

A Population-Based Prospective Study of Carcinogenic Human Papillomavirus Variant Lineages, Viral Persistence, and Cervical Neoplasia

Mark Schiffman¹, Ana Cecilia Rodriguez³, Zigui Chen⁴, Sholom Wacholder¹, Rolando Herrero³, Allan Hildesheim¹, Rob Desalle⁵, Brian Befano⁶, Kai Yu¹, Mahboobeh Safaeian¹, Mark E. Sherman¹, Jorge Morales³, Diego Guillen³, Mario Alfaro³, Martha Hutchinson⁷, Diane Solomon², Philip E. Castle¹, and Robert D. Burk⁴

Abstract

Human papillomavirus (HPV) types differ profoundly in cervical carcinogenicity. For the most carcinogenic type HPV16, variant lineages representing further evolutionary divergence also differ in cancer risk. Variants of the remaining 10 to 15 carcinogenic HPV types have not been well studied. In the first prospective, population-based study of HPV variants, we explored whether, on average, the oldest evolutionary branches within each carcinogenic type predicted different risks of >2-year viral persistence and/or precancer and cancer [cervical intraepithelial neoplasia grade 3+ (CIN3+)]. We examined the natural history of HPV variants in the 7-year, 10,049-woman Guanacaste Cohort Study, using a nested case-control design. Infections were assigned to a variant lineage determined by phylogenetic parsimony methods based on URR/E6 sequences. We used the Fisher's combination test to evaluate significance of the risk associations, cumulating evidence across types. Globally, for HPV types including HPV16, the *P* value was 0.01 for persistence and 0.07 for CIN3+. Excluding HPV16, the *P* values were 0.04 and 0.37, respectively. For HPV16, non-European viral variants were significantly more likely than European variants to cause persistence [odds ratio (OR), 2.6; *P* = 0.01] and CIN3+ (OR, 2.4; *P* = 0.004). HPV35 and HPV51 variant lineages also predicted CIN3+. HPV variants generally differ in risk of persistence. For some HPV types, especially HPV16, variant lineages differ in risk of CIN3+. The findings indicate that continued evolution of HPV types has led to even finer genetic discrimination linked to HPV natural history and cervical cancer risk. Larger viral genomic studies are warranted, especially to identify the genetic basis for HPV16's unique carcinogenicity. *Cancer Res*; 70(8); 3159–69. ©2010 AACR.

Introduction

The evolutionary biology of papillomaviruses, as understood through analyses of phylogenetic trees, is tightly linked to host and tissue tropisms, biological behavior, and neoplastic potential. Papillomaviruses have evolved over millions of years from common ancestors into specific host ecosystems. Among human papillomaviruses (HPV), the α genus is associated with infections of the anogenital and oral mucosa (Fig. 1).⁸ The topology of the

α HPV phylogenetic tree based on complete genome information or restricted to the early region of the papillomavirus genome strongly predicts viral behavior, including highly specific tissue tropisms (e.g., vagina versus cervix), induced cytomorphic abnormalities, and persistence related to carcinogenicity (1–3).

A subset of mucosal HPV genotypes ("types") in the α genus (Fig. 1) causes virtually all cases of cervical cancer (and some other anogenital and extragenital cancers not discussed here). More specifically, there are three main branches (clades) of the α genus; all established carcinogenic HPV types are in a single high-risk clade composed of five species groups ("species"): α -5, α -6, α -7, α -9, and α -11.

The five species in the high-risk clade have different risk profiles. α -9 is the most important species, consisting almost entirely of carcinogenic types.

Even infections with HPV types in the high-risk clade typically "clear," defined as becoming undetectable even using sensitive HPV DNA test methods (4, 5). Many infections are undetectable by 3 to 6 months later, and >90% eventually clear within a few years. Persistent HPV infections with

Authors' Affiliations: Divisions of ¹Cancer Epidemiology and Genetics and ²Cancer Prevention, National Cancer Institute, Bethesda, Maryland; ³Proyecto Epidemiológico Guanacaste, San Jose, Costa Rica; ⁴Albert Einstein College of Medicine, Yeshiva University, Bronx, New York; ⁵American Museum of Natural History, Manhattan, New York; ⁶Information Management Services, Silver Spring, Maryland; and ⁷Women and Infants Hospital, Providence, Rhode Island

Corresponding Author: Mark Schiffman, Clinical Genetics Branch, National Cancer Institute, 6120 Executive Boulevard, Room 7012, Rockville, MD 20852. Phone: 301-451-4562; Fax: 301-496-1854; E-mail: schiffmm@mail.nih.gov.

doi: 10.1158/0008-5472.CAN-09-4179

©2010 American Association for Cancer Research.

⁸ Z. Chen and R.D. Burk, personal communication; manuscript in preparation.

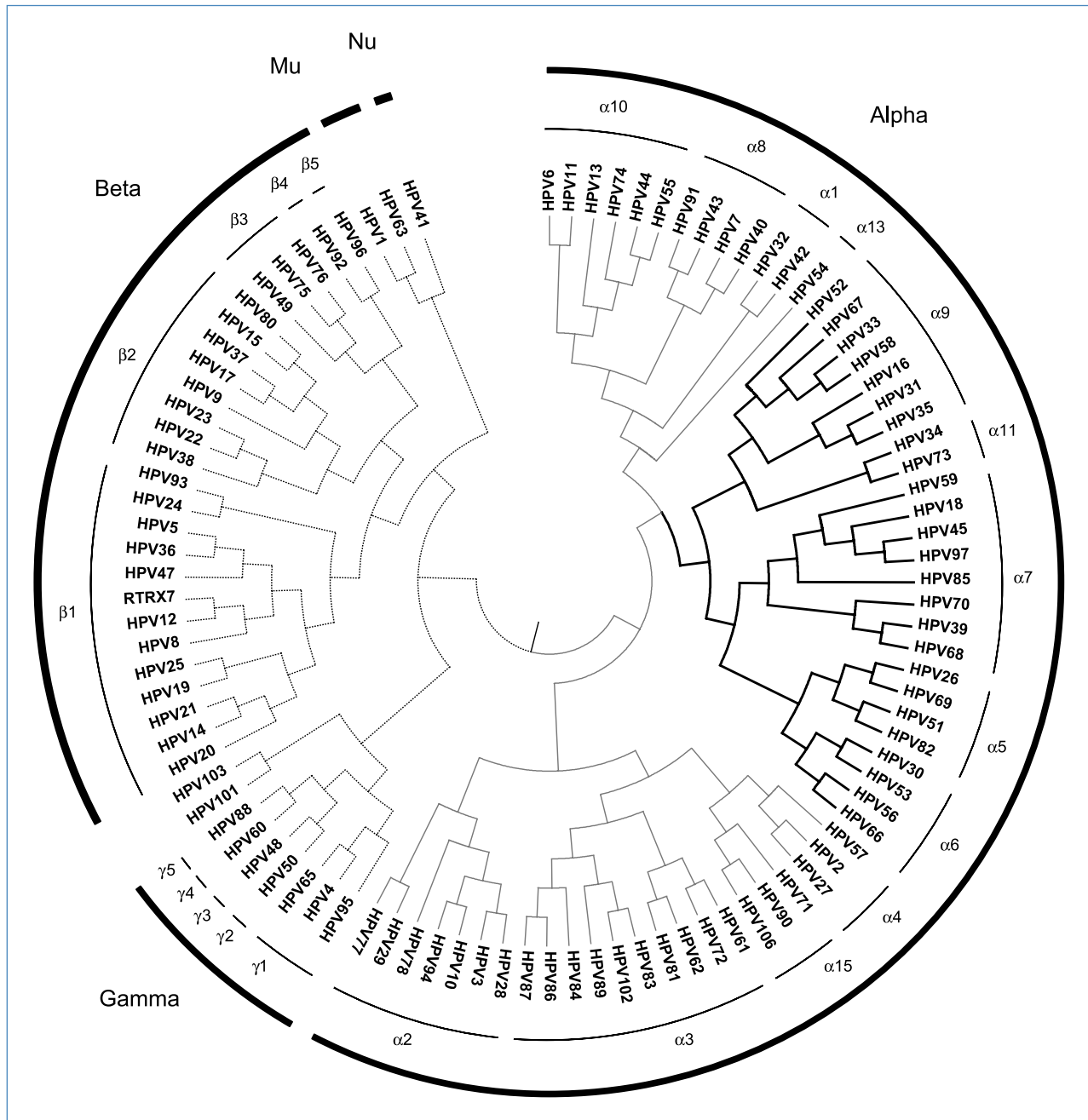


Figure 1. Bayesian tree of 99 human papillomavirus types inferred from the concatenated amino acids and nucleotide sequences of six ORFs (E6, E7, E1, E2, L2, and L1). MrBayes v3.1.2 (44, 45) with 10,000,000 cycles for the Markov chain Monte Carlo algorithm was used to generate phylogenetic trees. For Bayesian tree construction, the computer program ModelTest v3.7 (46) was used to identify the best evolutionary model; the identified γ model was set for among-site rate variation and allowed all substitution rates of aligned sequence to be different. The ORF sequence was left blank if unavailable. HPV species groups were generally classified according to the classification system for PVs by de Villiers and colleagues (6). All high-risk mucosal HPV types cluster together within the α papillomavirus genus, highlighted with bold lines. The broken lines represent skin HPV types.

carcinogenic types are linked to a high absolute risk of subsequent precancer [cervical intraepithelial neoplasia grade 3 (CIN3)] and an elevated risk of eventual cancer.

At an even finer level, within each species of the high-risk clade, HPV type-defined viral behaviors of persistence and

carcinogenicity have characteristics of a quantifiable genetic trait. An HPV type is >10% different than all other characterized papillomavirus types in the L1 open reading frame (ORF; ref. 6). Individual types in the highest risk α -9 species show very large differences in risk of precancer (associated with

viral persistence) and eventual cancer. At the extremes, HPV16 is responsible for half of the cases of cervical cancer globally, whereas HPV67 has very rarely been found alone in cancer (7, 8).

After the emergence of distinct HPV types within species, population expansion, geographic dispersion, and time have contributed to the further evolution of HPV types into variant lineages ("variants"). Variants of each type have evolved primarily through single-nucleotide variations with a pattern of lineage fixation of these variations (9, 10). HPV variants can be classified based on their nucleotide sequence differences with most variants differing by <3%; however, there are a few examples of more distantly related variants that have also been termed subtypes (e.g., HPV68 variants; refs. 11, 12).

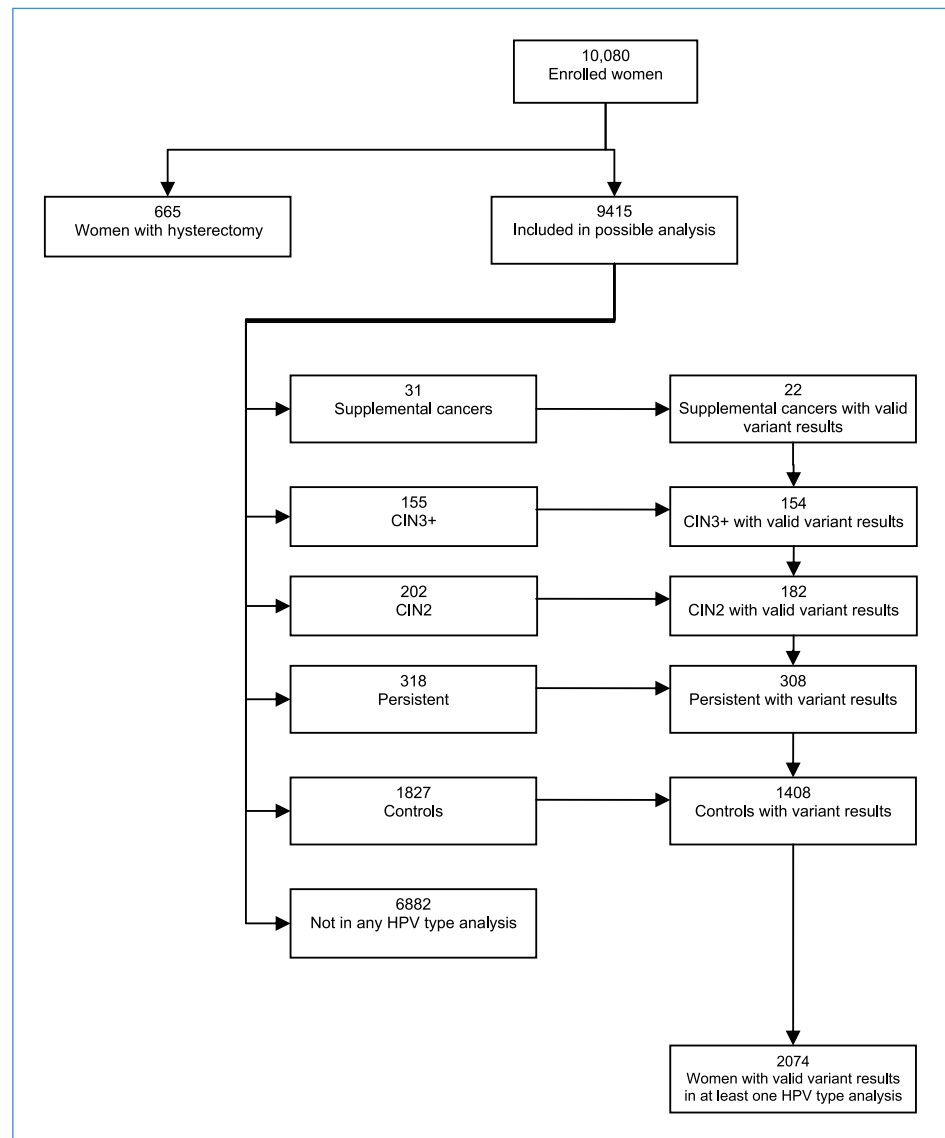
Despite evolutionary relatedness, HPV variants can differ in carcinogenicity. As stated above, the variants have co-evolved with humans to correspond, at least for HPV16, with

human geographic dispersion (13). Asian-American or African HPV16 variants have been associated in case-control studies with 5-fold or greater risks of cancer compared with European variants. Of note, European variants of HPV16 are still very high risk, albeit at lower risk than other HPV16 variants.

There is some suggestion that variants of other HPV types also confer different risks of viral persistence and/or progression to precancer. However, except for some work on HPV18 in α -7 and on HPV variants in HIV-infected individuals, there is little known about the natural history of variants of types other than HPV16 among immunocompetent women.

This research gap leaves important questions unanswered. For example, HPV16 variants are all carcinogenic although they differ in risk. But variants in a moderately carcinogenic type like HPV31 could represent a homogeneous set of moderately carcinogenic variants or a heterogeneous set of

Figure 2. Study population. The cases and controls for the analysis were drawn from the population-based prospective Proyecto Epidemiológico Guanacaste, which included a random sample of approximately one of six of the adult women in the province. During the enrollment phase, we attempted to enroll all women with invasive cancer diagnosed in the province during the same time period. As such, this project represents a true population-wide study of the different outcomes of type-specific HPV infections and variants. The analysis is based on individual viral types, and therefore, women could contribute to more than one type-specific analysis.



variants ranging from HPV16-like behavior to much weaker behavior. HPV types with only borderline carcinogenicity could have a single or few minor variants that are definitely carcinogenic.

In a previous case-control study of HPV16 variants in Guanacaste Province, Costa Rica, we observed differences in carcinogenicity (14). To address systematically the role of HPV variants in viral natural history and cervical carcinogenicity, we have now performed a comprehensive population-based prospective study of variants of all the HPV types in the high-risk α clade. We assessed whether variants present in Guanacaste were associated with persistence and/or risk of CIN3 and cancer (CIN3+). As a logical starting point, we studied whether the oldest (i.e., most divergent) evolutionary branching of each carcinogenic type affected viral natural history. The ultimate goal is to identify sequence-specific differences that provide clues about the biological basis of HPV carcinogenicity and eventually therapeutic antiviral targets.

Materials and Methods

Study population

The Proyecto Epidemiológico Guanacaste study population is outlined in Fig. 2. The project consented and enrolled a random sample of 10,049 (93.6% of eligible) women residing in a high-risk Costa Rican province in 1993 to 1994. The study was approved by Costa Rican and U.S. ethical boards (15, 16). To supplement the random sample, we attempted to recruit all women with invasive cervical cancer diagnosed in Guanacaste during the enrollment phase of the study and successfully enrolled 31.

A few, well-trained nurses screened adult women (age range, 18–97) using visual assessment, an early pooled HPV test (Hybrid Capture Tube Test, Digene), three kinds of cytology, and cervicography (magnified cervical images). We did not screen virgins ($n = 583$) until they became sexually active. Women without hysterectomies were categorized into several groups based on the results of the enrollment screening examination (which did not include HPV genotyping) and followed differently according to predicted risk of developing CIN3+. Women with any abnormal screening result were re-screened at 6 to 12 mo of intervals for 7 y. Those without any abnormal screening test were re-screened at the 5-y anniversary of their enrollment.

Throughout the study, women with possible CIN2 (equivocal precancer) or worse (CIN3, cancer), suggested by any screening technique including nurse concern on gross examination, were referred to colposcopy with either guided biopsy of visible lesions or immediate loop electrical excision procedure depending on referral reason, colposcopic impression, age, and parity. These women were followed individually for safety, with in-patient surgery if needed but no longer studied in the cohort.

Cases and controls

A case-control study was nested within this prospective cohort analysis. For each woman and viral type, we defined disease outcomes in a hierarchical manner as the most se-

vere diagnosis during the study (baseline and follow-up combined). The original case categories were invasive cancer, else histologic CIN3, else CIN2, else high-grade squamous intraepithelial lesion (HSIL) cytology by liquid-based cytology that did not lead to histologic confirmation of CIN2 or worse, else definite long-term viral persistence of 2 y or longer (median, 46 mo), else short-term viral persistence for at least two visits not observed to last 2 y (median, 8 mo). For each HPV type, the original control group was chosen as a random sample of women with a single detection of that HPV type, frequency matched ~2:1 to the original cases.

We condensed the original case-control groups to achieve more statistical power. Because infections with short-term persistence did not vary from single detections in terms of variant distribution (data not shown), we combined the two to form a single control group (Fig. 2). We compared this single control group to three case groups: CIN3+, CIN2, and long-term persistence (a relatively rare outcome used as an end point in some HPV vaccine trials). We combined the few cases of invasive cancer with CIN3 (CIN3+) but examined invasive cancer separately for HPV16, because we had adequate power.

The women in this nested case-control analysis had a mean of 5.0 visits (SD, 3.0 visits), median of 4.0 visits, and a range of 1 to 14 visits, including enrollment and follow-up specimen collection visits. The mean follow-up was 6.5 ± 2.3 y (median, 7.0 y).

Histopathology

Histology used for clinical purposes in Costa Rica was reviewed by U.S. pathologists to define study disease outcomes. The few cases of glandular lesions were combined with comparably severe squamous *in situ* or invasive carcinoma. The final assignment of cases as invasive cancer, CIN3, CIN2, or <CIN2 was made by an algorithm of the Costa Rican and U.S. reviews (17). A few difficult cases were adjudicated by joint review, occasionally with consideration of cytologic slides as well as histology.

CIN1 was considered a morphologic consequence of HPV infection, not serious neoplasia, and was not treated during follow-up. Although CIN2 is a widely accepted treatment threshold, it is a suboptimal surrogate end point for cancer risk because of its poor reproducibility and diagnostic heterogeneity (17, 18). Thus, we kept histologic CIN2 and even more equivocal outcomes, like histologically unconfirmed cytologic HSILs, separate from the primary CIN3/cancer analytic case group as diagnosed by biopsy, LEEP, or surgical specimen.

HPV detection and genotyping

Typing methods. Cervical specimens were collected by nurses using broom devices and were placed into Digene specimen transport medium (STM). HPV DNA was detected using a MY09/M11 L1 primer PCR system (MY09/11 PCR) with AmpliTaq Gold polymerase as previously described (19–21). Briefly, an aliquot of the STM specimen was lysed, and the specimen DNA was precipitated by ammonium acetate/ethanol solution, pelleted by centrifugation, and resuspended in TE buffer. After PCR, the amplicons were

analyzed by gel electrophoresis and Southern blot hybridization as previously described. Filters were individually hybridized to biotinylated, type-specific oligonucleotide probes. Dot-blot hybridization was used for typing (probes: 2, 6, 11, 13, 16, 18, 26, 31–35, 39, 40, 42–45, 51–59, 61, 62, 64, 66–74, 81–85, and 89). A specimen was considered HPV positive but uncharacterized if it tested positive for HPV DNA by the radiolabeled generic probe mix but was not positive for any type-specific probe.

Testing with additional PCR primer sets. To verify the sensitivity of the MY09/11-based PCR method and to rule out false negative results, we tested ~2,000 initially negative specimens using other primer sets (GP5+ and FAP; refs. 21, 22). The specimens were derived from initially HPV-negative women with cytologic or histologic abnormalities (low grade or worse) at enrollment or during follow-up ($n = 850$), women with five or more sexual partners in their lifetime ($n = 187$), and a random selection of >1,000 women with no reason to suspect false negativity ($n = 955$). Few specimens retested as positive for any HPV type by any of the other primers ($n = 69$); we included specimens positive for carcinogenic types in variant testing.

HPV variant selection and testing methods

Choice of HPV types for variant testing. For this analysis, we were interested in the possible relationship of HPV variants to risk of CIN3+. Thus, we considered only HPV types in the high-risk clade (Fig. 1), which includes all types considered carcinogenic or at least possibly carcinogenic according to the latest IARC classification (7). These include all types in genus α -9 (HPV31, HPV33, HPV35, HPV52, HPV58, and HPV67), α -11 (HPV73), α -7 (HPV18, HPV39, HPV45, HPV70, and HPV68), α -5 (HPV51 and HPV69), and α -6 (HPV53, HPV56, and HPV66). We excluded those types in the high-risk clade that were extremely rare in our population.

Choice of which specimen to sequence from each woman. In this longitudinal analysis, most women had several specimens available for analysis. Initial comparison of variants from multiple samples from a single participant confirmed that we rarely detected more than one variant or a change in variant per subject over time. Specifically, to assess the intraindividual variability of variants in women infected with the same type more than once, we tested two specimens from different visits for a convenience sample of 268 women (mean time difference, 3.0 ± 2.2 y; range, 0.7–7.7 y). Only six women had discordant variant results (mean time, 4.6 ± 0.9 y).

We had begun by choosing the last positive specimen for cases before diagnosis of CIN2+ and (arbitrarily) the first positive for controls. Infections after CIN2+ diagnosis were not considered because of possible treatment effects. Some specimens were depleted, and when we confirmed that HPV variants rarely varied for a woman, we switched to selecting the available sample with the highest signal strength on PCR testing for that HPV type.

Determination of variant lineages. We have been studying variant lineages from specimens tested in our laboratory; they are derived from Guanacaste, Costa Rica, and many other countries. To classify variant lineages of carcinogenic HPV types we designed type-specific primer sets to amplify

a partial fragment of the URR region and/or E6 gene using a one-tube nested PCR method (23). We test E6 regions for specimens that do not yield data for the URR region. PCR product size is confirmed by gel electrophoresis, purified using QuickStep 2 PCR Purification kit (Edge BioSystems) or QIAquick Gel Extraction kit (Qiagen), and submitted for sequencing of both strands.

We consider a specimen to contain a possibly unique variant if the amplified URR and/or E6 fragment sequences, in more than one specimen, differ by more than two single-nucleotide polymorphisms (SNP) compared with the prototype variant lineage or other available complete genome sequence. To pursue the possibility of a new variant, we amplify and sequence the complete genomes (9, 10). Type-specific variant lineages are defined based on a complete genome nucleotide sequence having >1% dissimilarity with the prototype and are named sequentially as A (prototype), B, C, etc., according to the phylogenetic tree topologies for each individual type.⁸ If the dissimilarity is <1%, we sometimes designate A1 versus A2, etc. The nomenclatures of HPV16, HPV18, and HPV45 variant lineages are based on previous publications (10, 23, 24).

Lineage groupings. For this project testing Guanacaste specimens, masked to case-control status, we assigned each HPV type detected to a specific variant lineage. The phylogenetic trees for the vast majority of the α papillomavirus variants have a root bifurcation (two main evolutionary branches), similar to that seen for HPV16, HPV18, and HPV45 (9, 10). The average percentage of the complete genomic variation between the dichotomized lineage groups varies by type. For example, for HPV68, the two variant lineages differ by ~8% across the complete genomes, indicating that these are technically subtypes of HPV68. For HPV35, the two lineages found in Guanacaste differed by <1% in the complete genomes; as they are minimally divergent, they were named A1 and A2 rather than assigning them separate letters.

Data analysis

This nested case-control analysis was performed at the level of individual infections, not individual women. Multiple HPV infections, whether concurrent or not, are common. Each infection outcome was considered separately, and the women in the cohort could contribute to multiple type analyses and case-control outcomes. For example, an individual with HPV X linked to development of CIN3 or CIN2 (it could not be both because of censoring for treatment) and HPV Y that cleared quickly would contribute both a case infection for the HPV X analysis and a control infection for the HPV Y analysis. Moreover, a case of CIN3+, CIN2, persistence, or a control could be associated with two or more HPV types and be counted more than once. This occurred in 38.4% of the 2,074 women in the study; on average, each woman contributed 1.6 infections to the analysis. Because we have not found confounding between multiple infections in previous HPV natural studies in which we used GEE statistical methods (25), we checked for possible confounding more simply in this analysis by restriction to women with single infections. This decreased statistical power and widened confidence intervals

Table 1. Variant lineages of HPV types in the high-risk clade and outcome of infection in the prospective Guanacaste, Costa Rica, cohort study

| Species | HPV type | Test variant | n | % | Reference | n | % | Case status | OR | P | |
|---------|----------|--------------|-----|---------|-----------|---------|---------|-------------|------------|-------|-------|
| α-9 | 52 | A, B, C | 16 | 6.78% | D | 0 | 0.00% | CIN3+ | ∞ | 0.78 | |
| | | | 20 | 8.47% | | 0 | 0.00% | CIN2 | ∞ | 0.42 | |
| | | | 16 | 6.78% | | 2 | 16.67% | Persistent | 0.43 | 0.17 | |
| | | | 184 | 77.97% | | 10 | 83.33% | Control | 1.00 | | |
| | | | 236 | 100.00% | | 12 | 100.00% | Total | | | |
| | 67 | A | | 1 | 8.33% | B | 1 | 5.88% | CIN3+ | 1.36 | |
| | | | | 0 | 0.00% | | 1 | 5.88% | CIN2 | 0.00 | |
| | | | | 11 | 91.67% | | 15 | 88.24% | Control | 1.00 | |
| | | | | 12 | 100.00% | | 17 | 100.00% | Total | | |
| | 33 | A, B | | 9 | 9.78% | C | 0 | 0.00% | CIN3+ | ∞ | 0.80 |
| | | | | 11 | 11.96% | | 1 | 9.09% | CIN2 | 1.24 | 0.80 |
| | | | | 10 | 10.87% | | 3 | 27.27% | Persistent | 0.38 | 0.12 |
| | | | | 62 | 67.39% | | 7 | 63.64% | Control | 1.00 | |
| | 58 | A | | 92 | 100.00% | B, C, D | 11 | 100.00% | Total | | |
| | | | | 21 | 7.87% | | 0 | 0.00% | CIN3+ | ∞ | 0.47 |
| | | | | 36 | 13.48% | | 0 | 0.00% | CIN2 | ∞ | 0.16 |
| | | | | 28 | 10.49% | | 5 | 29.41% | Persistent | 0.37 | 0.050 |
| | 16 | Non-E | | 182 | 68.16% | E | 12 | 70.59% | Control | 1.00 | |
| | | | | 267 | 100.00% | | 17 | 100.00% | Total | | |
| | | | | 21 | 29.58% | | 68 | 15.74% | CIN3+ | 2.43 | 0.004 |
| 5 | | | | 7.04% | 63 | | 14.58% | CIN2 | 0.63 | 0.45 | |
| 31 | A,B | | 11 | 15.49% | C | 33 | 7.64% | Persistent | 2.63 | 0.010 | |
| | | | 34 | 47.89% | | 268 | 62.04% | Control | 1.00 | | |
| | | | 71 | 100.00% | | 432 | 100.00% | Total | | | |
| | | | 18 | 10.23% | | 4 | 5.71% | CIN3+ | 1.81 | 0.38 | |
| 35 | A1 | | 20 | 11.36% | A2 | 6 | 8.57% | CIN2 | 1.34 | 0.56 | |
| | | | 16 | 9.09% | | 11 | 15.71% | Persistent | 0.58 | 0.22 | |
| | | | 122 | 69.32% | | 49 | 70.00% | Control | 1.00 | | |
| | | | 176 | 100.00% | | 70 | 100.00% | Total | | | |
| 73 | A | | 4 | 13.33% | A2 | 2 | 2.99% | CIN3+ | 6.35 | | |
| | | | 2 | 6.67% | | 5 | 7.46% | CIN2 | 1.27 | | |
| | | | 7 | 23.33% | | 6 | 8.96% | Persistent | 3.71 | | |
| | | | 17 | 56.67% | | 54 | 80.60% | Control | 1.00 | | |
| α-11 | 73 | A | 30 | 100.00% | B | 67 | 100.00% | Total | | | |
| | | | 1 | 4.00% | | 0 | 0.00% | CIN3+ | ∞ | | |
| | | | 1 | 4.00% | | 2 | 7.69% | CIN2 | 0.50 | | |
| | | | 1 | 4.00% | | 2 | 7.69% | Persistent | 0.50 | | |
| α-7 | 18 | AA,E | 22 | 88.00% | Af | 22 | 84.62% | Control | 1.00 | | |
| | | | 25 | 100.00% | | 26 | 100.00% | Total | | | |
| | | | 17 | 9.94% | | 4 | 8.00% | CIN3+ | 1.23 | 0.89 | |
| | | | 12 | 7.02% | | 0 | 0.00% | CIN2 | ∞ | 0.05 | |
| 45 | B | | 18 | 10.53% | A | 10 | 20.00% | Persistent | 0.52 | 0.12 | |
| | | | 124 | 72.51% | | 36 | 72.00% | Control | 1.00 | | |
| | | | 171 | 100.00% | | 50 | 100.00% | Total | | | |
| | | | 5 | 6.41% | | 2 | 3.17% | CIN3+ | 2.25 | 0.35 | |
| 78 | | | 3 | 3.85% | A | 4 | 6.35% | CIN2 | 0.68 | 0.58 | |
| | | | 10 | 12.82% | | 3 | 4.76% | Persistent | 3.00 | 0.11 | |
| | | | 60 | 76.92% | | 54 | 85.71% | Control | 1.00 | | |
| | | | 78 | 100.00% | | 63 | 100.00% | Total | | | |

(Continued on the following page)

Table 1. Variant lineages of HPV types in the high-risk clade and outcome of infection in the prospective Guanacaste, Costa Rica, cohort study (Cont'd)

| Species | HPV type | Test variant | <i>n</i> | % | Reference | <i>n</i> | % | Case status | OR | <i>P</i> |
|---------|----------|--------------|----------|---------|-----------|----------|------------|-------------|------|----------|
| α-5 | 70 | B | 3 | 3.33% | A | 3 | 1.81% | CIN3+ | 1.77 | |
| | | | 5 | 5.56% | | 7 | 4.22% | CIN2 | 1.26 | |
| | | | 13 | 14.44% | | 34 | 20.48% | Persistent | 0.68 | |
| | | | 69 | 76.67% | | 122 | 73.49% | Control | 1.00 | |
| | | | 90 | 100.00% | | 166 | 100.00% | total | | |
| | 39 | A2 | 1 | 2.86% | A1 | 2 | 2.94% | CIN3+ | 1.09 | |
| | | | 3 | 8.57% | | 3 | 4.41% | CIN2 | 2.19 | |
| | | | 4 | 11.43% | | 4 | 5.88% | Persistent | 2.19 | |
| | | | 27 | 77.14% | | 59 | 86.76% | Control | 1.00 | |
| | | | 35 | 100.00% | | 68 | 100.00% | Total | | |
| | 68 | A | 1 | 5.00% | B | 2 | 4.26% | CIN3+ | 1.33 | |
| | | | 3 | 15.00% | | 0 | 0.00% | CIN2 | ∞ | |
| | | | 1 | 5.00% | | 5 | 10.64% | Persistent | 0.53 | |
| | | | 15 | 75.00% | | 40 | 85.11% | Control | 1.00 | |
| | | | 20 | 100.00% | | 47 | 100.00% | Total | | |
| 69 | A | 1 | 11.11% | B | 1 | 9.09% | CIN2 | 1.50 | | |
| | | 2 | 22.22% | | 1 | 9.09% | Persistent | 3.00 | | |
| | | 6 | 66.67% | | 9 | 81.82% | Control | 1.00 | | |
| | | 9 | 100.00% | | 11 | 100.00% | Total | | | |
| | | | | | | | | | | |
| 51 | B | 7 | 16.28% | A | 3 | 3.19% | CIN3+ | 5.19 | 0.02 | |
| | | 1 | 2.33% | | 13 | 13.83% | CIN2 | 0.17 | 0.09 | |
| | | 4 | 9.30% | | 9 | 9.57% | Persistent | 0.99 | 0.88 | |
| | | 31 | 72.09% | | 69 | 73.40% | Control | 1.00 | | |
| | | 43 | 100.00% | | 94 | 100.00% | Total | | | |
| 53 | A | 2 | 3.57% | B | 7 | 3.11% | CIN3+ | 1.07 | 0.84 | |
| | | 2 | 3.57% | | 8 | 3.56% | CIN2 | 0.94 | 0.85 | |
| | | 4 | 7.14% | | 30 | 13.33% | Persistent | 0.50 | 0.21 | |
| | | 48 | 85.71% | | 180 | 80.00% | Control | 1.00 | | |
| | | 56 | 100.00% | | 225 | 100.00% | Total | | | |
| 56 | A, B | 6 | 4.38% | C | 0 | 0.00% | CIN3+ | ∞ | 0.12 | |
| | | 10 | 7.30% | | 1 | 2.04% | CIN2 | 4.17 | 0.13 | |
| | | 18 | 13.14% | | 5 | 10.20% | Persistent | 1.50 | 0.54 | |
| | | 103 | 75.18% | | 43 | 87.76% | Control | 1.00 | | |
| | | 137 | 100.00% | | 49 | 100.00% | Total | | | |
| 66 | A | 4 | 6.56% | B | 1 | 1.85% | CIN3+ | 3.45 | 0.27 | |
| | | 4 | 6.56% | | 3 | 5.56% | CIN2 | 1.15 | 0.85 | |
| | | 2 | 3.28% | | 6 | 11.11% | Persistent | 0.29 | 0.11 | |
| | | 51 | 83.61% | | 44 | 81.48% | Control | 1.00 | | |
| | | 61 | 100.00% | | 54 | 100.00% | Total | | | |

NOTE: The order of the species groups and the types parallel Fig. 1 to highlight phylogenetic relationships. For ease of interpretation, the choices of “test” and “reference” variant lineage for each type were made to produce OR values for CIN3+ versus control of >1. This does not affect the *P* value for each type or the combined *P* values. *P* values were computed only when the number of cases of CIN3+ for test and reference lineages combined was 5 or more.

For HPV types including HPV16, the combined mid-*P* value was 0.07 for CIN3+ versus controls, 0.27 for CIN2 versus controls, and <0.01 for long-term persistence. When we excluded HPV16, the combined mid-*P* values were 0.37, 0.25, and 0.04, respectively.

but did not change the conclusions (data not shown). Thus, the full results are presented.

Statistical testing. We wished to know whether the major bifurcation, i.e., the division into dichotomized variant

lineages (i.e., arbitrarily defined as lineage “A” or “B,” etc.) affected the probability of different case-control outcomes for each HPV type within the high-risk clade individually and globally combining over all types. To assess each type

individually, we compared the distribution of the major variant lineages among the cases and controls within that type. For ease of presentation, we defined the “test” and “reference” variants arbitrarily so that the odds ratio (OR) was positive for the risk of CIN3+. We calculated OR as a measure of the strength of association for each lineage and case group (CIN3, CIN2, persistence) compared with controls.

We did not have large enough numbers of individual types (except for HPV16) to calculate confidence intervals around the OR estimates; instead, we calculated a *P* value by Fisher's exact test (26). Specifically, we calculated a mid-*P* statistic for each type because the standard *P* values for exact tests are too conservative (27). To test whether variant lineages, on average, were associated with probability of case-control status, we combined the *P* values across the types (analogous to a meta-analysis) using the Fisher's combination test (28). For extra rigor, we computed the Fisher's combination test statistics twice, including and excluding HPV16 (because we had prior reason to expect that risk associations exist for HPV16 variants).

To assure that we highlighted evidence only from types for which we had reasonable statistical power, we excluded from statistical testing any case-control comparison for which there were fewer than five case outcomes. We also excluded variant lineages for types having limited variability within the Guanacaste population (i.e., <1% variability), reasoning that these variants might be too minor to show phenotypic differences. The results of this agnostic decision in the special and interesting case of HPV35 are discussed in the results.

Results

The results are shown in Table 1. The order of the species (α -9, 11, 7, 5, 6) and types within groups corresponds to Fig. 1, to highlight phylogenetic relatedness.

For HPV types including HPV16, the combined mid-*P* value for CIN3+ versus controls was 0.07, for CIN2 versus controls it was 0.27, and for long-term persistence, it was <0.01. When we excluded HPV16, the combined mid-*P* values were 0.37, 0.25, and 0.04, respectively. Thus, the risk associations for HPV16 contributed greatly to the global statistic for CIN3+ when the predominant, powerfully carcinogenic type was included. Nevertheless, the association of variants with persistence remained significant even when HPV16 was excluded.

In confirmatory analyses, we recreated Table 1 combining all women without CIN2+ as the controls and CIN3+ as the main case group (because CIN2 is unreliable). There were isolated small differences for a few types but no change in overall interpretation. We also restricted the analysis to incident infections. The statistical power was much reduced, but the risk relationships were unchanged within the limitations of the data.

Most type-specific analyses had limited power, except for HPV16 for which we were able to compute reliable confidence intervals. We previously published in a preliminary, cross-sectional examination of HPV16 during the enrollment phase of the Proyecto Epidemiológico Guanacaste that the non-European variant lineage was associated with an increased risk of histologic CIN2+ (14). We confirmed in the current analysis that the risk of the non-European lineage

was especially high for the 10 screen-detected cases of invasive cervix cancer [OR, 6.3; 95% confidence interval (95% CI), 1.6–24.6] and was significantly elevated among the 70 cases of CIN3 (OR, 1.8; 95% CI, 0.9–3.6). Addition of 10 HPV16-associated supplementary cases of invasive cervical cancer diagnosed during the time of the Guanacaste Cohort Study, outside of the random sample of the population that comprised our cohort, supported the association for invasive cancer (OR, 5.25; 95% CI, 1.4–19.6). Viral persistence of HPV16 for 2 or more years was linked to the same non-European variants as CIN3 and cancer (OR, 2.6; 95% CI, 1.2–5.7). It should be noted that the Asian American HPV16 variants comprised 94% of the non-European variants detected in the Guanacaste population.

The more equivocal precancerous diagnoses of histologic CIN2 or cytologic HSIL were not associated with HPV16 variant status (or any other type with the possible exception of HPV18).

α -9, which contains HPV16, is the most important HPV species with regard to carcinogenicity; all other types in this species with the exception of HPV67 are also clearly carcinogenic. Nevertheless, HPV35 infections in this population had limited genetic variability precluding the characterization into major variant lineages at the >1% difference level. Although HPV35 variants were excluded from the main calculation of *P* values, because A1 and A2 vary by <1% in the whole genome, A1 was associated with a substantial and statistically significantly elevated risk of both CIN3+ (OR = 6.4) and long-term viral persistence (OR = 3.7) compared with A2.

There are three lineages of HPV31 of nearly equal phylogenetic distance;⁸ comparing A and B with C resulted in an OR of 1.8, but the *P* value was nonsignificant. No other grouping was linked to risk. For several of the carcinogenic HPV16-related types in α -9, all cases of CIN3 were linked to one of the two variant lineage groups (e.g., HPV33, HPV52, and HPV58), resulting in OR estimates of infinity. However, these findings were highly unstable and not statistically significant.

For many of the remaining HPV types in other species, one of the variants was associated with >1.5 — twice the risk of CIN3+ compared with the other arbitrarily chosen reference. The B variant of HPV51 was significantly associated with increased risk of CIN3+, although there were few cases.

Discussion

We designed this population-based case-control study to extend our previous analysis in Guanacaste that showed strong differences in risk of viral persistence and CIN3+ diagnosis for different HPV types in the α genus (1). In that analysis combining phylogenetics and epidemiology, we observed that all carcinogenic and possibly carcinogenic types belonged to one evolutionary branching or clade. This high-risk clade was generated in large part by the early gene-containing segment of the HPV genome (11, 29), which contains the HPV oncogenes. Nevertheless, the types in the high-risk clade showed strong differences in their carcinogenic potential. HPV carcinogenicity may resemble a quantitative trait that peaks with the genome of HPV16 and nearly completely

dissipates with other types, such as the evolutionarily related α -9 species member HPV67.

We examined whether more recent evolutionary forces resulting in variant lineages of specific types might yield even better prediction of viral natural history. For each HPV type emerging from the high-risk α clade, we determined whether the earliest, most divergent variant lineages predicted risk of viral persistence or CIN3+ compared with viral clearance. To our knowledge, this is the first complete analysis of HPV variants and cervical carcinogenicity.

Our results show that variants of HPV16 have evolved clearly different propensities for both viral persistence and carcinogenicity. In Guanacaste, the higher risk, non-European HPV16 variants were virtually all Asian-American. Other investigators have shown that, compared with European lineages, African variants of HPV16 are associated with a higher risk of viral persistence, CIN3, and cancer (13). The underlying genetic details that make non-European variants of HPV16 more carcinogenic are not known, and this important clue definitely deserves pursuit to identify the underlying genetic basis of this association.

Even the large Guanacaste cohort was too small to generate enough cases to evaluate conclusively the association of variants of other HPV types and risk of CIN3+. We had hypothesized that some HPV types with intermediate (like HPV52 or 56) or even equivocal carcinogenicity (like HPV68) might be a mixture of strongly carcinogenic and noncarcinogenic variants. If true, such heterogeneity would have shown itself as large variation in risk between lineages. However, there were too few cases to find all but extreme risk differences. From a genetic epidemiologic point of view, this is reminiscent of the large sample sizes required to identify SNPs with a modest risk for breast, colon, or prostate cancer.

We found substantial and significant associations HPV35 variant A1 compared with A2 for both CIN3+ (OR 6.4) and long-term persistence (OR 3.7). The finding for HPV35 is especially interesting, because it is so closely related to HPV16. Variants of HPV35 have not been studied previously in relation to carcinogenicity. The number of SNP changes differentiating these HPV35 variants is only \sim 100. If we had included HPV35 in the mid-*P* calculations, the global estimates would have strengthened particularly for a summary (albeit *a posteriori*) examination of α -9 types (including HPV16, mid-*P* would be 0.001 for persistence, 0.03 for CIN3+). We intend to explore whether higher-risk HPV35 variants resemble HPV16 in pending full genome studies.

There have been only a few previous studies of variants of types related to HPV16 in immunocompetent women, including two previous reports of HPV33 variants associated with high-grade cytology/histology (30, 31) and one report of HPV58 variants associated with increasing severity of neoplasia (32). Another study of HPV58 and HPV52 variants was null (30). A needed standard nomenclature of variants of all the carcinogenic HPV types accepted by the International Classification of Tumor Viruses Committee is forthcoming, which will promote comparisons and combinations of data (available from the authors⁸).

Regarding HPV species other than α -9, the numbers of cases of CIN3 for individual types were very small; some of the pos-

itive and negative results could have resulted by chance. For example, we observed a significantly elevated risk of CIN3+ for HPV51 variant B compared with variant A. There is no published work on HPV51 variants to which we can refer.

We observed no differences for CIN2 compared with controls for HPV16 variants or any other type, with the possible exception of HPV18. The lack of associations with risk of CIN2 or HSIL is consistent with the belief that many CIN2 histologic lesions (and HSIL cytology interpretations) are misclassified as precancer even when diagnosed by an expert pathology panel (17). This further supports the critical importance of having the correct phenotype (end point definition) for analysis to avoid dilution of true cancer surrogate end points. This also applies to using long-term type-specific persistence of HPV as an outcome. Accurate repeated characterization of HPV types can be complicated by the presence of multiple types within a clinical sample; the methods we used are generally robust for the detection of multiple types (21).

The aggregate data showed that variants of other HPV types in the high-risk clade, even excluding HPV16, significantly influenced risk of persistence. We observed that viral variants influenced long-term viral persistence (2+ years) even when CIN3+ had not (yet?) been diagnosed. Viral persistence is an uncommon viral outcome (i.e., clearance is typical) and a necessary part of progression to precancer. The persistence of HPV with interruption of p53 and pRB pathways links HPV carcinogenesis to the mechanism of other tumor viruses (33). We had more power to study long-term viral persistence than CIN3+. Our finding that HPV variants assessed globally influence persistence is in line with a small literature including some studies in HIV-infected populations (34–38). It is not known whether observations among immunosuppressed individuals apply to immunocompetent populations in which most HPV evolution has presumably taken place.

Our analyses were based on viral lineages determined from URR sequence alignment, with examination of E6 when necessary. Lineage assignment based on this hypervariable region is very efficient; the evolution of HPV is slow with very little evidence for recombination (13, 39, 40). Thus, lineages are stable and predictive of most viral SNPs. However, URR-based lineages are not sufficient for all purposes, and our data might have missed important variability in other regions of the HPV genome that predicts persistence and/or carcinogenicity.

Using whole-genome sequencing and sufficiently large numbers of cases (i.e., thousands), we could make substantial progress in understanding the viral genetic basis of HPV carcinogenesis (41). Because the HPV genome is relatively small (\sim 8,000 bases) with only eight genes and because variants of each of the HPV types differ by only \sim 200 bp, we believe we will be able to identify viral genetic changes associated with carcinogenicity. We will be able to identify targets and or biochemical functions within the host, which are perturbed in the cascade to cancer.

Although much more difficult, we envision eventually implementing systems biology approaches to approach the vast networks of changes required for an infected cervical cell to become frankly malignant. It might prove fruitful to pursue why a given woman becomes a case of CIN3+ caused by one HPV type or variant but successfully clears other HPV types

and variants. Consideration of viral gene variation in relation to relevant human variation in immune recognition or tumor suppressor genes might permit an initial, focused look at systems biology (42, 43). If successful, understanding the multiple networks, pathways, and interactions could lead to a new understanding of cancer biology and innovative strategies to combat HPV carcinogenesis.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

- Schiffman M, Herrero R, Desalle R, et al. The carcinogenicity of human papillomavirus types reflects viral evolution. *Virology* 2005; 337:76–84.
- Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. *Lancet* 2007;370:890–907.
- Kovacic MB, Castle PE, Herrero R, et al. Relationships of human papillomavirus type, qualitative viral load, and age with cytologic abnormality. *Cancer Res* 2006;66:10112–9.
- Ho GY, Bierman R, Beardsley L, Chang CJ, Burk RD. Natural history of cervicovaginal papillomavirus infection in young women. *N Engl J Med* 1998;338:423–8.
- Rodriguez AC, Schiffman M, Herrero R, et al. Rapid clearance of human papillomavirus and implications for clinical focus on persistent infections. *J Natl Cancer Inst* 2008;100:513–7.
- de Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen H. Classification of papillomaviruses. *Virology* 2004;324:17–27.
- Bouvard V, Baan R, Straif K, et al. A review of human carcinogens: Part B. Biological agents. *Lancet Oncol* 2009;10:321–2.
- Schiffman M, Clifford G, Buonaguro FM. Classification of weakly carcinogenic human papillomavirus types: addressing the limits of epidemiology at the borderline. *Infect Agent Cancer* 2009;4:8.
- Chen Z, Terai M, Fu L, Herrero R, DeSalle R, Burk RD. Diversifying selection in human papillomavirus type 16 lineages based on complete genome analyses. *J Virol* 2005;79:7014–23.
- Chen Z, DeSalle R, Schiffman M, Herrero R, Burk RD. Evolutionary dynamics of variant genomes of human papillomavirus types 18, 45, and 97. *J Virol* 2009;83:1443–55.
- Narechania A, Chen Z, DeSalle R, Burk RD. Phylogenetic incongruence among oncogenic genital α human papillomaviruses. *J Virol* 2005;79:15503–10.
- Hazard K, Andersson K, Dillner J, Forslund O. Human papillomavirus subtypes are not uncommon. *Virology* 2007;362:6–9.
- Bernard HU, Calleja-Macias IE, Dunn ST. Genome variation of human papillomavirus types: phylogenetic and medical implications. *Int J Cancer* 2006;118:1071–6.
- Hildesheim A, Schiffman M, Bromley C, et al. Human papillomavirus type 16 variants and risk of cervical cancer. *J Natl Cancer Inst* 2001; 93:315–8.
- Bratti MC, Rodriguez AC, Schiffman M, et al. Description of a seven-year prospective study of human papillomavirus infection and cervical neoplasia among 10000 women in Guanacaste, Costa Rica. *Rev Panam Salud Publica* 2004;15:75–89.
- Herrero R, Schiffman MH, Bratti C, et al. Design and methods of a population-based natural history study of cervical neoplasia in a rural province of Costa Rica: the Guanacaste Project. *Rev Panam Salud Publica* 1997;1:362–75.
- Carreon JD, Sherman ME, Guillen D, et al. CIN2 is a much less reproducible and less valid diagnosis than CIN3: results from a histological review of population-based cervical samples. *Int J Gynecol Pathol* 2007;26:441–6.
- Stoler MH, Schiffman M. Atypical squamous cells of undetermined significance-low-grade squamous intraepithelial lesion triage study (ALTS) group. Interobserver reproducibility of cervical cytologic and histologic interpretations: realistic estimates from the ASCUS-LSIL Triage Study. *JAMA* 2001;285:1500–5.
- Castle PE, Schiffman M, Gravitt PE, et al. Comparisons of HPV DNA detection by MY09/11 PCR methods. *J Med Virol* 2002;68:417–23.
- Herrero R, Hildesheim A, Bratti C, et al. Population-based study of human papillomavirus infection and cervical neoplasia in rural Costa Rica. *J Natl Cancer Inst* 2000;92:464–74.
- Qu W, Jiang G, Cruz Y, et al. PCR detection of human papillomavirus: comparison between MY09/MY11 and GP5+/GP6+ primer systems. *J Clin Microbiol* 1997;35:1304–10.
- Chen Z, Schiffman M, Herrero R, Desalle R, Burk RD. Human papillomavirus (HPV) types 101 and 103 isolated from cervicovaginal cells lack an E6 open reading frame (ORF) and are related to γ -papillomaviruses. *Virology* 2007;360:447–53.
- Wheeler CM, Yamada T, Hildesheim A, Jenison SA. Human papillomavirus type 16 sequence variants: identification by E6 and L1 lineage-specific hybridization. *J Clin Microbiol* 1997;35:11–9.
- Ong CK, Chan SY, Campo MS, et al. Evolution of human papillomavirus type 18: an ancient phylogenetic root in Africa and intratype diversity reflect coevolution with human ethnic groups. *J Virol* 1993;67:6424–31.
- Plummer M, Schiffman M, Castle PE, Maucort-Boulch D, Wheeler CM, ALTS Group. A 2-year prospective study of human papillomavirus persistence among women with a cytological diagnosis of atypical squamous cells of undetermined significance or low-grade squamous intraepithelial lesion. *J Infect Dis* 2007;195:1582–9.
- Fisher RA. On the interpretation of χ^2 from contingency tables, and the calculation of P. *J R Stat Soc* 1922;85:87–94.
- Lancaster HO. Significance tests in discrete distributions. *J Am Statistical Assoc* 1961;56:223–34.
- Fisher RA. *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd; 1954.
- Burk RD, Chen Z, Van Doorslaer K. Human papillomaviruses: genetic basis of carcinogenicity. *Public Health Genomics* 2009;12:281–90.
- Xin CY, Matsumoto K, Yoshikawa H, et al. Analysis of E6 variants of human papillomavirus type 33, 52 and 58 in Japanese women with cervical intraepithelial neoplasia/cervical cancer in relation to their carcinogenic potential. *Cancer Lett* 2001;170:19–24.
- Khoudri S, Villa LL, Gagnon S, et al. Human papillomavirus type 33 polymorphisms and high-grade squamous intraepithelial lesions of the uterine cervix. *J Infect Dis* 2006;194:886–94.
- Chan PK, Lam CW, Cheung TH, et al. Association of human papillomavirus type 58 variant with the risk of cervical cancer. *J Natl Cancer Inst* 2002;94:1249–53.
- zur Hausen H. Papillomaviruses and cancer: from basic studies to clinical application. *Nat Rev Cancer* 2002;2:342–50.
- Gagnon S, Hankins C, Tremblay C, et al. Polymorphism of human papillomavirus type 31 isolates infecting the genital tract of HIV-seropositive and HIV-seronegative women at risk for HIV infection. *J Med Virol* 2005;75:213–21.
- Gagnon S, Hankins C, Tremblay C, et al. Viral polymorphism in

Grant Support

The Guanacaste Project was supported by National Cancer Institute, NIH, Department of Health and Human Services contracts NO1-CP-21081, NO1-CP-33061, NO1-CP-40542, and NO1-CP-506535. R.D. Burk was supported by National Cancer Institute grant CA78527 and used the facilities available through the Einstein Cancer Research Center and CFAR.

We thank the Costa Rican health authorities for the support and the superb project staff for the decade-long dedication.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received 11/17/2009; revised 01/19/2010; accepted 01/31/2010; published OnlineFirst 03/30/2010.

- human papillomavirus types 33 and 35 and persistent and transient infection in the genital tract of women. *J Infect Dis* 2004;190:1575–85.
36. Aho J, Hankins C, Tremblay C, et al. Genomic polymorphism of human papillomavirus type 52 predisposes toward persistent infection in sexually active women. *J Infect Dis* 2004;190:46–52.
 37. Schlecht NF, Burk RD, Palefsky JM, et al. Variants of human papillomaviruses 16 and 18 and their natural history in human immunodeficiency virus-positive women. *J Gen Virol* 2005;86:2709–20.
 38. Xi LF, Kiviat NB, Hildesheim A, et al. Human papillomavirus type 16 and 18 variants: race-related distribution and persistence. *J Natl Cancer Inst* 2006;98:1045–52.
 39. Prado JC, Calleja-Macias IE, Bernard HU, et al. Worldwide genomic diversity of the human papillomaviruses-53, 56, and 66, a group of high-risk HPVs unrelated to HPV-16 and HPV-18. *Virology* 2005;340:95–104.
 40. Calleja-Macias IE, Villa LL, Prado JC, et al. Worldwide genomic diversity of the high-risk human papillomavirus types 31, 35, 52, and 58, four close relatives of human papillomavirus type 16. *J Virol* 2005;79:13630–40.
 41. Diallo AB, Badescu D, Blanchette M, Makarenkov V. A whole genome study and identification of specific carcinogenic regions of the human papilloma viruses. *J Comput Biol* 2009;16:1461–73.
 42. de Araujo Souza PS, Maciag PC, Ribeiro KB, Petzl-Erler ML, Franco EL, Villa LL. Interaction between polymorphisms of the human leukocyte antigen and HPV-16 variants on the risk of invasive cervical cancer. *BMC Cancer* 2008;8:246.
 43. Koshiol J, Hildesheim A, Gonzalez P, et al. Common genetic variation in TP53 and risk of human papillomavirus persistence and progression to CIN3/cancer revisited. *Cancer Epidemiol Biomarkers Prev* 2009;18:1631–7.
 44. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 2001;17:754–5.
 45. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 2003;19:1572–4.
 46. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 1998;14:817–8.