Prevalence and Clustering Patterns of Human Papillomavirus Genotypes in Multiple Infections

Anil K. Chaturvedi,[1] Leann Myers,[2] Ansley F. Hammons,[3] Rebecca A. Clark,[3] Kathleen Dunlap,[4] Patricia J. Kissinger,[1] and Michael E. Hagensee[1,3]

Departments of [1]Epidemiology and [2]Biostatistics, Tulane University School of Public Health and Tropical Medicine; and Departments of [3]Medicine, [4]Obstetrics and Gynecology, Louisiana State University Health Sciences Center, New Orleans, Louisiana

Abstract

Prevalence of multiple human papillomavirus (HPV) infections, involvement of specific HPV phylogenetic clades in multiple infections, and clustering patterns of multiple infections at the clade level were assessed in 854 HIV (−) and 275 HIV (+) women cross-sectionally. Reverse line blot assay was used to detect 27 HPV genotypes. Involvement of specific clades in coinfections and clustering patterns were assessed using HPV clade/genotype as the unit of analyses. Expected frequencies assuming independence for all possible clade combinations in two-genotype infections were derived using a multinomial expansion and comparisons of observed and expected frequencies were done using a composite goodness-of-fit test. In all, 100 two-genotype infections were detected; 61 in HIV (−) and 39 in HIV (+) women. Clade A9 (HPV types 16, 31, 33, 35, 52, and 58) was significantly less likely to be involved in multiple infections compared with all other clades (55.2% versus 64.6%; adjusted odds ratios, 0.68; 95% confidence interval, 0.48-0.95). Observed patterns for all possible clade combinations (among HPV clades A3, A5, A6, A7, A9, and A10) in two-genotype infections did not significantly differ from those expected in the entire sample, across HIV, Pap smear, and age strata (all goodness-of-fit exact $P > 0.20$). These results indicate that clade A9 is less likely to be involved in multiple infections and that HPV genotypes predominantly establish multiple infections at random, with little positive/negative clustering for either phylogenetically related or unrelated types. The current method of analysis affords the opportunity to test clustering of a large number of HPV genotype/clade combinations at nominal alpha levels. (Cancer Epidemiol Biomarkers Prev 2005;14(10):2439–45)

Introduction

Among women infected with human papillomaviruses (HPV), concurrent infection with more than one HPV type is observed in 20% to 50% of HIV (−) (refs. 1, 2) and 50% to 80% of HIV (+) populations (3, 4). Recently, multiple HPV infections have gained increasing attention owing to reports of successful prophylactic vaccination trials (5, 6). The molecular and epidemiologic significance of infection with multiple HPV genotypes is as yet unclear. Although immunity towards HPV is predominantly type-specific (7), theoretical concerns exist regarding the impact of the removal of certain genotypes from human populations through vaccination (8). The removal of certain genotypes by type-specific vaccination could result in positive selective pressures on untargeted genotypes, thus increasing their prevalence; or type-specific vaccination could confer protective immunity against phylogenetically related genotypes (9). These concerns mandate a solid understanding of the equilibrium and clustering patterns among various HPV clades/genotypes in concurrent and sequential coinfections. Studying the clustering patterns in naturally occurring multiple infections may aid in better characterizing this equilibrium and in assessing the impact that type-specific vaccination may have on untargeted genotypes. Furthermore, studying clustering patterns may reflect on the existence of separate transmission patterns at the HPV clade/genotype level (2).

Few studies have addressed the issue of clustering patterns of HPV clades or genotypes involved in concurrent or sequential multiple infections. DNA-based follow-up studies indicate that preexisting HPV infection at baseline increases the risk of acquisition of both phylogenetically related and unrelated genotypes (1, 8, 10). These studies also indicate the lack of the existence of any significant patterns over and beyond chance expectations in the clustering of genotypes involved in coinfections (1, 8-10). Previous studies addressing clustering patterns have focused on either HPV 16 or a select few genotypes. Owing to the vast heterogeneity of HPVs, these analyses focusing on HPV clades or specific genotype combinations are affected by problems arising from restrictive sample sizes and multiple statistical testing. These problems preclude a comprehensive evaluation of all possible clustering patterns involving a vast range of genotypes.

In the current study, we assessed the clustering patterns at the phylogenetic clade level of 27 HPV genotypes that are involved in concurrent multiple infections. In order to protect the results from inflated type I error levels arising from multiple testing, a parallel was drawn with the Hardy-Weinberg equilibrium test (11), and observed patterns of coinfections were compared with those expected assuming independence of HPV clades using a composite goodness-of-fit test. The goals of the current study were to (*a*) describe the prevalence of multiple HPV infections; (*b*) assess if the distribution of number of coinfecting HPV types follows a random Poisson distribution; (*c*) assess if particular HPV clades were more or less likely to be involved in multiple infections; and (*d*) assess the clustering patterns of coinfections at the phylogenetic clade level, in a cross-sectional study of HIV (−) and HIV (+) women.

Materials and Methods

Study Subjects. The sample (n = 1,379) for this cross-sectional study consisted of a convenience sample of women seeking gynecologic care at the Medical Center of Louisiana in

Received 6/21/05; revised 7/29/05; accepted 8/12/05.

Grant support: Doris Duke clinical research grant, National Cancer Institute (grant NCI-1 R03 CA86378), and Health Excellency Fund of Louisiana.

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.

Requests for reprints: Anil K. Chaturvedi, 6120 Executive Boulevard, EPS 8015, Rockville, MD-20852. Phone: 301-451-2495; Fax: 301-402-0817. E-mail: chaturva@mail.nih.gov

doi:10.1158/1055-9965.EPI-05-0465

New Orleans associated with the Louisiana State University Health Sciences Center system. Women referred to the colposcopy clinic due to a previous history of cytologic abnormalities (n = 1,060) constituted the HIV (−) sample. Women attending the HIV outpatient clinic (n = 319) constituted the HIV (+) sample. Women ages 18 to 78 years attending the aforementioned clinics between January 1999 and June 2003 were included in the study. This study was approved by the Institutional Review Board and research review committees of Louisiana State University Health Sciences Center, Medical Center of Louisiana in New Orleans, and HIV outpatient clinic and exempted by the Institutional Review Board of Tulane University. Participation in the study was contingent upon obtaining signed informed consent.

Sample Collection and HPV Genotyping. Details of specimen and data collection, HPV PCR analysis, and HPV genotyping from this study have been previously reported (12). In brief, DNA from cervical/vaginal swabs was extracted and HPV DNA was detected using the PGMY09/11 biotinylated primer system (13). Genotyping was done by reverse line blot hybridization (Roche Molecular Systems, Alameda, CA; ref. 13). Pap smears were collected from each study subject prior to the collection of cervical/vaginal swabs. Abnormal cells were classified according to the Bethesda recommendations (14).

Statistical Analyses. Postrecruitment, 206 (19.4%) women from the HIV (−) group and 44 (13.7%) women from the HIV (+) group were excluded from analyses owing to missing/unsatisfactory cytology results (130 of 1,379; 9.43%) or nonamplifiability of the specimen (158 of 1,379; 11.45%) for HPV detection and genotyping. The final analyses included 1,129 women with 854 women from the high-risk HIV (−) group, and 275 HIV (+) women. No significant differences were observed in the age and race of women excluded from the study and women included in the study, either in the entire study sample or across HIV strata (data not shown).

Age was categorized as a binary variable (≤25 and >25 years); this categorization was based on previous research indicating a decline in HPV prevalence in women older than 25 years of age (15). Pap smear status was classified into a binary variable as normal and abnormal smears (abnormal category included ASCUS, LSIL and HSIL). HPV types were classified into high-oncogenic risk (HR) and low-oncogenic risk (LR) types based on the system adopted by Roche Molecular Systems (13).

Comparison of Observed and Expected Number of Coinfections. Poisson expected frequencies for the number of coinfecting genotypes were calculated using the mean number of infections and compared with observed frequencies using a one-way χ^2 goodness-of-fit test. Ratios of observed and expected frequencies were calculated and a 95% Poisson confidence interval (CI) was constructed around the ratio to test statistical significance. This procedure was done in the entire study sample and across HIV strata.

Involvement of Clades in Multiple Infections and Clustering of HPV Clades. For all analyses aimed at assessing if particular HPV clades were more or less likely to be involved in multiple infections and to assess the clustering patterns of coinfections at the phylogenetic clade level, the HPV genotype/clade was used as the unit of statistical analyses. Table 1 describes the phylogenetic clade assignments of the 27 HPV genotypes detected by the reverse line blot assay (16). HPV clades containing only one detected genotype were excluded from all analyses. These clades were A1, A4, A8, and A11 (corresponding HPV types were 42, 57, 40, and 73, respectively); furthermore, HPV 54 was excluded because it is yet to be assigned to a clade. This exclusion resulted in six phylogenetic clades for final analyses: A3, A5, A6, A7, A9, and A10. From the 1,129 subjects included in the analyses, 724 genotypes were detected; exclusion of HPV type 54 and clades A1, A4, and A8 resulted in 647 genotypes. This number, 647, was used as the denominator for the calculation of relative frequencies for each of the phylogenetic clades (A3, A5, A6, A7, A9, and A10). Consistent with the use of HPV genotype/clade as the unit of analysis, the relative frequency of each clade was used for all analyses instead of prevalence.

Table 1. Phylogenetic clade assignments for 27 HPV genotypes

HPV Clade	HPV Genotypes
A1	42
A2	—
A3	MM7 (83)*, MM8 (84)
A4	57
A5	26, 51, MM4 (82)
A6	53, 56, 66
A7	18, 39, 45, 59, 68
A8	40
A9	16, 31, 33, 35, 52, 58
A10	6, 11, 55
A11	MM9 (73)
Unclassified A	54

*Where appropriate, numerical assignments of HPV genotypes are indicated in parentheses.

In order to test heterogeneity in the proportion of multiple infections for particular HPV clades when compared with other clades, the presence/absence of multiple infections was used as the dependent variable. The independent variable of primary interest was the phylogenetic clade under question. The strength of association was assessed using odds ratios (OR) and 95% CI, where appropriate, exact 95% CIs were calculated. For multivariate analyses, based on clinical data, age, HIV status, and Pap smear status were considered as confounders for the association between each clade and multiple infection status. To derive adjusted OR and 95% CIs, individual binary logistic regression models were fit for each clade simultaneously adjusting for age, HIV status, and Pap smear status.

In order to test the hypothesis of random clustering of coinfecting HPV genotypes in two-genotype infections at the phylogenetic clade level, all possible observed coinfection pattern frequencies for the six clades (n = 21) were compared with null pattern frequencies of coinfections. The null pattern of coinfections was derived assuming independence of coinfecting HPV clades. Motivation for this method was derived from the statistical basis for Hardy-Weinberg equilibrium test (11). The Hardy-Weinberg equilibrium test is a binomial expansion of the form $(a + b)^2 = a^2 + 2\,ab + b^2$ where a and b are the allele frequencies, and a^2, $2\,ab$, and b^2 are the genotype frequencies. Drawing a parallel with the Hardy-Weinberg expected frequency calculation (in the current context, allele frequencies are equivalent to relative frequencies of each clade and genotype frequencies are equivalent to patterns of HPV coinfections), expected frequencies for two HPV type/clade infections, assuming statistical independence of clades, were calculated using a multinomial expansion of the form:

$$(\text{clade A3} + \text{clade A5} + \text{clade A6} + \text{clade A7} + \text{clade A9} + \text{clade A10})^2$$

The expanded multinomial includes 21 terms and represents a vector of probabilities for the 21 possible patterns of two-genotype infections. The product of this vector of probabilities with the number of subjects with two–HPV type infections represents a vector of expected frequencies for 21 patterns of two–HPV type infections.

In the expanded multinomial, clade-A9^2 represents the null probability of coinfection with two HPV types that belong to clade A9. This probability corresponding to clade A9^2 in turn equals the sum of cross-products of relative frequencies of all the HPV types that constitute clade A9 (16, 31, 33, 35, 52, and 58): 16*16, 16*31, 16*33, 16*35, 16*52, 16*58, and so on. A pattern of coinfections representing 16*16 (i.e., a coinfection of any HPV type with itself) is not observable and results in a structural zero. Although the patterns of coinfections of an HPV type with itself do not contribute to the observed frequencies, they contribute towards the calculation of expected frequencies. This results in an overestimation of expected frequencies for patterns that represent coinfections with two HPV types from the same clade. This bias of overestimation, however, owing to the absence of any structural zeros, does not exist for patterns which represent coinfections with two HPV types that belong to two different clades. In order to adjust analyses for this bias, expected probabilities for coinfection patterns involving two HPV types from the same clade (such as A3-A3, A5-A5, A6-A6, A7-A7, A9-A9, and A10-A10) were corrected by subtracting the sum of probabilities of all structural zeros within each pattern. Unlike the sum of uncorrected probabilities for the 21 coinfection patterns which equals 1.0, the sum of corrected probabilities is <1.0. In order to bring the sum of these corrected probabilities to 1.0, each of the corrected probability for the 21 patterns was divided by the sum of corrected probabilities of all 21 patterns. This process of normalization of each coinfection pattern probability with the sum of probabilities of all 21 patterns renders the use of relative frequency of an HPV clade equivalent to the use of prevalence of an HPV clade in the calculation of expected frequencies.

In order to test the goodness-of-fit of observed frequencies with expected frequencies, the absolute differences between observed and expected frequencies were calculated for each pattern of coinfections. A deviation score was then calculated by summation of the absolute differences of all 21 patterns. This deviation score from the study sample was compared with deviation scores generated from 10,000 Monte Carlo simulations. Exact P value for hypothesis testing was calculated as the number of simulations generating a deviation score as high as or higher than the study sample/10,000. In order to assess confounding effects of age (≤25 versus >25), HIV status, and Pap smear status (normal and abnormal), similar analyses (calculation of relative frequencies, expected frequencies for joint infection, and goodness-of-fit tests) were done stratifying across the confounders. Owing to sparse sample sizes, no tests of interaction were computed. The same procedure was extended to the situation of three genotype infections by raising the multinomial expansion to the power of three. The expanded multinomial for the three-genotype situation included 56 unique patterns. Because the total sample size of three genotype infections in the study (n = 30) was less than the number of unique patterns of infection (n = 56), no formal statistical testing was done.

All statistical analyses were completed in SAS version 9.1 (SAS Inc., Cary, NC). Significance of the clustering goodness-of-fit tests was assessed at $P < 0.20$; this decision was based on the composite nature of the null hypothesis which assumed that the observed data conformed to a random assortment of HPV clades for all patterns of two-genotype coinfections. The significance of all other hypothesis tests was assessed at $P < 0.05$ and adjustments for multiple comparisons were not done.

Results

Demographic and Clinical Characteristics of the Cohort. The demographic and clinical characteristics, HPV prevalence, and type-specific HPV prevalence of the HIV (−) and HIV (+) women included in the current study have been previously reported (12). In brief, the median age of the study subjects was 28 years (range, 18-78). The study sample was predominantly African-American (84.5%). Approximately 42% of the study sample presented with abnormal Pap smears; the majority of cytologic abnormalities observed in the study were ASCUS (42.7%) and LSIL (44.6%). HSIL was detected in only 5.3% of the study sample. Pap smear status was significantly different between HIV (+) and HIV (−) women (P value < 0.001); higher rates of ASCUS and HSIL were observed in the HIV (−) women and higher rates of LSIL were observed in HIV (+) women. When all grades of cytologic abnormalities were grouped together as ASCUS and above, no significant differences were observed in the rates of cytologic abnormalities between HIV (+) and HIV (−) women (45.5% and 41%, respectively, P value = 0.191). In all, 37.8% (427 of 1,129) of the study sample tested positive for any detectable HPV type. The prevalence of infection with HR and LR types was 32.3% and 12%, respectively. The majority of women with detectable HPV were infected with HR types (85.5%).

Multiple HPV Prevalence. Prevalence of multiple infections using a regular case definition (positive or negative for >1 HPV type) and enhanced case definition (based on epidemiologic classification of HPV), and unadjusted and age-adjusted OR and 95% CIs for comparisons between HIV (−) and HIV (+) women are shown in Table 2. Multiple HPV infections were observed in 15.1% (170 of 1,129) of the study sample. Both unadjusted and age-adjusted analyses revealed significantly higher prevalence of multiple infections in HIV (+) women (26.5%) compared with HIV (−) women (11.4%). When the HPV case-definition was enhanced using the epidemiologic classification, in unadjusted analyses, the prevalence of multiple infections involving both HR and LR infections was significantly higher in HIV (+) women (17.4%) compared with HIV (−) women (3.9%). After age-adjustment, the prevalence of multiple HR and multiple infections involving both HR and LR types was significantly higher in HIV (+) women [age-adjusted $OR_{\text{multiple HR}}$ = 2.98 (1.70-5.25); and age-adjusted $OR_{\text{multiple HR and LR}}$ = 10.08 (5.76-17.65)]. No significant differences were observed between HIV (−) and HIV (+) women when the prevalence of single HR and single LR infections were considered separately.

Table 2 also shows the frequency distribution of the number of coinfecting genotypes detected in the entire study sample, HIV (−), and HIV (+) women. The distribution of number of coinfecting genotypes was significantly different between the HIV (−) and HIV (+) women ($P < 0.05$). In 854 HIV (−) women 442 genotypes were detected; in 275 HIV (+) women, 282 genotypes were detected. The mean number of HPV infections was significantly higher in HIV (+) women (mean, 1.02; SD, 1.45) compared with HIV (−) women (mean, 0.51; SD, 0.86; $P <$ 0.05).

Comparison of Observed Number of Coinfecting Genotypes with Poisson Expected Frequencies. In the overall study sample, the observed frequencies significantly deviated from Poisson expectations ($P < 0.001$; Table 3). The observed number of zero infections was significantly higher than those expected. The ratio of observed and expected frequencies for infection with one and two HPV genotypes indicated that the observed counts were significantly less than expected. For infection with three and four or more genotypes, the observed counts were significantly higher than expected.

In both the HIV (−) and HIV (+) women, the observed frequencies significantly deviated from expectations derived from a random Poisson distribution ($P < 0.001$). In HIV (−) women, the observed number of null infections was significantly more than those expected. The observed number of single infections was significantly less than expected, and the observed numbers for three- and four-genotype infections

Table 2. Prevalence of multiple HPV infections in the study sample, HIV (−), and HIV (+) women (*n* = 1,129)

HPV	Overall, *n* = 1,129, *n* (%)	HIV (−), *n* = 854, *n* (%)	HIV (+), *n* = 275, *n* (%)	Unadjusted OR (95% CI)	Age-adjusted OR (95% CI)
Multiple HPV					
Negative	959 (84.9)	757 (88.6)	202 (73.5)	1.00	1.00
Positive	170 (15.1)	97 (11.4)	73 (26.5)	**2.82 (2.01-3.96)**	**4.87 (3.25-7.29)**
No HPV	702 (62.2)	561 (65.7)	141 (51.3)	1.00	1.00
Single HR	203 (18.0)	156 (18.3)	47 (17.1)	1.19 (0.82-1.74)	1.45 (0.97-2.16)
Single LR	54 (4.8)	40 (4.7)	14 (5.1)	1.39 (0.73-2.63)	1.96 (0.98-3.90)
Multiple HR	88 (7.8)	63 (7.4)	25 (9.1)	1.57 (0.95-2.60)	**2.98 (1.70-5.25)**
Multiple LR+					
Multiple HR and LR	82 (7.2)	34 (3.9)	48 (17.4)	**5.61 (3.48-9.04)**	**10.08 (5.76-17.65)**
Number of HPV genotypes	Overall *n* (n)*	HIV (−) *n* (n)*	HIV (+) *n* (n)*	*P*	
Mean (SD)	0.64 (1.06)	0.51 (0.86)	1.02 (1.45)	<0.001	
0	702 (723)	561 (574)	141 (149)	<0.001	
1	257 (253)	196 (189)	61 (64)		
2	96 (100)	59 (61)	37 (39)		
3	41 (30)	25 (23)	16 (7)		
≥4	33 (23)	13 (7)	20 (16)		

NOTE: Values in boldface are statistically significant at $P < 0.05$.
Abbreviations: HR, high-risk HPV; LR, low-risk HPV.
*n, the frequency distribution after exclusion of clades A1, A4, A8, A11 and HPV54.

were significantly more than expected. Similarly, in HIV (+) women, the observed number of null infections was significantly more than expected. Observed numbers for single and double infections were significantly less than expected and the number of four or more genotype coinfections were significantly more than expected.

Relative Frequency of HPV Clades in Single and Multiple Infections. Table 4 presents the distribution of the clades in single and multiple infections and unadjusted and adjusted comparisons for detecting differences in the proportions involved in multiple infections for each clade when compared with all other clades grouped together. The proportions for involvement in multiple infections ranged from 55.2% (clade A9) to 67.2% (clade A5). In unadjusted analyses, no significant differences were observed for the proportion of multiple infections for clades A3, A5, A6, A7, and A10. The proportion of multiple infections for clade A9 was significantly different from all other clades considered together (55.2% versus 64.6%; OR, 0.67; 95% CI, 0.48-0.93), indicating that clade A9 was significantly less likely to be involved in multiple infections. Of the genotypes constituting clade A9, the lowest proportion of multiple infections was observed for HPV 16 (43.6%) and the highest proportion was observed for HPV 33 (85.7%).

The proportions of both multiple infections and relative frequency of each clade may significantly differ across age, HIV status, and Pap smear status. Therefore, adjusted OR and 95% CI were calculated for each clade adjusting simultaneously for age, HIV status, and Pap smear status from six different binary logistic regression models. Postadjustment, the ORs for each clade did not differ significantly, and clades A3, A5, A6, A7, and A10 remained statistically insignificant. Clade A9 was significantly less likely to be involved in multiple infections (adjusted OR, 0.68; 95% CI, 0.48-0.95).

In order to further investigate if HPV 16 was driving this result for clade A9, sensitivity analyses were done re-analyzing the data excluding HPV 16. Clade A9 did not retain statistical significance in either unadjusted or adjusted analyses (OR_{unadj}, 0.83; 95% CI, 0.56-1.22 and OR_{adj}, 0.82; 95% CI, 0.57-1.19). Sensitivity analyses were also done to investigate if the exclusion of HPV types 42, 57, 40, 73, and 54 explained the significantly lower involvement in multiple infections for clade A9. Both unadjusted and adjusted results from this sensitivity analyses were very similar to analyses excluding the aforementioned types. In regression models for each clade, age, HIV status, and Pap smear status were all significantly associated with the outcome of multiple infections. Older age was significantly associated with decreased multiple infections and HIV (+) status and abnormal Pap smear status were associated with significantly increased multiple infections (data not shown).

Clustering of HPV Clades in Multiple Infections. The observed and expected frequencies, absolute differences for each HPV clade coinfection pattern in two-genotype infections, and deviation score for all 21 patterns of two genotype infections in the entire study sample, across HIV strata, and

Table 3. Comparison of observed numbers of coinfections with Poisson expected frequencies (*n* = 1,129)

Number of Infections	Overall		HIV (−)		HIV (+)	
	Observed (expected)	O/E (95% CI)*	Observed (expected)	O/E (95% CI)*	Observed (expected)	O/E (95% CI)*
0	702 (594.53)	**1.18 (1.09-1.27)**	561 (508.95)	**1.10 (1.01-1.19)**	141 (98.62)	**1.42 (1.21-1.68)**
1	257 (381.27)	**0.67 (0.59-0.76)**	196 (263.41)	**0.74 (0.64-0.85)**	61 (101.13)	**0.60 (0.49-0.77)**
2	96 (122.25)	**0.78 (0.64-0.95)**	59 (68.19)	0.86 (0.67-1.11)	37 (51.85)	**0.71 (0.52-0.98)**
3	41 (26.13)	**1.56 (1.15-2.13)**	25 (11.75)	**2.12 (1.43-3.14)**	16 (17.72)	0.90 (0.55-1.47)
≥4	33 (4.77)	**6.91 (4.76-9.71)**†	13 (1.67)	**7.78 (4.14-13.31)**†	20 (5.64)	**3.54 (2.16-5.47)**†
Total	1,129 (1,129)		854 (854)		275 (275)	
χ^2 (4 *df*)	786.836		115.630	—	589.277	
P	$P < 0.001$		$P < 0.001$		$P < 0.001$	

NOTE: Values in boldface are statistically significant at $P < 0.05$.
*Poisson 95% confidence intervals.
†Exact Poisson 95% confidence intervals.

Table 4. Relative frequencies of HPV clades and genotypes in single and multiple infections (n = 647)

HPV clade/type	Single infections n (%)	Multiple Infections n (%)	Unadjusted OR (95% CI)	Adjusted OR (95% CI)*
Clade A3	24 (39.3)	37 (60.7)	0.98 (0.57-1.69)	0.98 (0.55-1.75)
83	14 (35.9)	25 (64.1)		
84	10 (45.5)	12 (54.5)		
Clade A5	19 (32.8)	39 (67.2)	1.35 (0.76-2.39)	1.34 (0.74-2.42)
26	4 (50.0)	4 (50.0)		
51	14 (34.1)	27 (65.9)		
82	1 (11.1)	8 (88.9)		
Clade A6	28 (33.0)	55 (67.0)	1.30 (0.80-2.11)	1.15 (0.69-1.92)
53	14 (32.6)	29 (67.4)		
56	4 (30.8)	9 (69.2)		
66	10 (37.0)	17 (63.0)		
Clade A7	53 (34.9)	99 (65.1)	1.26 (0.86-1.85)	1.36 (0.91-2.02)
18	16 (40.0)	24 (60.0)		
39	8 (47.1)	9 (52.9)		
45	13 (41.9)	18 (58.1)		
59	10 (27.8)	26 (72.2)		
68	6 (21.4)	22 (78.6)		
Clade A9	113 (44.8)	139 (55.2)	**0.67 (0.48-0.93)**	**0.68 (0.48-0.95)**
16	44 (56.4)	34 (43.6)		
31	9 (50.0)	9 (50.0)		
33	2 (14.3)	12 (85.7)		
35	18 (45.0)	22 (55.0)		
52	24 (41.4)	34 (58.6)		
58	16 (36.4)	28 (63.6)		
Clade A10	16 (39.0)	25 (61.0)	1.00 (0.52-1.91)	0.96 (0.48-1.90)
6	6 (40.0)	9 (60.0)		
11	2 (28.6)	5 (71.4)		
55	8 (42.1)	11 (57.9)		

NOTE: Values in boldface are statistically significant at $P < 0.05$.
*Calculated from binary logistic regression models simultaneously adjusting for clade, age, HIV status, and Pap smear status.

across age strata are presented in Table 5. In the entire study sample, results from 10,000 simulations revealed no significant differences between the observed frequencies and theoretical expected frequencies (deviation score = 34.56; exact P = 0.29). Because no significant differences were observed, no further statistical tests were computed for comparison of observed and expected frequencies for each combinatorial pattern. As may be noted from Table 5, the observed and theoretical expected frequencies were very similar.

In order to assess the confounding effects of age, HIV status, and Pap smear status on conformity of observed frequencies to theoretical expected frequencies, similar computations were done across strata formed by age, HIV status, and Pap smear status. The total number of genotypes detected in each stratum

Table 5. Comparison of observed and expected frequencies in two genotype infections

Pattern of coinfection	Overall, observed (expected)	Overall [O-E]*	HIV (−) Observed (expected)	HIV (−) [O-E]*	HIV (+) Observed (expected)	HIV (+) [O-E]*	≤25 years Observed (expected)	≤25 [O-E]*	>25 years observed (expected)	>25 [O-E]*
A3-A3	0 (0.43)	0.43	0 (0.17)	0.17	0 (0.29)	0.29	0 (0.11)	0.11	0 (0.32)	0.32
A5-A5	1 (0.40)	0.60	1 (0.21)	0.79	0 (0.15)	0.15	1 (0.78)	0.78	0 (0.16)	0.16
A6-A6	2 (1.04)	0.96	1 (0.44)	0.56	1 (0.65)	0.35	1 (0.42)	0.58	1 (0.56)	0.44
A7-A7	6 (4.63)	1.37	5 (2.95)	2.05	1 (1.64)	0.64	4 (2.44)	1.56	2 (2.15)	0.15
A9-A9	14 (12.68)	1.31	12 (9.58)	2.42	2 (4.24)	2.24	11 (8.00)	3.00	3 (5.04)	2.04
A10-A10	0 (0.27)	0.27	0 (0.12)	0.12	0 (0.15)	0.15	0 (0.09)	0.09	0 (0.15)	0.15
A3-A5	2 (1.80)	0.20	0 (0.89)	0.89	2 (0.91)	1.09	0 (0.69)	0.69	2 (1.03)	0.97
A3-A6	3 (2.56)	0.44	0 (1.0)	1.00	3 (1.81)	1.19	0 (0.78)	0.78	3 (1.81)	1.19
A3-A7	2 (4.70)	2.70	1 (2.3)	1.3	1 (2.45)	1.45	0 (1.69)	1.69	2 (2.92)	0.92
A3-A9	6 (7.79)	1.79	3 (4.15)	1.15	3 (3.80)	0.80	0 (3.08)	3.08	6 (4.41)	1.59
A3-A10	1 (1.28)	0.28	0 (0.54)	0.54	1 (0.81)	0.19	0 (0.39)	0.39	1 (0.89)	0.11
A5-A6	1 (2.45)	1.45	0 (1.27)	1.27	1 (1.17)	0.17	1 (1.18)	0.18	0 (1.24)	1.24
A5-A7	6 (4.50)	1.50	4 (2.90)	1.10	2 (1.58)	0.42	4 (2.54)	1.46	2 (2.0)	0.00
A5-A9	10 (7.45)	2.55	9 (5.25)	3.75	1 (2.45)	1.45	8 (4.62)	3.37	2 (3.01)	1.01
A5-A10	0 (1.23)	1.23	0 (0.68)	0.68	0 (0.52)	0.52	0 (0.59)	0.59	0 (0.61)	0.61
A6-A7	8 (6.40)	1.60	3 (3.29)	0.29	5 (3.16)	1.83	3 (2.87)	0.13	5 (3.50)	1.50
A6-A9	7 (10.60)	3.60	4 (5.94)	1.94	3 (4.90)	1.90	4 (5.24)	1.24	3 (5.27)	2.27
A6-A10	6 (1.75)	4.25	3 (0.77)	2.23	3 (1.05)	1.95	0 (0.67)	0.67	6 (1.06)	4.93
A7-A9	22 (19.48)	2.52	14 (13.54)	0.46	8 (6.62)	1.37	12 (11.25)	0.74	10 (8.5)	1.5
A7-A10	1 (3.20)	2.20	0 (1.76)	1.76	1 (1.42)	0.42	1 (1.43)	0.43	0 (1.71)	1.71
A9-A10	2 (5.30)	3.30	1 (3.18)	2.18	1 (2.20)	1.20	1 (2.62)	1.62	1 (2.58)	1.58
Total	100 (100)		61 (61)		39 (39)		51 (51)		49 (49)	
Deviation score		34.56		26.61		19.80		23.21		24.41
Exact P †		0.29		0.23		0.72		0.33		0.29

*Absolute difference between observed and expected frequencies.
†Exact P calculated from 10,000 Monte Carlo simulations.

constituted the denominator for the purpose of calculating the relative frequency of each clade. No significant differences were observed between the observed and expected frequencies either in HIV (−) women (deviation score = 26.61; exact P = 0.23) or HIV (+) women (deviation score = 19.80; exact P = 0.72). No significant differences were observed between the observed and expected frequencies either in normal Pap smears (deviation score = 12.98; exact P = 0.85) or abnormal Pap smears (deviation score = 29.21; exact P = 0.30; data not shown).

Comparison of observed and expected frequencies in women ≤25 years of age (n = 51) and women >25 years of age (n = 49) are also presented in Table 5. In younger women, no significant differences were observed between the observed frequencies and expected frequencies (deviation score = 23.21; exact P = 0.33). Similarly, in women older than 25 years of age, no significant differences were observed between the observed and expected frequencies (deviation score = 24.41; exact P = 0.29).

Discussion

In this study, multiple infections were significantly higher in HIV (+) women compared with HIV (−) women. These results are consistent with several previous reports of increased prevalence of multiple infections in HIV (+) women (3, 4, 17). This increased prevalence in HIV (+) women may be attributed to the common mode of transmission of HPV and HIV, and/or reactivation of latent HPV infections owing to CD4 T cell depletion.

The frequency distribution of the number of coinfections in the entire study sample, HIV (+), and HIV (−) women did not conform to Poisson expected frequencies. The lack of conformity with a Poisson distribution was not surprising because HPV infections share common risk factors and a common mode of transmission (2). Additionally, previous follow-up studies have reported that women testing positive for any HPV type at baseline were significantly more likely to acquire another HPV infection (1, 8, 10). These results indicate that the distribution of the number of HPV genotypes is clumped and that when infected with at least one type, women were more likely to acquire other HPV infections.

Compared with all other clades grouped together, clade A9 (constituted by HPV types 16, 31, 33, 35, 52, and 58) was significantly less likely to be involved in multiple infections. In sensitivity analyses excluding HPV 16, clade A9 failed to retain statistical significance. It must however be noted that with HPV 16 being the most prevalent genotype, excluding it may have resulted in a loss of power to detect significant differences. The reasons for the observation of significantly lower involvement of clade A9 in multiple infections are unclear. It is possible that the risk factor and transmission patterns may be different for different HPV clades. Indeed, epidemiologic studies using a risk-based classification of HPV infection (as high-risk or low-risk) have revealed significant differences in risk factor profiles and transmission patterns between infection with high-risk HPV and low-risk HPV (18, 19). Admittedly, these estimates were not adjusted for classic risk factors of HPV infection such as lifetime and recent sexual behavior, and age at sexual debut (20). It remains to be seen if these results would arise from other studies when a similar analysis is done and adjustments are made for the range of confounders.

Few studies have addressed the issue of clustering of HPV either at the phylogenetic clade level or the genotype level. Liaw et al., from the Portland-Kaiser Permanente cohort reported that preexisting HPV 16 infection was generally associated with an increased risk for subsequent acquisition of other HPV genotypes (both phylogenetically related and unrelated; ref. 1). Using a cohort of college-aged women in Seattle, Thomas et al. reported that concurrent acquisition of multiple HPV types occurred more often than expected, and that no two HPV genotypes (among types 6, 11, 16, 18, 31, and 45) were more or less likely to be acquired concurrently than any two other types (10). Pertaining to sequential acquisition, this group reported that the risk of acquisition of a new type was not decreased for either phylogenetically related or unrelated genotypes (10). Similarly, Rousseau et al. from the Ludwig-McGill cohort concluded that the presence of preexisting HPV infection increased the risk for acquisition of other HPV genotypes (8). Franco et al., reported that observed counts were significantly higher than expected for coinfections involving type 16 with types 18, 31, 56, and MM8; type 31 with types 51 and 53; type 45 with 51; and type 53 with 45, 51, 52 and 58 (2). Rousseau et al., in an analysis restricted to women positive for at least one HPV type, reported that observed counts were significantly less than expected for HPV type 16 with 18 and types 16 and 18 with 6 or 11 (9).

Several issues regarding the unit of analysis and computational methods used in the current study need discussion. In all the previous studies, the study subject was used as the unit of statistical analysis, whereas in this study, the HPV clade was used as the unit of analysis. Although the HPV clade was used as the unit of analysis for the calculation of expected frequencies, it must be noted that the unit of analyses becomes the study subject for the goodness-of-fit test. This is because observed frequencies (pairs of infections) used in the goodness-of-fit test are derived from study subjects (number of study subjects with a particular pattern of coinfection). Consistent with the goal of characterizing the equilibrium that exists among the various HPV types in multiple (two-genotype) infections, all clustering analyses were restricted to n = 100 subjects with two-genotype infections. In effect, instead of a standard Poisson sampling analysis, a conditioned Poisson (same as a multinomial) approach was used (21). The use of relative frequencies of clades from the n = 100 two-genotype infections may introduce a bias. This bias arises from the fact that any genotypes/clades that may have led to cross-protection or competitive exclusion of other genotypes may occur more often in single infections. Therefore, relative frequencies from the entire study sample or the pertinent stratum were used to calculate the expected frequency.

Several disadvantages of the multinomial expansion method need special mention. As the number of genotypes involved in multiple infections increases, the number of unique patterns derived from expanding the multinomial expressions also increases. Thus, greater sample sizes are required for detecting any differences between observed and expected frequencies for infections with three or more genotypes. Another disadvantage of the multinomial method is its ability to adjust for potential confounders. Similar to Mantel-Haenszel methods in classical epidemiology, this method of assessment of confounding is not amenable to continuous confounders and larger sample sizes are required for the assessment of statistical interactions (22).

For multiple infections involving two HPV types, no significant differences were detected between the observed frequencies and the expected frequencies for the 21 different combinatorial patterns in the entire study group, across HIV strata, across Pap smear strata, and across age strata. These results indicate that HPV types involved in multiple infections are established at random and that the relative frequency of a particular clade/genotype seems the only determinant of being involved in multiple infections. These results must be interpreted with caution because they were based on sparse sample sizes and need to be confirmed in larger studies. Additionally, the grouping of HPV types at the phylogenetic clade level may not be biologically appropriate and may have masked any differences in clustering that exist at the HPV type level.

Several limitations of the current study mandate discussion. The current study used a cross-sectional design, with a convenience sampling scheme. The cross-sectional nature of the study precluded an assessment of persistence of single/multiple infections. Implicit in the interpretation of the results is the assumption that the observed multiple infections are established infections. In this study, 41% of the HIV (−) and 45% of the HIV (+) women presented with cytologic abnormalities. The corresponding numbers in women positive for any detectable HPV type were 67.2% and 59.0%, respectively. These observations, coupled with research demonstrating increased persistence of HPV in HIV (+) women, indicate that at least a majority, if not all, of the infections detected in the current study may represent established infections. The recruitment of HIV-positive and -negative women from different sites may have resulted in selection bias arising from differences in demographic/behavioral correlates of HPV infection and differences in genotypic prevalence.

Another limitation of the study was the lack of adjustment for important confounders such as age at sexual debut and correlates of lifetime and recent sexual behavior. The majority of analyses presented in the current study were restricted to women positive for at least one HPV type. A recent study showed several similarities in risk factor profiles for infection with single and multiple types (23). Therefore, the impact of residual confounding on the current results may have been minimal, however, confounding as an explanation for the current observations may not be completely ruled out.

In conclusion, results from this study indicate that clade A9 is significantly less likely to be involved in coinfections compared with all other clades. Results also indicate that HPV types predominantly establish multiple infections at random, with little positive or negative clustering for either phylogenetically related or unrelated types. These results may have important implications for vaccine design strategies. The reasons behind clade A9 being less involved in multiple infections need to be further investigated. Although clade A9 was less likely to be involved in multiple infections, no systematic patterns of selective inclusion or exclusion were observed. The observation that the relative frequency of a clade is the only determinant of being involved in coinfections is reassuring and indicates that type-specific prophylactic vaccination may have little impact on untargeted genotypes. Results from this study need further validation in larger studies with prospective follow-up.

Acknowledgments

We thank Dr. Janet Kornegay and Roche Molecular Systems for graciously donating the reagents for HPV typing; Jeanne Dumestre and Tracy Beckel for their help in specimen collection; Drs. Jennifer E. Cameron for critical reading of the manuscript; and Abhijit Dasgupta and other investigators at the U.S. National Cancer Institute for valuable suggestions.

References

1. Liaw KL, Hildesheim A, Burk RD, et al. A prospective study of human papillomavirus (HPV) type 16 DNA detection by polymerase chain reaction and its association with acquisition and persistence of other HPV types. J Infect Dis 2001;183:8–15.
2. Franco EL, Villa LL, Sobrinho JP, et al. Epidemiology of acquisition and clearance of cervical human papillomavirus infection in women from a high-risk area for cervical cancer. J Infect Dis 1999;180:1415–23.
3. Palefsky JM, Minkoff H, Kalish LA, et al. Cervicovaginal human papillomavirus infection in human immunodeficiency virus-1 (HIV)-positive and high-risk HIV-negative women. J Natl Cancer Inst 1999;91:226–36.
4. Levi JE, Kleter B, Quint WG, et al. High prevalence of human papillomavirus (HPV) infections and high frequency of multiple HPV genotypes in human immunodeficiency virus-infected women in Brazil. J Clin Microbiol 2002;40:3341–5.
5. Koutsky LA, Ault KA, Wheeler CM, et al. A controlled trial of a human papillomavirus type 16 vaccine. N Engl J Med 2002;347:1645–51.
6. Harper DM, Franco EL, Wheeler C, et al. Efficacy of a bivalent L1 virus-like particle vaccine in prevention of infection with human papillomavirus types 16 and 18 in young women: a randomised controlled trial. Lancet 2004;364:1757–65.
7. Carter JJ, Galloway DA. Humoral immune response to human papillomavirus infection. Clin Dermatol 1997;15:249–59.
8. Rousseau MC, Pereira JS, Prado JC, et al. Cervical coinfection with human papillomavirus (HPV) types as a predictor of acquisition and persistence of HPV infection. J Infect Dis 2001;184:1508–17.
9. Rousseau MC, Villa LL, Costa MC, et al. Occurrence of cervical infection with multiple human papillomavirus types is associated with age and cytologic abnormalities. Sex Transm Dis 2003;30:581–7.
10. Thomas KK, Hughes JP, Kuypers JM, et al. Concurrent and sequential acquisition of different genital human papillomavirus types. J Infect Dis 2000;182:1097–102.
11. Li CC. First course in population genetics. Pacific Grove, (CA.): The Boxwood Press; 1975.
12. Chaturvedi AK, Dumestre J, Gaffga AM, et al. Prevalence of human papillomavirus genotypes in women from three clinical settings. J Med Virol 2005;75:105–13.
13. Gravitt PE, Peyton CL, Alessi TQ, et al. Improved amplification of genital human papillomaviruses. J Clin Microbiol 2000;38:357–61.
14. The 1988 Bethesda System for reporting cervical/vaginal cytological diagnoses. National Cancer Institute Workshop. JAMA 1989;262:931–4.
15. Koutsky L. Epidemiology of genital human papillomavirus infection. Am J Med 1997;102:3–8.
16. de Villiers EM. Papillomavirus and HPV typing. Clin Dermatol 1997;15:199–206.
17. Levi JE, Fernandes S, Tateno AF, et al. Presence of multiple human papillomavirus types in cervical samples from HIV-infected women. Gynecol Oncol 2004;92:225–31.
18. Franco EL, Villa LL, Ruiz A, Costa MC. Transmission of cervical human papillomavirus infection by sexual activity: differences between low and high oncogenic risk types. J Infect Dis 1995;172:756–63.
19. Kjaer SK, van den Brule AJ, Bock JE, et al. Determinants for genital human papillomavirus (HPV) infection in 1000 randomly chosen young Danish women with normal Pap smear: are there different risk profiles for oncogenic and nononcogenic HPV types? Cancer Epidemiol Biomarkers Prev 1997;6:799–805.
20. Peyton CL, Gravitt PE, Hunt WC, et al. Determinants of genital human papillomavirus detection in a US population. J Infect Dis 2001;183:1554–64.
21. Agresti A. Categorical data analysis. 2nd ed. New York: Wiley; 2002.
22. Rothman KJ, Greenland S. Modern epidemiology. 2nd ed. Philadelphia: (Lippincott): Williams and Wilkins; 1998.
23. Rousseau MC, Abrahamowicz M, Villa LL, et al. Predictors of cervical coinfection with multiple human papillomavirus types. Cancer Epidemiol Biomarkers Prev 2003;12:1029–37.