Journal of Medical Virology 88:1279–1287 (2016)

Genetic Diversity of HPV16 and HPV18 in Brazilian Patients With Invasive Cervical Cancer

Joao Paulo C.B. Vidal,[1] Shayany Pinto Felix,[1] Cláudia B.P. Chaves,[1] Patrícia Patury,[2] Vanessa F. Franco,[2] Evaneide A. de Morais,[2] Neile A. de Carvalho,[3] Aurenice C.L. Carvalho,[3] Olimpio F. Almeida Neto,[2] Lina Maria T.M. Vieira,[2] Flavia Miranda Correa,[2] Luís Felipe Leite Martins,[3] Antonio Negrão,[3] Liz Maria de Almeida,[3] and Miguel Angelo Martins Moreira[1]*

[1]*Genetics Program, National Cancer Institute José de Alencar Gomes da Silva, Brazil*
[2]*Gynecologic Oncology Department, Cancer Hospital II, National Cancer Institute José de Alencar Gomes da Silva, Brazil*
[3]*Populational Research Department, National Cancer Institute José de Alencar Gomes da Silva, Brazil*

Cervical cancer is the fourth most common cancer among women, and ~70–80% of these cancers are associated with two human papillomavirus types: HPV16 and HPV18. Several studies have reported that intra-type diversity is associated with the progression of infection to invasive cancer. Herein, we report the genetic diversity of HPV16 and HPV18 in a cohort of 594 Brazilian women with invasive cervical cancer and describe the prevalence of lineages and intra-type diversity prior to the implementation of the public immunization program in Brazil. HPV detection and genotyping were performed using PCR, PGMY/GP primers, and DNA extracted from fresh tumors. The HPV16 (378 women) and HPV18 (80 women) lineages were identified by PCR and sequencing of the *LCR* and *E6* fragments, followed by SNV comparison and phylogenetic analysis. In our cohort, was found a higher frequency of the lineage A (in 217 women), followed by lineage D (in 97 women) and lineages B and C (in 10 women each) for HPV16; and a higher frequency of lineage A (in 56 women) followed by lineage B (in 15 women) in HPV18. The genetic diversity of HPV16 indicated a recent expansion of specific variants or a selective advantage that is associated with invasive cancer; this pattern was not observed for HPV18. ***J. Med. Virol. 88:1279–1287, 2016***.

KEY WORDS: HPV genetic diversity; human papillomavirus type 16; human papillomavirus type 18; HPV lineages

INTRODUCTION

Cervical cancer (CC) is the fourth most common cancer among women worldwide; 527,624 new cases occur each year, and this cancer exhibits a mortality rate of 265,672/year. In developing countries, this cancer ranks second after breast cancer in women (Globocan 2012, IARC, http://globocan.iarc.fr). In Brazil, ~15,590 new cases of invasive CC were expected in 2014, with an estimated incidence of 15.33 cases per 100,000 women (INCA, 2014).

Human papillomavirus (Family: *Papillomaviridae*) (HPV) presence is considered necessary but not sufficient for cervical cancer development [Bosch et al., 2002]. Papillomaviruses are classified based on *L1* gene divergence. More than 200 HPV genomes have been completely sequenced (Human Papillomavirus Reference Center, http://www.hpvcenter.se/html/refclones.html), and papillomavirus type is defined by sequence differences in the L1 region of greater than 10% [De Villiers, 2013]. Burk et al. [2013] proposed a classification and a nomenclature (adopted herein) for intratype diversity based on the divergence of complete HPV genome sequences into

Grant sponsor: INCT for Cancer Control; Grant sponsor: Pan-American Health Organization (PAHO/OMS); Grant sponsor: Ministry of Health; Grant sponsor: The National Council for Scientific and Technological Development (CNPq, Brazil); Grant sponsor: Carlos Chagas Filho Research Support Foundation (FAPERJ, Brazil)

Conflicts of interest: None.

*Correspondence to: Miguel Angelo Martins Moreira, Genetics Program, National Cancer Institute José de Alencar Gomes da Silva, André Cavalcanti 37, Rio de Janeiro, RJ, Brazil, 20231-050 E-mail: miguelm@inca.gov.br

Accepted 19 December 2015

DOI 10.1002/jmv.24458
Published online 12 January 2016 in Wiley Online Library (wileyonlinelibrary.com).

lineages (with 1–10% divergence) and sublineages (with 0.5–1% divergence).

An epidemiological classification has been proposed based on the association of different HPV types with the risk of cervical cancer development. This classification groups HPVs as follows: (i) oncogenic (or high risk HPV), including types that are more frequently associated with cancer; (ii) probably/possibly oncogenic (or low risk HPV), including types that are more frequently associated with benign epithelial proliferations in the genital area and less frequently associated with cancer; and (iii) undefined risk HPV, including viruses with incomplete epidemiological data for defining risk with respect to association with cervical cancer [Muñoz et al., 2003]. Fifteen genotypes are classified as oncogenic. Among these types, HPV16 *(*Family: *Pappilomaviridae*; Genus: *Alphapapillomavirus*; Species: 7*)* and HPV18 *(*Family: *Pappilomaviridae*; Genus: *Alphapapillomavirus*; Species: 9*)* are found in ~70% of all cervical cancer cases worldwide [Bosch et al., 2008]. Oncogenic HPV types 16 and 18 are highly prevalent not only in cancer patients but also in women with normal cervical cytology and cervical lesions [Bruni et al., 2010].

A small proportion of oncogenic infections are persistent and progress to invasive cancer; however, the factors associated with progression are poorly understood. One important hypothesis suggested by several studies relates to intra-type diversity [Sichero et al., 2007; Gheit et al., 2011; Smith et al., 2011; Arroyo et al., 2012; Cornet et al., 2013; Chen et al., 2014]. These studies suggested that intra-type lineages differ in their biological characteristics and pathogenicity. For example, the sublineage B2 of HPV45 is associated with a higher risk for cervical cancer development than other sublineages [Chen et al., 2014].

Previous studies have suggested the existence of four major lineages of HPV16 and HPV18, which were initially named based on their frequencies in different human populations [Ho et al., 1993; Ong et al., 1993; Yamada et al., 1997]. Burk et al. [2013] modified this nomenclature by grouping HPV16 variants into four major lineages: A, formerly named the European and Asian (EAS) lineage; B, formerly named the African 1 (AF1) lineage; C, formerly named the African 2 (AF2) lineage; and D, encompassing the formerly named North-American and Asian-American (NA and AA, respectively) lineages. Burk et al. [2013] grouped HPV18 variants into three lineages: A, encompassing the previously named Asian-American and European lineages; and B, and C (both including the formerly African lineage, Af).

Few Brazilian studies have studied HPV16 and HPV18 intra-type variants [Villa et al., 2000; Cruz et al., 2004; Sichero et al., 2007; Cerqueira et al., 2008; Junes-Gill et al., 2008; Sichero et al., 2012; Freitas et al., 2014], and almost no studies have investigated samples obtained from invasive cancer [Junes-Gill et al., 2008]. Interest in HPV16 and HPV18 intra-type diversity has been increasing in recent years due to reports suggesting an interaction between host and viral characteristics influencing the natural history of infection and the prognosis of lesions at different stages [de Araujo Souza et al., 2008, 2009; Lopera et al., 2014].

It is important to underline that the papillomaviruses are a double-strand DNA viruses with a lower mutation rate when compared to RNA viruses, which makes the occurrence of vaccine escape mutations improbable [Schiller and Lowy, 2012]. Additionally, Harari et al. [2015], in a study carried out in Costa Rica, did not found differences in the bivalent vaccine efficacy in respect to different lineages of HPV16 and HPV18. However, the same authors ponder that this conclusion did not take into account the presence of rare variants and the HPV diversity found in other populations. Based on this perspective and the advent of prophylactic vaccines against HPV16/18 and cervical cancer, it has become necessary to identify the HPV genetic variants that are present in populations of interest to verify in the coming years if vaccination will be effective against all lineages and variants of HPV16 and HPV18. This is particularly true in developing countries, where cervical cancer is a serious public health problem, and limited data are available regarding HPV genetic intra-type diversity.

Here, we studied the genetic diversity of HPV16 and HPV18 associated with invasive cervical cancer. We describe the prevalence of lineages and intra-type diversity based on the genetic variability of two genomic regions: *E6* and the long control region (LCR). The study examined a cohort of 594 women who were diagnosed with cervical cancer and enrolled between July 2011 and April 2014 for treatment at the Instituto Nacional de Cancer (Brazil) before the implementation of the HPV vaccination program by the Brazilian Public Health System in March 2014.

METHODS

Subjects and Samples

This study was approved by the local Ethics Committee, and all participants signed an informed consent form. Patients included in this study comprised a cohort of 594 women who were diagnosed with invasive cervical cancer and enrolled consecutively at the Instituto Nacional de Cancer, Rio de Janeiro, Brazil, between July 2011 and March 2014. Of these women, 456 (76%) were identified as being infected with HPV16 and HPV18. In the HPV16+ group, 37% of the patients were residents of the City of Rio de Janeiro and 63% lived in adjacent regions within the state; the mean age of the patients was 48 years (range 19–93 years). In the HPV18+ group, 31.6% were residents of the City of Rio de Janeiro and 68.4% lived in adjacent regions; the mean age was 46 years (range 21–77 years).

Fresh tumor biopsies (stages 1B to 4) were stored in RNA-Later at −80°C. Total DNA was isolated using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany)

according to the manufacturer's instructions, eluted in 200 μl AE buffer, and stored at −80°C.

HPV Detection and Genotyping

HPV DNA presence was detected using the polymerase chain reaction (PCR) with the primer set PGMY 09/11 [Gravitt et al., 2000]. When this reaction yielded negative results, HPV DNA presence was detected using nested PCR with the primer sets PGMY09/11 (first round PCR) and GP5+/GP6+ (second round PCR, Fuessel Haws et al. [2004]); these primers amplify a segment of the *L1* gene. DNA samples from Caski and HeLa cells were used as positive controls. PCR products amplified using PGMY09/11 or GP5+/GP6+ were purified using the GFX PCR and DNA Band Purification kit (GE Healthcare), labeled using the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA), and analyzed using an ABI Prism 3130XL Genetic Analyzer (Applied Biosystems). All sequences were edited and analyzed using 4Peaks Software (Nucleobytes, Amsterdam, Netherlands). HPV genotypes were identified using the BLASTn (Basic Local Alignment Search Tool, http://blast.ncbi.nlm.nih.gov/).

LCR and *E6* Amplification and DNA Sequencing

To characterize the HPV16 and HPV18 lineages, the contiguous *LCR* and *E6* genomic regions of the HPV genome were analyzed using overlapping PCR products along a region of ~1,300 bp. DNA samples were subjected to PCR amplification of the *LCR* and *E6* regions using HPV type-specific primers for each region (Table I). PCR reactions were performed in 25 μl mixtures containing 1 × PCR buffer, 2.5 mM $MgCl_2$, 0.25 μM of each DNTP, 100 pmol/l of each primer, 50–100 ng of DNA, and 2.5 U of Platinum *Taq* DNA Polymerase (Life-Technologies). Amplifications were performed as follows: 95°C for 10 min; followed by 40 cycles of 1 min at 95°C, 1 min at the annealing temperature (Tm, see Table I), and 1 min at 72°C; followed by a final extension at 72°C for 15 min. The resulting PCR products were subsequently purified and sequenced as described above. The consensus sequences were assembled using Geneious software (Biomatters Ltd.), and all sequences were aligned with lineage-specific reference sequences [Burk et al., 2013] using MEGA (version 6: www.megasoftware.net). All sequences were deposited in GenBank under accession numbers KP965018–KP965162 for HPV16 and KP965163–KP965195 for HPV18.

Lineage Identification, Phylogenetic Analysis, and Haplotype Characterization

The HPV16 and HPV18 lineages were identified based on sequences obtained from high-quality *LCR* and *E6* electropherograms. Two strategies were employed. First, HPV16 lineages were identified based on single nucleotide variants (SNVs) at specific *LCR* and *E6* sites according to the method of Cornet et al. [2012], and HPV18 lineages were characterized by identifying lineage-distinctive SNVs [Arias-Pulido et al., 2005; Arroyo et al., 2012]. Second, a phylogenetic reconstruction was performed using a dataset of 1,300 bp *LCR* + *E6* sequences, including the reference sequences proposed by Burk et al. [2013]. Haplotype identification was performed using DNAsp version 5, maximum likelihood analysis was carried out using PHYML 3.0 [Guindon et al., 2010], and the general time reversible model GTR + I + G was inferred using the software package ModelGenerator v.0.85 [Keane et al., 2006]. Group support was estimated by bootstrap estimates based on 10,000 replicates. The reference sequences were as follows: K02718 (REF SEQ, lineage A), AF536179 (lineage A), HQ644236 (lineage A), AF534061 (lineage A), AF536180 (lineage B), HQ644298 (lineage B), AF472509 (lineage C), HQ644257 (lineage D), AY686579 (lineage D), and AF402678 (lineage D) for HPV16; and AY262282 (REF-SEQ, lineage A), EF202146 (lineage A), EF202147 (lineage A), EF202151 (lineage A), GQ180787 (lineage A), EF202152 (lineage B), EF202155 (lineage B), KC470225 (lineage B), and KC470229 (lineage C) for HPV18.

Median-joining (MJ) network topologies were constructed using NETWORK 4.6.1.1 [Bandelt et al., 1999]. Molecular diversity indices (haplotype diversity and nucleotide diversity) estimates and Fu's Fs test for selective neutrality were performed using Arlequin 3.5 [Excoffier and Lischer, 2010].

RESULTS

The study population comprised 594 women who were diagnosed with invasive cervical carcinoma and

TABLE I. Primers Used for PCR Amplification of *LCR* and *E6* Regions of HPV16 and HPV18, the Respective Annealing Temperatures, and Annealing Positions at the Genomic Reference Sequences

Primer	Sequence (5′-3′)	Tm	Ref. Seq	Genomic position
LCR F HPV 16	CACCCACCACCTCATCTACC	56°C	K02718.1 HPV16	7100–7120
LCR R HPV 16	CACACACCCATGTGCAGTTT			7835–7855
E6 F HPV 16	CACATATTTTTGGCTTGTT	50°C		7701–7720
E6 R HPV 16	GGAGATACACCTACATTGCATGAA			570–592
LCR F HPV 18	TCTAAACCTGCCAAGCGTGT	56°C	AY262282 HPV18	7095–7115
LCR R HPV 18	ATGTGATGCCCAACCTATTT			7825–7845
E6 F HPV 18	GTTGCCTTTGGCTTATGTCTG	56°C		7468–7488
E6 R HPV 18	TTGCCTTTAGGTCCATGCATAC			587–607

attending the Department of Gynecology of the Instituto Nacional de Cancer, Rio de Janeiro, Brazil, between June 2011 and March 2014. After the patients signed informed consent forms and answered an epidemiological questionnaire, tumoral biopsies were collected and identified with respect to HPV type using PCR and *L1* genomic region sequencing. HPV16 was the most frequent viral genotype and was identified in 378 samples (63.6%); the next most frequent genotype was HPV18 (80 samples, 13.4%). HPV45, HPV35, HPV58, HPV52, HPV73, HPV31, HPV33, HPV59, HPV39, HPV26, HPV51, HPV56, HPV68, and HPV83 were also identified in the cohort. Multiple infections were detected in 4% of cases, being 13 samples co-infected with both HPV16 and HPV18. These samples were not used in the present study and these data will be published elsewhere.

HPV16 Variant Characterization

To identify HPV16 variants, a 1,310 bp region of the HPV16 genome was sequenced based on two overlapping PCR products (from nucleotides 7,157 to 559) corresponding to *LCR* and *E6*. The *LCR* and *E6* regions were completely or partially sequenced in 334 samples. Coinfection by two HPV16 lineages was detected by the overlapping of sequence peaks at specific nucleotide positions (SNV signatures) in two samples (both co-infected with lineages A and D), and PCR amplification was unsuccessful for one or both regions in 42 samples, probably due to failed primer annealing or to a small proportion of HPV DNA in the sample. Data from these 44 samples were excluded from subsequent analyses.

Identification of the HPV16 lineages was based on nucleotide signatures resulting from single nucleotide variants (SNVs) at 46 specific genomic sites as proposed by Cornet et al. [2012]. Of the 334 samples, 306 fell within the nine 46-nucleotide signatures proposed by Cornet et al. [2012]. However, 28 samples were not strictly concordant with any nucleotide signature proposed by these authors and exhibited a minimal amount of discordance ranging from 1/46 to 6/46 positions. Despite these discordances, 334 samples were allocated to one of the four HPV16 major lineages based on the nucleotides present at the diagnostic positions and on the overall similarity with any of the nine signatures. Lineage A was the most frequent lineage (217 samples, or ~57% of the HPV16 samples), followed by lineages D (97 samples, or ~25%), and B and C (10 samples each, or ~5%) (Table II).

Patients sharing HPV16 haplotypes were identified by haplotyping 314 of the 334 samples. HPV16 sequences from 20 patients were excluded because data regarding the ends of *LCR* and/or *E6* were unreliable. A total of 125 haplotypes were identified (Supplemental Table SIa), including 96 variants present in single patients and 29 variants shared by at least two patients. The four most frequent haplotypes, which were present in approximately 40% of the samples, were Hap09 (from lineage A; 68 patients), Hap11 (from lineage D; 40 patients), Hap47 (from lineage A; 12 patients), and Hap04 (from lineage D; 11 patients) (Supplemental Table SIa).

TABLE II. Number of Samples Identified for Lineages of HPV16 and HPV18 in Respect to Tumor Histological Type

LINEAGES	SCC	ADC	Others	Total (%)
HPV16				
A	175	24	18	217 (57.4)
B	10	0	0	10 (2.6)
C	9	1	0	10 (2.6)
D	78	15	4	97 (25.5)
ND	31	7	6	45 (11.9)
HPV18				
A	33	17	5	55 (70)
B	7	5	4	16 (18.7)
ND	6	2	1	9 (11.3)

SCC, squamous cell carcinoma; ADC, adenocarcinoma; Others, other histological types; ND, lineages not determined.

A maximum likelihood (ML) topology tree was constructed based on the 125 haplotypes (Supplemental Fig. S1a) and the reference sequences for each lineage proposed by Burk et al. [2013]. Most haplotypes were grouped consistently with their lineage identification based on the SNV signatures proposed by Cornet et al. [2012]; the exceptions were the haplotypes allocated in lineages B and C. These 20 haplotypes do not formed two monophyletic groups respective to lineages B and C.

The network arrangement (Fig. 1) was also in agreement with the ML topology and SNV signatures and showed that the three most frequent haplotypes occupied a central position with a star-like pattern in lineages A (Hap09 and Hap47) and D (Hap11). Most of the other, less frequent haplotypes diverged from these central haplotypes by one and two nucleotide substitutions. Haplotypes from lineages B and C did not exhibit a star-like pattern, indicating a larger divergence between haplotypes. The molecular diversity indices (haplotype diversity and nucleotide diversity) and Fu's Fs [Fu, 1997] test based on 314 samples are shown in Table III. The Fu's Fs test (-23.85, $P = 0.004$) suggests that the HPV16 viral population associated with invasive cancer had undergone a recent population expansion or that specific haplotypes were subject to positive selection.

E6 contained 19 amino acid changes with respect to the prototype sequence (Supplemental Table SIc), seven of which have previously been described. The twelve novel amino acid substitutions were D25G, D25A, E29A, D56Y, D64N, S71C, D98H, H126P, N127K, R131L, S138F, and R144T. The variant L83V of HPV16 *E6* was shared by all haplotypes of lineage D and by 31 haplotypes of lineage A (14%).

HPV18 Variant Characterization

Of the 80 HPV18 samples, nine were excluded from the analyses due to the unsuccessful PCR amplification of *LCR* and/or *E6*. For the remaining 71 samples,

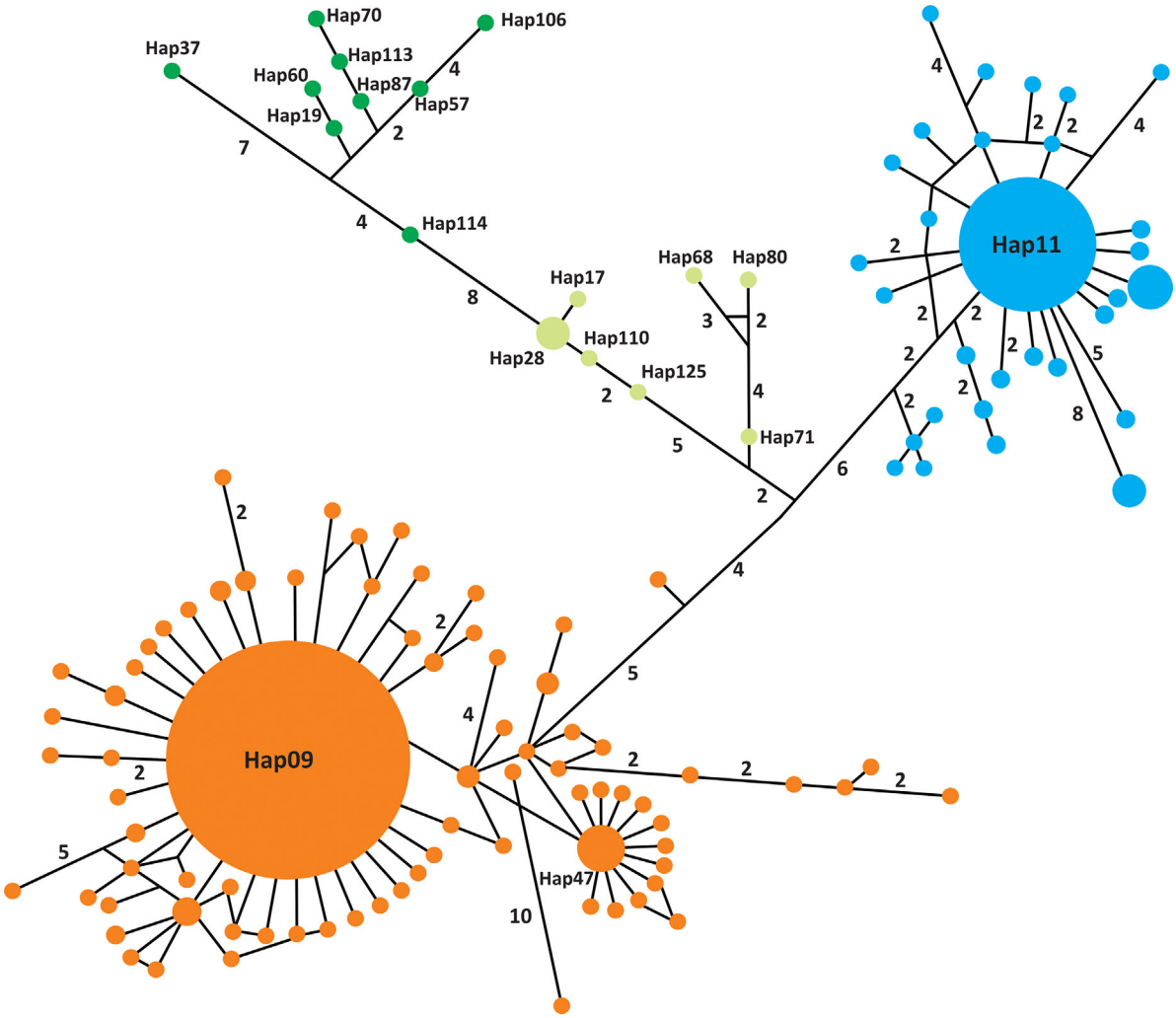

Fig. 1. Network topology of the 125 HPV16 haplotypes. Lineages are indicated using the following colors: orange, lineage **A**; light green, lineage **B**; dark green, lineage **C**; and blue, lineage **D**. Each circle corresponds to one of the 125 haplotypes, and the circle diameter is proportional to the number of samples sharing each haplotype. The lines connecting the circles indicate at least one nucleotide substitution. Haplotypes that differ by more than one substitution are indicated by a number (corresponding to the number of substitutions) written beside the connecting line. The three more frequent haplotypes are indicated: Hap09 (n = 68), Hap11 (n = 40), and Hap47 (n = 12). All haplotypes from lineages **B** and **C** were indicated.

lineage identification was based on an approach similar to that used for HPV16; that is, based on sequencing a region encompassing *LCR* and *E6* and identifying nucleotide signatures based on the SNVs described by Arias-Pulido et al. [2005] and Arroyo et al. [2012]. Fifty-five samples (~70 % of the HPV18 samples) were identified as belonging to lineage A and 16 samples (~19 % of the HPV18 samples) were identified as belonging to lineage B (Table II). The HPV18-distinctive SNVs for lineages A, B, and C affected nine nucleotide positions in *LCR* (A7152G, C7161T, C7164G, C7185T, C7496G, G7512A, T7530C, T7651C, and T7704C) and three nucleotide positions in *E6* (T251C, G374A, and A548G). Additionally, a 7-bp deletion in the *LCR* region was shared by all haplotypes of lineages B and C (Supplemental Table SIb).

TABLE III. Molecular Diversity Indexes Estimated for HPV16 and HPV18

	HPV16	HPV18
Number of samples	314	71
Number of haplotypes	125	33
Haplotype diversity	0.932	0.958
Nucleotyde diversity	0.084	0.01
Fu's Fs test	−23.86 ($P = 0.004$)	−4.11 ($P = 0.156$)

Thirty-three haplotypes were identified among the 71 samples; of these, 15 occurred in at least two patients (n = 53 patients) and 18 occurred in 18 patients. The most frequent haplotype was Hap22 (from HPV18 lineage A; 10 patients), followed by Hap12 (from HPV18 lineage A; 6 patients), and Hap33 (from HPV18 lineage B; 6 patients). These

three haplotypes were present in ~31% of the patients (Supplemental Table SIb).

The ML phylogenetic analysis was consistent with the identification based on SNV signatures and showed haplotype groupings in two lineages (Supplemental Fig. S1b): lineages A and B. This analysis indicated that lineages A and B formed two well-supported groups, each showing a maximum divergence of 2% between haplotypes.

The network median-joining topology (Fig. 2) showed that the most frequent haplotypes occupied a central, star-like position surrounded by less frequent haplotypes, but in a configuration that was not as evident as that observed for HPV16. The molecular diversity indices (haplotype diversity and nucleotide diversity) and Fu's Fs test based on the 71 samples are shown in Table III. The Fu's Fs test was negative but not significant (Fu's Fs = −4.11; $P = 0.156$) and provided no evidence of recent viral population expansion or positive selection.

E6 contained two previously described amino acid substitutions with respect to the prototype sequence: Y72H in one haplotype [Wang et al., 2009] and N129K in all of the B and C haplotypes [De Boer et al., 2004]. Additionally, four previously undescribed amino acid substitutions were found: H80Y (in one haplotype), H133P (in two haplotypes), Y134F (in one haplotype), and R144Q (in one haplotype) (Supplemental Table SId).

DISCUSSION

Several reports have associated population origin, persistence of infection, risk of cervical intraepithelial neoplasia, and risk for cervical invasive cancer with specific HPV16 and HPV18 lineages [Berumen et al., 2001; Burk et al., 2003; Xi et al., 2006; Sichero et al.,

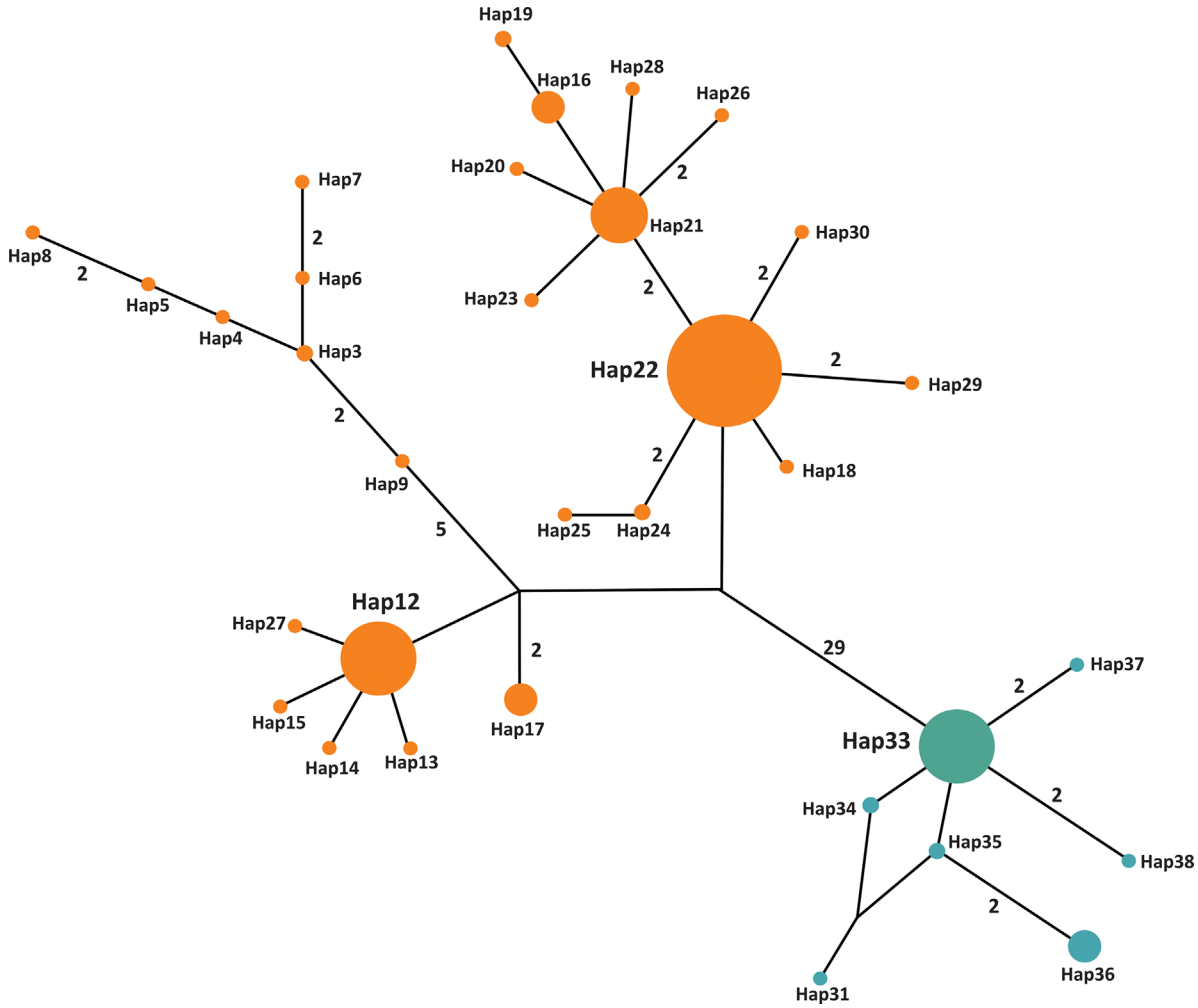

Fig. 2. Network topology of the 33 HPV18 haplotypes. Lineages are indicated using the following colors: orange, lineage **A** and green, lineage **B**. Each circle corresponds to one of the 33 haplotypes, and the circle diameter is proportional to the number of samples sharing each haplotype. The lines connecting the circles indicate at least one nucleotide substitution. Haplotypes that differ by more than one substitution are indicated by a number (corresponding to the number of substitutions) written beside the connecting line. The three more frequent haplotypes are indicated: Hap22 (n = 10), Hap12 (n = 6), and Hap33 (n = 6).

2012], indicating different prevalences of these lineages among populations. In this study, HPV16 lineages A and D and HPV18 lineage A were the most prevalent. Previous reports showed that these HPV lineages are the most prevalent lineages in European populations (HPV16 lineage A and HPV18 lineage A) and in populations of Amerindian ancestry (HPV16 lineage D), being lineages B and C from HPV16 and HPV18 most prevalent in African continent [Ho et al., 1993; Ong et al., 1993; Yamada et al., 1997]. The genetic background of the Rio de Janeiro metropolitan region population is highly mixed, mostly of European and African ancestry (55% and 31%, respectively) with a lower proportion of Amerindian ancestry (14% based on autosomal genetic markers, Manta et al. [2013]). Taking into account the contribution of these population-groups to the formation of the current population of the Rio de Janeiro metropolitan region, we expected a larger prevalence of HPV16 B and C lineages and HPV18 B and C lineages. However, it is important to take into account that a direct association between the human ancestry and HPV lineage is not necessarily expected in a population with ethnic admixture and with individuals having different proportions of genetic ancestry. This fact was reported by Lopera et al. [2014] analyzing women with cervical cancer that presented an inverse correlation between infections by HPV16 lineage D (common in Native American population) with the higher proportion of informative genetic markers (AIMs) from Native American ancestry. To better understand this finding it is required to analyze the prevalence of these lineages in women without cancer, because the current prevalence in invasive cancer might be associated with differences of their carcinogenic potential in the studied population.

The prevalence of HPV16 lineages in invasive cervical cancer in Brazil was also analyzed by Junes-Gill et al. [2008] in 63 women from Belém in the Brazilian Northern Region (Amazonia). These authors found lineage D to be the most prevalent (46%), followed by lineage A (41.3%) and B/C (3.2%), not considering samples with co-infection (9.5%). The higher prevalence of lineage D in this region than in Rio de Janeiro (46% in Junes-Gill et al. vs. 29% reported in the present work) might reflect the influence of the Amerindian ethnic background in Belém [Alves-Silva et al., 2000; Pereira et al., 2012]. The only other study reporting the prevalence of HPV16 lineages in women with invasive cervical cancer in Latin America (based on data from eight countries, Yamada et al. [1997]) showed results similar to those reported here; that is, a higher prevalence of lineage A (77%), followed by lineages D (19%) and B/C (~4%). Other studies performed in Brazil focusing on HPV16 intra-type diversity analyzed the prevalence of lineages in women with precancerous lesions [Villa et al., 2000] or with lesions of different grades (including invasive cancers) [Cruz et al., 2004; Freitas et al., 2014]. These studies also reported a higher prevalence of HPV16 lineage A, ranging from 65.8% [Freitas et al., 2014] to 41% [Cruz et al., 2004].

HPV18 lineage A was the most frequent in the studied cohort, followed by lineage B; lineage C was not detected. The paucity of studies on HPV18 lineages makes the drawing of comparisons difficult. Most studies included mainly women with normal cervical cells, LSIL, and HSIL, including two studies performed in Brazil [Villa et al., 2000; Sichero et al., 2007] that showed a higher prevalence of lineage A (80–90%) and a lower prevalence of lineage B. Lineage C was reported by only Sichero et al. [2007], at a frequency of 6.6%. Studies performed in other countries also reported a higher frequency of HPV18 lineage A. Xi et al. [2006] found HPV18 lineage A in 71.4% of women with normal cervical cells, LSIL, and HSIL followed by lineages B/C (28.6%) in the US. In Spain, studies performed by Arroyo et al. [2012] and Pérez et al. [2014] based on normal cervical cells, cervical lesions, and invasive cancer also found a higher frequency of HPV18 lineage A, followed by lineage B; lineage C was only reported by Arroyo et al. [2012] at a frequency of ~11%. In Costa Rica, Schiffman et al. [2010] evaluated women with normal cervical cells and women with cervical lesions of various grades and showed a lower frequency of HPV18 lineage B (22.6%) than that of lineages A and C (77.4%).

Hap09, Hap11, Hap47, and Hap04 were the most frequent HPV16 haplotypes in the 314 samples analyzed, altogether accounting for 41% of the total. The network arrangement showed that three of these haplotypes (Hap09, Hap11, and Hap47) were associated with less frequent haplotypes in a star-like configuration, most of which differed by one nucleotide substitution. This pattern suggests a rapid population expansion of HPV16 haplotypes associated with invasive cancer due to historical-neutral events or exposure to selective factors favoring specific haplotypes, with selection acting over *LCR* or *E6* regions or over another HPV genomic region resulting in genetic-hitchhiking. These possibilities were supported by the negative and significant value obtained using Fu's Fs test (-23.86; $P = 0.004$). This result might also indicate the association of specific variants with a higher risk of cervical cancer, similar to what was observed for HPV45 sub-lineage B2 variants [Chen et al., 2014]. Analysis of the prevalence of HPV16 variants in women without cervical lesions might enable us to evaluate these proposals. The Fu's Fs test result (-4.11; $P = 0.156$) was not significant for HPV18, suggesting a different evolutionary history. The non-significance found, indicating that the Fs value did not differ from 0 (zero), implies that the HPV18 genomes in the studied population are evolving in accord to a selective neutral model, and that no recent event affecting viral population size (and consequently HPV18 diversity) occurred.

It is necessary to underline that the haplotypes placed in lineage B of HPV16 by SNV analysis did

not form a monophyletic group in the ML topology (Supplemental Fig. S1) and in the Network arrangement (Fig. 1), in spite of the two reference sequences of lineage B were grouped together in the ML topology. The arrangement observed for these haplotypes in the ML topology were low supported by the boostrap analysis, pointing to the requirement of a larger genomic region to be analyzed to increase the support for the presented topology or for an alternative topology.

In conclusion, the results obtained here show that women diagnosed with invasive cervical cancer associated with HPV16 and HPV18 in the Brazilian Southwestern Region showed a higher frequency of the HPV lineages that are most prevalent in Europe and North America (HPV16 lineage A and HPV18 lineage A). The genetic diversity of HPV16 indicated a recent expansion of specific variants or a selective advantage associated with invasive cancer; this pattern was not observed for HPV18. HPV diversity data are important for the development of HPV diagnostics, vaccines, and therapeutic approaches to monitoring virus-induced diseases. The major limitation of this study was the absence of a control group with women infected with HPV but without invasive cancer. This is the largest study to investigate the genetic diversity of HPV16 and HPV18 related to invasive cervical cancer conducted in Brazil prior to the implementation of the immunization program against cervical cancer by the Brazilian Public Health System (initiated in March 2014). All patients included between July 2011 and March 2014 resided in the second largest metropolitan region of the country (Rio de Janeiro State); this population is economically and educationally diverse. These data provide a baseline for future studies evaluating the genetic diversity of these HPV lineages after implementation of the immunization program.

ACKNOWLEDGMENTS

We thank Dr Hector Seuanez for critically revising the manuscript. We are also indebted to Luiz Figueiredo Mathias (MD), Chief of the Gynecologic Oncology Department, Instituto Nacional de Câncer, deceased in 2014.

REFERENCES

Alves-Silva J, da Silva Santos M, Guimarães PE, Ferreira AC, Bandelt HJ, Pena SD, Prado VF. 2000. The ancestry of Brazilian mtDNA lineages. Am J Hum Genet 67:444–461.

Arias-Pulido H, Peyton CL, Torrez-Martínez N, Anderson DN, Wheeler CM. 2005. Human papillomavirus type 18 variant lineages in United States populations characterized by sequence analysis of LCR-E6, E2, and L1 regions. Virology 338:22–34.

Arroyo SL, Basaras M, Arrese E, Hernáez S, Andía D, Esteban V, Garcia-Etxebarria K, Jugo BM, Cisterna R. 2012. Human papillomavirus (HPV) genotype 18 variants in patients with clinical manifestations of HPV related infections in Bilbao, Spain. Virol J 9:258.

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48.

Berumen J, Ordoñez RM, Lazcano E, Salmeron J, Galvan SC, Estrada RA, Yunes E, Garcia A, Gonzalez-lira G, Madrigal-de la Campa A. 2001. Asian-American variants of human papillomavirus 16 and risk for cervical cancer: A case-control study. J Natl Cancer Inst 93:1325–1330.

Bosch FX, Burchell AN, Schiffman M, Giuliano AR, de Sanjose S, Bruni L, Tortolero-Luna G, Kjaer SK, Muñoz N. 2008. Epidemiology and natural history of human papillomavirus infections and type-specific implications in cervical neoplasia. Vaccine 26:K1–K16.

Bosch FX, Lorincz A, Muñoz N, Meijer CJLM, Shah KV. 2002. The causal relation between human papillomavirus and cervical cancer. J Clin Pathol 55:244–265.

Bruni L, Diaz M, Castellsagué X, Ferrer E, Bosch FX, de Sanjosé S. 2010. Cervical human papillomavirus prevalence in 5 continents: Meta-analysis of 1 million women with normal cytological findings. J Infect Dis 202:1789–1799.

Burk RD, Terai M, Gravitt PE, Brinton LA, Kurman RJ, Barnes WA, Greenberg MD, Hadjimichael OC, Fu L, McGowan L, Mortel R, Schwartz PE, Hildesheim A. 2003. Distribution of human papillomavirus types 16 and 18 variants in squamous cell carcinomas and adenocarcinomas of the cervix. Cancer Res 63:7215–7220.

Burk RD, Harari A, Chen Z. 2013. Human papillomavirus genome variants. Virology 445:232–243.

Cerqueira DM, Raiol T, Véras NM, von Gal Milanezi N, Amaral FA, de Macedo Brígido M, Martins CR. 2008. New variants of human papillomavirus type 18 identified in central Brazil. Virus Genes 37:282–287.

Chen A, Heideman DAM, Boon D, Gheit T, Snijders PJF, Tommasino M, Franceschi S, Clifford GM. 2014. Human papillomavirus 45 genetic variation and cervical cancer risk worldwide. J Virol 88:4514–4521.

Cornet I, Gheit T, Franceschi S, Vignat J, Burk RD, Sylla BS, Tommasino M, Clifford GM. 2012. Human papillomavirus type 16 genetic variants: Phylogeny and classification based on E6 and LCR. J Virol 86:6855–6861.

Cornet I, Gheit T, Iannacone MR, Vignat J, Sylla BS, Del Mistro A, Franceschi S, Tommasino M, Clifford GM. 2013. HPV16 genetic variation and the development of cervical cancer worldwide. Br J Cancer 108:240–244.

Cruz MR, Cerqueira DM, Cruz WB, Camara GNL, Brígido MM, Silva EO, Carvalho LGS, Martins CRF. 2004. Prevalence of human papillomavirus type 16 variants in the Federal District, Central Brazil. Mem Inst Oswaldo Cruz 99:281–282.

de Araujo Souza PS, Maciag PC, Ribeiro KB, Petzl-Erler ML, Franco EL, Villa LL. 2008. Interaction between polymorphisms of the human leukocyte antigen and HPV-16 variants on the risk of invasive cervical cancer. BMC Cancer 8:246.

de Araujo Souza PS, Sichero L, Maciag PC. 2009. HPV variants and HLA polymorphisms: The role of variability on the risk of cervical cancer. Future Oncol 5:359–370.

De Boer MA, Peters LAW, Aziz MF, Siregar B, Cornain S, Vrede MA, Jordanova ES, Kolkman-Uljee S, Fleuren GJ. 2004. Human papillomavirus type 16 E6, E7, and L1 variants in cervical cancer in Indonesia, Suriname, and The Netherlands. Gynecol Oncol 94:488–494.

De Villiers EM. 2013. Cross-roads in the classification of papillomaviruses. Virology 445:2–10.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resourc 10:564–567.

Freitas LB, Chen Z, Muqui EF, Boldrini NAT, Miranda AE, Spano LC, Burk RD. 2014. Human papillomavirus 16 non-European variants are preferentially associated with high-grade cervical lesions. PLoS ONE 9:e100746.

Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915–925.

Fuessel Haws AL, He Q, Rady PL, Zhang L, Grady J, Hughes TK, Stisser K, Konig R, Tyring SK. 2004. Nested PCR with the PGMY09/11 and GP5(+)/6(+) primer sets improves detection of HPV DNA in cervical samples. J Virol Methods 122:87–93.

Gheit T, Cornet I, Clifford GM, Iftner T, Munk C, Tommasino M, Kjaer SK. 2011. Risks for persistence and progression by human papillomavirus type 16 variant lineages among a population-based sample of Danish women. Cancer Epidemiol Biomarkers Prev 20:1315–1321.

Gravitt PE, Peyton CL, Alessi TQ, Wheeler CM, Coutlée F, Hildesheim A, Schiffman MH, Scott DR, Apple RJ. 2000. Improved amplification of genital human papillomaviruses. J Clin Microbiol 38:357–361.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst Biol 59:307–321.

Harari A, Chen Z, Rodríguez AC, Hildesheim A, Porras C, Herrero R, Wacholder S, Panagiotou OA, Befano B, Burk RD, Schiffman M. Costa Rica HPV Vaccine Trial (CVT) group 2015. Cross-protection of the bivalent human papillomavirus (HPV) vaccine against variants of genetically related high-risk HPV infections. J Infect Dis Epub ahead of print; DOI: 10.1093/infdis/jiv519

Ho L, Chan S, Burk R, Das B. 1993. The genetic drift of human papillomavirus type 16 is a means of reconstructing prehistoric viral spread and the movement of ancient human populations. J Virol 16:6413–6423.

INCA-Instituto Nacional de Cancer Jose Alencar Gomes da Silva 2014. Coordenação de prevenção e vigilância. Estimativa 2014: Incidência de câncer no Brasil. Rio de Janeiro: Instituto Nacional de Câncer José Alencar Gomes da Silva, p 124.

Junes-Gill K, Sichero L, Maciag PC, Mello W, Noronha V, Villa LL. 2008. Human papillomavirus type 16 variants in cervical cancer from an admixtured population in Brazil. J Med Virol 80:1639–1645.

Keane TM, Creevey CJ, Pentony MM, Naughton TJ, Mclnerney JO. 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol Biol 6:29.

Lopera EA, Baena A, Florez V, Montiel J, Duque C, Ramirez T, Borrero M, Cordoba CM, Rojas F, Pareja R, Bedoya AM, Bedoya G, Sanchez GI. 2014. Unexpected inverse correlation between Native American ancestry and Asian American variants of HPV16 in admixed Colombian cervical cancer cases. Infect Genet Evol 28:339–348.

Manta FSN, Pereira R, Caiafa A, Silva DA, Gusmao L, Carvalho EF. 2013. Analysis of genetic ancestry in the admixed Brazilian population from Rio de Janeiro using 46 autosomal ancestry-informative indel markers. Ann Hum Biol 40:94–98.

Muñoz N, Bosch FX, de Sanjosé S, Herrero R, Castellsagué X, Shah KV, Snijders PJF, Meijer CJLM. 2003. Epidemiologic classification of human papillomavirus types associated with cervical cancer. N Engl J Med 348:518–527.

Ong CK, Chan SY, Campo MS, Fujinaga K, Mavromara-Nazos P, Labropoulou V, Pfister H, Tay SK, Meulen J, Villa LL. 1993. Evolution of human papillomavirus type 18: An ancient phylogenetic root in Africa and intratype diversity reflect coevolution with human ethnic groups. J Virol 67:6424–6431.

Pereira R, Phillips C, Pinto N, Santos C, dos Santos SEB, Amorim A, Carracedo A, Gusmão L. 2012. Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. PLoS ONE 7:e29684.

Pérez S, Cid A, Iñarrea A, Pato M, Lamas MJ, Couso B, Gil M, Álvarez MJ, Rey S, López-Miragaya I, Melon S, de Ona M. 2014. Prevalence of HPV 16 and HPV 18 lineages in Galicia, Spain. PLoS ONE 9:1–10.

Schiffman M, Rodriguez AC, Chen Z, Wacholder S, Hildesheim A, Desalle R, Befano B, Yu K, Safaeian M, Sherman ME, Morales J, Guillen D, Alfaro M, Hutchinson M, Solomon D, Castle PE, Burk RD. 2010. A population-based prospective study of carcinogenic human papillomavirus (HPV) variant lineages, viral persistence, and cervical neoplasia. Cancer Res 70:3159–3169.

Schiller JT, Lowy DR. 2012. Understanding and learning from the success of prophylactic human papillomavirus vaccines. Nat Rev Microbiol 10:681–692.

Sichero L, Ferreira S, Trottier H, Duarte-Franco E, Ferenczy A, Franco EL, Villa LL. 2007. High grade cervical lesions are caused preferentially by non-European variants of HPVs 16 and 18. Int J Cancer 120:1763–1768.

Sichero L, Sobrinho JS, Villa LL. 2012. Oncogenic potential diverge among human papillomavirus type 16 natural variants. Virology 432:127–132.

Smith B, Chen Z, Reimers L, van Doorslaer K, Schiffman M, Desalle R, Herrero R, Yu K, Wacholder S, Wang T, Burk RD. 2011. Sequence imputation of HPV16 genomes for genetic association studies. PLoS ONE 6:e21375.

Villa LL, Sichero L, Rahal P, Caballero O, Ferenczy A, Rohan T, Franco EL. 2000. Molecular variants of human papillomavirus types 16 and 18 preferentially associated with cervical neoplasia. J Gen Virol 81:2959–2968.

Wang SS, Bratti MC, Rodríguez AC, Herrero R, Burk RD, Porras C, González P, Sherman ME, Wacholder S, Lan ZE, Schiffman M, Chanock SJ, Hildesheim A. 2009. Common variants in immune and DNA repair genes and risk for human papillomavirus persistence and progression to cervical cancer. J Infec Dis 199:20–30.

Xi LF, Kiviat NB, Hildesheim A, Galloway DA, Wheeler CM, Ho J, Koutsky L. 2006. Human papillomavirus type 16 and 18 variants: Race-related distribution and persistence. J Natl Cancer Inst 98:1045–1052.

Yamada T, Manos MM, Peto J, Greer CE, Munoz N, Bosch FX, Wheeler CM. 1997. Human papillomavirus type 16 sequence variation in cervical cancers: A worldwide perspective. J Virol 71:2463–2472.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.