

Phylogenetic Incongruence among Oncogenic Genital Alpha Human Papillomaviruses

Apurva Narechania,¹ Zigui Chen,¹ Rob DeSalle,⁵ and Robert D. Burk^{1,2,3,4*}

Departments of Microbiology and Immunology,¹ Pediatrics,² Obstetrics, Gynecology and Women's Health,³ and Epidemiology and Population Health,⁴ Albert Einstein Cancer Center, Albert Einstein College of Medicine, 1300 Morris Park Ave., Bronx, New York 10461, and Division of Invertebrate Zoology, American Museum of Natural History, New York, New York 10024⁵

Received 7 June 2005/Accepted 12 September 2005

The human papillomaviruses (HPVs) have long been thought to follow a monophyletic pattern of evolution with little if any evidence for recombination between genomes. On the basis of this model, both oncogenicity and tissue tropism appear to have evolved once. Still, no systematic statistical analyses have shown whether monophyly is the rule across all HPV open reading frames (ORFs). We conducted a taxonomic analysis of 59 mucosal/genital HPVs using whole-genome and sliding-window similarity measures; maximum-parsimony, neighbor-joining, and Bayesian phylogenetic analyses; and localized incongruence length difference (LILD) analyses. The algorithm for the LILD analyses localized incongruence by calculating the tree length differences between constrained and unconstrained nodes in a total-evidence tree across all HPV ORFs. The process allows statistical evaluation of every ORF/node pair in the total-evidence tree. The most significant incongruence was observed at the putative high-risk (i.e., cancer-associated) node, the common oncogenic ancestor for alpha HPV species 9 (e.g., HPV type 16 [HPV16]), 11, 7 (e.g., HPV18), 5, and 6. Although these groups share early-gene homology, including high degrees of similarity among E6 and E7, groups 9 and 11 diverge from groups 7, 5, and 6 with respect to L2 and L1. The HPV species groups primarily associated with cervical and anogenital cancers appear to follow two distinct evolutionary paths, one conferred by the early genes and another by the late genes. The incongruence in the genital HPV phylogeny could have occurred from an early recombination event, an ecological niche change, and/or asymmetric genome convergence driven by intense selection. These data indicate that the phylogeny of the oncogenic HPVs is complex and that their evolution may not be monophyletic across all genes.

Over the past 30 years, intense systematic sequencing efforts have resulted in the availability of full genomes for over 100 human papillomavirus (HPV) types. Several phylogenetic studies have emerged at various points during the accumulation of these data (4, 5, 15, 27). Classification of types into gross categories, such as cancer-causing and non-cancer-causing or cutaneous and mucosal viruses, is possible, since most oncogenic types share a common ancestor, and most common ancestors radiate within an ecological niche (e.g., infection of skin) (27), but genomic classifications are not always straightforward. The evolutionary histories of genes can be incongruent, confounding sequence-based taxonomies, which can be discrepant with the paleontological history of the taxa (10, 11). Despite efforts spent on minimizing its effects (19), incongruence is also a sign of a gene's break with organismal evolutionary history, sometimes revealing recombination or periods of intense selection (10, 20). The considerable collection of papillomavirus (PV) sequence data, coupled to its inherent genomic stability, provides an ideal system for analysis of the evolutionary history of individual PV open reading frames (ORFs).

The typical HPV genome contains seven or eight open reading frames. The late genes, L1 and L2, encode structural proteins that comprise the viral capsid, while two of the early

genes, E1 and E2, mediate viral genome replication (7). The ability of high-risk HPV infection to progress to malignancy is due in large part to the activities of two other early ORFs, E6 and E7. High-risk E6 and E7 are the primary oncoproteins, influencing the activities of cell cycle protein p53 and pRB, respectively (6). Many animal PVs and nearly all cutaneous and low-risk HPVs also contain E6 and E7. Though other proteins, such as E4 and E5, are thought to complement their activity by modulating viral late functions (9), high-risk E6 and E7 are important cell cycle regulators, and their expression is often sufficient to immortalize primary human keratinocytes in tissue culture systems (13).

A striking feature of PV genome organization is that the ORFs have remained largely intact through millions of years of evolution extending through avian and mammalian PV lineages (22). As double-stranded DNA viruses, PVs maintain high fidelity in viral genome replication by usurping the host cell proofreading polymerases. Evolution proceeds clonally through mutation and selection or genetic drift, and remarkably, no examples of recombination have been documented.

The overall stability of PV genomes suggests a simple monophyletic model of evolution. Indeed, several traits, including oncogenicity and tissue tropism appear to have evolved only once (15, 27). Bravo and Alonso (2) have shown that this monophyly breaks down when the phylogenies of early and late genes are compared qualitatively. However, no studies have yet shown the extent of this breakdown in statistical terms. To date, the approach of most classification studies has conflated

* Corresponding author. Mailing address: Albert Einstein College of Medicine, Ullman Building, Room 515, 1300 Morris Park Ave., Bronx, NY 10461. Phone: (718) 430-3720. Fax: (718) 430-8975. E-mail: burk@acom.yu.edu.

the phylogenetic histories of the papillomavirus genes to serve a taxonomic end. For example, because the PV L1 gene is known to be highly conserved, it has been regarded as the analysis standard of PV evolution: genera, species, types, subtypes, and variants have been classified by degrees of divergence in L1, and the L1 phylogeny is often taken as representative of the virus (5). In this report, we provide quantitative statistical evidence that phylogenies created from several alpha HPV ORFs are in fact incongruent with the recently proposed taxonomy.

MATERIALS AND METHODS

Viral genome sequences. The majority of the 59 genital HPV genomes compiled for this analysis were retrieved using GenBank's taxonomy browser using the keywords *Papillomavirus* and *Papillomaviridae*. Exceptions to this are HPV type 18R (HPV18R) (GenBank accession number AY262282), HPV54 (AF436129), HPV68 (DQ080079), HPV71 (AY330621), and HPV82 (AF293961), all of which were resequenced in our lab. Additionally, two newly sequenced, novel genital HPV genomes, HPV106 and candHPV102, were included in this analysis.

Sequence alignment and phylogenetic analysis. Multiple-sequence alignments were created using CLUSTAL W with gap cost 10.0 and either the Gonnet cost matrix for protein or the IUB matrix for DNA (25). Phylogenies were created using only equal weighted characters from the coding regions E6, E7, E1, E2, L2, and L1. E5 was not used because homologous ORFs were not found throughout this data set. E4, embedded within E2, is also excluded from the analysis, but it is treated in depth elsewhere (16). The total-evidence tree shown in Fig. 1 was generated with both protein and nucleotide sequences of all six coding regions (a rationale for using concatenated nucleotide and amino acid sequences is provided in reference 1), while the early- and late-gene trees shown in Fig. 2 contain E6, E7, E1, and E2 nucleotide sequences and L2 and L1 nucleotide sequences, respectively. Parsimony and neighbor-joining trees were generated with PAUP version 4.10 (24). Before these trees were constructed, alignment gaps were coded as missing. In both the parsimony and distance analyses, 100 bootstrap replicates were performed to assess robustness at each node. Bayesian trees were constructed using the Markov-chain Monte Carlo technique in MRBAYES sampling every 100th generation in a series of 100,000 cycles (8).

Full open reading frame and sliding-window analyses. Whole-genome open reading frame distance scans were created using software built around Pearson's global alignment program ALIGN, part of the FASTA package (17). The sliding-window distance analysis was performed in the Recombination Detection Program (RDP) version 1.08 using whole-genome alignments (12).

Localized incongruence. To locate incongruence and to assess the relative degrees of incongruence across different node/partition pairs, we developed a computational algorithm modeled after the localized incongruence length difference (LILD). LILD measures the tree length difference between a parsimony search wherein a node optimal in the total-evidence phylogeny constrains tree topology and a parsimony search where that constraint is lifted (26). For analysis of the genital HPVs, we treated E6, E7, E1, E2, L2, and L1 as separate partitions and calculated the LILD for each of these partitions at every node in the total-evidence phylogeny shown in Fig. 1. For the 59 internal nodes in a phylogeny of 60 PVs, this translates into 354 unique LILD experiments. To determine the nonparametric statistical significance of these LILD calculations, we evaluated the difference relative to a distribution of 100 random partitions created from the original total-evidence data matrix. Constrained and unconstrained tree lengths were therefore calculated for 202 trees at every node for every partition. For the test partition, we employed a search of 100 bootstrap replicates for both the constrained and unconstrained conditions. Random partitions were searched using 20 bootstrap replicates. Coupled to the number of unique node/partition pairs, this analysis was computationally intensive, requiring nearly 1.5 million tree reconstructions. To facilitate the analyses, the LILD pipeline algorithm was adapted to run in parallel. The PERL programs designed for this task employed both Bioperl (23) and PAUP* (24) and are available from the authors.

Nucleotide sequence accession numbers. The sequences of two newly sequenced, novel genital HPV genomes were deposited in GenBank under the following accession numbers: HPV106, DQ080082; and candHPV102, DQ080083.

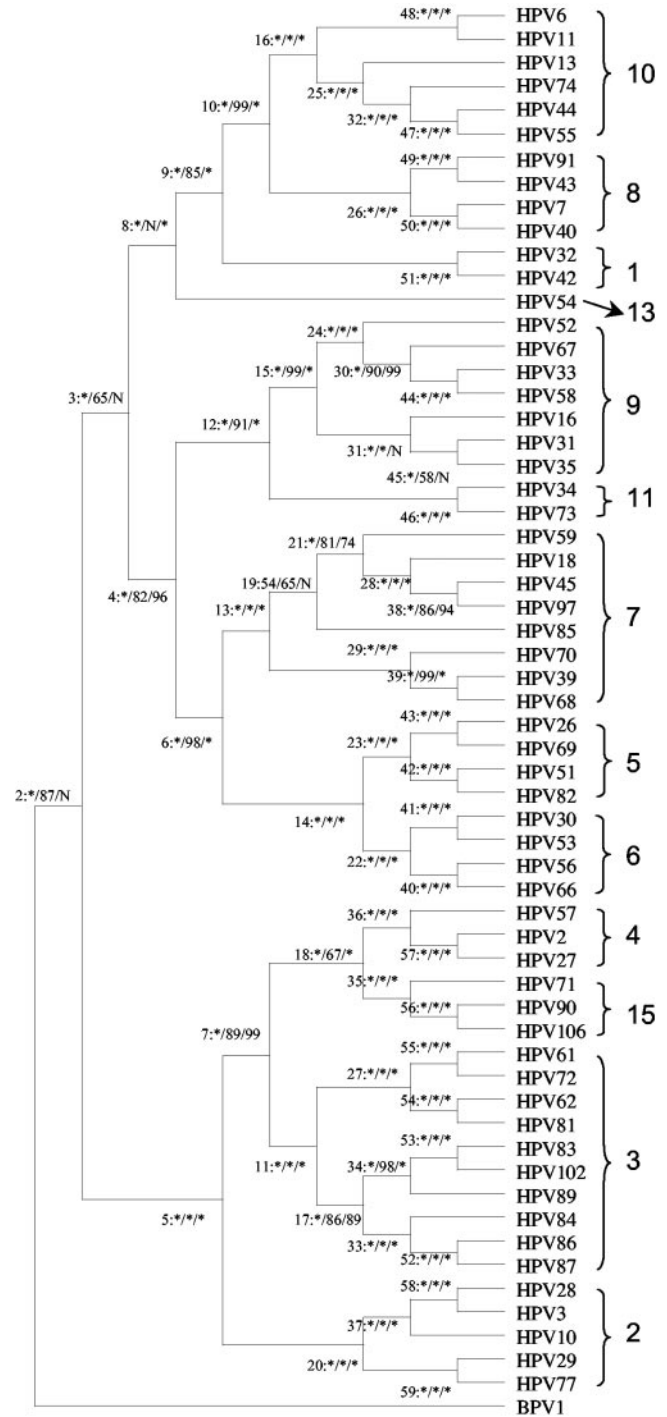


FIG. 1. Total-evidence phylogeny. A phylogenetic tree was inferred from maximum-parsimony, neighbor-joining, and Bayesian methods. The tree shown is from the Bayesian analysis inferred from alignment of protein and nucleotide sequences of six concatenated ORFs (E6, E7, E1, E2, L2, and L1). Numbers on or near branches indicate first the node number and after the colon support indices from methods in the following order: Bayesian credibility value, maximum-parsimony bootstrap percentage, and neighbor-joining bootstrap percentage. Methods that show 100% support are represented with an asterisk. Any conflict between the Bayesian tree and either of the other two methods is indicated by an "N" at the appropriate node. Only informative sites were kept for the analyses. Bovine PV type 1 (BPV1) was used as the outgroup taxon.

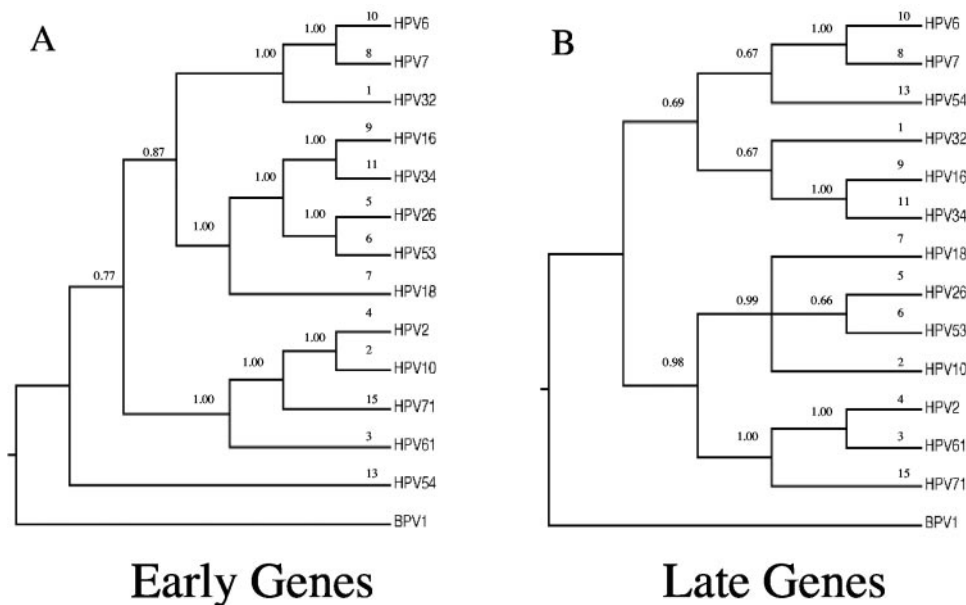


FIG. 2. Early- and late-gene phylogenies. Phylogenetic trees were inferred using Bayesian methods. The early tree (A) was calculated from E6, E7, E1, and E2 concatenated nucleotide alignments, while the late tree (B) was derived from combined L2 and L1 nucleotide sequence data. Bayesian credibility values are provided near the appropriate nodes, and alpha papillomavirus group designations are shown on their respective leaf branches. A representative virus was chosen from each of 13 alpha HPV species groups (5).

RESULTS

Phylogenetic relationship of alpha HPV genomes. The total-evidence phylogeny is shown in Fig. 1. The tree topology displayed is from a Bayesian analysis of concatenated nucleotide and protein sequences. Maximum-parsimony and neighbor-joining analyses were also performed on both protein and nucleotide sequences. The bootstrap support values for each method are shown at each node. All three analyses displayed strong support for a common ancestor of the high-risk clade of PVs (node 4) that includes species groups 9, 11, 7, 5, and 6. Lack of congruence between topologies generated by the three methods is most evident at the deeper nodes where the neighbor-joining tree failed to yield nodes 2 and 3 and where the maximum-parsimony tree was in conflict with the Bayesian analysis at node 8. Though most nodes that lack robustness are clustered around the root, nodes 19, 31, and 45 also show some conflict between the three methods.

Separate Bayesian trees for the early genes and late genes were computed, and the results are displayed in Fig. 2. Since members (i.e., types) within a given species group (60 to 70% sequence identity) almost always cluster together (data not shown) independent of constituent types chosen to construct the tree, representative viruses were chosen as “type species” as suggested by de Villiers et al. (5). The deep branching patterns of the early and late trees constructed with representative type species are markedly different. Though the tree constructed using early-gene sequences mirrors the total-evidence phylogeny, the late-gene tree is incongruent at several points, including the common ancestor of the high-risk clade (species 5, 6, 7, 9, and 11). The late-gene tree subdivides this clade, splitting groups 9 and 11 from groups 5, 6, and 7. Individual trees were also generated for E6, E7, E1, E2, L2, and L1. In all cases, using the early genes resulted in trees that

maintain an oncogenic node, while L1 and L2 trees split this node consistent with two lineages of high-risk types (data not shown).

Visualization of early- and late-gene sequence divergence. The topological incongruence seen between trees generated from the early and late genes was further characterized by performing pairwise open reading frame analyses as shown in representative scans for E6 and L2 in Fig. 3. These graphs display global similarity for each ORF of viruses from each species group compared with each HPV type used to construct the tree shown in Fig. 1. This graphic approach facilitated identification of ORF regions that showed exaggerated divergence from others. We used the total-evidence phylogeny generated by concatenated HPV ORFs to guide the profiles scanned. The difference between early- and late-gene trees suggested that each ORF should show a distinct open reading frame scan; however, only the E6 and L2 scans showed patterns consistent with the incongruence noted between trees generated from the early and late genes. The plateau effect evident in the E6 trace (>70% identity relative to a 60% baseline) is indicative of a deep, shared homology in all high-risk types, highlighting the relationship between E6 function and cancer association. The L2 trace, however, splits this plateau and instead displays a shouldering phenomenon (~75% identity relative to a 70% baseline in groups 10, 8, 1, 13, 9, and 11) that unites groups 7, 5, and 6 with groups 4, 15, 3, and 2. This is consistent with the differences in branching patterns observed in the early- and late-gene trees shown in Fig. 2 and suggests the position of incongruence may be around the presumed oncogenic node. While all early genes seem to retain the oncogenic node and both late genes seem to break it, the open reading frame scans suggested that the incongruence is most pronounced for E6 and L2.

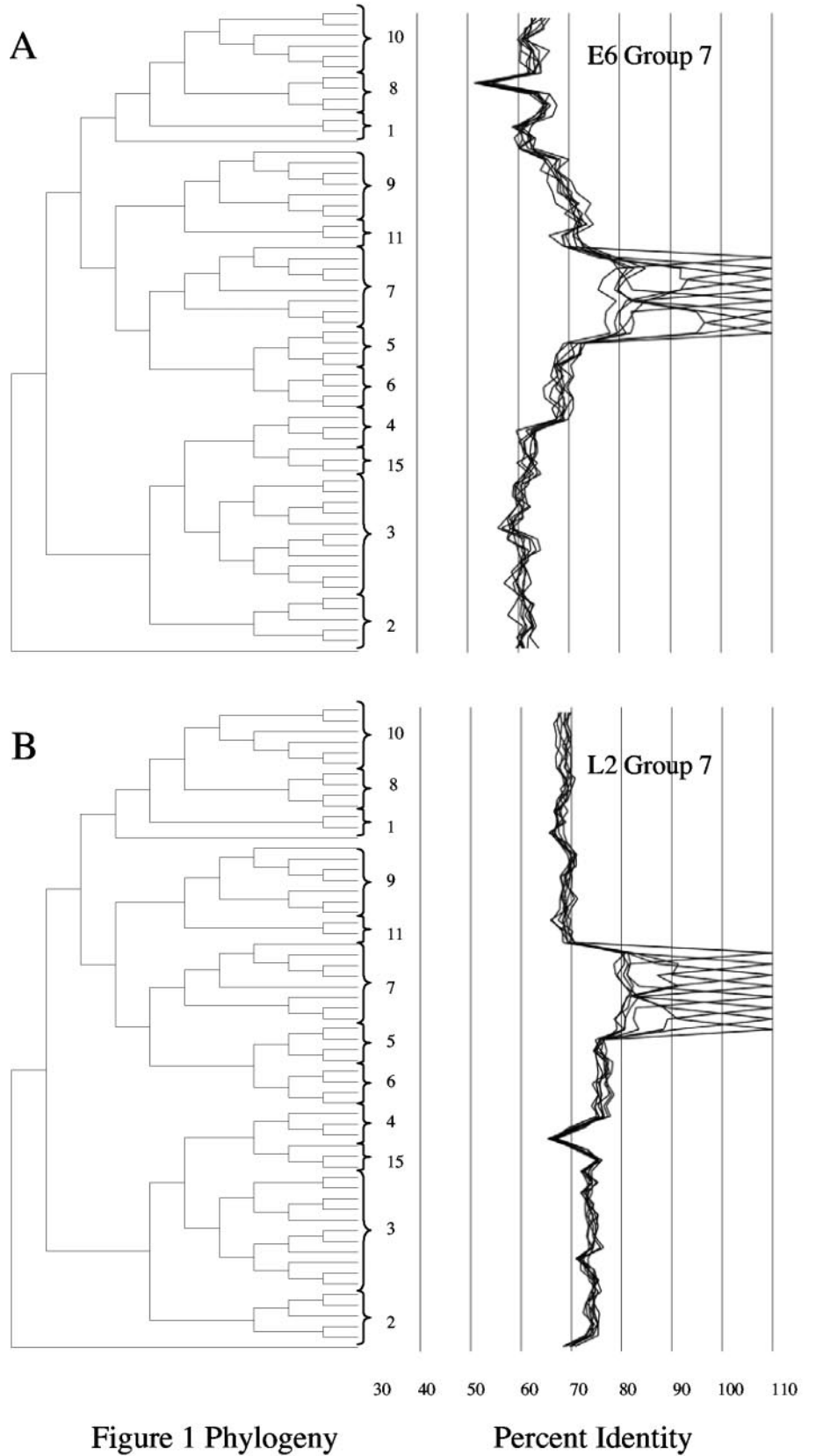


Figure 1 Phylogeny

Percent Identity

FIG. 3. Open reading frame scans. The frame of the phylogenetic tree and its species groups displayed in Fig. 1 are presented alongside plots of percent identity calculated from global pairwise alignments. The behavior of species group 7 viruses differs with respect to E6 (A) and L2 (B). A common shared homology in E6 appears to be broken in L2.

To determine whether the early-gene/late-gene incongruence is in fact being driven by E6 and L2, we performed a whole-genome sliding-window similarity analysis. HPV type 18 (HPV18) and HPV16, representative of species 7 and 9, respectively, were examined by aligning each with a whole-genome multiple-sequence alignment of representative genital HPVs used for the analyses shown in Fig. 2. Percent similarities are displayed along the topological positions of ORFs within the HPV genome. Figure 4 shows that the E1 or L1 ORFs were similar between all clades. E2 displayed divergence across all species, especially in the hinge region (16). Two distinct populations are apparent in the E6/E7 region, corresponding to the high- and low-risk species. The L2 region also shows two distinct populations predominantly localized to the N-terminal half of the ORF. However, these populations are not correlated with oncogenic risk. To confirm that the L2 N-terminal portion drives both the shouldering phenomenon observed in the open reading frame scans and the similarity discrepancy in the sliding window, we split the L2 open reading frame in half and subjected both segments to an open reading frame scan. The N-terminal sequence actually enhances the shouldering phenomenon (>70% identity relative to a 55% baseline); however, the shoulder disappears completely when using the L2 C terminus (data not shown). In fact, Fig. 4 reveals that HPV18 shows high similarity to groups 5, 6, 4, 15, 3, and 2 but significantly lower similarity to groups 10, 8, 1, 9, and 11. The same trend is not observed in the HPV16 analysis where similarity measures in the L2 N terminus do not split the viruses into two distinct populations. The sliding-window analysis reinforces the incongruence hinted at in the tree topologies and the open reading frame scans and strengthens the view that E6 and L2 significantly contribute to the differing evolutionary pathways.

Localizing gene divergence to specific viral species groups.

The trees and scans provide evidence for two distinct patterns of alpha PV evolution. To quantify and localize this incongruence and to evaluate its statistical significance, we developed a computational system to measure the length differences between constrained and unconstrained trees at each internal node (the localized incongruence length difference) (26). The algorithm allows for the determination of nodes where potential recombination or selection events occurred in the evolutionary history of the viral genomes.

Figure 5 shows a distribution of those node/partition combinations deemed significant with a *P* of 0.01 using a nonparametric test (26). The plot displays relative incongruence across ORF/node pairs. Of the 11 statistically significant node/partition pair LILDs, 9 involve either L1 or L2. And remarkably, significant LILD is concentrated at the oncogenic node, corresponding to node 4 in Fig. 1 and 5, the common ancestor of species 9, 11, 7, 5, and 6. Both L2 and L1 show incongruence at this node: the tree length difference for L2 is 60 steps, while that for L1 is 33 steps. The L2 LILD far outstrips any other node/ORF pairs in terms of relative significance, as indicated by arrow 1 in Fig. 5. Node 3 (the common ancestor to the oncogenic types and groups 10, 8, 1, and 13) also shows significant relative incongruence with respect to L2. However, in this case, poor bootstrap support coupled to a diffuse distribution of random partitions suggests a generally unstable node (Fig. 5). Only two of the significant node/ORF pairs involve the

early genes, E7 (5/E7) and E6 (6/E6), which is not totally unexpected, given that the total-evidence phylogeny was relatively consistent with the early ORF tree (compare Fig. 1 and Fig. 2A).

DISCUSSION

Early indications of conflicting evolutionary patterns in alpha HPV gene regions. The clinical relevance of HPV taxonomy is supported by both phylogenetic and epidemiological studies. Van Ranst et al. (27) originally proposed that HPV types associated with high-grade cervical intraepithelial neoplasia and/or invasive cervical cancer were contained within a monophyletic clade of human papillomaviruses, including HPV16 and HPV18. This observation was based on phylogenies built from E6 alignments. Subsequent inclusion of E7, E1, E2, L2, and L1 served to confirm subdivision into mucosal high- and low-risk clades even in broader, genome-wide treatments (21). Additional support for these predictions recently highlighted the remarkable agreement between phylogenetic and epidemiological data sets (14).

With the growth of HPV sequence data, a number of groups have attempted phylogenetic classification of novel new HPV types (4, 15). These studies all employed L1, primarily because the field's most commonly used primers (e.g., MY09/MY11 and GP5+/GP6+) amplify a fragment in this ORF that was readily available for sequencing, but also because the L1 ORF is highly conserved in PV genomes. The L1 phylogeny has since become the PV standard. de Villiers et al. used the L1 ORF in the most recent classification of PVs, delimiting, for the first time, sequence similarity cutoffs for genus and species designations (5). However, previous phylogenies built from either the short ~300-base-pair L1 segments or complete L1 sequences were incongruent with E6 phylogenies (5, 15). Whereas E6 phylogenies cluster high-risk genital HPVs, suggesting a common ancestor (Fig. 1), the L1 trees split this oncogenic clade, grouping species 9 and 11 with 1, 8, and 10, and species 7, 5, and 6 with 4, 15, 3, and 2 (Fig. 2). Bravo and Alonso also highlight the differing topologies of early- and late-gene trees (2).

In light of the connection between clinical manifestations and HPV evolution, the discrepancies in tree topology is an important one and may have significant implications for future studies on the biology of HPV genomes. Though trees from both E6 and L1 have been generated in numerous classification studies thus far, this fundamental incongruence has not been rigorously examined and subjected to statistical analysis. Previous studies aimed to categorize the complete body of PV sequences, concentrating more on the relationships between genus supergroups rather than the shifting branching pattern between species.

Whole-genome HPV scanning allows visualization of divergent evolutionary history. The first indication of a conflict in the evolutionary histories of the genital HPV ORFs was in the E6 and L2 open reading frame scans (Fig. 3). The E6 plateau, uniting the high-risk types, overlapped with the L2 shoulder, hinting that specific regions were driving the deep incongruence characteristic of the discordant early- and late-gene phylogenies. Extending the simple similarity measures used in the open reading frame scan to a similarity analysis employing a

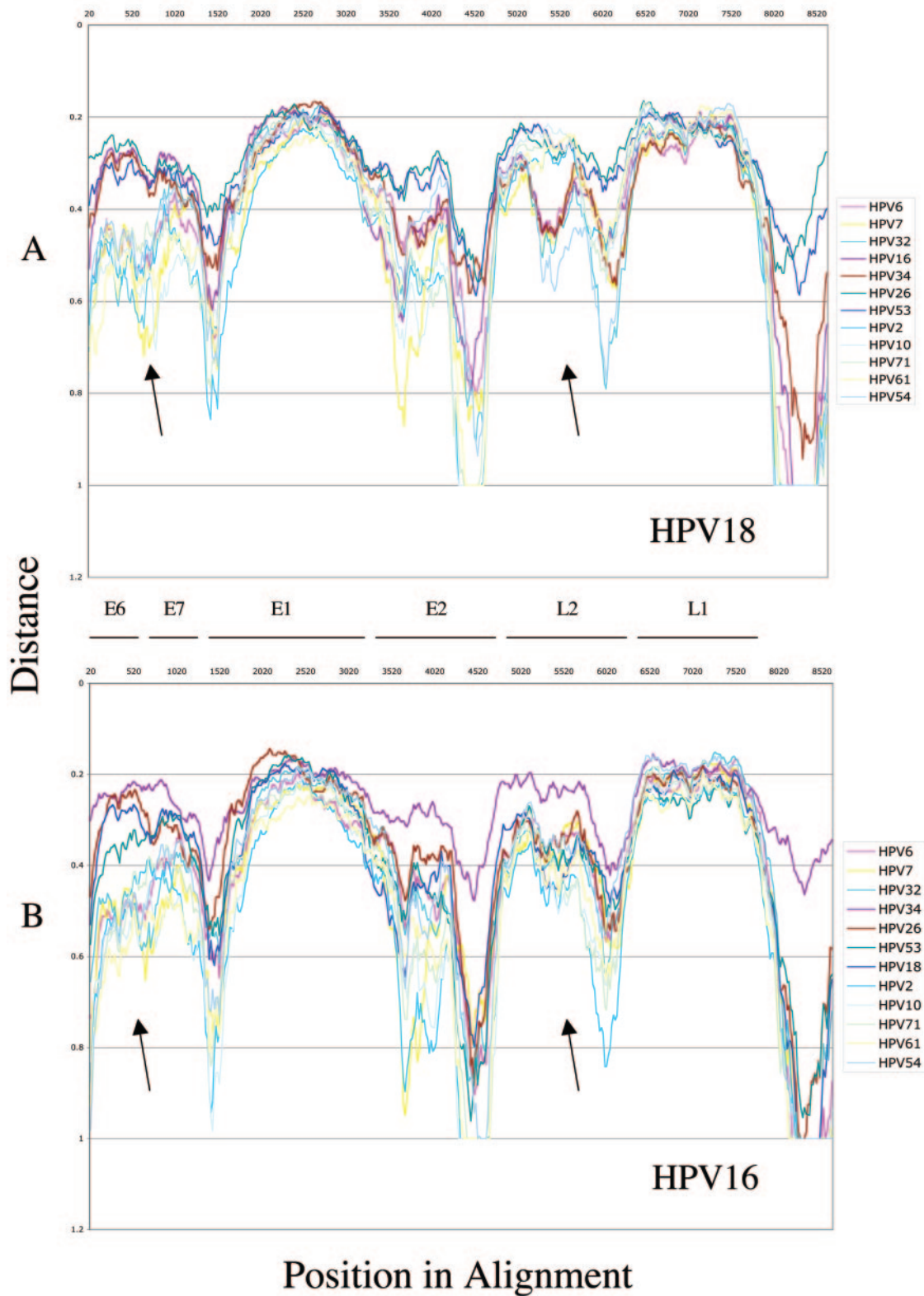


FIG. 4. Sliding-window analysis. The pairwise distances of HPV18 (A) and HPV16 (B) relative to HPVs selected for Fig. 2 are plotted for comparison. The analysis employed a 500-nucleotide window sliding 20 positions at a time. Arrows emphasize positions of interest, indicating two E6/E7 populations of viruses corresponding to high- and low-risk viruses and the L2 N terminus that displays two populations in the HPV18 scan, but not in the HPV16 scan.

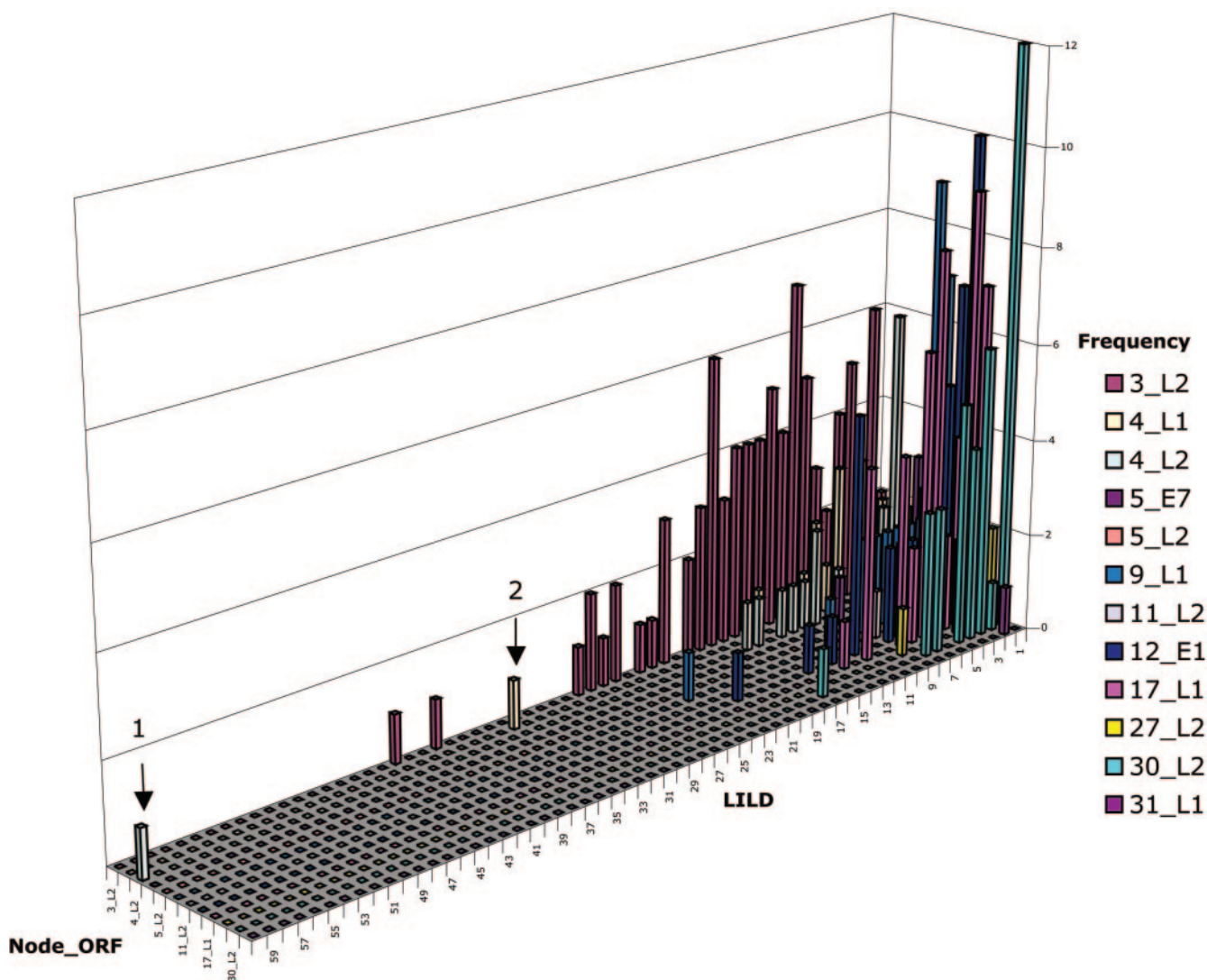


FIG. 5. Localized incongruence length difference analysis. The LILD distributions for all node/partition pairs considered significant at a P of 0.01 are shown. Node numbers correspond to those given in Fig. 1. The full distribution (both the test partition and random partitions greater than zero) for each significant pair is given in order to visualize the relative extent of incongruence from each calculation. The node/partition pairs of particular interest are highlighted with arrows: 4/L2 (arrow 1) and 4/L1 (arrow 2).

sliding window reinforced and extended these observations, implicating the amino half of L2 (Fig. 4). The alpha papillomavirus species groups 7, 5, and 6 are at the intersection of the E6 plateau and the L2 shoulder, appearing to share two distinct phylogenetic histories.

Determining incongruent branches and ORFs used to reconstruct the HPV phylogeny. To pinpoint the branches in the phylogenetic tree that demonstrate incongruence with statistical confidence, we developed a high-volume, high-throughput system for localizing tree length differences between constrained and unconstrained nodes in the total-evidence tree. In effect, the process is an extension of the localized incongruence length difference test (26), designed to dissect the competing influences of component genes across every node of a total-evidence phylogeny. For the alpha HPVs, the results of this analysis are striking (Fig. 5). Incongruence is firmly focused on the oncogenic node, the putative common ancestor for all

high-risk (i.e., cancer-associated) genital mucosal types. L2 was particularly significant, implying that the evolutionary histories of the high-risk viruses follow two distinct paths, one conferred by the early genes and another by the late genes. This conclusion is also supported by the oncogenic node incongruence observed in L1. Though its incongruence was not evident in the pairwise similarity measures employed in the open reading frame scans and sliding-window analyses, L1 clearly contributed to the late-gene incongruence.

Localization of incongruence has been used in studies of domain shuffling of steroid receptors. In that study, detection of incongruence length differences localized to a particular node allowed the inference of domain shuffling (26). It has also been used in explaining a genomic island contributing to the diversity of *Actinobacillus actinomycetemcomitans* (18). In the data presented here for whole HPV genomes, a localized length difference could indicate an early recombination event

or strong divergence due to very different selective constraints on particular genes or gene regions. What is clear from these analyses is that no single ORF provides a definitive evolutionary history of the alpha PVs. When genes are pooled, it is possible to resolve the incongruence and favor a single point of origin for high-risk types. However, this convenient resolution blurs the ORFs' competing effects and masks the important evolutionary differences in the paths taken by the early and late genes.

Perhaps the most important observation from these experiments is that the high-risk types are not necessarily monophyletic. From an evolutionary point of view, there appear to be two distinct populations of high-risk PVs: the HPV16-related groups 9 and 11 and the HPV18-related groups 7, 5, and 6. The two groups share homology in their early genes, particularly the oncogenic genes E6 and E7, but diverge with respect to L1 and L2. A number of histological studies comparing infections with HPV16 and HPV18 show that the two viruses differ in terms of their biological niche distribution (e.g., association with squamous or glandular cancer) (3). A progenitor of the oncogenic types related to HPV18 (alpha PV groups 7, 5, and 6) may have adapted characteristics of the L2 and L1 genes of alpha PV groups 4, 15, 3, and 2. Alternatively, the E6 and E7 genes of that same progenitor may have evolved features that mimic the oncogenicity of alpha PV groups 9 and 11, suggesting two distinct evolutionary instances of the high-risk phenotype. Although an ancient recombination event is possible, lack of evidence for recombination in present-day PVs suggests that natural selection appears to have converged on either a single solution for the late genes or a single solution for oncogenicity.

ACKNOWLEDGMENTS

We thank Andras Fiser for the use of his computational cluster. This work is supported in part by a grant from the National Cancer Institute, National Institutes of Health to R.D.B. (CA78527).

REFERENCES

1. Agosti, D., D. Jacobs, and R. DeSalle. 1996. On combining protein sequences and nucleic acid sequences in phylogenetic analysis: the homeobox protein case. *Cladistics* **12**:65–82.
2. Bravo, L., and A. Alonso. 2004. Mucosal human papillomaviruses encode four different E5 proteins whose chemistry and phylogeny correlate with malignant or benign growth. *J. Virol.* **78**:13613–13626.
3. 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.
4. Chan, S.-Y., H. Delius, A. L. Halpern, and H.-U. Bernard. 1995. Analysis of genomic sequences of 95 papillomavirus types: uniting typing, phylogeny, and taxonomy. *J. Virol.* **69**:3074–3083.
5. de Villiers, E. M., C. Fauquet, T. R. Broker, H. U. Bernard, and H. Zur Hausen. 2004. Classification of papillomaviruses. *Virology* **324**:17–27.
6. Fehrmann, F., and L. A. Laimins. 2003. Human papillomaviruses: targeting differentiating epithelial cells for malignant transformation. *Oncogene* **22**:5201–5207.
7. Fields, B., P. Howley, D. Griffin, R. Lamb, M. Martin, B. Roizman, S. Straus, and D. Knipe. 2001. *Virology*, 4th ed. Lippincott Williams & Wilkins Publishers, Baltimore, Md.
8. Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**:754–755.
9. Longworth, M. S., and L. A. Laimins. 2004. Pathogenesis of human papillomaviruses in differentiating epithelia. *Microbiol. Mol. Biol. Rev.* **68**:362–372.
10. Maddison, W. P. 1997. Gene trees in species trees. *Syst. Biol.* **46**:523–536.
11. Martin, A. P., and T. M. Burg. 2002. Perils of paralogy: using HSP70 genes for inferring organismal phylogenies. *Syst. Biol.* **51**:570–587.
12. Martin, D., and E. Rybicki. 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16**:562–563.
13. Munger, K., A. Baldwin, K. M. Edwards, H. Hayakawa, C. L. Nguyen, M. Owens, M. Grace, and K. Huh. 2004. Mechanisms of human papillomavirus-induced oncogenesis. *J. Virol.* **78**:11451–11460.
14. 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.
15. Myers, G., F. Sverdrup, C. Baker, A. McBride, K. Munger, H. Bernard, and J. D. Meissner (ed.). 1997. Human papillomaviruses 1997: a compilation and analysis of nucleic acid and amino acid sequences. Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, N.Mex.
16. Narechania, A., M. Terai, and R. D. Burk. 2005. Overlapping reading frames in closely related human papillomaviruses result in modular rates of selection within E2. *J. Gen. Virol.* **86**:1307–1313.
17. Pearson, W. R. 2000. Flexible sequence similarity searching with the FASTA3 program package. *Methods Mol. Biol.* **132**:185–219.
18. Planet, P. J., S. C. Kachlany, D. H. Fine, R. DeSalle, and D. H. Figurski. 2003. The widespread colonization island of *Actinobacillus actinomycetem-comitans*. *Nat. Genet.* **34**:193–198.
19. Rokas, A., B. L. Williams, N. King, and S. B. Carroll. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**:798–804.
20. Satta, Y., J. Klein, and N. Takahata. 2000. DNA archives and our nearest relative: the trichotomy problem revisited. *Mol. Phylogenet. Evol.* **14**:259–275.
21. Schiffman, M., R. Herrero, R. DeSalle, A. Hildesheim, S. Wacholder, A. C. Rodriguez, M. C. Bratti, M. E. Sherman, J. Morales, D. Guillen, M. Alfaro, M. Hutchinson, T. C. Wright, D. Solomon, Z. Chen, J. Schussler, P. E. Castle, and R. D. Burk. 2005. The carcinogenicity of human papillomavirus types reflects viral evolution. *Virology* **337**:76–84.
22. Shadan, F. F., and L. P. Villarreal. 1995. The evolution of small DNA viruses of eukaryotes: past and present considerations. *Virus Genes* **11**:239–257.
23. Stajich, J. E., D. Block, K. Boulez, S. E. Brenner, S. A. Chervitz, C. Dagdigan, G. Fuellen, J. G. Gilbert, I. Korf, H. Lapp, H. Lehvaslaiho, C. Matsalla, C. J. Mungall, B. I. Osborne, M. R. Pockock, P. Schattner, M. Senger, L. D. Stein, E. Stupka, M. D. Wilkinson, and E. Birney. 2002. The Bioperl toolkit: Perl modules for the life sciences. *Genome Res.* **12**:1611–1618.
24. Swofford, D. L. 1998. PAUP*. Phylogenetic analysis using parsimony (* and other methods), version 4. Sinauer Associates, Inc., Sunderland, Mass.
25. Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
26. Thornton, J. W., and R. DeSalle. 2000. A new method to localize and test the significance of incongruence: detecting domain shuffling in the nuclear receptor superfamily. *Syst. Biol.* **49**:183–201.
27. Van Ranst, M., J. B. Kaplan, and R. D. Burk. 1992. Phylogenetic classification of human papillomaviruses: correlation with clinical manifestations. *J. Gen. Virol.* **73**:2653–2660.