JOURNAL OF VIROLOGY, Feb. 2009, p. 1443–1455 Vol. 83, No. 3
0022-538X/09/$08.00+0 doi:10.1128/JVI.02068-08


# Evolutionary Dynamics of Variant Genomes of Human Papillomavirus Types 18, 45, and 97▽†

Zigui Chen,[1] Rob DeSalle,[2] Mark Schiffman,[3] Rolando Herrero,[4] and Robert D. Burk[1,5]*

*Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, New York[1]; Sackler Institute of Comparative Genomics, American Museum of Natural History, New York, New York 10024[2]; Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services, Bethesda, Maryland[3]; Proyecto Epidemiológico Guanacaste, Costa Rican Foundation for Health Sciences, San José, Costa Rica[4]; and Departments of Pediatrics, Epidemiology and Population Health, Obstetrics, Gynecology and Woman's Health, and Albert Einstein Cancer Center, Albert Einstein College of Medicine, Bronx, New York[5]*

Received 1 October 2008/Accepted 14 November 2008

**Human papillomavirus type 18 (HPV18) and HPV45 account for approximately 20% of all cervix cancers. We show that HPV18, HPV45, and the recently discovered HPV97 comprise a clade sharing a most recent common ancestor within HPV α7 species. Variant lineages of these HPV types were classified by sequence analysis of the upstream regulatory region/E6 region among cervical samples from a population-based study in Costa Rica, and 27 representative genomes from each major variant lineage were sequenced. Nucleotide variation within HPV18 and HPV45 was 3.82% and 2.39%, respectively, and amino acid variation was 4.73% and 2.87%, respectively. Only 18 nucleotide variations, of which 10 were nonsynonymous, were identified among three HPV97 genomes. Full-genome comparisons revealed maximal diversity between HPV18 African and non-African variants (2.6% dissimilarity), whereas HPV18 Asian-American [E1 (AA)] and European (E2) variants were closely related (less than 0.5% dissimilarity); HPV45 genomes had a maximal difference of 1.6% nucleotides. Using a Bayesian Markov chain Monte Carlo (MCMC) method, the divergence times of HPV18, -45, and -97 from their most recent common ancestors indicated that HPV18 diverged approximately 7.7 million years (Myr) ago, whereas HPV45 and HPV97 split off around 5.7 Myr ago, in a period encompassing the divergence of the great ape species. Variants within the HPV18/45/97 lineages were estimated to have diverged from their common ancestors in the genus *Homo* within the last 1 Myr (<0.7 Myr). To investigate the molecular basis of HPV18, HPV45, and HPV97 evolution, regression models of codon substitution were used to identify lineages and amino acid sites under selective pressure. The E5 open reading frame (ORF) of HPV18 and the E4 ORFs of HPV18, HPV45, and HPV18/45/97 had nonsynonymous/synonymous substitution rate ratios ($d_N/d_S$) over 1 indicative of positive Darwinian selection. The L1 ORF of HPV18 genomes had an increased proportion of nonsynonymous substitutions (4.93%; average $d_N/d_S$ ratio [M3] = 0.3356) compared to HPV45 (1.86%; M3 = 0.1268) and HPV16 (2.26%; M3 = 0.1330) L1 ORFs. In contrast, HPV18 and HPV16 genomes had similar amino acid substitution rates within the E1 ORF (2.89% and 3.24%, respectively), while HPV45 E1 was highly conserved (amino acid substitution rate was 0.77%). These data provide an evolutionary history of this medically important clade of HPVs and identify an unexpected divergence of the L1 gene of HPV18 that may have clinical implications for the long-term use of an L1-virus-like particle-based prophylactic vaccine.**

Papillomaviruses (PVs) are a large family of related viruses with circular double-stranded DNA genomes 8 kb in size. Some PV types cause epithelial hyperplasias ranging from benign exophytic warts to premalignant lesions that can progress to invasive cancer. Among the 61 currently designated alpha human PVs (HPVs), the majority have been isolated from the mucosal surface of the genital or oral region (8, 14). Of these, a select group have oncogenic potential and are associated with cervical cancer (11). Specifically, HPV type 16 (HPV16) and HPV18 have been identified in approximately two-thirds of cervical cancers, this tumor is the second most common cancer in women, and it is the principal cancer of women in developing countries (5, 24, 25, 30, 37).

To date, studies of HPV18 variants have identified three lineages corresponding to the continental locations where the viral samples were obtained: European (E), Asian-American (AA), and African (Af) (29). The phylogeny of HPV18 variants is reflective of the migration patterns of *Homo sapiens* and suggests that HPV18 variant lineages might have diverged through genetic isolation at approximately the same time as *Homo sapiens* began establishing residence in different continental regions. Previous HPV18 intratypic phylogenetic analyses were limited to partial regions of the genome (3, 7, 29). Nevertheless, studies also suggest that HPV18 variants are associated with different levels of oncogenic potential and persistence and histological tumor types (1, 6, 35, 36, 46).

HPV45 and HPV97 are the viral types most closely re-

* Corresponding author. Mailing address: Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461. Phone: (718) 430-3720. Fax: (718) 430-8975. E-mail: burk@aecom.yu.edu.

† Supplemental material for this article may be found at http://jvi.asm.org/.

▽ Published ahead of print on 26 November 2008.

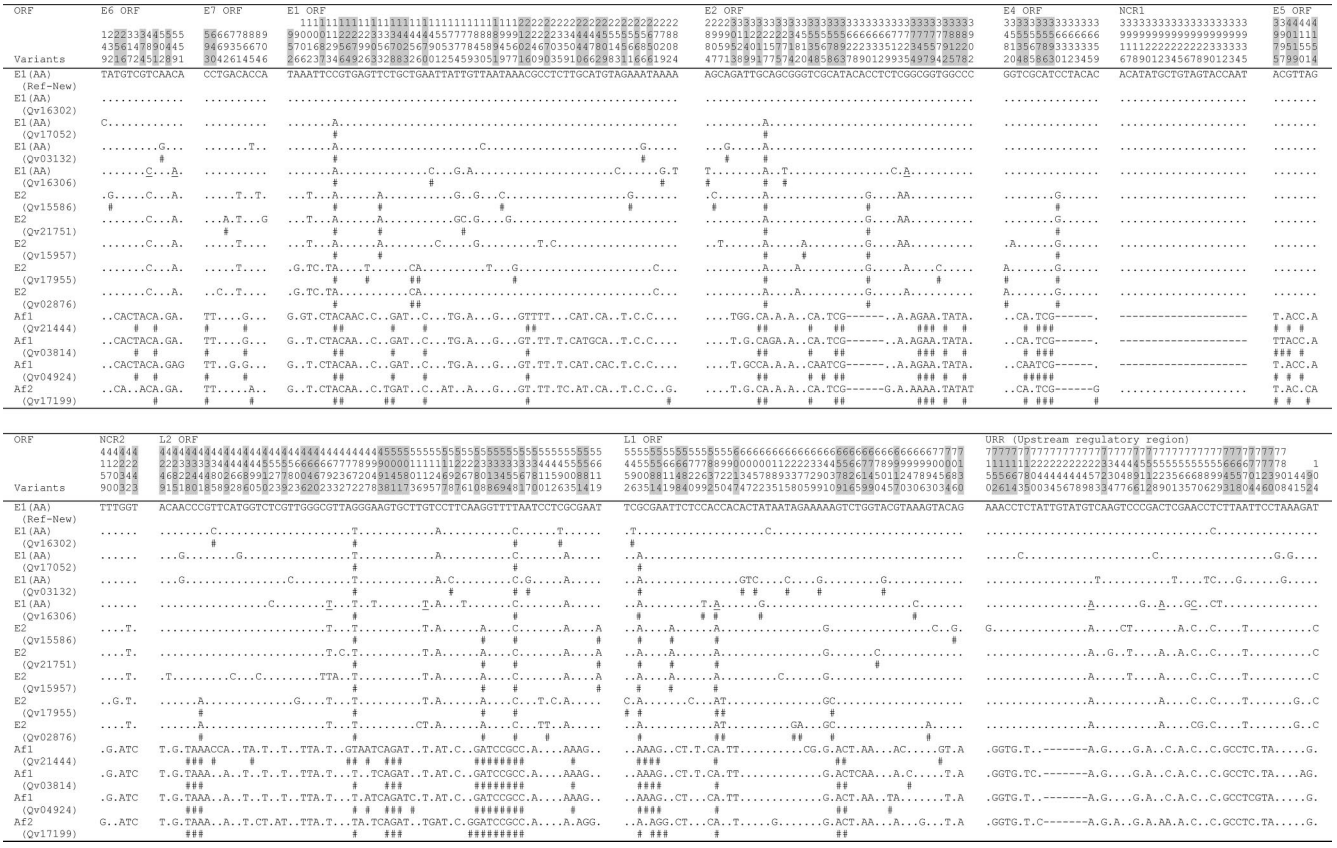


FIG. 1. Nucleotide sequence variations within the complete genome of HPV18 variants. Number signs (#), nucleotide changes resulting in amino acid changes (nonsynonymous); dots, sites matched with HPV18R reference sequence; dashes, indel events. HPV18 Af and non-Af lineage-specific variable positions are highlighted in gray. The underlined nucleotide variations indicate sites of Qv16306 [HPV18 E1 (AA) variant] that are identical to the HPV18 E2 variant. NCR1, noncoding region between E2 and E5 ORFs; NCR2, noncoding region between E5 and L2 ORFs.

lated to HPV18 and taken together form a clade and share a most recent common ancestor (MRCA). HPV97 is a recently described rare type (8, 17). HPV18 and HPV45 account for approximately 20% of all cervix cancers (25). Although HPV45 is a common type found in cervical cancer, its evolutionary history and sequence variability have not been extensively studied.

In this report, 27 complete genomes representing the major

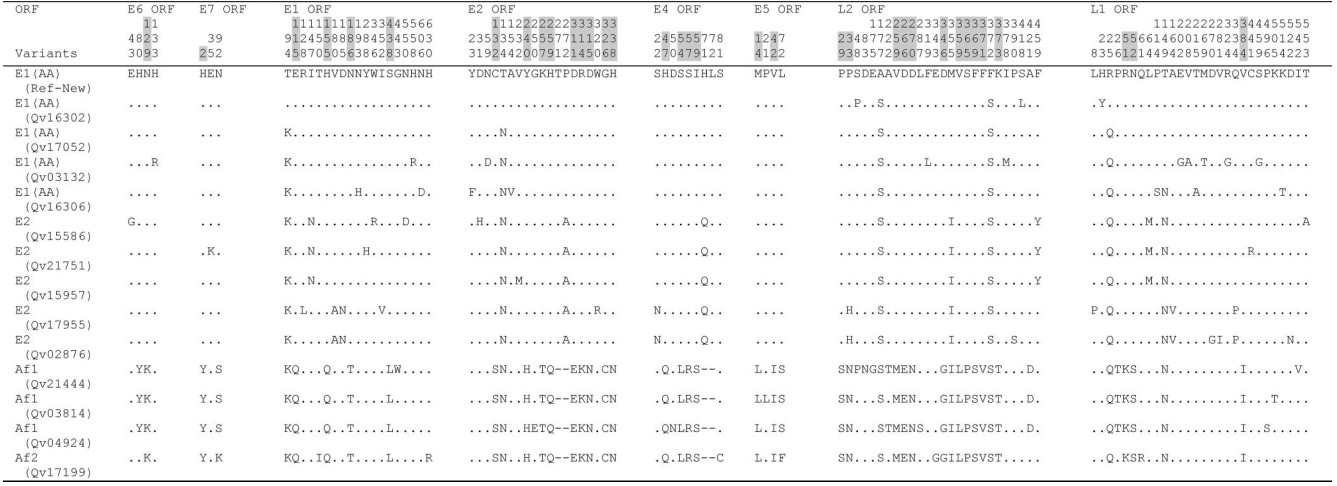


FIG. 2. Amino acid variations within eight ORFs of HPV18 variants. Dots, sites matched with HPV18R reference sequence; dashes, indel events. HPV18 Af and non-Af lineage-specific variable positions are highlighted in gray.

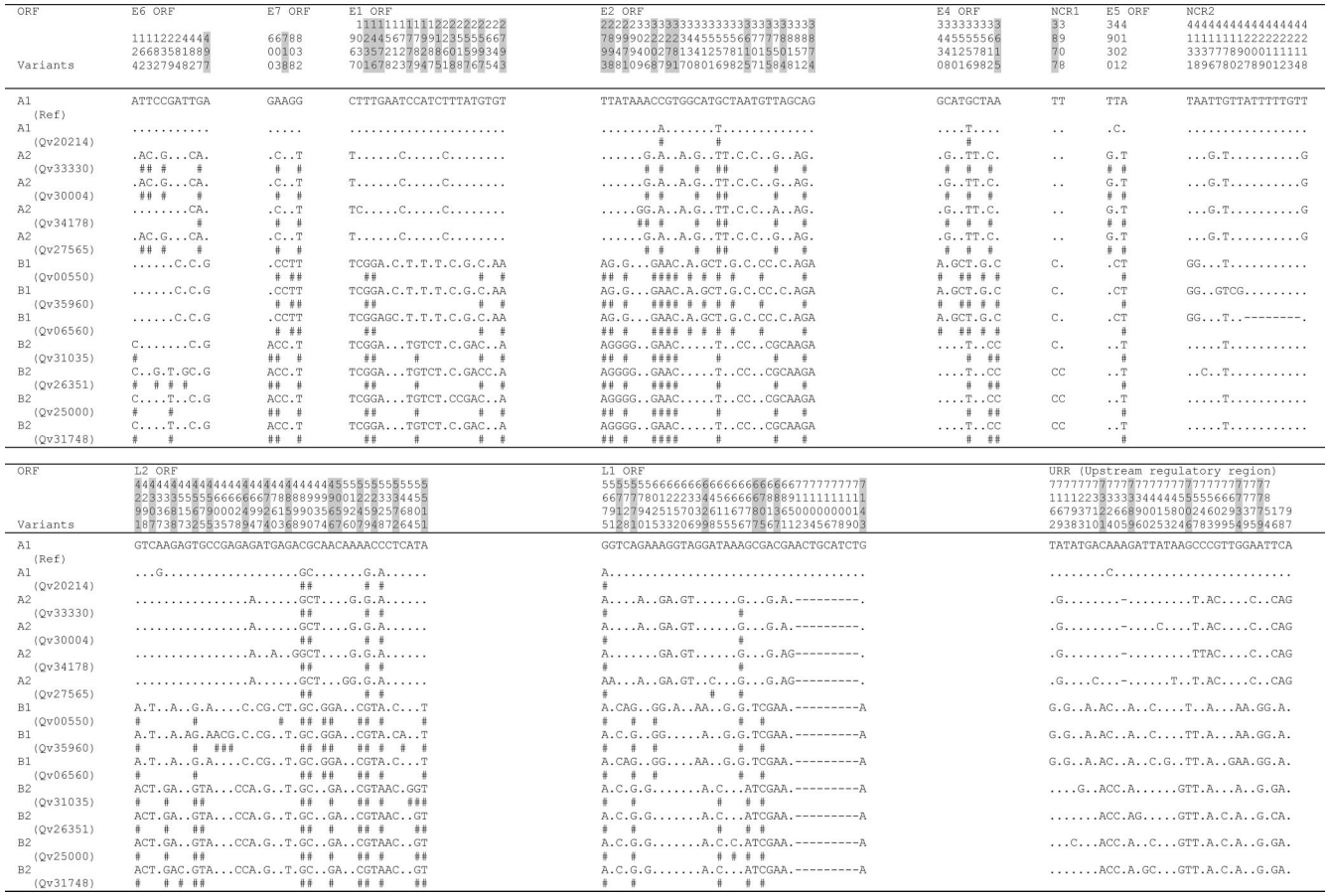



FIG. 3. Nucleotide sequence variations within the complete genome of HPV45 variants. Number signs (#), nucleotide changes resulting in amino acid changes (nonsynonymous); dots, sites matched with HPV45 reference sequence; dashes, indel events. HPV45 A and B lineage-specific variable positions are highlighted in gray. NCR1, noncoding region between E2 and E5 ORFs; NCR2, noncoding region between E5 and L2 ORFs.

variant lineages of HPV18, HPV45, and HPV97 were cloned and/or sequenced from clinical samples. Based on full genomes, the intratype/intertype evolutionary trees of HPV18, HPV45, and HPV97 were constructed. By examining the rate ratio of nonsynonymous ($d_N$) to synonymous ($d_S$) substitutions per site, diversifying selection acting on each of the eight protein-encoding regions of HPV18, HPV45, and HPV97 was evaluated. In addition, the times of divergence of HPV18/45/97 variants from their

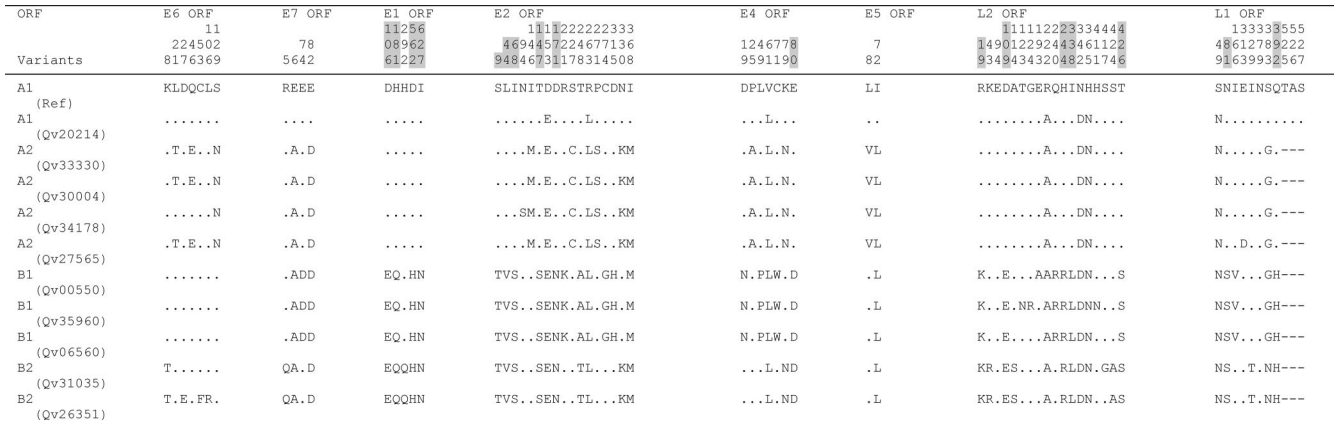



FIG. 4. Amino acid variations within eight ORFs of HPV45 variants. Dots, sites matched with HPV45 reference sequence; dashes, indel events. HPV45 A and B lineage-specific variable positions are highlighted in gray.

| ORF/Region | E7 | E1 | E2 | E4 | L2 | L1 | URR |
|---|---|---|---|---|---|---|---|
| | | 1 1 2 | 3 3 3 3 3 | 3 | 5 5 | 5 5 6 6 | 77 |
| | 5 | 31 3 1 | 21 42 63 6 73 | 4 | 23 4 | 4 5 33 3 | 12 |
| | 52 | 47 7 9 | 07 65 00 0 55 | 65 | 05 2 | 2 3 92 9 | 98 |
| nt/aa position | 64 | 17 3 2 | 11 18 25 3 46 | 16 | 28 3 | 33 2 78 9 | 40 |
| HPV97.Qv28597 | CA | CH A G | GM GS GR G AN | GV | GD G | GR G AT A | CG |
| HPV97.W15189 | TV | AN G A | AI TI .. A GD | TL | AN A | AH A TS G | T. |
| | # | # | # # # | # | # | # # | |
| HPV97.624 | TV | AN . A | AI TI AQ . GD | TL | AN A | AH . TS . | TC |
| | # | # | # # # # | # | # | # # | |

FIG. 5. Sequence variations of the complete genome of HPV97 variants. Number signs (#), nucleotide changes resulting in amino acid changes (nonsynonymous); dots, sites matched with HPV97 reference sequence. Positions highlighted in gray indicate amino acid variations due to the nonsynonymous nucleotide changes on the left.

MRCA were investigated. These data provide an evolutionary history of this medically important clade of HPVs.

## MATERIALS AND METHODS

**Clinical specimens and HPV sequencing.** Cervicovaginal cells were obtained from women participating in a population-based study of cervical neoplasia in Costa Rica (19), except for one sample from the Women's Interagency HIV Study (39). Samples determined to have HPV18, HPV45, and/or HPV97 by MY09/11 PCR and dot blot analyses were further subclassified into intratypic lineages by sequencing the upstream regulatory region (URR) and/or E6 region from PCR products (19, 35, 39).

Type-specific primer sets were designed based on the prototype sequences to amplify the complete genomes of HPV18, HPV45, and HPV97 isolates in two to three overlapping fragments (8, 9, 41, 42). Oligonucleotide primer sequences used in this study are available from the authors. Each PCR product was purified (Qiagen gel extraction kit; Qiagen, Valencia, CA) after confirmation of appropriate product size, ligated into the pGEM-T Easy vector (Promega, Madison, WI), and sequenced by the Einstein Sequencing Facility, New York. Subsequent sequencing was performed using primer walking. HPV genome sequences and the NCBI/GenBank accession numbers are listed in Table S1 in the supplemental material.

**Phylogenetic analyses and tree construction.** The amino acid of each predicted open reading frame (ORF) was aligned using Cluster X (43). Codon Align (version 1.0) (available from Sinauer Associates) was used to align the nucleotide sequences of each coding region corresponding to the aligned amino acid sequence.

Phylogenetic trees were constructed to assess the evolutionary histories of HPV18, HPV45, and HPV97 variants. MrBayes v3.1.2 (20) was used to generate a tree from the alignment of concatenated amino acid and nucleotide sequences of eight ORFs (E6, E7, E1, E2, E4, E5, L2, and L1). The computer program ModelTest v3.06 (32) was used to identify the best evolutionary model; the identified gamma model was set for among-site rate variation and allowed all substitution rates of aligned sequences to be different. Maximum parsimony (MP) and neighbor joining (NJ) trees were calculated by a heuristic search in PAUP* v4.0b10 (40). For MP analyses, amino acid and nucleotide sequences were reduced to phylogenetically informative sites. Data were bootstrap resampled 1,000 times. The prototype sequences of HPV39, HPV59, HPV68, HPV70, and HPV85 within genital HPV species α7 were obtained from the NCBI/GenBank database (10, 16, 27, 34, 45). HPV56 (GenBank accession no. X74483) and HPV66 (GenBank accession no. U31794) were selected as the outgroup taxa. Separate Bayesian trees were inferred from nucleotide sequences of "early genes" (E6, E7, E1, E2, and E5), "late genes" (L2 and L1), and the URRs of HPV18 and HPV45 variants.

**Positive selection estimation.** The nonsynonymous/synonymous rate ratio ($\omega = d_N/d_S$) is an indicator of natural selection, with $\omega = 1$ representing neutral variation, $\omega < 1$ representing purifying selection, and $\omega > 1$ representing diversifying positive selection. Amino acid sites in a protein are expected to be under different selective pressures and have different underlying ω ratios (50, 51). Six codon substitution models were used to investigate whether positive selection could be identified within the eight ORFs of HPV18 and HPV45: M0 (one-ratio), M1 (neutral), M2 (selection), M3 (discrete), M7 (beta), and M8 (beta and ω). These models view the codon as the fundamental unit of evolutionary change and take into account genealogic history when calculating scores. Log likelihood scores evaluate the quality of the fit of the input data to the conditions of the model. In these models, $\omega = d_N/d_S$ was estimated for separate classes of codons that are assumed to evolve independently of one another. The six models used for the ω distribution were implemented in the CODEML program in the PAML package (49, 50). MP within PAUP* v4.0b10 (40) was used for tree reconstruction.

Three likelihood ratio tests (LRTs) were performed to assess the influence of positive selection on a particular coding region, which compared M1 with M2, M0 with M3, and M7 with M8. When alternative models (M2, M3, and M8) suggest the presence of sites with $\omega > 1$, all three tests taken together are considered evidence of positive selection (28, 50).

**Molecular divergence estimates.** A Bayesian Markov chain Monte Carlo (MCMC) method was used to predict the divergence times of HPV18, -45, and -97 by selecting variant genomes representing each intratype variant lineage for analysis (18). We assumed a general time-reversible model of nucleotide substitution with gamma-distributed rate heterogeneity among sites and a proportion of invariant sites. In addition, we assumed an uncorrected lognormal distribution molecular clock model of rate variation among branches in the tree. A fixed (known) mean substitution rate of HPV genes, 1.95E−08 (95% confidence interval, 1.32E−08 to 2.47E−08) substitutions per site per year, was set for the times to the MRCA of variants of HPV18, HPV45, and HPV97 based on previous work (33). In addition, models using different rates of substitution of each ORF over time were also used (33): 1.84E−08 (95% confidence interval, 1.27E−08 to 2.35E−08) substitutions per site per year for the L1 ORF, 2.13E−08 (95% confidence interval, 1.46E−08 to 2.76E−08) for the L2 ORF, 1.76E−08 (95% confidence interval, 1.20E−08 to 2.31E−08) for the E1 ORF, 2.11E−08 (95% confidence interval, 1.52E−08 to 2.81E−08) for the E2 ORF, 2.39E−08 (95% confidence interval, 1.70E−08 to 3.26E−08) for the E6 ORF, and 1.44E−08 (95% confidence interval, 0.97E−08 to 2.00E−08) for the E7 ORF. The MCMC analysis was run for 10,000,000 steps. Calculations were performed in BEAST v1.4.7 (15). Results were displayed using Tracer v1.4 (A. Rambaut and A. J. Drummond, 2007 [http://beast.bio.ed.ac.uk/Tracer]).

## RESULTS

**Sequence variation of the complete genomes of HPV18, HPV45, and HPV97.** HPV18 ($n = 299$) and HPV45 ($n = 207$) genomes were initially classified based on sequence analysis of the URRs/E6 regions. Since HPV97 was detected in only two samples, both were selected for sequencing. The URR/E6 region sequences were aligned, and trees were generated. To study the extent of intratypic diversity and evolution among closely related HPV genomes, samples containing HPV18 and HPV45 from each clade of the URR/E6 tree were selected for complete genome analyses (data not shown).

A summary of nucleotide and amino acid sequence variation throughout the genomes of HPV18, HPV45, and HPV97 is shown in Fig. 1 to 5. Measures of variability for each ORF, noncoding region, and complete genomes of HPV18 and HPV45 are shown in Fig. 6 and 7, respectively. An insertion/deletion (indel) was considered as a single event irrespective of the number of nucleotides disrupted.

| ORF | Number of nucleotide sequences | Number variable nuc. positions [a] | Total nucleotide variations [b] | Number of amino acids | Number of variable codons [c] | Number of nonsynonymous changes [d] | Number of synonymous changes [e] | Nonsynonymous /synonymous changes [f] |
|---|---|---|---|---|---|---|---|---|
| E6 | 477 | 13 (2.73%) | 13 (2.73%) | 158 | 13 (8.23%) | 4 (2.53%) | 9 (5.70%) | 0.44 |
| E7 | 318 | 10 (3.14%) | **11** (3.46%) | 105 | 10 (9.52%) | 4 (3.81%) | 6 (5.71%) | 0.67 |
| E1 | 1974 | 61 (3.09%) | 61 (3.09%) | 657 | 60 (9.13%) | 19 (2.89%) | 41 (6.24%) | 0.46 |
| E2 | 1098 | 38 (3.46%) | **40** (3.64%) | 365 | 41 (11.23%) | 20 (5.48%) | 21 (5.75%) | 0.95 |
| E4 | 267 | 11 (4.12%) | **12** (4.49%) | 88 | 12 (13.64%) | 10 (11.36%) | 2 (2.27%) | 5.00 |
| NCR1 | 21 | 1 (4.76%) | 1 (4.76%) | | | | | |
| E5 | 222 | 7 (3.15%) | 7 (3.15%) | 73 | 7 (9.59%) | 5 (6.85%) | 2 (2.74%) | 2.50 |
| NCR2 | 86 | 6 (6.98%) | 6 (6.98%) | | | | | |
| L2 | 1389 | 69 (4.97%) | 69 (4.97%) | 462 | 69 (14.94%) | 27 (5.84%) | 42 (9.09%) | 0.64 |
| L1 | 1707 | 53 (3.10%) | 53 (3.10%) | 568 | 53 (9.33%) | 28 (4.93%) | 25 (4.72%) | 1.12 |
| URR | 825 | 46 (5.58%) | **48** (5.82%) | | | | | |
| *Aggregate* [g] | 7857 | 295 (3.75%) | 300 (3.82%) | 2476 | 265 (10.70%) | 117 (4.73%) | 149 (6.01%) | 0.79 |

FIG. 6. Comparison of nucleotide and amino acid sequence variability within HPV18 genes and the URR. a, numbers and percentages of positions showing nucleotide variations; b, numbers of nucleotide variations, including multiple changes per position (the genes containing multiple variation changes are highlighted in bold); c, numbers and percentages of variable codons; d, numbers and percentages of changed amino acid positions per total amino acid/codon in each ORF (nonsynonymous changes); e, numbers and percentages of synonymous/silent changes per total amino acid/codon in each ORF; f, ratios of the numbers of nonsynonymous changes to the numbers of synonymous changes per codon; g, each nucleotide position is only counted once. nuc., nucleotide.

Of 7,857 and 7,858 nucleotide positions in HPV18 and HPV45, 295 (3.82%) and 186 (2.39%) were variable, respectively ($P < 0.01$, $\chi^2$). Within the 2,476 amino acids (aa) comprising eight ORFs of HPV18 and HPV45, 117 (4.73%) and 71 (2.87%) positions were also variable, respectively ($P < 0.01$) (Fig. 6 and 7). The noncoding regions between E2 and E5 and between E5 and L2 were the most variable, followed by the URR. The absolute ratio of nonsynonymous to synonymous changes was over 1.0 in the E4, E5, and L1 ORFs among HPV18 genomes and the E6, E7, E2, E4, and E5 ORFs among HPV45 variants. This ratio is different than the $d_N/d_S$ rate ratio, which adjusts for numbers of possible changes (9, 50). The HPV18 L1 ORF had over twice as many nonsynonymous changes (28, 4.93%) as did HPV45 L1 (10, 1.86%) ($P < 0.01$) and HPV16 L1 (12, 2.26%) ($P = 0.02$) (9). However, HPV18 and HPV16 genomes had similar amino acid substitution rates within the E1 ORF (2.89% and 3.24%, respectively), indicating that the changes in L1 do not represent global genomic differences, while the HPV45 E1 was highly conserved (amino acid substitution rate, 0.77%). Three HPV97 genomes were analyzed and revealed only 18 nucleotide variations, of which 10 are nonsynonymous (Fig. 5) (8, 17).

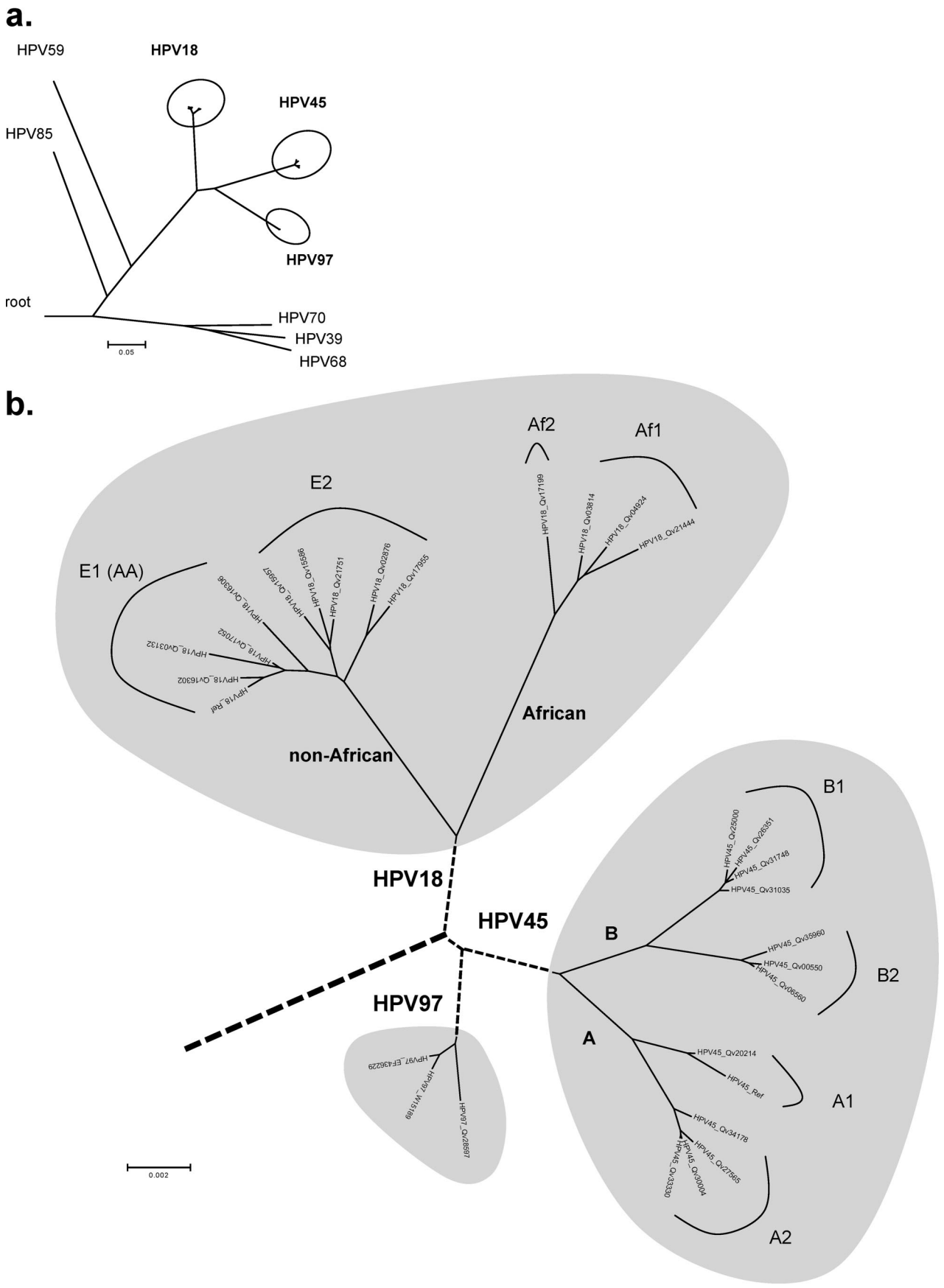
Indel events were rare. Within HPV18 genomes, three different indel events were detected within E2/E4 (6 bp), noncoding region 1 (NCR1 is the region between the stop codon of the E2 ORF and the start codon of the E5 ORF) (19 to 20 bp), and the URR (7 bp) (Fig. 1). Within HPV45 genomes, indel events were detected within the L1 ORF (9 bp) and the noncoding region 2 (NCR2 is the region between the stop codon of the E5 ORF and the start codon of the L2 ORF) (8 bp) and a single base indel was detected within the URR (Fig. 3).

**Phylogeny of HPV18, HPV45, and HPV97 variants.** Multiple algorithms including Bayesian, MP, and NJ were used to predict the relationships of HPV18, HPV45, and HPV97 within HPV species α7. Phylogenetic trees generated from complete genome sequences of these viruses confirmed that HPV18, HPV45, and HPV97 form a strongly supported clade distinct from the other types within the α7 species (Fig. 8) (8). This implies that they share an MRCA.

Phylogenetic analyses of HPV18 Af and non-Af isolates based on complete genomes and the L1 ORFs indicated maximal sequence diversities of 2.6% and 1.8%, respectively (Fig. 9a). The previously termed E (E2) and AA (E1) variants formed two closely related clades that are 0.8% dissimilar to each other. Although the E2 sublineage variant differs from the E1 (AA) sublineage variants at 37 positions across the genome (e.g., nucleotides A976G, A1012T, T1353A, and C3630G), 19 of these sites distinguish the E2 variants into two groups: Qv02876/Qv17955 and Qv15957/Qv15586/Qv21751

| ORF | Number of nucleotide sequences | Number variable nuc. positions [a] | Total nucleotide variations [b] | Number of amino acids | Number of variable codons [c] | Number of nonsynonymous changes [d] | Number of synonymous changes [e] | Nonsynonymous /synonymous changes [f] |
|---|---|---|---|---|---|---|---|---|
| E6 | 477 | 11 (2.31%) | 11 (2.31%) | 158 | 10 (6.33%) | 7 (4.43%) | 3 (1.90%) | 2.33 |
| E7 | 321 | 5 (1.56%) | 5 (1.56%) | 106 | 5 (4.72%) | 4 (3.77%) | 1 (0.94%) | 4.00 |
| E1 | 1932 | 22 (1.14%) | 22 (1.14%) | 643 | 22 (3.42%) | 5 (0.77%) | 17 (2.64%) | 0.29 |
| E2 | 1107 | 30 (2.71%) | **31** (2.80%) | 368 | 29 (7.88%) | 18 (4.89%) | 11 (2.99%) | 1.64 |
| E4 | 273 | 9 (3.30%) | 9 (3.30%) | 90 | 9 (10.00%) | 7 (7.78%) | 2 (2.22%) | 3.50 |
| NCR1 | 33 | 2 (6.06%) | 2 (6.06%) | | | | | |
| E5 | 222 | 3 (1.35%) | 3 (1.35%) | 73 | 3 (4.11%) | 2 (2.74%) | 1 (1.37%) | 2.00 |
| NCR2 | 105 | 10 (9.52%) | 10 (9.52%) | | | | | |
| L2 | 1392 | 41 (2.95%) | 41 (2.95%) | 463 | 39 (8.42%) | 18 (3.89%) | 21 (4.54%) | 0.86 |
| L1 | 1620 | 29 (1.79%) | 29 (1.79%) | 539 | 28 (5.19%) | 10 (1.86%) | 18 (3.34%) | 0.56 |
| URR | 810 | 34 (4.20%) | **35** (4.32%) | | | | | |
| *Aggregate* [g] | 7858 | 186 (2.37%) | 188 (2.39%) | 2476 | 145 (5.86%) | 71 (2.87%) | 74 (2.99%) | 0.96 |

FIG. 7. Comparison of nucleotide and amino acid sequence variability within HPV45 genes and the URR. a, numbers and percentages of positions showing nucleotide variations; b, numbers of nucleotide variations, including multiple changes per position (the genes containing multiple variation changes are highlighted in bold); c, numbers and percentages of variable codons; d, numbers and percentages of changed amino acid positions per total amino acid/codon in each ORF (nonsynonymous changes); e, numbers and percentages of synonymous/silent changes per total amino acid/codon in each ORF; f, ratios of the numbers of nonsynonymous changes to the numbers of synonymous changes per codon; g, each nucleotide position is only counted once. nuc., nucleotide.


a.
HPV59
HPV18
HPV45
HPV85
HPV97
root
HPV70
HPV39
HPV68
0.05
b.
Af2
Af1
E2
E1 (AA)
HPV18_Qv17199
HPV18_Qv03814
HPV18_Qv04924
HPV18_Qv21444
HPV18_Qv02876
HPV18_Qv17955
HPV18_Qv21751
HPV18_QV15586
HPV18_QV17957
HPV18_QV16306
HPV18_QV17052
HPV18_QV03132
HPV18_QV16302
HPV18_Ref
African
non-African
HPV18
HPV45
HPV97
B1
B
B2
A
A1
A2
HPV45_Qv25000
HPV45_Qv26351
HPV45_Qv31748
HPV45_Qv31035
HPV45_Qv35960
HPV45_Qv00550
HPV45_Qv06560
HPV45_Qv20214
HPV45_Ref
HPV45_Qv34178
HPV45_Qv27565
HPV45_Qv00004
HPV45_Qv33290
HPV97_EF436229
HPV97_JN151189
HPV97_Qv28597
0.002

(e.g., T1843G and A2701C), differing by 0.5% between their genomes (Fig. 1).

Qv16306 [an HPV18 E1 (AA) variant] is the most basal E1 (AA) variant and could be considered a "bridge variant" between the E1 and E2 sublineages; this variant shares 9 of 34 (26.5%) nucleotide changes found in E2 but not E1 sublineages (underlined in Fig. 1). Similarly, the HPV18 Af variant Qv17199 is found basally in the Af lineage. Its genome shows 0.8 to 0.9% nucleotide sequence dissimilarity to that of other Af variants; this difference is equivalent to that calculated between HPV18 E2 and E1 sublineages (0.5 to 0.8%); thus, Qv17199 constitutes the Af2 sublineage. However, when the L1 genes were compared, the variant Qv17199 showed only 0.4 to 0.6% difference in nucleotide sequence from other Af variants (Fig. 9a). This suggests that complete genome analyses reveal more genomic diversity within the HPV18 Af lineage, of which there appear to be two sublineages (Af1 and Af2). In summary, these data support the empirical classification of HPV18 into two lineages that are further divided into sublineages.

Two deeply separated lineages of HPV45 variants were identified from genome comparisons and phylogenetic analyses, arbitrarily termed A and B. They are ~1.6% dissimilar to each other and contain two sublineages. The A1 sublineage is 0.8 to 0.9% dissimilar to the A2 sublineage; the B1 sublineage is 0.7 to 0.9% dissimilar to the B2 sublineage. The HPV45 prototype is clustered into the A1 sublineage. Since all HPV45 variants were sampled from admixed Hispanic females in Costa Rica, it was not possible to define the geographic origins of HPV45 lineages from this data set. All three HPV97 variants clustered together with only 18 of 7,843 (0.2% difference) nucleotides changed across the complete genome (Fig. 5 and 8b).

**Lineage fixation among different regions of HPV18 and HPV45.** Nucleotide variations in PV genomes and other rarely recombining genomes are fixed within lineages akin to linkage disequilibrium in organisms with recombining genomes known as haplotypes. Among the 295 and 186 variable nucleotide positions identified within the HPV18 and HPV45 genomes, 109 and 50 were lineage specific, respectively (Fig. 1 and 3, highlighted in gray). For instance, HPV18 E6 nucleotide changes T251C, G266A, G374A, C491A, and A548G were specific to the Af lineage, while HPV45 E1 nucleotide changes T1231G, T1456G, and G1477A differentiate the A lineage of HPV45 from the B lineage. Since HPV genomic recombination is very rare, if it occurs at all among HPVs, sequence changes in one region (e.g., E6) are highly correlated with and inseparable from changes in other regions (e.g., E1) within genomes from the same lineages, as revealed in previous analyses of HPV16 complete genomes (9). Lineage fixation of correlated genetic changes was observed throughout all regions of HPV18 and HPV45 variant genomes. For example, amino acid changes in HPV18 at E6 aa 129; E7 aa 2; and E1 aa 115, 155, 186, and 438 and 29 additional positions in the E2, E4, E5, L2, and L1 ORFs all segregated together, representing ancestral changes between the Af and non-Af taxa of HPV18 (Fig. 2). Fixed changes in HPV45 at E1 aa 106, 181, 562, and 627 and E2 aa 9, 44, 68, 147, and 171 and eight additional variations in the E4, L2, and L1 ORFs also represent ancestral changes between HPV45 A and B lineages (Fig. 4).

**Molecular clock predictions of genital HPV α7 species.** To calculate the approximate divergence times of HPV18, HPV45, and HPV97 variant lineages from their MRCAs, a Bayesian MCMC method was employed. Based on nucleotide sequence alignments of E6, E7, E1, E2, L2, and L1 ORFs, variants representing the main lineages were selected for analyses. As shown in Fig. 10 (top dendrogram) using the combined ORFs, HPV18, HPV45, and HPV97 shared an MRCA that evolved from a common ancestor with HPV59 around 14.6 million years (Myr) ago (95% highest posterior density [HPD], 8.7 to 23.4 Myr). This period of time overlaps the timing of the appearance of the common ancestor of the great ape species (2). Approximately 7.7 Myr ago (95% HPD, 4.5 to 10.5), the HPV18/45/97 MRCA began to diverge into distinct types; the HPV18 lineage diverged first, followed by the HPV45 and HPV97 lineages (≈5.7 Myr; 95% HPD, 3.1 to 7.8). This time period encompasses the era when the great ape species diverged. Variants within the HPV18, HPV45, and HPV97 lineages diverged from their common ancestors within the last million years (<0.7 Myr), which corresponds to a period of time when several genus *Homo* species including *H. sapiens* diverged and migrated across the globe. When the evolutionary rate was assumed to be fixed and equal within all HPV genes (data not shown), the L1 ORF identified the earliest divergence time of HPV18, HPV45, and HPV97 splitting from their common ancestor (≈11.8 Myr ago). Analysis using the L2 ORF indicated a slightly later emergence (≈10.4 Myr). However, analysis of early genes, particularly E1 and E2, indicated that these three types separated from their MRCA more recently (E6 gene, ≈10.4 Myr; E7 gene, ≈6.8 Myr; E2 gene, ≈6.7 Myr; E1 gene, ≈6.0 Myr). The result clearly reveals different evolutionary patterns and/or rates specific to each ORF.

**Natural selection among different genes of HPV18, HPV45, and HPV97.** To determine whether positive selection has been a force in the evolution of HPV18, HPV45, and HPV97 variants, nonsynonymous/synonymous rate ratios ($\omega = d_N/d_S$) were estimated. The likelihood analyses, including parameter esti-

FIG. 8. Intertypic and intratypic relationships of HPV18, HPV45, and HPV97 variants. Using a Bayesian MCMC method, phylogenetic trees were constructed based on concatenated amino acids and nucleotide sequences of eight ORFs (E6, E7, E1, E2, E4, E5, L2, and L1). (a) Phylogenetic relationships of HPV18, HPV45, and HPV97 within the genital PV α7 species. (b) Intratypic relationships of HPV18, HPV45, and HPV97 variants. HPV18 Af and non-Af variants formed two closely related clades/sublineages, arbitrarily termed Af1 and Af2 and termed E1 (AA) and E2, respectively. Phylogenetic evidence identified two deeply separated lineages of HPV45 variants, arbitrarily termed A and B, both of which contain two intratypic sublineages (A1 and A2 and B1 and B2). No lineage was determined for HPV97 variants. The bar indicates percent variation per unit length. The reference sequences of the α7 HPV types used are listed in Table S1 in the supplemental material, with the NCBI/GenBank database accession numbers indicated in parentheses: HPV97_624 (EF436229), HPV39 (NC_001535), HPV59 (NC_001635), HPV68 (DQ080079), HPV70 (NC_001711), and HPV85 (NC_004762).

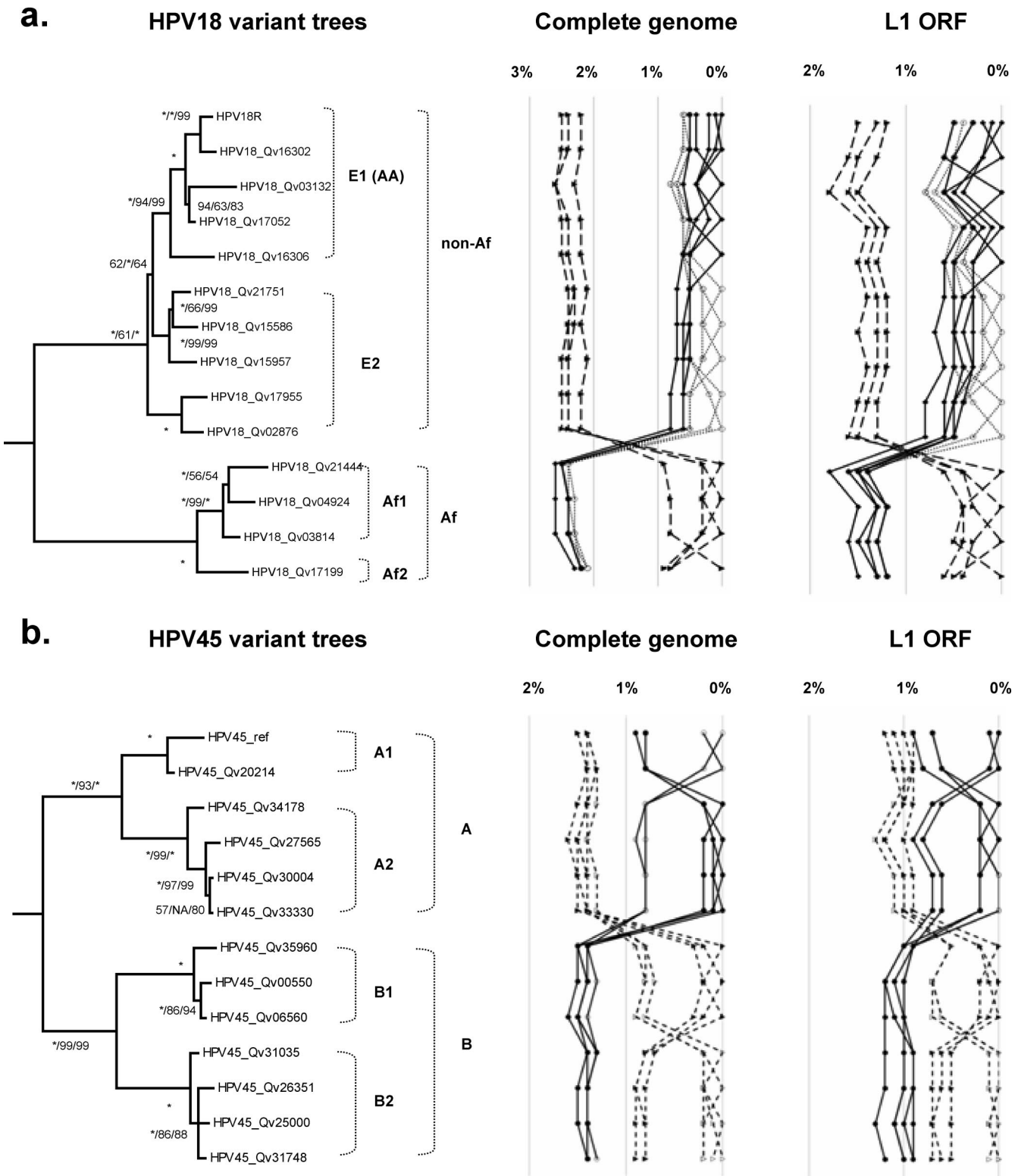



FIG. 9. Phylogeny and distance plots of HPV18 and HPV45. (a) Phylogenetic trees indicating the intratypic relationships of HPV18 variants. The nucleotide sequence dissimilarities of HPV18 variants inferred from the complete genomes and the L1 ORFs are shown in the panels to the right of the trees. (b) Phylogenetic trees indicating the intratypic relationships of HPV45 variants. The nucleotide sequence dissimilarities of HPV18 and HPV45 variants are calculated from the global pairwise alignments of the complete genomes and the L1 ORFs and are displayed in the panels to the right of the trees. Trees were inferred from the concatenated amino acids and nucleotide sequences of eight ORFs (E6, E7, E1, E2, E4, E5, L2, and L1). Numbers on or near branches indicate support indices in the following order: Bayesian credibility value using MrBayes v3.1.2 (20), MP bootstrap percentage, and NJ bootstrap percentage using PAUP* v4.0b10 (40). An asterisk indicates 100% agreement between methods. NA, disagreement between a method and the reference Bayesian tree at a given node.

mates for different models, are shown in Tables S1 to S3 in the supplemental material.

For each ORF, six models employing different assumptions about selection (ω) were used, and the model with the largest log likelihood value was used as the "best" model. In essentially all ORFs, the M3 (discrete) model was optimal. The $d_N/d_S$ ratio (ω) is an average over all sites in an ORF. For instance, the HPV18 E5 ORF had the largest average, ω = 1.9 by M3, with about 3.7% of sites under diversifying/Darwinian selection with ω = 36.7. The most statistically significant site in HPV18 E5 was identified as aa 72L (see Table S1a in the supplemental material). Similarly, using M3 for the HPV45 E6 ORF, the average $d_N/d_S$ ratio was 0.7. The majority (98.5%) of sites were under purifying selection with ω = <1, but 1.5% of sites were under positive selection with ω = 22.2, driven by HPV45 E6 aa 21L (see Table S2a in the supplemental material). The HPV45 L1 ORF had a site (L1 aa 383S) detected to be under diversifying selection with ω = 11.3 by the M3 model. Although these sites above were potentially under positive selection, they did not meet criteria for positive selection using the LRT as suggested by Yang et al. (50, 51) (see Tables S1b and S2b in the supplemental material). The HPV18 E4 (ω = 1.7) and E5 (ω = 1.9) ORFs and the HPV45 E4 ORF (ω = 1.6) had the highest average $d_N/d_S$ ratios, suggesting that these genes, as whole units, may be evolving under positive Darwinian selection (Fig. 11). When HPV18, HPV45, and HPV97 were considered as an "evolutionary unit," the E4 gene also had an average $d_N/d_S$ ratio slightly greater than 1 (ω = 1.1) (see Table S3 in the supplemental material); however, no specific amino acid site was identified to be under positive selection using the LRT (Fig. 6).

**Relationship of HPV18 and HPV45 variants.** It has been suggested based on URR sequence analyses that HPV45 is most closely related to the HPV18 Af lineage (29). To assess the relationships of HPV45 to HPV18 Af and non-Af variants, different Bayesian trees inferred from the nucleotide sequence alignments of early genes (E6, E7, E1, E2, and E5), late genes (L2 and L1), and the URR were constructed (Fig. 12). Within a 793-bp fragment of the URR (indel considered as a single event), 213 variant positions (26.9%) were detected among HPV18 and HPV45 isolates. The HPV18 Af variants showed 12 nucleotide changes identical to the HPV45 variants, whereas the HPV18 non-Af variants showed two nucleotide changes shared with the HPV45 variants. Similarly, the tree topology inferred from the URR sequences indicated a phylogenetic root of HPV18 and HPV45 in Africa. However, among 848 (19.8%) and 699 (23.4%) nucleotide variations identified within the early genes and the late genes, respectively, the HPV18 Af variants showed fewer nucleotide changes identical to HPV45 than did the HPV18 non-Af variants (early genes, 20 versus 28; late genes, 19 versus 23). Thus, the phylogenetic root of the HPV45 variant cluster was determined to be closer to the HPV18 non-Af origins in both "early gene" and "late gene" trees. Taken together, these analyses do not support an emergence of HPV45 from the HPV18 Af variant MRCA but suggest a more ancient origin, consistent with the molecular clock data. These results also indicate the need to consider the full viral genome sequence when assessing evolutionary history.

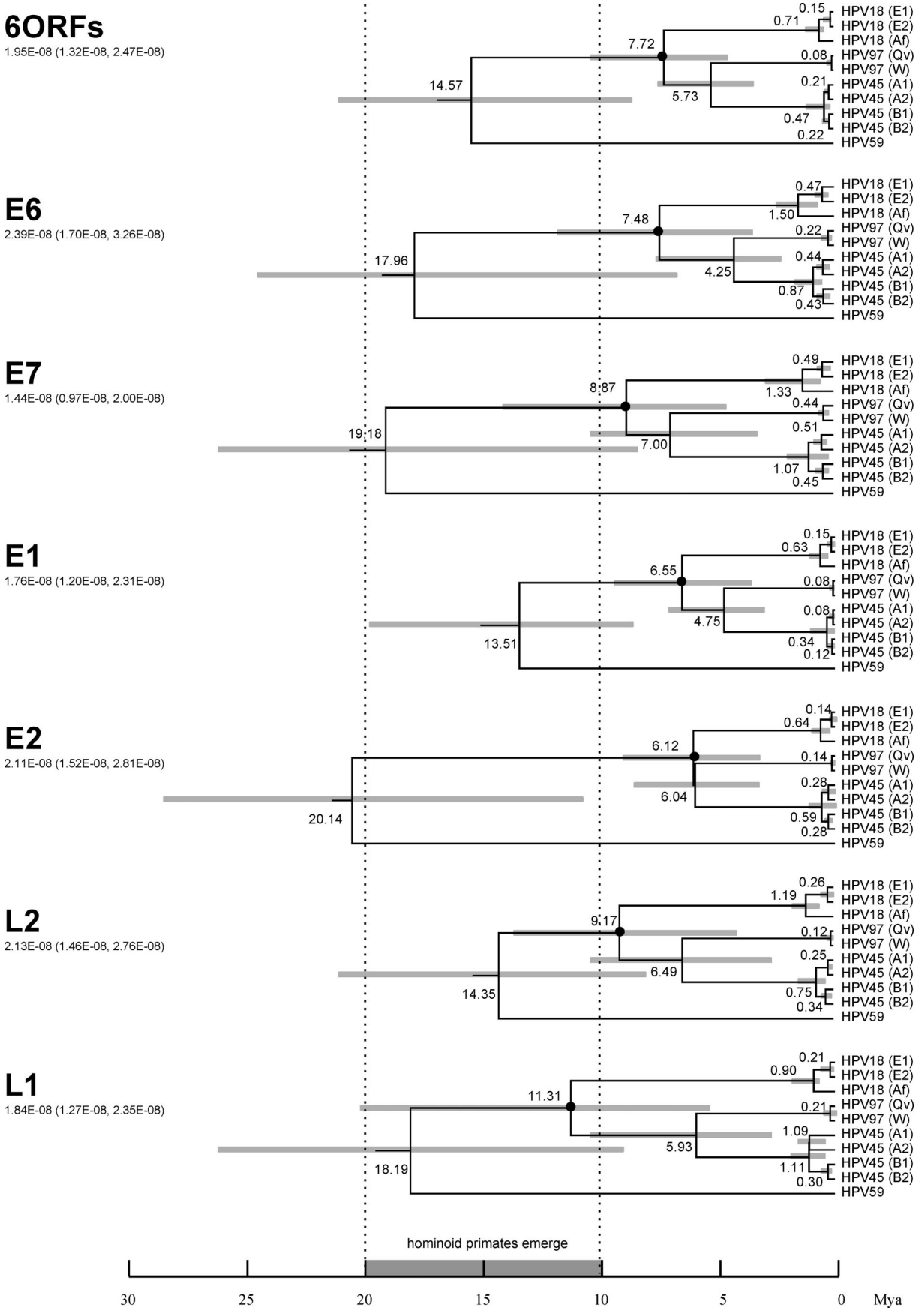
## DISCUSSION

**Evolution and divergence of HPV18, HPV45, and HPV97 coevolving with *Homo sapiens*.** These data suggest that HPV18/45/97 have expanded relatively recently with the divergence of *Homo sapiens* and the subsequent population growth. Since recombinant genomes of HPV have not been identified even after thorough and extensive characterization in humans, the evolution of HPV types is thought to be vertical; nevertheless, recombination events cannot be excluded. Although details are controversial among paleontologists, a widely accepted viewpoint regarding human evolution and global migration contends that *H. sapiens* evolved in Africa about 200,000 years ago, spreading from there into southern Asia and Australia from 80,000 to 60,000 years ago, replacing earlier genus *Homo* species, and thereafter reaching northern Asia (55,000 to 45,000 years ago) and Europe (35,000 years ago) (4, 13, 21, 22).

Based on the combination of six HPV genes (E6, E7, E1, E2, L2, and L1 ORFs), the MRCA of present-day HPV18, HPV45, and HPV97 viruses most likely appeared approximately 14.6 Myr ago, the time at which a variety of great ape common ancestors began to emerge (Fig. 10). It is also the time that most HPV α7 species types separated from their common ancestor. The HPV18, HPV45, and HPV97 group are more closely related to each other than to other α7 species types and subsequently diverged from their MRCA 7.7 Myr ago. Interestingly, when the evolutionary rate was fixed across all genes, the late genes showed earlier divergent times of HPV18/45/97 splitting from their common ancestor than did the early genes (i.e., L1 > L2 ≈ E6 > E7 ≈ E2 > E1). Given a lack of ORF recombination between HPV genomes, all ORFs should show similar divergence times if evolutionary pressures are equivalent across the PV genome. Differences in estimates of divergence times for different genes thus indicate different evolutionary rates and/or selective pressures across distinct HPV regions. In addition, intratypic variants of these types also diverged, at least in part, through genetic drift when human groups migrated to various geographical regions. Speciation of HPV18 and HPV45 occurred substantially before the development of intratype heterogeneity.

**Evolutionary selection within HPV18 and HPV45 genes.** The average $d_N/d_S$ rate ratios of the HPV18 E4 and E5 ORFs, the HPV45 E4 ORF, and the HPV18/45/97 E4 ORF were greater than 1, suggesting that these genes are under positive selection pressure. However, no specific sites were identified using the LRT (Fig. 6). The low overall nonsynonymous/synonymous substitution rate ratios (i.e., <1) observed in HPVs suggest that HPVs are under strong purifying selective pressure. Moreover, the low rate of change can also be attributed to the fact that PVs use the host cell DNA replication machinery, characterized by high fidelity, proofreading capacity, and postreplication repair mechanisms. In addition, many core functions of HPV-encoded proteins are required for the vegetative viral life cycle. These functions (e.g., viral capsid structure) result in purifying selection limiting the actual number of possible evolutionary events. Nevertheless, it is possible that other modes of genome evolution are in action such as codon usage or noncoding changes that were not measured in the current analyses.

Previous reports have observed six codon sites that are


6ORFs
1.95E-08 (1.32E-08, 2.47E-08)
E6
2.39E-08 (1.70E-08, 3.26E-08)
E7
1.44E-08 (0.97E-08, 2.00E-08)
E1
1.76E-08 (1.20E-08, 2.31E-08)
E2
2.11E-08 (1.52E-08, 2.81E-08)
L2
2.13E-08 (1.46E-08, 2.76E-08)
L1
1.84E-08 (1.27E-08, 2.35E-08)
HPV18 (E1)
HPV18 (E2)
HPV18 (Af)
HPV97 (Qv)
HPV97 (W)
HPV45 (A1)
HPV45 (A2)
HPV45 (B1)
HPV45 (B2)
HPV59
hominoid primates emerge
30
25
20
15
10
5
0
Mya

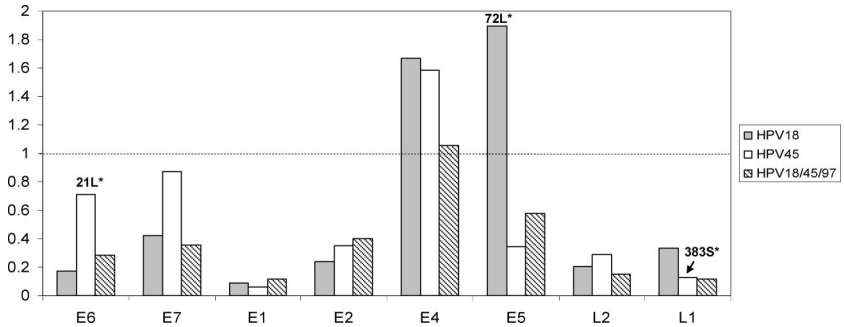



FIG. 11. Average $d_N/d_S$ rate ratios of each ORF of HPV18, HPV45, and HPV18/45/97. The M3 (discrete) model was used as the "best" one with the CODEML program in the PAML package (50, 51). The calculated values are taken from Tables S1 to S3 in the supplemental material. The dotted line marks neutral evolution with values below 1 suggesting purifying selection and values more than 1 suggesting Darwinian selection.

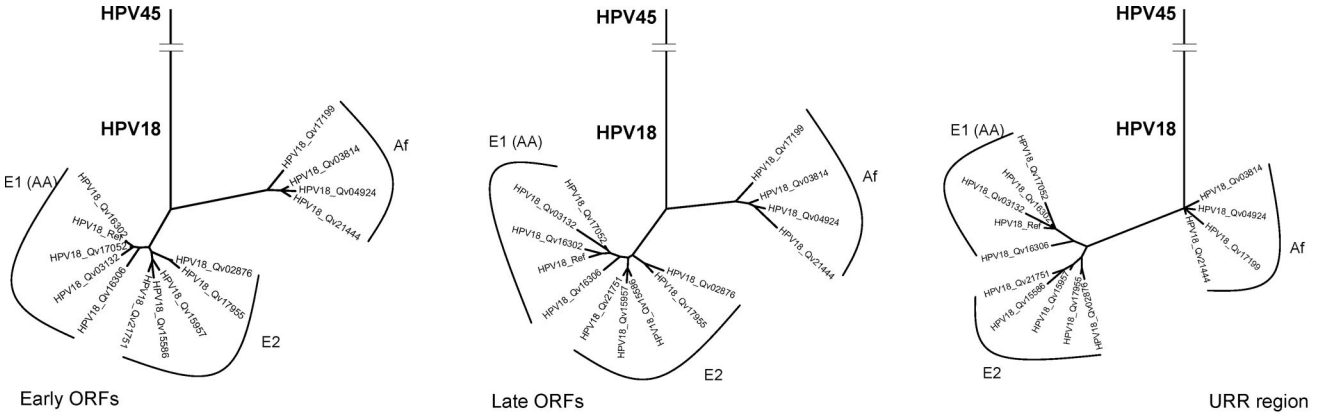
evolving under the influence of diversifying selection within the E6 and E5 genes of HPV16 (9, 12). We did not detect diversifying selection in the E5/E6 ORFs of the HPV18/45/97 clade. This suggests that different species and/or types of genital (alpha) PVs may be under different selective forces, such as avoiding the host immune response, adapting to specific epithelial tissues, and/or regulating the differentiation program of epithelial cells to facilitate viral replication.

Twelve amino acid variations (2.26%) within the HPV16 L1 (9) and 10 (1.86%) within the HPV45 L1 were observed (Fig. 7). Surprisingly, the L1 protein of HPV18 has at least 28 (4.93%) amino acid variations, over twice as many amino acid changes as HPV16 and HPV45 L1 ORFs. Amino acid alterations of L1 could affect efficiency of infection or alter viral antigenicity. Although amino acid variations within the HPV16 L1 capsid proteins contain cross-reacting epitopes (31, 44), it has also been reported that natural L1 protein mutations can affect virus-like particle assembly in vitro and negatively interfere with immunogenicity in mice (48). Antigenic diversity related to genomic variation has also been reported for the minor capsid protein L2 (47). Since the PV capsid structure is fixed, it was not surprising that changes within the HPV18 L1 ORF did not alter the predicted three-dimensional protein structure of the HPV18 L1 model (data not shown) (23). However, the substantial L1 amino acid variation could affect surface epitopes with the possibility of further changes favoring development of resistance to the current vaccine, which uses virus-like particles produced from a single HPV18 genome.

**Classification and phylogeny of HPV18 and HPV45 variants.** Although isolates from the same HPV type are referred to as "variants" when their L1 genes contain 1 to 2% nucleotide sequence diversity (14), this variability tends to differ by ORF due to different evolutionary rates and/or selective pressures. For instance, the HPV16 non-E lineages (Af1, Af2, and AA) differ by 0.3 to 0.9% within their L1 ORF, whereas the total differences increased to 1.1 to 1.5% when whole genomes were used ($P < 0.001$) (9). Similarly, the differences between HPV18 non-Af and Af variants and HPV45 A and B variants increased with the use of sequence data from whole genomes (Fig. 9). Importantly, the L1 ORF did not represent the full genome diversity. The classification of viral variant lineages should be based on the topology of a phylogeny using maximum sequence information. There appears to be a deep bifurcation of most HPV types ranging from variant differences of <1% for recently evolved HPV types (e.g., HPV97) to HPV types with differences of >5% (e.g., HPV68 forming subtypes).

Although conserved variations of HPV45 Af variants based on partial L1 ORF sequences have been described (38), there are few data on HPV45 variant lineages and their geographic distribution due to small sample sizes and limited sequence information. Based on the data in this report, HPV45 variants were divided into two well-separated lineages, each of which contained two monophyletic sublineages. The HPV45 prototype, initially isolated from a 26-year-old white female, was clustered into the A1 sublineage (26). Since all HPV45 variants in this work were sampled from admixed Hispanic females in Costa Rica, there was not sufficient information to define the geographic origins of HPV45 lineages. However, HPV characterization in a population from Rwanda revealed that 80% of 10 HPV45 isolates were classified in the B2 sublineage (R. D. Burk, personal communication). Since Rwanda is in east-central Africa, the HPV45 B2 variants might be classified as of "African" origin (Fig. 8). However, HPV45 isolates from Zambia in southern Africa gave a more complicated picture,

FIG. 10. Dating of phylogenetic nodes. Variants representing each intratype variant lineage of HPV18, HPV45, and HPV97 were used to generate trees based on nucleotide sequence alignments of six ORFs (E6, E7, E1, E2, L2, and L1) and each individual ORF. Branch lengths are proportional to divergence times. Numbers above the nodes are the mean estimated divergence times (in Myr). The bars in gray represent the 95% HPD interval for the divergence times. A Bayesian MCMC method was used to calculate divergence times in the BEAST program v1.4.7 by selecting the uncorrelated lognormal distribution model (15). The node of the MRCA of HPV18, HPV45, and HPV97 is denoted with a black circle. The ORF used to generate each tree is shown on the left. The number under the ORF is the mutation rate (nucleotide substitutions per site per year, with 95% confidence interval) used to calculate the divergence times (33). The scale at the bottom is in Myr with the period of hominoid primate emergence shown in gray with broken vertical lines highlighted. Variants representing each intratype variant are HPV18R for E1 (AA), HPV18_Qv15586 for E2, HPV18_Qv04924 for Af, HPV97_Qv28597, HPV97_W15189, HPV45 ref for A1, HPV45_Qv27565 for A2, HPV45_Qv00550 for B1, and HPV45_Qv25000 for B2.




Nucleotide sequence alignment of HPV18 non-African and African variants and HPV45 variants

| Regions | Early genes | | | | | Late genes | | URR | Lineage |
|---|---|---|---|---|---|---|---|---|---|
| | 11111111111111111111222222222222 | 33333333333333333333 | 34444 | 44444444444444444455555555555555555555 | 55555556666666677 | 77777777777777777777 | | | |
| | 223334455 5678 | 99000112223334444477788122234444555 67 | 01222345555567777777888 | 90111 | 22333344556666789999000111223333333345555 6 | 67889904455677901 | 11111234455555566667777 1 | | |
| | 561478944 9456 | 57168295670570256737849026730448056 50 | 54117183578932234559122 | 75155 | 48244428170469720491458146280134556810080 1 | 48237221037821416 3 | 55668528912366945570234 90 | | |
| | 167245189 3014 | 26237346493388326054931169051062916 61 | 39175424886303549792578 | 59904 | 95801896096203227838117977708869481003514 9 | 84925047109169974 0 | 26145836629037231804601 24 | | |
| Variants | E6 E7 | E1 | E2 | E5 | L2 | L1 | | | |
| HPV18R_ref | TGTCGTCAC CCCA | TAATTCCGTGTTTGCTGAATGTAACCCCTGCTGGATA | GTGCGGGTCCATCCGGCGGGGCC | AGTTG | AACCCGCGCGGGGGGGGAAGTGCGCCTAGGTTTTATCGCGAT | TCCACCACAAGTCGGGCG | AACCCTACCGCTGATTTAATCCAAT | | E1(AA) |
| HPV18_Qv03132 | ..... .... | .....A..... | ...A..... | ..... | .....T.....A.....C...A.... | ..... | .....C..G..... | | E1(AA) |
| HPV18_Qv17052 | ..... .... | .....A..... | ...A..... | ..... | .....T.....A.....C...A.... | ..... | .....C..G..... | | E1(AA) |
| HPV18_Qv16302 | ..... .... | ..... | ..... | ..... | .....T.....A.....C..... | ..... | .....C..... | | E1(AA) |
| HPV18_Qv16306 | .....C..A .... | .....A.....GA..... | ...A.....A..... | ..... | .....TT..T....TA.....C...A... | ....A..... | .....A....A..CC..... | | E1(AA) |
| HPV18_Qv21751 | .....C..A ..T. | ..T...A....A.....G.G..... | ...A.....GAA..... | ..... | .....TT.....TA...A....C...A..A | ....A...G..... | .....A.T..A.ACC...T....C | | E2 |
| HPV18_Qv15586 | .....C..A ..T. | ..T...A....A.....G.G..... | ...A.....GAA..... | ..... | .....TT.....TA...A....C...A..A | ....A...G..... | .....A.T....ACC...T....C | | E2 |
| HPV18_Qv15957 | .....C..A ..T. | ..T...A....A.....G..T..... | ...A.....GAA..... | ..... | .....TT.....TA...A....C...A..A | ....A...G..... | .....A.T..A..CC...T....C | | E2 |
| HPV18_Qv17955 | .....C..A ..T. | .GTC.TA.....CA..... | .....C | ....A..T....G.A..... | ....A.....TT.....TA...A....C..TA... | ....AT..GC..... | .....A....A..CC...T...G.C | | E2 |
| HPV18_Qv02876 | .....C..A ..T. | .GTC.TA.....CA..... | .....C | ....A..T....G.A..... | ....A.....TT.....TA...A....C..TA... | ....AT..GC..... | .....A....A..CC...T...G.C | | E2 |
| HPV18_Qv21444 | CACTACAGA TT.G | G.T.CTACAAC.GAT..CTGA.GGTTT.CATCATCC. | TGCAAA.CATCG-A.AGAATATA | TACCA | TGTAAAATTTTTTATTAATCAGATTATCGATCCGCCA.AAAG. | CTTCA.TTG.ACTAAATA | GGTGTAG.GA.CACCGCCTCTA.G. | | Af1 |
| HPV18_Qv04924 | CACTACAGA TT.G | G.T.CTACAAC.GAT..CTGA.GGTTTTCATCATCC. | TGCAAA.CATCG-A.AGAATATA | TACCA | TGTAAAATTTTTTATT.ATCAGATTATCGATCCGCCA.AAAG. | CT.CA..TTG.ACTAAATA | GGTGTAG.GA.CACCGCCTCTA.G. | | Af1 |
| HPV18_Qv03814 | CACTACAGA TT.G | G.T.CTACAAC.GAT..CTGA.GGTTTTCATCATCC. | TGCAAA.CATCG-A.AGAATATA | TACCA | TGTAAAATTTTTTATT..TCAGATTATCGATCCGCCA.AAAG. | CTTCA.TTG.ACTAAATA | GGTGTAG.GA.CACCGCCTCTA.G. | | Af1 |
| HPV18_Qv17199 | CA..ACAGA TT.. | G.T.CTACAAC.GAT..CT.A.GGTTTT.ATCATCC. | TGCAAA.CATCG-A.AAAATATA | TAC.A | TGTAAAATTTTTTATTA.TCAGATTATCGATCCGCCA.A.AG. | CT.CA..TG.ACTAAATA | GGTGTAG.GAAAACCGCCTCTA.G. | | Af2 |
| HPV45_ref | .A.T....A .T.G | G.....A.AC.C-A-G.-.AAA.....A.AA.AAT.. | .G.AAAAC..-CAAAA.C.TATA | C...A | ...AA..TATATTTT.CT.CGATCAT..AT...C..ACAA.. | .TT..TTTGCA....TT. | GGTGTCTGAA--AT.GC.TCTA..G | | A1 |
| HPV45_Qv20214 | .A.T....A .T.G | G.....A.AC.C-A-G.-.AAA.....A.AA.AAT.. | .G.AAAAC..-CAAAA.C.TATA | C...A | ...AA..TATATTTT..T.CGATCAT..AT...C..ACAA.. | .TT..TTTGCA....TT. | GGTGTCTGAA--AT.GC.TCTA..G | | A1 |
| HPV45_Qv34178 | .A.T.C..A .T.G | G.....A.AC.C-A-G.-.AAA.....A.AA.AAT.. | .G.AAAAC..-CAAAA.C.TATA | C...A | ...AA..TATATTTT..T.CGATCAT..AT...C..ACAA.. | .TT..TTTGCA....TT. | GGTGTCTGAA--AT.GC.TCTA..G | | A2 |
| HPV45_Qv27565 | .A.T.C..A .T.G | G.....A.AC.C-A-G.-.AAA.....A.AA.AAT.. | .G.AAAAC..-CAAAA.C.TATA | C...A | ...AA..TATATTTT..T.CGATCAT..AT...C..ACAA.. | .TT..TTTGCA....TT. | GGTGTCTGAA--AT.GC.TCTA..G | | A2 |
| HPV45_Qw30004 | .A.T.C..A .T.G | G.....A.AC.C-A-G.-.AAA.....A.AA.AAT.. | .G.AAAAC..-CAAAA.C.TATA | C...A | ...AA..TATATTTT..T.CGATCAT..AT...C..ACAA.. | .TT..TTTGCA....TT. | GGTGTCTGAA--AT.GC.TCTA..G | | A2 |
| HPV45_Qw33330 | .A.T.C..A .T.G | G.....A.AC.C-A-G.-.AAA.....A.AA.AAT.. | .G.AAAAC..-CAAAA.C.TATA | C...A | ...AA..TATATTTT..T.CGATCAT..AT...C..ACAA.. | .TT..TTTGCA....TT. | GGTGTCTGAA--AT.GC.TCTA..G | | A2 |
| HPV45_Qw35960 | .A.T.C..A .T.G | G.....A.AC.C-A-G.-.AAA.....A.AA.AAT.. | .G.AAAAC..-CAAAA.C.TATA | C...A | ...AA..TATATTTT..T.CGATCAT..AT...C..ACAA.. | .TT..TTTGCA....TTA | GGTGTCTTAA--AT.GC.TCTA..G | | B1 |
| HPV45_Qw00550 | .A.T.C..A .T.G | G.....A.AC.C-A-G.-.AAA.....A.AA.AAT.. | .G.AAAAC..-CAAAA.C.TATA | C...A | ...AA..TATATTTT..T.CGATCAT..AT...C..ACAA.. | .TT..TTTGCA....TTA | GGTGTCTTAA--AT.GC.TCTA..G | | B1 |
| HPV45_Qw06560 | .A.T.C..A .T.G | G.....A.AC.C-A-G.-.AAA.....A.AA.AAT.. | .G.AAAAC..-CAAAA.C.TATA | C...A | ...AA..TATATTTT..T.CGATCAT..AT...C..ACAA.. | .TT..TTTGCA....TTA | GGTGTCTTAA--AT.GC.TCTA..G | | B1 |
| HPV45_Qw31035 | .A.T.C..A .T.G | G.....A.AC.C-A-G.-.AAA.....A.AA.AAT.. | .G.AAAAC..-CAAAA.C.TATA | C...A | ...AA..TATATTTT..T.CGATCAT..AT...C..ACAA.. | .TT..TTTGCA....TTA | GGTGTCTTAA--AT.GC.TCTA..G | | B2 |
| HPV45_Qw26351 | .A.T.C..A .T.G | G.....A.AC.C-A-G.-.AAA.....A.AA.AAT.. | .G.AAAAC..-CAAAA.C.TATA | C...A | ...AA..TATATTTT..T.CGATCAT..AT...C..ACAA.. | .TT..TTTGCA....TTA | GGTGTCTTAA--AT.GC.TCTA..G | | B2 |
| HPV45_Qw25000 | .A.T.C..A .T.G | G.....A.AC.C-A-G.-.AAA.....A.AA.AAT.. | .G.AAAAC..-CAAAA.C.TATA | C...A | ...AA..TATATTTT..T.CGATCAT..AT...C..ACAA.. | .TT..TTTGCA....TTA | GGTGTCTTAA--AT.GC.TCTA..G | | B2 |
| HPV45_Qv31748 | .A.T.C..A .T.G | G.....A.AC.C-A-G.-.AAA.....A.AA.AAT.. | .G.AAAAC..-CAAAA.C.TATA | C...A | ...AA..TATATTTT..T.CGATCAT..AT...C..ACAA.. | .TT..TTTGCA....TTA | GGTGTCTTAA--AT.GC.TCTA..G | | B2 |
| *No. of nt changes* | HPV18 Non-Af: 28 | | | | HPV18 Non-Af: 23 | | HPV18 Non-Af: 2 | | |
| *Identical to HPV45* | HPV18 Af: 20 | | | | HPV18 Af: 19 | | HPV18 Af: 12 | | |

FIG. 12. Phylogenetic relationship of HPV18 and HPV45 variants. Bayesian trees were inferred from the nucleotide sequences of early genes (E6, E7, E1, E2, and E5), late genes (L2 and L1), and the URR.

with 8/22 (36%) being from the B2 lineage and 10 (46%) being from the A1 lineage (R. D. Burk, personal communication). Thus, both main variant lineages of HPV45 have geographic origins in Africa and suggest that the recent evolution of HPV45 differs from that of HPV16 and HPV18; additional studies are under way to assess the geographic history of HPV45 variants.

HPV18, HPV45, and HPV97 comprise a clade sharing an MRCA within HPV $\alpha$7 species. This report describes a large set of PV genomes representing the major variant lineages of HPV18, HPV45, and HPV97. These genomes capture the heterogeneity within a group of $\alpha$7 HPV types responsible for 20% of cervix cancers. In this work, intratypic and intertypic relationships reveal a deep division between variants of HPV18 and HPV45 resulting from historical events of early primate evolution. Lack of evidence for recombination and the low overall nonsynonymous/synonymous substitution rate ratios observed suggest that HPVs are under strong purifying selective pressure. The data indicate that other evolutionary mechanisms not measured may be important for HPV evolution. Using a Bayesian MCMC method, the divergence times of HPV18, HPV45, and HPV97 from their MRCA predicted a timescale strongly supporting virus-host coevolution. Understanding the heterogeneity of present-day PVs can serve as a model for monophyletic evolution of double-stranded DNA viral genomes over long periods of time.

## ACKNOWLEDGMENTS

We acknowledge Anne Breheny, Amy Razukiewicz, and Andrew Prior for performing HPV DNA genotyping analyses.

This work was supported in part by Public Health Service award CA78527 from the National Cancer Institute. R.D. thanks the Louis and Dorothy Cullman Program in Molecular Systematics at the American Museum of Natural History for support.

## REFERENCES

1. **Altekruse, S. F., J. V. Lacey, L. A. Brinton, P. E. Gravitt, S. G. Silverberg, W. A. Barnes, M. D. Greenberg, O. C. Hadjimichael, L. McGowan, R. Mortel, P. E. Schwartz, and A. Hildesheim.** 2003. Comparison of human papillomavirus genotypes, sexual, and reproductive risk factors of cervical adenocarcinoma and squamous cell carcinoma: northeastern United States. Am. J. Obstet. Gynecol. **188:**657–663.
2. **Andrews, P.** 1992. Evolution and environment in the Hominoidea. Nature **360:**641–646.
3. **Arias-Pulido, H., C. L. Peyton, N. Torrez-Martinez, D. N. Anderson, and C. M. Wheeler.** 2005. Human papillomavirus type 18 variant lineages in United States populations characterized by sequence analysis of LCR-E6, E2, and L1 regions. Virology **338:**22–34.
4. **Behar, D. M., R. Villems, H. Soodyall, J. Blue-Smith, L. Pereira, E. Metspalu, R. Scozzari, H. Makkan, S. Tzur, D. Comas, J. Bertranpetit, L. Quintana-Murci, C. Tyler-Smith, R. S. Wells, and S. Rosset.** 2008. The dawn of human matrilineal diversity. Am. J. Hum. Genet. **82:**1130–1140.

5. **Bosch, F. X., M. M. Manos, N. Munoz, M. Sherman, A. Jansen, J. Peto, M. Schiffman, V. Moreno, R. Kurman, K. Shah, et al.** 1995. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. J. Natl. Cancer Inst. **87:**796–802.
6. **Burk, R. D., M. Terai, P. E. Gravitt, L. A. Brinton, R. J. Kurman, W. A. Barnes, M. D. Greenberg, O. C. Hadjimichael, L. Fu, L. McGowan, R. Mortel, P. E. Schwartz, and A. Hildesheim.** 2003. Distribution of human papillomavirus types 16 and 18 variants in squamous cell carcinomas and adenocarcinomas of the cervix. Cancer Res. **63:**7215–7220.
7. **Calleja-Macias, I. E., M. Kalantari, J. Huh, R. Ortiz-Lopez, A. Rojas-Martinez, J. F. Gonzalez-Guerrero, A. L. Williamson, B. Hagmar, D. J. Wiley, L. Villarreal, H. U. Bernard, and H. A. Barrera-Saldana.** 2004. Genomic diversity of human papillomavirus-16, 18, 31, and 35 isolates in a Mexican population and relationship to European, African, and Native American variants. Virology **319:**315–323.
8. **Chen, Z., L. Fu, R. Herrero, M. Schiffman, and R. D. Burk.** 2007. Identification of a novel human papillomavirus (HPV97) related to HPV18 and HPV45. Int. J. Cancer **121:**2947–2952.
9. **Chen, Z., M. Terai, L. Fu, R. Herrero, R. DeSalle, and R. D. Burk.** 2005. Diversifying selection in human papillomavirus type 16 lineages based on complete genome analyses. J. Virol. **79:**7014–7023.
10. **Chow, V. T., and P. W. Leong.** 1999. Complete nucleotide sequence, genomic organization and phylogenetic analysis of a novel genital human papillomavirus type, HLT7474-S. J. Gen. Virol. **80:**2923–2929.
11. **Cogliano, V., R. Baan, K. Straif, Y. Grosse, B. Secretan, and F. El Ghissassi.** 2005. Carcinogenicity of human papillomaviruses. Lancet Oncol. **6:**204.
12. **DeFilippis, V. R., F. J. Ayala, and L. P. Villarreal.** 2002. Evidence of diversifying selection in human papillomavirus type 16 E6 but not E7 oncogenes. J. Mol. Evol. **55:**491–499.
13. **DeSalle, R., and I. Tattersall.** 2008. Human origins. Texas A&M University Press, College Station.
14. **de Villiers, E. M., C. Fauquet, T. R. Broker, H. U. Bernard, and H. zur Hausen.** 2004. Classification of papillomaviruses. Virology **324:**17–27.
15. **Drummond, A. J., and A. Rambaut.** 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. **7:**214.
16. **Forslund, O., and B. G. Hansson.** 1996. Human papillomavirus type 70 genome cloned from overlapping PCR products: complete nucleotide sequence and genomic organization. J. Clin. Microbiol. **34:**802–809.
17. **Gorska-Flipot, I., J. Sawick, L. A. Gaboury, M. Krajinovic, D. Labuda, I. Brukner, D. Rouleau, G. Ghattas, E. L. Franco, and F. Coutlee.** 2008. Newly-isolated HPV97, related to HPV18 and 45 is frequently detected in HIV-positive men from the Montreal area. Int. J. Cancer **122:**1195–1197.
18. **Hastings, W.** 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika **57:**97–109.
19. **Herrero, R., P. E. Castle, M. Schiffman, M. C. Bratti, A. Hildesheim, J. Morales, M. Alfaro, M. E. Sherman, S. Wacholder, S. Chen, A. C. Rodriguez, and R. D. Burk.** 2005. Epidemiologic profile of type-specific human papillomavirus infection and cervical neoplasia in Guanacaste, Costa Rica. J. Infect. Dis. **191:**1796–1807.
20. **Huelsenbeck, J. P., and F. Ronquist.** 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17:**754–755.
21. **Macaulay, V., C. Hill, A. Achilli, C. Rengo, D. Clarke, W. Meehan, J. Blackburn, O. Semino, R. Scozzari, F. Cruciani, A. Taha, N. K. Shaari, J. M. Raja, P. Ismail, Z. Zainuddin, W. Goodwin, D. Bulbeck, H. J. Bandelt, S. Oppenheimer, A. Torroni, and M. Richards.** 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. Science **308:**1034–1036.
22. **Mellars, P.** 2006. Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. Proc. Natl. Acad. Sci. USA **103:**9381–9386.
23. **Modis, Y., B. L. Trus, and S. C. Harrison.** 2002. Atomic model of the papillomavirus capsid. EMBO J. **21:**4754–4762.
24. **Moscicki, A. B., M. Schiffman, S. Kjaer, and L. L. Villa.** 2006. Chapter 5: updating the natural history of HPV and anogenital cancer. Vaccine **24**(Suppl. 3)**:**S42–S51.
25. **Munoz, N., F. X. Bosch, S. de Sanjose, R. Herrero, X. Castellsague, K. V. Shah, P. J. Snijders, and C. J. Meijer.** 2003. Epidemiologic classification of human papillomavirus types associated with cervical cancer. N. Engl. J. Med. **348:**518–527.
26. **Naghashfar, Z. S., N. B. Rosenshein, A. T. Lorincz, J. Buscema, and K. V. Shah.** 1987. Characterization of human papillomavirus type 45, a new type 18-related virus of the genital tract. J. Gen. Virol. **68:**3073–3079.
27. **Narechania, A., Z. Chen, R. DeSalle, and R. D. Burk.** 2005. Phylogenetic incongruence among oncogenic genital alpha human papillomaviruses. J. Virol. **79:**15503–15510.
28. **Nielsen, R., and Z. Yang.** 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics **148:**929–936.
29. **Ong, C. K., S. Y. Chan, M. S. Campo, K. Fujinaga, P. Mavromara-Nazos, 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.
30. **Parkin, D. M., F. Bray, J. Ferlay, and P. Pisani.** 2005. Global cancer statistics, 2002. CA Cancer J. Clin. **55:**74–108.
31. **Pastrana, D. V., W. C. Vass, D. R. Lowy, and J. T. Schiller.** 2001. NHPV16 VLP vaccine induces human antibodies that neutralize divergent variants of HPV16. Virology **279:**361–369.
32. **Posada, D., and K. A. Crandall.** 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14:**817–818.
33. **Rector, A., P. Lemey, R. Tachezy, S. Mostmans, S. J. Ghim, K. Van Doorslaer, M. Roelke, M. Bush, R. J. Montali, J. Joslin, R. D. Burk, A. B. Jenson, J. P. Sundberg, B. Shapiro, and M. Van Ranst.** 2007. Ancient papillomavirus-host co-speciation in Felidae. Genome Biol. **8:**R57.
34. **Rho, J., A. Roy-Burman, H. Kim, E. M. de Villiers, T. Matsukura, and J. Choe.** 1994. Nucleotide sequence and phylogenetic classification of human papillomavirus type 59. Virology **203:**158–161.
35. **Schlecht, N. F., R. D. Burk, J. M. Palefsky, H. Minkoff, X. Xue, L. S. Massad, M. Bacon, A. M. Levine, K. Anastos, S. J. Gange, D. H. Watts, M. M. Da Costa, Z. Chen, J. Y. Bang, M. Fazzari, C. Hall, and H. D. Strickler.** 2005. Variants of human papillomaviruses 16 and 18 and their natural history in human immunodeficiency virus-positive women. J. Gen. Virol. **86:**2709–2720.
36. **Sichero, L., S. Ferreira, H. Trottier, E. Duarte-Franco, A. Ferenczy, E. L. Franco, and L. L. Villa.** 2007. High grade cervical lesions are caused preferentially by non-European variants of HPVs 16 and 18. Int. J. Cancer **120:**1763–1768.
37. **Smith, J. S., L. Lindsay, B. Hoots, J. Keys, S. Franceschi, R. Winer, and G. M. Clifford.** 2007. Human papillomavirus type distribution in invasive cervical cancer and high-grade cervical lesions: a meta-analysis update. Int. J. Cancer **121:**621–632.
38. **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.
39. **Strickler, H. D., R. D. Burk, M. Fazzari, K. Anastos, H. Minkoff, L. S. Massad, C. Hall, M. Bacon, A. M. Levine, D. H. Watts, M. J. Silverberg, X. Xue, N. F. Schlecht, S. Melnick, and J. M. Palefsky.** 2005. Natural history and possible reactivation of human papillomavirus in human immunodeficiency virus-positive women. J. Natl. Cancer Inst. **97:**577–586.
40. **Swofford, D. L.** 1998. PAUP*. Phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer Associates, Sunderland, MA.
41. **Terai, M., and R. D. Burk.** 2001. Characterization of a novel genital human papillomavirus by overlapping PCR: candHPV86 identified in cervicovaginal cells of a woman with cervical neoplasia. J. Gen. Virol. **82:**2035–2040.
42. **Terai, M., and R. D. Burk.** 2002. Identification and characterization of 3 novel genital human papillomaviruses by overlapping polymerase chain reaction: candHPV89, candHPV90, and candHPV91. J. Infect. Dis. **185:**1794–1797.
43. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:**4876–4882.
44. **Touze, A., S. El Mehdaoui, P. Y. Sizaret, C. Mougin, N. Munoz, and P. Coursaget.** 1998. The L1 major capsid protein of human papillomavirus type 16 variants affects yield of virus-like particles produced in an insect cell expression system. J. Clin. Microbiol. **36:**2046–2051.
45. **Volpers, C., and R. E. Streeck.** 1991. Genome organization and nucleotide sequence of human papillomavirus type 39. Virology **181:**419–423.
46. **Xi, L. F., L. A. Koutsky, A. Hildesheim, D. A. Galloway, C. M. Wheeler, R. L. Winer, J. Ho, and N. B. Kiviat.** 2007. Risk for high-grade cervical intraepithelial neoplasia associated with variants of human papillomavirus types 16 and 18. Cancer Epidemiol. Biomarkers Prev. **16:**4–10.
47. **Yaegashi, N., L. Xi, M. Batra, and D. A. Galloway.** 1993. Sequence and antigenic diversity in two immunodominant regions of the L2 protein of human papillomavirus types 6 and 16. J. Infect. Dis. **168:**743–747.
48. **Yang, R., C. M. Wheeler, X. Chen, S. Uematsu, K. Takeda, S. Akira, D. V. Pastrana, R. P. Viscidi, and R. B. Roden.** 2005. Papillomavirus capsid mutation to escape dendritic cell-dependent innate immunity in cervical cancer. J. Virol. **79:**6741–6750.
49. **Yang, Z.** 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13:**555–556.
50. **Yang, Z., R. Nielsen, N. Goldman, and A. M. Pedersen.** 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155:**431–449.
51. **Yang, Z., R. Nielsen, and M. Hasegawa.** 1998. Models of amino acid substitution and applications to mitochondrial protein evolution. Mol. Biol. Evol. **15:**1600–1611.