,[2] BERND DRESCHER,[3] MATTHIAS DÜRST,[4] JAN TER MEULEN,[4] LUISA VILLA,[5] JEFF LUANDE,[6] HANS N. MGAYA,[7] AND HANS-ULRICH BERNARD[1]*

Institute of Molecular and Cell Biology[1] and Department of Microbiology,[2] National University of Singapore, Singapore 0511, Singapore; Department of Molecular Biophysics[3] and Institute for Virus Research,[4] German Cancer Research Center, Heidelberg, Germany; Ludwig Institute for Cancer Research, São Paulo, Brazil[5]; and Tanzania Tumor Center[6] and Muhimbili Medical Center,[7] Dar es Salaam, Tanzania

Received 13 September 1991/Accepted 18 December 1991

We have amplified by the polymerase chain reaction, cloned, and sequenced genomic segments of 118 human papillomavirus type 16 (HPV-16) isolates from 76 cervical biopsy, 14 cervical smear, 3 vulval biopsy, 2 penile biopsy, 2 anal biopsy, and 1 vaginal biopsy sample and two cell lines. The specimens were taken from patients in four countries—Singapore, Brazil, Tanzania, and Germany. The sequence of a 364-bp fragment of the long control region of the virus revealed 38 variants, most of which differed by one or several point mutations. Phylogenetic trees were constructed by distance matrix methods and a transformation series approach. The trees based on the long control region were supported by another set based on the complete E5 protein-coding region. Both sets had two main branches. Nearly all of the variants from Tanzania were assigned to one (African) branch, and all of the German and most of the Singaporean variants were assigned to the other (Eurasian) branch. While some German and Singaporean variants were identical, each group also contained variants that formed unique branches. In contrast to the group-internal homogeneity of the Singaporean, German, and Tanzanian variants, the Brazilian variants were clearly divided between the two branches. Exceptions to this were the seven Singaporean isolates with mutational patterns typical of the Tanzanian isolates. The data suggest that HPV-16 evolved separately for a long period in Africa and Eurasia. Representatives of both branches may have been transferred to Brazil via past colonial immigration. The comparable efficiencies of transfer of the African and the Eurasian variants to the New World suggest pandemic spread of HPV-16 in past centuries. Representatives of the African branch were possibly transferred to the Far East along old Arab and Indonesian sailing routes. Our data also support the view that HPV-16 is a well-defined virus type, since the variants show only a maximal genomic divergence of about 5%. The small amount of divergence in any one geographic location and the lack of marked divergence between the Tanzanian and Brazilian African genome variants two centuries after their likely introduction into the New World suggest a very slow rate of viral evolution. The phylogenetic tree therefore probably represents a minimum of several centuries of evolution, if not an age equal to that of the respective human races.

Human papillomavirus type 16 (HPV-16) is the most frequently encountered member of a group of papillomaviruses that are associated with neoplasias of the genital epithelia, particularly with benign and malignant lesions of the cervix uteri (64). The virus is believed to be casually associated for several reasons. Its genome or that of related types is present in most if not all cervical lesions (57). Epidemiological data indicate the involvement of an infectious agent (44). Finally, the products of the viral genes E5, E6, and E7 disrupt cellular homeostasis in a manner typical of various tumor viruses (22, 35, 38, 61).

Cervical cancer is one of the most frequent cancers in women, although its incidence rate shows regional differences of nearly one order of magnitude (31, 60). Genital papillomaviruses have been found in all populations analyzed, and HPV-16 is usually the predominant type (57). Genital neoplasias have been recorded as much as 2,000 years ago (58), and thus a similar age may be inferred for the commonest likely causal agent. Therefore, HPV-16 and related genital papillomaviruses may have been present in humans throughout recorded history, if not throughout the period of human evolution.

Such a scenario has to be compared with the alternative, namely, evolution of the virus in animals and recent transfer across the species barrier. Some viruses are believed to cross the human-animal species barrier frequently. This may involve the creation of new genotypes, as for influenza viruses (39). Others may have done so only once or rarely. It is well documented that the human immunodeficiency virus types 1 and 2 (HIV-1 and HIV-2) originated in two distinct regions of Africa and probably appeared in humans only recently (53). Measles and smallpox, probably needing human populations of a certain minimal size, seem to have appeared in Europe or Western Asia about 1,500 years ago (36, 41). They first became widespread in the Old World, from which they were carried by Europeans to the New World. The writings of Hippocrates indicate the presence of herpesvirus in humans 2,500 years ago (41), and Egyptian relics suggest an even older age for poliovirus (1, 32).

* Corresponding author.

It is not yet known whether pathogens like poliovirus, herpesvirus, and papillomaviruses coevolved with the human species, but dormancy or long latencies in the case of herpesviruses and papillomaviruses could be a prerequisite. This would ensure their continued presence in isolated populations without eradication by tribal immunity. In order to reconstruct the pathways of spread of any particular pathogen, the disease should have an acute onset and its signs should make it recognizable in historical descriptions. Both conditions are met for diseases like smallpox and the plague but not for HPV-16 infections.

Comparison of genomic sequences of viral isolates has become a powerful epidemiological and phylogenetic tool. Its increasing use is beginning to bring the fields of retrospective epidemiology and phylogenetics closer together. Different isolates frequently differ in some detail of their genomic sequence. Mutational changes may occur so rapidly in RNA viruses that isolates from one individual may already differ greatly from isolates from another who was the source of the infection. Despite this limitation, researchers have been able to trace the origin of HIV-1 infections in local populations (4), reconstruct the phylogeny of HIV-1, HIV-2, and the simian immunodeficiency virus (53, 55), and follow the spread of dengue virus (46).

All mammal, bird, and reptile species that have been carefully studied carry species-specific papillomaviruses (57). Furthermore, most species are infected by several papillomavirus types, e.g., more than 60 in the case of humans. Many of these viruses are specific for particular epithelial cell subtypes. Sequence analysis shows that virus types that infect the same or similar epithelial subtypes (e.g., HPV-16 compared with HPV-18) are more closely related than those that infect different epithelia (e.g., HPV-16 compared with HPV-1) (13). Consequently, viral evolution was linked not only to host species but also to a particular target tissue. We believe that a comparison of papillomavirus genomes worldwide is likely to shed light on its epidemiology and phylogeny. A detailed gene phylogeny is an important first step towards elucidating the viral phylogeny and the nature of the host-parasite relationship. Toward this end, we have attempted to study the evolution of HPV-16 with humans as reflected in the sequence variation and geographic distribution of present-day isolates.

MATERIALS AND METHODS

Plasmids and cell lines. The prototype sequence of HPV-16 was derived from the original viral isolate (15) and cloned in the form of a *Bam*HI fragment into pSP64. The sequencing of this plasmid revealed a mistake in the original published sequence (52). The result is that the sequence 3′ of position 7761 (TCTAGG) becomes TCTAAGG (26). Nucleotide numbers 3′ of position 7764 were adjusted accordingly, e.g., position 7769 became position 7770. In addition, several groups (3, 9, 23, 47) have published corrections to the original sequence that were outside the HPV-16 long control region (LCR). To avoid confusion, our numbering, with one exception, was not adjusted to reflect these changes. One of these corrections (23) led to the repositioning of the E5 open reading frame, and we have followed the numbering system of that reference. The cell lines SiHa and CaSki, with endogenous HPV-16 copies, have been cultured in our laboratory for several years.

Biopsy and smear samples. (i) Singapore. Smear samples from asymptomatic patients (Ss-x) and biopsy samples from lesions of various degrees of neoplastic progression (Sb-x and Sv-x) have been listed previously except for Sb-23, Sb-24, Sb-25, and Sv-1. All except two samples were from Chinese patients, who are often second- to fourth-generation descendants of migrants from China. Sb-6 was from a Malay patient, and Sb-15 was from a Japanese patient.

(ii) Brazil. Smear samples from asymptomatic patients (Bs-x) and biopsy samples from lesions (Bb-x) were from northeast Brazil (Recife and João Pessoa), except for Bb-1 to Bb-7 and Bb-13 to Bb-15, which were from São Paulo. All biopsy samples were from invasive cervical carcinomas with the exception of one vaginal (Bb-20), two vulvar (Bb-1 and Bb-2), two penile (Bb-12 and Bb-14), and two anal (Bb-4 and Bb-10) carcinoma samples.

(iii) Tanzania. Smear samples from asymptomatic patients (Ts-x) and biopsy samples from invasive cervical carcinomas (Tb-x) were taken at various hospitals in Dar es Salaam from patients who originated from different geographic regions of Tanzania.

(iv) Germany. Biopsy samples from invasive cervical carcinomas (Gb-x) were obtained from patients living in the cities of Düsseldorf, Frankfurt, Freiburg, Hamburg, and Ulm.

PCR and DNA sequencing. The polymerase chain reaction (PCR) was performed as described previously (12, 54). Amplification of a 364-bp HPV-16 segment (viral transcriptional enhancer) from positions 7478 to 7841 was done with previously described primers (26). This permitted cloning of the amplified fragment with *Asp* 718 and *Xba*I ends into pUC19. A genomic segment flanking the E5 open reading frame was amplified with two primers, (A) 5′-GGGGATCCC AGTGTCTACTGGATTTGTGTC-3′ and (B) 5′-GGAAGCT TCGATGCACGTTTTGTGCGTT-3′. These corresponded to positions 3825 to 3846 (upper strand) and 4278 to 4259 (lower strand) of HPV-16, respectively. Artificial *Bam*HI and *Hin*dIII sites were positioned at the respective 5′ ends of the primers and used to clone the amplified fragment into pUC18. For purposes other than those addressed in this study, primer A also contained an A to G mutation at position 3842. Sequence information was obtained and analyzed only for positions 3850 to 4101, which is the coding sequence of the E5 gene. Both strands of the supercoiled template were sequenced as described previously (26), using a modified dideoxynucleotide sequencing protocol, T7 DNA polymerase (Pharmacia), and M13/pUC forward and reverse sequencing primers (Boehringer Mannheim).

Phylogeny construction and evaluation. Phylogenies for both the enhancer and E5 segments were constructed by several methods.

(i) Distance matrix analysis. Sequences between all pairs of variants were aligned by the method of Wilbur and Lipman (62). The resultant similarity scores were transformed into a distance matrix by the Kimura two-parameter model (30). This was then used to construct phylogenies by two methods: the unweighted pair-group method with arithmetic average (UPGMA) cluster analysis (56), which was executed with the CLUSTAL V multiple alignment package (24); and the neighbor-joining method (48), which was executed by the NEIGHBOR option in the PHYLIP 3.4 phylogeny inference package of Felsenstein (19) (data not shown).

(ii) Transformation series analysis. When a string of characters, e.g., nucleotide positions, may possess multiple states (G, A, T, or C), it may be possible to order them unambiguously in an undirected transformation series if there are few homoplasies (convergence and reversion events). For example, the three sequences GATCA, GATGA, and GTTGA would suggest the transformation series GATCA→GATGA→GTTGA or GATCA←GATGA←GTTGA. However, unless one had independent data as to the ancestral states (e.g., an ancestral out-group sequence), the polarity of the series is undetermined. Such an approach was used to produce unrooted phylogenetic trees from third-chromosome rearrangements in *Drosophila pseudoobscura* and *D. persimilis* (2) and from human mitochondrial DNA restriction fragment length polymorphisms (16). In our unrooted tree, variants were serially linked to each other according to the parsimony criterion of the least number of changes required to transform one sequence into its nearest neighbor. Linking was done by inspection, and the tree presented is but one of several that are possible. Relative frequency of the variants was not taken into account. (The resultant phylogeny was identical to that of a maximum-likelihood phylogeny generated later by using the DNAML option in PHYLIP 3.4.)

The distinction between the African and Eurasian lineages in the phylogeny was tested for significance by bootstrap resampling (19) of 100 replicates and found to be significant at the 95% confidence level [SEQBOOT, DNADIST(K2), NEIGHBOR, and CONSENSE options in PHYLIP 3.4].

Nucleotide sequence accession numbers. The sequence variants were submitted to GenBank by using the AUTHORIN 2.1 (IntelliGenetics) package and assigned the accession numbers shown in parentheses.

LCR variants: CaSki (M83776/M83840), SiHa (M83777), Sb-5 (MM83778), Ss-1a (M83835), Sb-13 (M83836), Sb-14 (M83837), Sb-1 (M83838), Sb-7 (M83839), Sb-21 (M83841), Sb-16 (M83842), Sb-21a (M83843), Bb-8a (M83844), Bb-24 (M83834/M83845), Bb-3 (M83846), Bb-1 (M83847), Bb-14 (M83848), Bb-19 (M83849), Gb-13 (M83850), Gb-10 (M83851), Gb-21a (M83852), Tb-8 (M83853), Tb-13 (M83854), Tb-4 (M83855), Ts-6a (M83856), Ts-6b (M83857), Tb-16 (M83858), Tb-15 (M83859), Ts-5 (M83860), Ts-5a (M83861), Ts-5b (M83862/M83863), Ts-5c (M83864), Ts-3 (M83865), Ts-3a (M83866), Ts-3b (M83867), Ts-3c (M83868), and Tb-1 (M83885).

E5 variants: Sb-5 (M83869), Sb-2 (M83870), CaSki (M83871), Sb-13 (M83872/M83873), SiHa (M83874), Sb-7 (M83875), Sb-17 (M83876), Sb-10 (M83877), Sb-21a (M83878/M83879), Sb-16 (M83880), Tb-4 (M83881), Tb-16 (M83883), and Tb-13 (M83884).

RESULTS

Variability in the LCR of HPV-16. Recently, we observed that corresponding segments of the LCR of independent HPV-16 isolates differed by multiple point mutations (26). We speculated that this segment might be hypervariable, since genomic sequences generally diverge more in noncoding than in coding sequences (34), and that the substitutions would be largely neutral, since only a fraction of the base pairs were functionally constrained as *cis*-responsive elements (11, 21). In our diagnostic study, these findings had been useful in ruling out the possibility of exogenous contaminations with cloned viral DNA and proving multiple infections of a patient with the same HPV type.

The isolates for this previous study had been obtained from a mostly Chinese patient population in Singapore and from a mixed European and African population in São Paulo and northeast Brazil. Interestingly, comparison of the mutation patterns in most of these genomic variants suggested an ancestor-descendant relationship through successive single-nucleotide substitution events. A transformation series network seemed to place HPV-16 variants from both cohorts on all major branches, but some minor branches were suggestive of geographically distinct variants. To investigate this further, we extended our study in three ways. We collected HPV-16 isolates from two additional cohorts, sequenced characteristic LCR variants in another genomic region, and then applied several different methods to the analysis of our data.

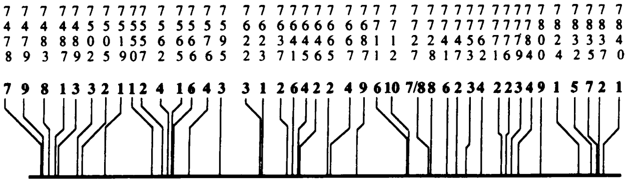

FIG. 1. Linear representation and identification of all point mutations in the 364-bp enhancer fragment of HPV-16 (positions 7478 to 7841). The four numbers set vertically above each column identify the position of the mutation. The single boldface number immediately above the vertical line is a numerical code for the base exchange identified in the sequence of the upper strand, as follows: 1, G to A; 2, A to G; 3, T to C; 4, C to T; 5, G to T; 6, T to G; 7, A to T; 8, A to C; 9, C to A; 10, T to A; 11, C to G.

In total, 40 isolates from Singapore, 29 from Brazil, 28 from Tanzania, and 19 from Germany were analyzed. As in our previous study, each sample was amplified, cloned, and sequenced twice independently. Rare cases of conflict between two sequences from the same sample were resolved by sequencing additional clones. Variants that were found at least twice were counted as "natural" variants, and those found only once were counted as potential PCR mutation artifacts. Two or more variants each identified at least twice in the same biopsy sample were scored as cases of multiple infection. These are identified by adding a lowercase letter as a suffix to the sample code, e.g., Sb-21a. In general, we considered a possible confounding effect by PCR-derived mutations unlikely because of the very low error rate that we had determined previously (26). This has since been supported by similar E5 and E7 gene control studies (data not shown).

Figure 1 summarizes the location and the nature of all 42 point mutations recorded in the 118 HPV-16 isolates. Figure 2 represents the combinations of these mutations and of four additional deletions and insertions that had occurred in individual isolates. There are several observations to be made. HPV-16 LCR segments that did not differ from that of the original isolate (15) (referred to as the prototype solely to indicate this fact) were found commonly in samples from Brazil and Singapore but only once in a Tanzanian and twice in a German sample. A mutation at position 7519 was especially common. Variants with no other changes but this were predominant in Germany and Singapore. Altogether, the prototype and variant 7519 represented 46 of the 118 isolates studied. The Singaporean and Brazilian isolates seem to fall into two groups, one having few (maximally four) and the other many (seven or more) mutations relative to the prototype. The group that has few mutations seems to be exclusively represented in Germany, and the one that is abundant in mutations is represented nearly exclusively in Tanzania.

Relationship among 38 HPV-16 LCR variants. Figure 3 shows the UPGMA phylogeny as inferred from the LCR variations. The most striking feature is the division of the tree into two main lineages. All Tanzanian variants except two, Tb-7 (prototype) and Tb-6 (variant 7519), are assigned

	7478	7479	7483	7487	7489	7502	7505	7519	7550	7557	7562	7565	7566	7576	7595	7622	7623	7637	7641	7645	7646	7665	7667	7687	7711	7712	7727	7728	7741	7747	7753	7762	7771	7776	7779	7784	7800	7824	7832	7835	7837	7840	MISC
CaSki								■																																			
SiHa								■																																			Δ
Sb-4, Sb-8, Sb-9, Sb-12, Ss-2, Ss-3, Ss-4, Ss-5, Ss-6, Ss-7																																											
Sb-5																						■																					
Ss-1a	■	■												■				■																									
Sb-2, Sb-3, Sb-8a, Sb-18, Sb-20, Ss1, Ss-7a								■																																			
Sb-13, Sb-22								■																																		■	
Sb-7, Sb-11, Sb-17, Sb-19								■																				■														■	
Sb-1, Sb-6, Sb-10, Sb-12a, Sb-15, Sb-24, Sb-25								■																				■							■							■	
Sb-14					■			■																													■					■	
Sb-23				■				■																		■						■				■			■				
Sb-21			■	■				■															■									■				■							
Sb-16, Ss-6a, Ss-8			■	■				■															■	■			■		■			■				■							
Sb-21a, Sv-1			■	■				■															■									■				■		■	■	■	■		
Bb-4, Bb-5, Bb-6, Bb-8, Bb-10, Bb-13, Bb-15, Bb-18, Bs-2																																											
Bb-8a																																											Δ
Bb-21																						■																					
Bb-24																															■			■									
Bb-12								■																																			
Bb-3								■					■																														
Bb-5a								■																																		■	
Bb-11, Bb-16			■	■				■															■	■			▨		■			■				■							
Bb-2, Bb-7, Bb-9, Bb-17, Bb-20, Bb-22, Bb-23, Bb-25, Bs-1			■	■				■															■	■			■		■			■				■							
Bb-1			■	■				■															■	■			■		■			■				■							I
Bb-14			■	■			■	■															■	■			■					■				■							
Bb-19			■	■			■	■															■	■			■		■			■				■							
Tb-7																																											
Tb-6, Tb-17, Ts-2								■																																			
Tb-8, Tb-11				■				■																■								■				■			■				
Tb-13				■				■																		■						■				■			■				
Tb-4, Tb-9, Tb-14, Tb-19, Ts-6				■				■																■		■						■				■			■				
Ts-6a				■				■		■														■		■						■				■			■				
Ts-6b				■				■													■			■		■						■				■			■				
Tb-16				■				■															■	■								■				■			■				
Tb-15				■				■															■	■								■				■			■				Δ
Ts-5			■	■				■											■					■								■				■		■	■	■			
Ts-5a			■	■				■											■				■	■								■				■		■	■	■			
Ts-5b			■	■				■								■			■				■	■								■				■		■	■	■			
Ts-5c			■	■				■											■	■			■	■								■				■		■	■	■			
Tb-1, Tb-2, Tb-5, Tb-18			■	■				■															■	■								■				■		■	■	■	■		
Ts-3			■	■				■				■											■	■								■				■		■	■	■	■		
Ts-3a			■	■		■		■							■								■	■								■				■		■	■	■	■		
Ts-3b			■	■				■															■	■						■		■				■		■	■	■	■		
Ts-3c			■	■				■			■												■	■								■				■		■	■	■	■		
Gb-11, Gb-15																																											
Gb-1, Gb-3, Gb-4, Gb-5, Gb-6, Gb-7, Gb-8, Gb-14, Gb-19, Gb-21, Gb-22, Gb-23								■																																			
Gb-13								■																	■																		
Gb-10, Gb-12								■	■																																		
Gb-21a, Gb-12a								■	■								■																■										

FIG. 2. Combined representation of the LCR segment from all 118 HPV-16 isolates that were studied. Each row represents one genomic variant. The first column specifies the alphanumeric code of the isolate. The numbers at the top identify the genomic position, where mutations as identified in Fig. 1 have occurred. Black squares indicate the presence and white squares indicate the absence of the mutation relative to the prototype sequence. The grey square at position 7727 identifies a change from A to T. Groupings separated by shaded bars are (top to bottom) cell lines and the isolates from Singapore, Brazil, Tanzania, and Germany. The symbols Δ and I refer to deletions and insertions, respectively, in addition to the point mutations present. The extent of these has been described previously (21).

to the upper lineage. Also assigned to a subbranch of this lineage are five variants (Bb-11 group, Bb-2 group, Bb-1, Bb-14, and Bb-19) from 14 Brazilian isolates. All of them have six mutations (positions 7483, 7487, 7667, 7687, 7762, and 7784) that are absent from all the other Brazilian isolates. Six Singaporean isolates (Sb-16, Sb-21, Sb-21a, Ss-6a, Ss-8, and Sv-1) have five of these mutations in common and are also assigned to this branch. The lower lineage contains all samples from Germany, 85% of the Singaporean samples, and those Brazilian samples that differed by only a small number of mutations from the prototype.

Figure 4 shows the transformation series phylogeny of the LCR variants. One can discern three branches, African, Brazilian, and Eurasian, because there is evidently more variation between than within populations. The two Tanzanian subgroups centered on Tb-1 and Tb-4 are on one branch, the Brazilian group centered on Bb-2 is on another, and the German and Singaporean variants are on the third. Generally, the need to accommodate homoplasious events (convergence and/or reversion) is common in sequence data sets and makes the resultant phylogenies unclear. In our data set, few homoplasies have to be postulated. One of the rare examples requiring a postulation of homoplasy comes from the comparison of Tb-8 (mutated at position 7687), Tb-13 (mutated at position 7712), and Tb-4 (mutated at both positions). If mutation 7687 had occurred only once and subsequent to 7712, one would not have found Tb-8; and if 7712 had occurred only once and subsequent to 7687, there would be no Tb-13. Consequently, during the evolution of Tb-8, Tb-13, and Tb-4, one of these two mutations must have occurred twice independently (a convergence), or a reversion at 7687 has occurred in Tb-13. Our parsimony criterion has favored reversion rather than convergence. An explanation is offered in the Discussion for the odd occurrences of one geographic lineage variant in another region, e.g., Sb-21a and Sb-22 in Singapore.

Variability in open reading frame E5 of HPV-16. We next asked whether sequence analysis of another genomic region would also reveal an informative degree of variability. Toward this end, we sequenced from selected HPV-16 isolates a 252-bp segment between positions 3850 and 4101,

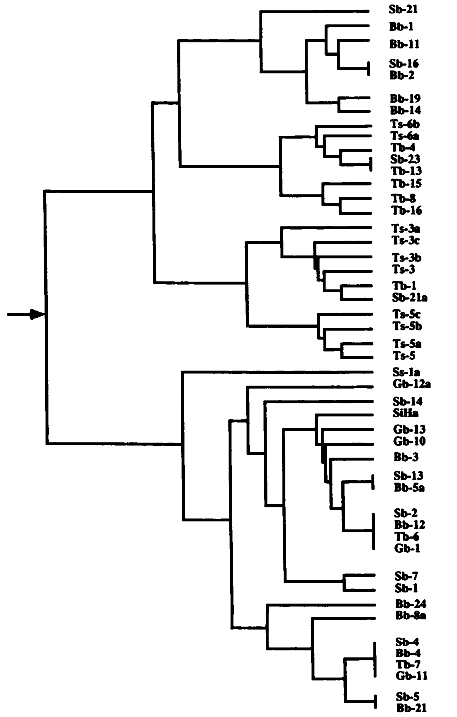

FIG. 3. UPGMA tree phylogeny of the 38 HPV-16 LCR variants that were identified in samples from Singapore (S samples), Brazil (B samples), Tanzania (T samples), and Germany (G samples). Distinct African (top) and Eurasian (bottom) lineages are seen bifurcating from the root (arrow), which is placed centrally between the two groups.

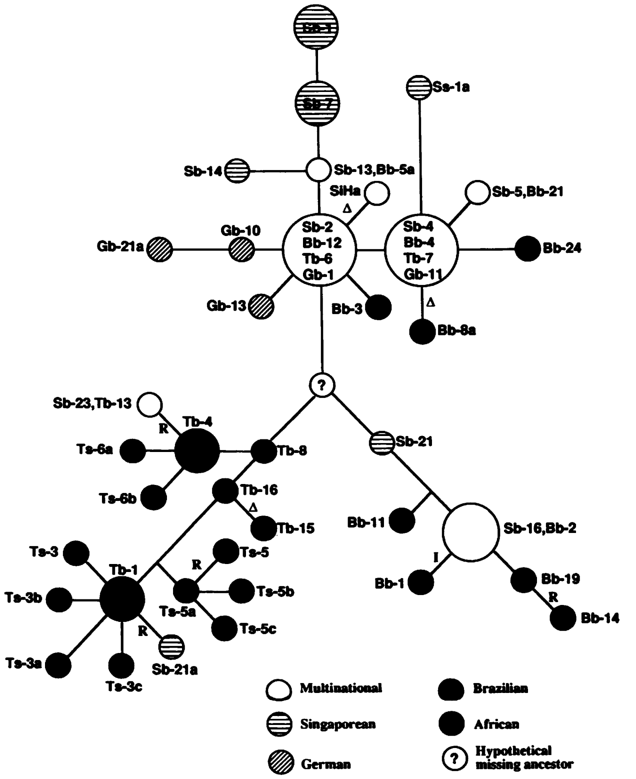

FIG. 4. Transformation series phylogeny of the 38 HPV-16 LCR variants. The length of the branches is proportional to the minimum number of nucleotide changes needed to transform one variant into its nearest neighbor. The four different sizes of the circles represent, in ascending order, variants with 1 to 3, 4 to 10, 11 to 20, and >20 members, respectively. The symbols Δ, I, and R refer to a deletion, an insertion, and a proposed reversion event, respectively.

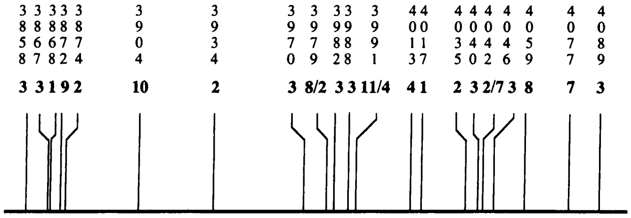

FIG. 5. Linear representation and identification of all point mutations identified in the 252-bp E5 open reading frame of HPV-16. The four numbers above each vertical line identify the position. The single boldface number immediately above each line identifies the exact mutational change. For details of this code, see the legend to Fig. 1.

	3858	3867	3868	3872	3874	3904	3934	3970	3979	3982	3988	3991	4013	4017	4035	4040	4042	4046	4059	4077	4089
prototype, Sb-4, Bb-4, Gb-10																					
Sb-5				■				■	■								■				
Sb-2									■						■		■				
Caski				■					■								■				
Siha						■			■			■					■				
Sb-13									■	■						■	■			■	
Sb-7, Sb-19									■								■			■	
Sb-17									■								■	■		■	
Sb-10					■		■		■								■			■	
Sb-21a, Sv-1, Tb-1, Ts-3	■								■			■					■		■		■
Sb-16, Bb-2, Bb-11	■								■			■		■			■				■
Tb-4	■		■						■			▨					▨				■
Tb-16	■	■	■						■			▨					▨				■
Tb-13	■		■						▨		■	▨	■				▨				■

FIG. 6. Combined representation of the E5 sequence variants found in 23 HPV-16 isolates selected from those shown in Fig. 2. For details of the presentation, see the legend to Fig. 2. The shaded squares identify the following changes: at position 3979, from A to G; at 3991, from C to T; and at 4042, from A to T.

encompassing the complete E5 open reading frame. We analyzed selected samples to address the following questions. Would isolates with identical HPV-16 sequences in the LCR also have identical E5 coding sequences? Would isolates that are closely related but different in their LCR sequences show a similar close relationship in their E5 sequences? Would isolates that were assigned to either of the two major branches of the LCR tree also be arranged similarly in an E5 phylogeny?

Figure 5 summarizes the mutations that were found during the analysis of this E5 segment from 23 HPV-16 variants. Evidently, the E5 segment is as rich in mutations as the LCR. Figure 6 shows the combinations of point mutations that were found in individual isolates. The designations of these isolates are identical to those in Fig. 2. The E5 clone Sb-21a was amplified from the same DNA that yielded clone Sb-21. It was given this designation because the LCR variant Sb-21a was far more abundant than the variant Sb-21, implying that the corresponding HPV-16 genome was also more abundant in the original DNA preparation. Consequently, after finding only E5 clones with the pattern shown in Fig. 6, we concluded that these were derived from the same viral genomes as the more abundant LCR type.

Four clones show no mutations relative to the prototype. The prototype plasmid, Sb-4, and Bb-4 also have identical sequences in the LCR, while Gb-10 has two mutations in the LCR. All remaining clones had common mutations in two positions, namely 3979 and 4042. All of these also had the mutation in the LCR at position 7519 with the exception of Sb-5, which had the prototype sequence. However, Tb-4, Tb-16, and Tb-13 had at position 4042 a different mutation than the other 16 variants. Tb-13 also had a different mutation at position 3979 than the other 18 variants.

There were two interesting observations. Isolates from Singapore (Sb-7, Sb-10, Sb-13, Sb-17, and Sb-19) with the common mutation at position 7840 in the LCR also had the same mutation at position 4077 in the E5 gene. All five Tanzanian isolates (Tb-1, Tb-4, Tb-13, Tb-16, and Ts-3) and six isolates from Singapore and Brazil (Sb-16, Sb-21a, Ss-6a, Sv-1, Bb-2, and Bb-11) had two unique mutations in common at position 3858 and at 4089 and were also assigned to the same branch of the LCR tree.

Relationship between HPV-16 E5 variants. Figure 7 shows the UPGMA phylogenies of selected isolates based on the E5 and LCR variations. The E5 phylogenetic tree has two major lineages, like the LCR tree. All variants that were assigned to the upper lineage of this E5 tree are identical with those that were assigned to one of the two major lineages (the upper lineage in Fig. 3 and Fig. 7B) of the LCR tree. Similarly, all variants that were assigned to the lower lineage were also on the same major lineage of the LCR tree. The compositions of the terminal groups in both trees support each other; c.f. the Sb-16, Bb-2, Bb-11, the Sb-21a, Sv-1, Tb-1, Ts-3, and the Sb-7, Sb-17, Sb-19, Sb-1, Sb-10 groupings. SiHa was excluded from the LCR tree (Fig. 7B) because it was grouped alone on a distant third branch due to the position of its deletion, which penalized it heavily in the UPGMA analysis.

The transformation series method groups the Singaporean isolates together and continues to recognize as one subgroup (Sb-7, Sb-10, Sb-13, Sb-1, and Sb-19) those that were distinguished by the LCR mutation at position 7840 (data not shown). The two Tanzanian subgroups centered on Tb-4 and Tb-1 are also distinguished, as are the Brazilian African-type genomes (Bb-2 and Bb-11). The tree, however, groups the Brazilian variants more closely with the Tb-1 subgroup. The identification by LCR patterns of Sb-16 as a Brazilian-lineage variant and Sb-21a and Sv-1 as African-lineage variants (found in Singapore) is supported by the E5 data. As in the case of the LCR, the number of homoplasies to be accommodated is small.

DISCUSSION

The combination of two techniques, DNA sequencing and amplification by PCR, has become an extraordinarily pow-

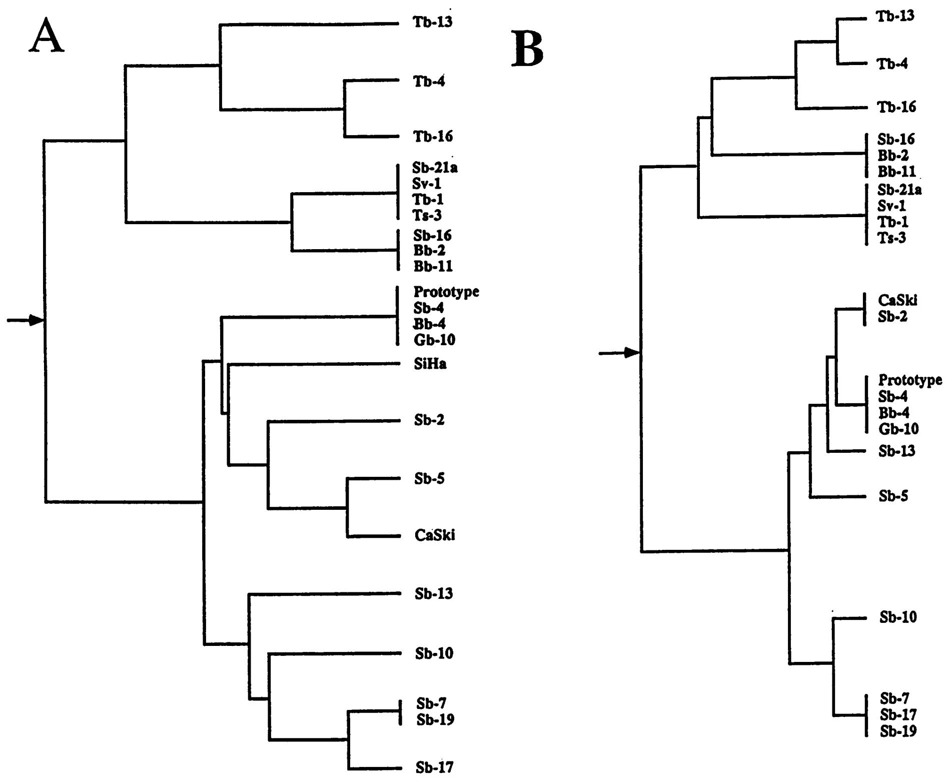

FIG. 7. UPGMA tree phylogeny of the E5 gene (A) for 23 HPV-16 isolates. For the purposes of comparison, panel B shows a simplified version of Fig. 3, namely, the UPGMA tree of the LCR mutations for the same 23 isolates (except SiHa). Only particular isolates, selected on the basis of the LCR data, were sequenced in the E5 region. The groupings are generally consistent (see also Fig. 3 and 4) and support the hypothesis of two separately evolving African and Eurasian lineages. The midpoint root (arrow), as in Fig. 3, represents the hypothetical common ancestor.

erful tool for studying molecular evolution. This discipline had to rely, until recently, on methods that were extremely laborious, like protein sequencing, or that yielded limited information, like restriction fragment length polymorphism analysis, electromorph, hybridization, and immunological studies (10, 25). The PCR technique allows us to obtain practical quantities of DNA from minuscule amounts of starting material (28), and sequencing allows us to examine the raw material of evolution at the highest resolution. A variety of numerical methods and evolutionary models (17, 18, 40) have been developed that derive unrooted phylogenetic trees from these molecular data. No single method is clearly superior for all data sets, and it seems better to apply several methods while also considering the results of classical comparative studies and the fossil record, when applicable.

Examples of this kind of research in virology include discovering the origin of HIV-1 from the degree of sequence diversity (53) and following the local spread of HIV-1 by genomic similarities (4). The genome of herpes simplex virus type 1 shows geographic variation (48), and this may shed light on its spread. The patterns of regional and pandemic spread have already been described for vesicular stomatitis virus isolates (6), different reovirus strains (14), and dengue virus subtypes (46). In contrast, the evolution and spread of papillomaviruses have not been studied much.

Our data (Fig. 3, 4, and 7) support the view that HPV-16 genomes sampled from geographically extreme locations have been evolving in two separate lineages. One contains nearly all present-day Tanzanian variants and a distinct Brazilian subgroup (Bb-1, Bb-2, Bb-11, Bb-14, and Bb-19), and the other is composed of very similar variants found in Singapore and Germany. Several different methods applied to the analysis of two different genomic regions (LCR and E5) give the same phylogenetic groupings. Uncertainty as to the rooting (due to the absence of any justifiable out-group) of the phylogeny does not alter this conclusion. Furthermore, we would argue that the UPGMA rooting of the trees (indicated by the arrow in the figures) is highly plausible when one considers the correlation with geographic origin, the mode of transmission, and the circumstantial historical evidence.

The transformation of character data, namely, the mutational changes, into a measure of evolutionary distance (as the distance matrix methods do) leads to some loss of information (43). We propose that determining phylogenetic trees directly from character data (parsimony and maximum likelihood) may more closely reflect the evolutionary events. For example, an inspection of the LCR sequences of Sb-13, Sb-7, Sb-17, Sb-19, and Sb-10 suggests successive occurrence of mutations 7519, 7840, 7728, and 7779. In the UPGMA LCR tree, some of these variants are separated on relatively remote minor branches (e.g., Sb-13 relative to Sb-7 and Sb-1) and arranged closer to variants like SiHa that do not have the distinctive mutation at position 7840. The E5 tree actually supports this suspected relationship: Sb-13, Sb-10, Sb-7, Sb-19, and Sb-17 are found in Fig. 7 on the same minor branch, remote from SiHa. A similar error may have

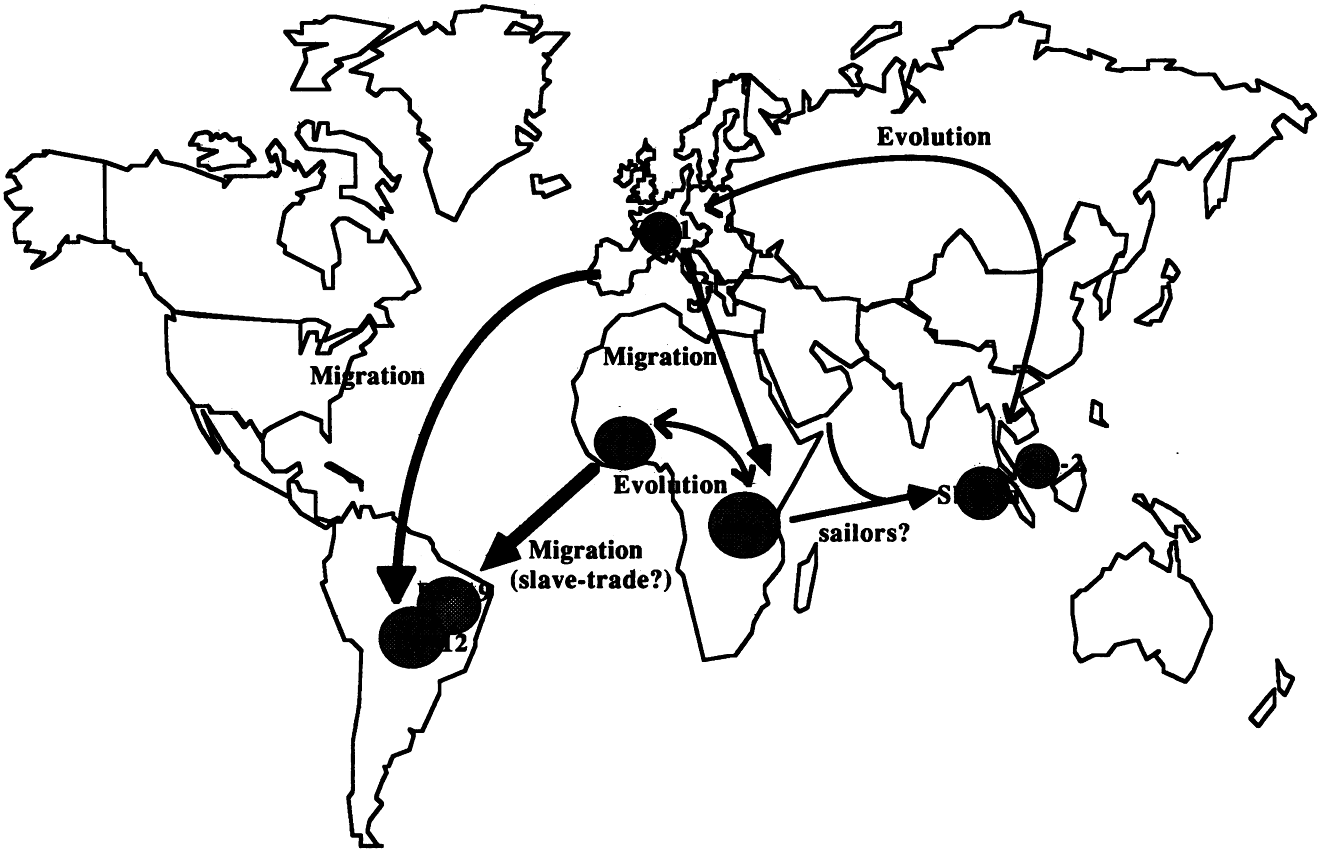

FIG. 8. Molecular evolution and migratory spread of HPV-16 variants. African and Eurasian variants are depicted as evolving separately in their respective continents after their unknown common origin. The presence of anomalous geographic variants in another region is explained by specific gene flow events, for example, the transfer of Eurasian variants to Africa and Brazil by European and Spanish colonists, respectively, and the transfer of African variants to Brazil by the slave trade and to Singapore by a combination of Arab and Indonesian slave trading, commerce, and colonizations. See Discussion for details.

occurred in the placement of some German variants. Gb-12a and Gb-21a seem to be derived from Gb-10, although they are placed closer to variants that do not share the distinctive mutation 7550.

An interesting assumption of the transition series approach is that some present-day LCR variants are identical to the ancestors of other present-day variants. If a more complete study of the genome were done, one could probably synthesize all the resultant gene transformation networks into a phylogeny of the whole virus. This could predict the likely genomic composition of the ancestral viruses.

The tremendous species and type diversity of papillomaviruses suggests a well-adapted parasite that has been intimately associated with its host for a long time. The small degree of diversity within geographic regions compared with the large diversity between regions seems more consistent with a slow accumulation of mutations over a lengthy time period. But what of the large and seemingly unconnected variations between isolates within Singapore and Brazil? We believe that they are more reasonably explained by the geographic transfers summarized in Fig. 8. It can be speculated that the occurrence of Eurasian HPV-16 variants in Tanzania may be associated with the historically recent arrival of Europeans on the African continent. The presence of African HPV-16 variants in Singapore may likewise be associated with the ancient Arab slave trade and Indonesian maritime colonizers of East Africa (42). But the temporal and geographical details of these transfers will remain speculative unless one can sample extensively from source and target populations, and provided that no other gene flow events have occurred to obscure the picture.

However, we believe that the bipartite nature of the Brazilian HPV-16 isolates invites detailed interpretations. Some of these clearly belong to the same major phylogenetic branch as the Tanzanian variants, although the two mutations at positions 7727 and 7741 (Brazilian) and that at position 7832 (Tanzanian) set them apart. Today, there is little travel and exchange of population groups between Africa and South America, and it seems unlikely that a population with a high prevalence of either Eurasian or African HPV-16 variants can subsequently become infected to a high degree by the sporadic arrival of variants of another lineage. Consequently, we believe that the approximately equal abundance of each variant branch represents the historic colonization of Brazil by European settlers and the 10 million Africans abducted into slavery largely between 1700 and 1800 A.D. (5). A large proportion of these slaves were ethnolinguistically Bantus from eastern and southern Africa and Sudanese from West Africa (50, 51). Future research with material from patients in West Africa—the geographic origin of the most of the European slave trade (the Portuguese also used Mozambique in East Africa as a slave trade port)—will settle whether the minor differences between isolates from Brazil and Tanzania represent evolution in the respective continents over the last two centuries or existing branches of variation within the African continent.

The periodicity of amino acid substitutions has been noticed to be clocklike (63). The nucleotide substitution rate for various mammalian pseudogenes (33) has been estimated to be about 4.7×10^{-9} substitutions per site per year. This is virtually identical to the average synonymous substitution rate for many mammalian nuclear genes (37, 44). Brown et al. (7, 8), however, have reported a 10-times-higher rate for silent substitutions in mitochondrial genes. Extremely high mutation rates have been observed in RNA viruses (27), and influenza A virus mutation rates are on the order of 1×10^{-2} per site per year. It is believed that these high rates are associated with the lack of enzyme proofreading and/or DNA and RNA repair mechanisms. In microorganisms, the mutation rate is generally a function of both absolute time and replication rate (number of generations) (40). Papillomavirus replication is largely dependent on cellular proteins, including the cellular DNA polymerase (59), and its replication appears to be tightly coupled to the state of the differentiating epithelium. However, the virus has a much shorter generation time and is more fecund, so a range of 10^{-9} to 10^{-7} substitutions per site per year seems a reasonable estimate. Based on this, we can estimate the age of the HPV-16 phylogeny. The maximal difference seen in our study was 17 substitutions (between Ss-1a and Ts-3a) over a 364-bp fragment. With the Jukes-Cantor (24) correction, this is about 4.8×10^{-2} substitutions per site. Assuming an ancestral sequence rooted at the midpoint of the two lineages, we derive a chronological range of 240,000 to 24,000,000 years B.P. (before the present). Interestingly, this range brackets the estimated period of primate and human evolution (10, 45).

Our study shows that distinct human races in remote regions still harbor populations of a venereally transmitted virus that seem to reflect their geographic, sexual, and racial separation. The apparent large differences between African and Eurasian viral lineages may suggest that the European and Asiatic human races are genealogically closer to each other than either is to the African races. In addition, these differences allow one to determine the existence and estimate the rate of the molecular clock (assuming there is a clocklike divergence) in the viral LCR when given independent estimates of racial divergence times. Larger and more widespread sampling may unearth variants that could connect the two lineages and clarify the relationship of the Brazilian branch within the African lineage. It is also possible that the observed discontinuities are real and are a result of founder effects, bottlenecks, genetic drift in small populations, or natural selection that occurred in the past.

ACKNOWLEDGMENTS

We thank Shih-Ping Chan, Department of Mathematics, National University of Singapore, for helpful technical discussions and for obtaining, compiling, and configuring PHYLIP 3.4 for mainframe computing; and Joseph Felsenstein, Department of Genetics, University of Washington, Seattle, for the excellence and free availability of his software.

We acknowledge the support of the Science and Technology for Development Programme of the European Community, contract TS2-0009-D to J.t.M.

REFERENCES

1. **Almond, J. W.** 1987. The attenuation of poliovirus neurovirulence. Annu. Rev. Microbiol. **41:**153–180.
2. **Anderson, W. W., T. Dobzhansky, O. Pavolovsky, J. Powell, and D. Yardley.** 1975. Genetics of natural populations. XLII. Three decades of genetic change in *Drosophila pseudoobscura*. Evolution **29:**24–36.
3. **Baker, C. C., W. C. Phelps, V. Lindgren, M. J. Braun, M. A. Gonda, and P. M. Howley.** 1987. Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines. J. Virol. **61:**962–971.
4. **Balfe, P., P. Simmonds, C. A. Ludlum, J. O. Bishop, and A. Y. L. Brown.** 1990. Concurrent evolution of human immunodeficiency virus type 1 in patients infected from the same source: rate of sequence change and low frequency of inactivating mutations. J. Virol. **64:**6221–6223.
5. **Barraclough, G. (ed.).** 1982. The Times concise atlas of world history. Times Books Ltd., London.
6. **Bilsel, P. A., and S. T. Nichol.** 1990. Polymerase errors accumulating during natural evolution of the glycoprotein gene of vesicular stomatitis virus Indiana serotype isolates. J. Virol. **64:**4873–4883.
7. **Brown, W. M., M. George, Jr., and A. C. Wilson.** 1979. Rapid evolution of animal mitochondrial DNA. Proc. Natl. Acad. Sci. USA **76:**1967–1971.
8. **Brown, W. M., E. M. Prager, A. Wang, and A. C. Wilson.** 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol. **18:**225–239.
9. **Bubb, V., D. J. McCance, and R. Schlegel.** 1988. DNA sequence of the HPV 16 E5 ORF and the structural conservation of its encoded protein. Virology **163:**243–246.
10. **Cann, R. L., M. Stoneking, and A. C. Wilson.** 1987. Mitochondrial DNA and human evolution. Nature (London) **325:**31–36.
11. **Chong, T., W.-K. Chan, and H.-U. Bernard.** 1990. Transcriptional activation of human papillomavirus 16 by nuclear factor 1, AP1, steriod receptors and a possible novel transcription factor, PVF: a model for the composition of genital papillomavirus enhancers. Nucleic Acids Res. **18:**465–470.
12. **Chow, V. T.-K., K.-M. Tham, and H.-U. Bernard.** 1990. *Thermus aquaticus* DNA polymerase-catalyzed chain reaction for the detection of human papillomavirus. J. Virol. Methods **27:**101–112.
13. **Cole, S. T., and O. Danos.** 1987. Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome. J. Mol. Biol. **193:**599–608.
14. **Dermody, T. S., M. L. Nibert, R. Bassel-Duby, and B. N. Fields.** 1990. Sequence diversity in S1 genes and S1 translation products of 11 serotype 3 reovirus strains. J. Virol. **64:**4842–4850.
15. **Duerst, M., L. Gissman, 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.
16. **Excoffier, L., and A. Langaney.** 1989. Origins and differentiation of human mitochondrial DNA. Am. J. Hum. Genet. **44:**73–85.
17. **Farris, J. S., A. G. Kluge, and M. J. Eckardt.** 1970. A numerical approach to phylogenetic systematics. Syst. Zool. **19:**172–189.
18. **Felsenstein, J.** 1982. Numerical methods for inferring evolutionary trees. Q. Rev. Biol. **57:**379–404.
19. **Felsenstein, J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:**783–791.
20. **Felsenstein, J.** 1988. Phylogenies from molecular sequences: inference and reliability. Annu. Rev. Genet. **22:**521–565.
21. **Gloss, B., T. Chong, and H.-U. Bernard.** 1989. Numerous nuclear proteins bind the long control region of human papillomavirus type 16: a subset of 6 of 23 DNase I-protected sequences coincides with the location of the cell-type-specific enhancer. J. Virol. **63:**1142–1152.
22. **Goldstein, D. J., M. E. Finboe, T. Andresson, P. McLean, K. Smitz, V. Bubb, and R. Schlegel.** 1991. Bovine papillomavirus E5 oncoprotein binds to the 16 K component of the vacuolar ATPases. Nature (London) **352:**347–349.
23. **Halbert, C. L., and D. Galloway.** 1988. Identification of the E5

open reading frame of the human papillomavirus type 16. J. Virol. **62**:1071–1075.

24. **Higgins, D. G., A. J. Bleasby, and R. Fuchs.** Submitted for publication.
25. **Hillis, D. M., and C. Moritz (ed.).** 1990. Molecular systematics. Sinauer Associates, Sunderland, Mass.
26. **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.
27. **Holland, J., K. Splindler, F. Horodyski, E. Grabau, S. Nichol, and S. VandePol.** 1982. Rapid evolution of RNA genomes. Science **215**:1577–1585.
28. **Innis, M. A., D. H. Gelfand, J. J. Snisky, and T. J. White.** 1990. PCR protocols: a guide to methods and applications. Academic Press, Inc., San Diego.
29. **Jukes, T. H., and C. R. Cantor.** 1969. Evolution of protein molecules, p. 21–132. *In* H. N. Munro (ed.), Mammalian protein metabolism. Academic Press, Inc., New York.
30. **Kimura, M.** 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.
31. **Kjaaer, S. K., E. M. de Villiers, B. J. Hangaard, R. B. Chrstensen, C. Teisen, K. A. Moller, P. Poll, H. Jensen, B. P. Vestergaard, E. Lynge, and O. M. Jensen.** 1989. Human papillomavirus, herpes simplex virus, and cervical cancer incidence in Greenland and Denmark. A population based cross-sectional study. Int. J. Cancer **41**:518–524.
32. **Knight, C. A.** 1974. Molecular virology, p. 1–2. McGraw-Hill Book Company, New York.
33. **Li, W.-H., T. Gojobori, and M. Nei.** 1981. Pseudogenes as a paradigm of neutral evolution. Nature (London) **292**:237–239.
34. **Li, W.-H., C.-C. Luo, and C.-I. Wu.** 1985. Evolution of DNA sequences, p. 1–94. *In* R. J. MacIntyre (ed.), Molecular evolutionary genetics. Plenum Publishing Corp., New York.
35. **Martin, P., W. C. Vass, T. J. Schiller, D. R. Lowy, and T. J. Velu.** 1989. The bovine papillomavirus E5 transforming protein can stimulate the transforming activity of EGF and CSF-1 receptors. Cell **59**:21–32.
36. **McNeill, W. H.** 1976. Plagues and people: a natural history of infectious diseases. Anchor Press, Garden City, N.Y.
37. **Miyata, T., T. Yasunaga, and T. Nishida.** 1980. Nucleotide sequence divergence and functional constraint in mRNA evolution. Proc. Natl. Acad. Sci. USA **77**:7328–7332.
38. **Muenger, K., B. A. Werness, N. Dyson, W. C. Phelps, E. Harlow, and P. M. Howley.** 1989. Complex formation of human papillomavirus E7 proteins with the retinoblastoma tumor suppressor gene product. EMBO J. **8**:4099–4105.
39. **Murphy, B. R., and R. G. Webster.** 1990. Orthomyxoviruses, p. 1091–1152. *In* B. N. Fields and D. M. Knipe (ed.), Virology. Raven Press, New York.
40. **Nei, M.** 1987. Molecular evolutionary genetics. Columbia University Press, New York.
41. **Norrby, E., and H. N. Oxman.** 1990. Measles virus, p. 1013–1044. *In* B. N. Fields and D. M. Knipe (ed.), Virology. Raven Press, New York.
42. **Oliver, R., and G. Matthew (ed.).** 1963. History of East Africa. Oxford University Press, Oxford.
43. **Penny, D.** 1982. Towards a basis for classification: the incompleteness of distance measures, incompatibility analysis and phenetic classification. J. Theor. Biol. **96**:129–142.
44. **Peto, R., and H. zur Hausen (ed.).** 1986. Viral etiology of cervical cancer (Banbury report 21). Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
45. **Read, D. W.** 1975. Primate phylogeny, neutral mutations and "molecular clocks." Syst. Zool. **24**:209–221.
46. **Rico-Hesse, R.** 1990. Molecular evolution and distribution of dengue viruses type 1 and 2 in nature. Virology **174**:479–493.
47. **Romanczuk, H., F. Thierry, and P. M. Howley.** 1990. Mutational analysis of *cis* elements involved in E2 modulation of human papillomavirus type 16 P_{97} and type 18 P_{105} promoters. J. Virol. **64**:2849–2859.
48. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.
49. **Sakaoka, H., H. Saito, K. Sekine, T. Aomori, L. Grillner, G. Wadell, and K. Fujinaga.** 1987. Genomic comparison of herpes simplex virus type 1 isolates from Japan, Sweden and Kenya. J. Gen. Virol. **68**:749–764.
50. **Saldanha, P. H.** 1989. Mistura de raças, mistura de genes. Ciénc. Hoje **9**:48–54.
51. **Salzano, F. M.** 1986. Em busca das raizes. Ciénc. Hoje **5**:48–53.
52. **Seedorf, K., G. Kraemmer, M. Duerst, S. Suhai, and W. G. Roewekamp.** 1985. Human papillomavirus type 16 DNA sequence. Virology **145**:181–185.
53. **Sharp, P. M., and W.-H. Li.** 1988. Understanding the origins of AIDS viruses. Nature (London) **336**:315.
54. **Shibata, D. V., N. Arnheim, and W. J. Martin.** 1988. Detection of human papillomavirus in paraffin-embedded tissue using the polymerase chain reaction. J. Exp. Med. **167**:225–230.
55. **Smith, T. F., A. Srinivasan, G. Schochetman, M. Marcus, and G. Myers.** 1988. The phylogenetic history of immunodeficiency virus. Nature (London) **333**:573–575.
56. **Sneath, P. H., and R. R. Sokal.** 1973. Numerical taxonomy. Freeman Publications, San Francisco.
57. **Syrjaenen, K. J., L. Gismann, and L. G. Koss.** 1987. Papillomaviruses and human disease. Springer-Verlag, Berlin.
58. **Temkin, O.** 1956. Soranus' gynaecology. Johns Hopkins University Press, Baltimore, Md.
59. **Ustav, M., and A. Stenlund.** 1991. Transient replication of BPV-1 requires two viral polypeptides encoded by the E1 and E2 open reading frames. EMBO J. **10**:449–457.
60. **Villa, L. L., and E. L. F. Franco.** 1989. Epidemiologic correlates of cervical neoplasia and risk of human papillomavirus infection in asymptomatic women in Brazil. J. Natl. Cancer Inst. **81**:332–340.
61. **Werness, B. A., A. J. Levine, and P. M. Howley.** 1988. Association of human papillomavirus types 16 and 18 E6 proteins with p53. Science **248**:76–79.
62. **Wilbur, W. J., and D. J. Lipman.** 1983. Rapid similarity searches of nucleic acid data banks. Proc. Natl. Acad. Sci. USA **80**:726–730.
63. **Zukerkandl, E., and L. Pauling.** 1962. Molecular disease, evolution, and genetic heterogeneity, p. 189–225. *In* M. Kasha and B. Pullman (ed.), Horizons in biochemistry. Academic Press, Inc., New York.
64. **zur Hausen, H.** 1989. Papillomaviruses as carcinomaviruses. Adv. Viral. Oncol. **8**:1–26.