Geographical distribution and oncogenic risk association of human papillomavirus type 58 E6 and E7 sequence variations

Paul K. S. Chan1*, Chuqing Zhang1, Jong-Sup Park2, Karen K. Smith-McCune3,4, Joel M. Palefsky5, Lucia Giovannelli6, Francois Coutlée7, Samantha Hibbitts9, Ryo Konno9, Wannapa Settheetham-Ishida10, Tang-Yuan Chu11, Annabelle Ferrera12, María Alejandra Picconi13, Federico De Marco14, Yin-Ling Woo15, Tainá Raiol16, Patricia Piña-Sánchez17, Jeong-Hoon Bae2, Martin CS Wong18, Mike Z. Chirenje19, Tsitsi Magure19, Anna-Barbara Moscicki4,20, Alison N. Fiander8, Giuseppina Capra6, Eun Young Ki2, Yi Tan21, Zigui Chen22, Robert D Burk22, Martin CW Chan1, Tak-Hong Cheung23, David Pim24, and Lawrence Banks24

Department of Microbiology, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong Special Administrative Region, China1

Department of Obstetrics and Gynecology, Seoul St. Mary's Hospital, The Catholic University of Korea, Seoul, Korea2

Department of Obstetrics, Gynecology and Reproductive Sciences, University of California San Francisco, California, United States of America3

Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, California, United States of America4

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an ‘Accepted Article’, doi: 10.1002/ijc.27932
Department of Medicine, University of California San Francisco, California, United States of America

Sezione di Microbiologia, Dipartimento di Scienze per la Promozione della Salute, Azienda Ospedaliera Universitaria Policlinico P. Giaccone, Palermo, Italy

Département de Microbiologie et Infectiologie, Centre Hospitalier de l’Université de Montréal, Montréal, Québec, Canada

HPV Research Group, Cancer and Genetics Research Institute, School of Medicine, Cardiff University, United Kingdom

Department of Obstetrics and Gynecology, Jichi Medical University, Saitama Medical Center, Saitama, Japan

Department of Physiology, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand

Department of Obstetrics and Gynaecology, Buddhist Tzu Chi General Hospital, Institute of Medical Sciences, Tzu Chi University, Hualien, Taiwan

School of Microbiology, Science Faculty, Universidad Nacional Autónoma de Honduras, Tegucigalpa, Honduras

Servicio Virus Oncogénicos, Departamento Virología, Instituto Nacional de Enfermedades Infecciosas, ANLIS “Carlos G. Malbrán”, Buenos Aires, Argentina

Laboratory of Virology – The “Regina Elena” National Cancer Institute, Rome, Italy

Department of Obstetrics and Gynaecology, University of Malaya, Malaysia

Department of Cellular Biology, Institute of Biology, University of Brasilia, Brazil
Short running head: HPV58 E6 E7 sequence variation

Novelty and Impact Statements:

This study reports the sequence variation of HPV58 E6 and E7 genes from a global perspective. E6 was found to be more conserved, and thus a more appropriate target for type-specific detection, whereas the more variable E7 can be used for strain differentiation. HPV58 variant with T20I and
G63S substitutions in E7 conferred a higher oncogenic risk. This variant was more prevalent in Asia and may explain the higher disease burden observed in this region.

**Word count:** Abstract (230); main content (2498); tables (2); figures (4); references (31)

**Footnote:**

This study was supported by the International Centre for Genetic Engineering and Biotechnology (ICGEB) (project no. CRP/CHN08-03). Dr. Francois Coutlée’s research on HPV variants is supported by the Cancer Society of Canada, and he has received financial support for research projects and honoraria for oral presentations on HPV from Roche Diagnostics and Merck. Dr. Federico De Marco is partly supported by the Italian Ministry of Foreign Affairs, DGPC Uff V and by the Italian Ministry of Health. Dr. Karen Smith McCune is a member of the Scientific Advisory Board and owner of shares of stock in OncoHealth Inc, a startup company developing diagnostic assays for cervical cancer screening. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Conflicts of interest:** The authors have no conflict of interest to disclose.

**Correspondence:** Prof. Paul KS Chan
Department of Microbiology, The Chinese University of Hong Kong, 1/F Clinical Sciences Building, Prince of Wales Hospital, Shatin, New Territories, Hong Kong Special Administrative Region, People’s Republic of China

Tel: +852 2632 3333  Fax: +852 2647 3227  Email: paulkschan@cuhk.edu.hk

**KEYWORDS**

HPV, variant, cervical cancer, phylogeny, oncogenic risk
ABSTRACT

Human papillomavirus (HPV) 58 accounts for a notable proportion of cervical cancers in East Asia and parts of Latin America, but it is uncommon elsewhere. The reason for such ethnogeographical predilection is unknown. In this study, nucleotide sequences of E6 and E7 genes of 401 HPV58 isolates collected from 15 countries/cities across four continents were examined. Phylogenetic relationship, geographical distribution and risk association of nucleotide sequence variations were analyzed. We found that the E6 genes of HPV58 variants were more conserved than E7. Thus, E6 is a more appropriate target for type-specific detection, whereas E7 is more appropriate for strain differentiation. The frequency of sequence variation varied geographically. Africa had significantly more isolates with E6-367A (D86E), but significantly less isolates with E6-203G, -245G, -367C (prototype-like) than other regions (P ≤ 0.003). E7-632T, -760A (T20I, G63S) was more frequently found in Asia, and E7-793G (T74A) was more frequent in Africa (P < 0.001). Variants with T20I and G63S substitutions at E7 conferred a significantly higher risk for cervical intraepithelial neoplasia grade III and invasive cervical cancer compared to other HPV58 variants (odds ratio = 4.44, P = 0.007).

In conclusion, T20I and/or G63S substitution(s) at E7 of HPV58 is/are associated with a higher risk for cervical neoplasia. These substitutions are more commonly found in Asia and the Americas, which may account for the higher disease attribution of HPV58 in these areas.
INTRODUCTION

High-risk human papillomavirus (HPV) plays a necessary role in the development of cervical cancer, which is the third most common cancer in women worldwide. Among the 15 HPV types classified as high-risk, HPV16 and HPV18 are the two most important types accounting for about 70% of cervical cancers across the world. The remaining cancers are caused by HPV31, 33, 35, 39, 45, 51, 52, 56, 58 and 59. The distribution of these less common types shows some degree of ethnogeographical variation. HPV58 is an uncommon type with skewed ethnogeographical distribution. The prevalence of HPV58 is remarkably high among cervical squamous cell carcinomas in East Asia (33% in Shaanxi, 26% in Shanghai, 16% in South Korea, 10% in Hong Kong and Taiwan, and 8% in Japan), whereas it only accounts for about 2% of cervical cancers in other regions. In addition to East Asia, a higher prevalence of HPV58 has also been observed in a few places in Latin America. HPV58 was found in 12% of cervical cancers in Costa Rica, 21% in Southeastern Mexico, 9.5% in Northeastern Brazil, and was the second most common type (about 13%) found in precancerous lesions in Central and Southern Brazil.

The reason for an ethnogeographical predilection in the attribution of HPV58 to cervical cancer is not fully understood. HPV58 variants are classified into four lineages. These lineages exhibit a distribution pattern that varies across the world. Studies comparing the risk association of these variant lineages are very limited and a conclusion has not yet been reached.

The E6 and E7 proteins encoded by high-risk HPV play a key role in cellular transformation. Previous studies from Eastern and Southern Chinese populations suggested that sequence variations within these regions may be associated with a higher risk for cervical neoplasia. At present, information on sequence variation of the E6 and E7 open reading frames (ORF) of HPV58...
have been derived mainly from Chinese populations.\textsuperscript{18-20} The current study examined the sequence polymorphisms of HPV58 E6 and E7 ORF and assessed the geographical distribution and risk association with cervical neoplasia from a global perspective.

**MATERIALS AND METHODS**

**Study Samples**

A total of 401 samples including 295 cervical scrapes or swabs, 69 vaginal swabs and 37 anal samples collected from 15 countries and cities were examined in this study (Table 1). These samples have been used to establish a lineage classification system for HPV58 variants.\textsuperscript{16}

**Nucleotide Sequencing**

The full length of E6 and E7 ORF was amplified by polymerase chain reaction (PCR) using primers 5'-CGA AAA CGG TCT GAC CGA AA-3' and 5'-TAT CGT CTG CTG TTT CGT CC-3', followed by a second PCR using inner primers (5'-GAC CGA AAC CGG TGC ATA TA-3' and 5'-ACC GCT TCT ACC TCA AAC CA-3') when necessary. Briefly, the PCR was conducted in a 50-μL reaction mix containing 4 μL of extracted DNA, 200 μM of deoxynucleotide triphosphates (dNTPs), forward and reverse primers at 0.25 μM each, 1.25 unit of HotStar\textsuperscript{\textregistered} Taq Plus polymerase (Qiagen, Hilden, Germany). The cycling conditions were activation of polymerase at 95°C for 5 min, 40 cycles of denaturation at 94°C for 50 sec, annealing at 57°C (59°C for nested PCR) for 1 min and extension at 72°C for 1 min, followed by a final extension at 72°C for 8 min. Ten microliters of the purified PCR products were mixed with 2 μL of BigDye\textsuperscript{\textregistered} Terminator v.3.1 sequencing reaction mix (Applied Biosystems, Foster City, USA), 3 μL of 5X sequencing buffer, and 3.2 pmol of the sequencing primer which was the corresponding PCR primer as described above; and made up to a final volume of 20 μL. The cycling conditions for the labeling
PCR were 25 cycles at 95°C for 15 sec, 50°C for 15 sec, and 65°C for 75 sec. Fluorescence-labeled PCR products were purified with DyeEx (Qiagen, Hilden, Germany) and run on an ABI 3130 automated sequence analyzer (Applied Biosystems, Foster City, USA). Sequence data were obtained from both directions and analyzed with SeqScape v2.5 (Applied Biosystems, Foster City, USA). Variations that occurred only in one isolate were confirmed by repeating PCR-sequencing from the original sample.

Phylogenetic Analysis

The program MEGA5 was applied to identify the best evolutionary model and to construct the maximum-likelihood trees using the close-neighbor-interchange (CNI) search approach. The data were bootstrap resampled 1,000 times for tree topology evaluation. HPV33 was used as an outgroup sequence. The program Datamonkey (http://www.datamonkey.org) was used to detect positive or negative selection.

Statistical Analysis

All data analyses were performed by the Statistical Package for Social Sciences (version 18.0.0, IBM Corp. NY). The rate of detection of E6 and E7 sequence patterns was compared between the four regions by chi-squared test and using the Bonferroni method to account for multiple comparisons. For the purpose of risk association analysis, women with normal cytology or absence of lesion at colposcopy or biopsy were grouped as “normal cervices”. Using cervical intraepithelial neoplasia grade III (CIN3) and cervical cancer as an outcome variable and “normal cervices” as control, the risk association of non-synonymous substitutions was first assessed by univariate analysis using chi-squared test or Fisher’s exact test. Those mutations found to be significant for the association (p<0.05) were tested for interactions, followed by unconditional binary logistic regression analyses.
controlling for patients’ age and their country of origin. All P values <0.05 were regarded as statistically significant.

This study has included 37 anal samples collected from men in the United States of America. In our previous study on HPV58 variant lineage based on the same cohort of samples, no differences in the distribution of variants between these anal samples and cervical samples collected from the same region was observed. In this study, there were also no differences in the distribution of E6 and E7 variants among these two sources of samples. For these reasons, anal samples were pooled together with cervical samples for the purpose of analysis in this study.

RESULTS

E6 Sequence Variation

Altogether, 19 variants showing nucleotide sequence variations in the E6 ORF were identified (Figure 1). The maximum variation between any two variants was nine nucleotides, a variability of 2.0% over the 450-bp E6 ORF. The mean dN/dS was 0.916 and no site with significant positive or negative selection was found. Altogether, 16 nucleotide positions showing sequence polymorphism were identified, with 11 positions had non-synonymous substitutions leading to amino acid changes. Based on the phylogenetic relationship, the E6 sequence patterns could be classified into four groups designated as E6-A, -B, -C, and -D; and each group was associated with a signature nucleotide sequence variation as shown in Figure 2A. The maximum inter-group nucleotide sequence variability ranging from 0.9% (E6-A and -B) to 2.0% (E6-B and -D), whereas the maximum intra-group variability ranging from 0.7% (E6-A and -C) to 1.3% (E6-B).

E7 Sequence Variation
Twenty-one variants showing sequence variations in the E7 ORF were identified (Figure 3). The maximum nucleotide sequence variation between any two variants was nine, a variability of 3.0% over the 297-bp ORF. The mean dN/dS of E7 was 0.748 indicating a positive selection. Overall, 18 nucleotide positions showing sequence polymorphisms were identified, and 11 had non-synonymous substitutions (Figure 3). Four groups of E7 sequence patterns designated as E7-A with signature sequence variation of 632C, 694G, 793A; E7-B (694A); E7-C (632T); and E7-D (793G) were identified. The phylogenetic relationship and their associated signature nucleotide sequence variations are shown in Figure 2B.

**Geographical Distribution**

The frequencies of different E6 and E7 sequence variation patterns among isolates collected from different parts of the world are shown in Figure 4. E6-A was the most common pattern found overall (Figure 4A). In Asia, Americas and Europe, E6-A predominated and accounted for 86-96% of isolates. In Africa, E6-A (51%) co-circulated with E6-C (40%). E6-A was significantly less frequent in Africa compared to the other three regions (P = 0.003 for Americas, P < 0.001 for Asia and Europe); whereas E6-C was significantly more frequent in Africa than in other regions (P < 0.001 for all 3 regions).

The geographical distribution of E7 sequence patterns was more diverse (Figure 4B). In Asia, both E7-B and E7-C patterns were commonly found; whereas E7-B and E7-D patterns co-dominated in Africa. E7-C was significantly more frequent in Asia (P < 0.001 for all three regions); whereas E7-D was significantly more frequent in Africa (P < 0.001 for all three regions).

**Risk Association with Cervical Neoplasia**
Among the 401 samples examined, 33 were histology-confirmed invasive cervical cancer, 31 were CIN3, and 79 were normal cervices as defined by normal cytology or no lesion detected at colposcopy or biopsy. These cases were used to analyze the risk association of sequence variations. Two E7 substitutions (G41R and G63D) showed a significantly lower risk, and the other two E7 substitutions (G63S and T20I) showed a significantly higher risk based on univariate analysis (Table 2). All isolates with G41R also carried G63D, and a significant interaction was found (P = 0.023). Similarly, all isolates with T20I also carried G63S, and a significant interaction was found (P = 0.002). We therefore selected G41R and T20I into two separate final regression models, respectively, controlling for age and country of origin. It was found that T20I and/or G63S conferred an independent increase in risk for CIN3 and invasive cervical cancer (adjusted odds ratio 4.44, 95% CI. 1.51-13.05, P = 0.007). The risk association observed for G41R and/or G63D was no longer significant after multivariate analysis.

Discussion

The current study has delineated the sequence variations of two oncoprotein-encoding genes of HPV58. This data set provides essential information for further in depth studies to elucidate the epidemiology of this characteristic HPV type that exhibits a higher disease attribution in certain ethnogeographical groups. While the current study has examined a large series of samples covering a wide range of geographical regions, one should note that the number of samples available from certain areas was small and might not be representative of those areas.

The degree of nucleotide sequence variation among HPV58 isolates was found to be small (maximum 2% for E6, 3% for E7), as expected for HPV in general.\textsuperscript{15,25-27} Phylogenetically related E6 and E7 sequence patterns could be identified by distinct signatures of non-synonymous sequence
variations. These molecular signatures not only provide convenient markers for large-scale epidemiological studies, but the substitutions themselves may carry biological implications.

Overexpression of E6 and E7 proteins is required to maintain the transformed phenotype of keratinocytes, and thus they are constitutively retained in cancer cells and not affected by viral genome disruption resulting from integration. Therefore, E6 and E7 are good diagnostic targets, particularly for type-specific detection as E6 and E7 genes are diverse among HPV types. Our results show that the sequences of E7 were more variable than those of E6, and more heterogenously distributed across the world. This observation of HPV58 is in contrast to those reported for HPV16, 31, 33, 35 and 52 where the E7 ORF is less variable, and therefore E7 is often chosen as the target for diagnostic detection of HPV16. Based on the current findings, E6 rather than E7 is a better target for type-specific detection of HPV58.

The first clone of HPV58 was obtained from a Japanese patient with invasive cervical cancer. This strain is regarded as the prototype and commonly used for primer design. The E6-A2 and E7-A1 sequence pattern, as assigned by this study, corresponds respectively to the E6 and E7 sequence of the prototype. Our data showed that these “prototype-like” sequences were uncommon (10%) in Asia, and rare (< 3%) in other regions. Primers and probes designed based on the prototype sequence may have the problem of mismatch with the more commonly found strains.

Geographical predilection in the distribution of E6 and E7 sequence patterns was observed. E6-A predominated in most parts of the world except Africa, where it co-dominated with E6-C that was uncommon elsewhere. Two characteristics in the distribution of E7 sequence variation patterns were observed. Firstly, E7-D with a nucleotide signature of 793G was a common sequence pattern found in 51% of isolates from Africa, but uncommon (14%) in Americas and rare in Asia (4%) and Europe (7%). The second observation was that E7-C with an amino acid substitution signature of T20I
and G63S was commonly found in Asia (34%) and Americas (11%), but rare elsewhere. Geographical variation in the distribution of HPV58 E6 and E7 sequence patterns should be considered in designing diagnostic assays.

HPV58 is rare worldwide but attributes to a higher proportion of cervical cancers in East Asia and some parts of Latin America. This study attempted to identify HPV58 E6 and E7 sequence variations that might carry a higher oncogenic risk, and examined whether the distribution of these variations could account for the ethnogeographical predilection in disease burden. Isolates carrying G63S and/or T20I were found to have an independent increase in risk for CIN3 and invasive cervical cancer. Only one variant harboured G63S alone without T20I. This variant was only found in four samples precluding analysis on these two amino acid variations separately.

The high-risk mutations T20I and G63S were found in 33% of isolates in Asia and 10% in Americas, but rare in Europe (3%) and Africa (0%). This distribution is reminiscent of the geographical predilection in cervical cancers associated with HPV58. A previous study from southern China (Hong Kong) also revealed that substitutions T20I and/or G63S were associated with a 6.9-fold increase in odds ratio for developing invasive cervical cancer, and the age of patients were significantly younger than those infected with other variants of HPV58. Another study from eastern China (Zhejiang) also reported a significant positive trend of association between these substitutions and the severity of cervical lesions.

It has been shown that HPV58 variants can be divided into four lineages, which is mainly based on the sequence variations within the long control region (LCR). Although, naturally occurring LCR sequence variations have been shown to carry different promoter activity in-vitro, two previous studies that analyzed epidemiological risk association based on the LCR-lineage classification system did not reveal any significant differences among the lineages or sub-lineages.
This is special for HPV58 as an association between LCR polymorphisms and lesion has been reported for other types. The current study suggested that the nucleotide sequence variations at E7 are more useful for predicting the risk of HPV58 variants. In-vitro study to compare the transforming activities among E7 proteins carrying these amino acid substitutions is ongoing.

In conclusion, this study provides a comprehensive analysis on the sequence variation of E6 and E7 of HPV58 from a worldwide perspective. E6 is less variable and more appropriate as a target for type-specific diagnostic detection. Geographical variation in the distribution of HPV58 E6 and E7 sequence variations exists, which should be considered when designing diagnostic assays and therapeutic vaccines. Amino acid substitution(s) T20I and/or G63S at E7 is(are) associated with a higher risk for CIN3 and invasive cervical cancer. This high-risk sequence signature is more commonly found in isolates from Asia and the Americas than elsewhere, resembling the geographical predilection of cervical cancers attributed to HPV58.
Acknowledgements

We are grateful to the contribution of Dr. Sergio Andres Tonon and Dr. Cláudia Renata F. Martins who unfortunately have passed away before the completion of this study.
Legends for figures

Figure 1. Nucleotide sequence variations at E6 of 19 HPV58 isolates.

1 The group containing the prototype was assigned as E6-A. The HPV58 prototype (http://www.ncbi.nlm.nih.gov/genbank accession no. NC_001443) was used as reference.

2 No. (%) of each sequence pattern found among the 401 samples examined.

3 Nucleotide positions where variations were detected are written vertically across the top. Sites with no changes are marked with dashes. Capital letters indicate alternations result in an amino acid change. Lower case letters indicate synonymous mutations.

4 The one letter amino acid code is used. The position of amino acid change is stated numerically. The letter preceding this number refers to the reference amino acid, and the letter following refers to its substitution.

5 Lineage and sub-lineage classification based on full genome or concatenated E6-E7-E2-E5-L1-PCR sequences described in previous studies. Some E6 sequence patterns are not sub-lineage-specific. --, no sub-lineage assigned. Single nucleotide substitutions diagnostic for E6 variants are boxed. GenBank accession numbers for E6 sequences are ## (pending).

Figure 2. Phylogenetic trees of E6 and E7 of HPV58.

The maximum-likelihood trees were constructed by MEGA5 using the close-neighbor-interchange (CNI) search approach. Bootstrap values of ≥ 50% generated by 1,000 resamplings are shown.

A. Constructed from 450-bp E6 nucleotide sequences of 19 HPV58 isolates. The length of scale bar represents 0.004 substitutions per nucleotide position. The sample HK969 with E6 nucleotide sequence identical to the prototype (GenBank accession no. NC_001443) is indicated by an arrow.
The group containing the prototype sequence was assigned as E6-A. Sample CAN301 was grouped into sequence pattern E6-A.

**B. Constructed from 297-bp E7 nucleotide sequences of 21 HPV58 isolates.** The length of scale bar represents 0.004 substitutions per nucleotide position. The sample HK523 with nucleotide sequence identical to the prototype (GenBank accession no. NC_001443) is indicated by an arrow. The group containing the prototype sequence was assigned as E7-A.

**Figure 3. Nucleotide sequence variations at E7 of 21 HPV58 isolates.**

1. The group containing the prototype was assigned as E7-A. The HPV58 prototype (http://www.ncbi.nlm.nih.gov/genbank accession no. NC_001443) was used as reference.

2. No. (%) of each sequence pattern found among the 401 samples examined.

3. Nucleotide positions where variations were detected are written vertically across the top. Sites with no changes are marked with dashes. Capital letters indicate alternations result in an amino acid change. Lower case letters indicate synonymous mutations.

4. The one letter amino acid code is used. The position of amino acid change is stated numerically. The letter preceding this number refers to the reference amino acid, and the letter following refers to its substitution.

5. Lineage and sub-lineage classification based on full genomes or concatenated E6-E7-E2-E5-L1-PCR sequences described in previous studies. Some E7 sequence patterns are not lineage / sub-lineage-specific.

--, no sub-lineage assigned. NA, not applicable. Single nucleotide substitutions diagnostic for E7 variants are boxed.
GenBank accession numbers for E7 sequences are ## (pending).

Figure 4. Geographical distribution of HPV58 isolates according to E6 and E7 sequence variation patterns.

A. E6  B. E7
References


Table 1. Source and cervical pathology status of study samples

<table>
<thead>
<tr>
<th>Region</th>
<th>Country/city</th>
<th>Total no. of specimen</th>
<th>Normal</th>
<th>ASCUS ¹</th>
<th>LGSIL ²</th>
<th>HGSIL ³</th>
<th>Carcinoma</th>
<th>Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>Zimbabwe</td>
<td>69</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>69</td>
</tr>
<tr>
<td>Americas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>63</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>10</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Canada</td>
<td>10</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>United States of America</td>
<td></td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mexico</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Argentina</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Brazil</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Honduras</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Asia</td>
<td></td>
<td>238</td>
<td>57</td>
<td>23</td>
<td>57</td>
<td>66</td>
<td>31</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Mainland China</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Hong Kong</td>
<td>90</td>
<td>17</td>
<td>0</td>
<td>15</td>
<td>41</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Taiwan</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Korea</td>
<td>119</td>
<td>30</td>
<td>23</td>
<td>40</td>
<td>22</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Japan</td>
<td>14</td>
<td>9</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Thailand</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Europe</td>
<td></td>
<td>31</td>
<td>16</td>
<td>1</td>
<td>9</td>
<td>3</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Great Britain</td>
<td>14</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Italy</td>
<td>17</td>
<td>10</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

¹ ASCUS, atypical squamous cells of undetermined significance.
² LGSIL, low-grade squamous intraepithelial lesions, including cervical intraepithelial neoplasia grade 1 (CIN1), and mild dysplasia.
³ HGSIL, high-grade squamous intraepithelial lesions, including CIN2, CIN3, and severe dysplasia.
⁴ Cervical or vaginal swabs.
⁵ Anal swab samples.
Table 2. Univariate analysis on risk association of HPV58 E6 and E7 amino acid substitutions

<table>
<thead>
<tr>
<th>Amino acid substitution</th>
<th>Cervical pathology¹</th>
<th>P-value²</th>
<th>Odds ratio (95% CI)³</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal (N = 79)</td>
<td>CIN3/ICC (N = 64)</td>
<td></td>
</tr>
<tr>
<td>E6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E25K</td>
<td>0</td>
<td>1</td>
<td>0.448</td>
</tr>
<tr>
<td>E32Q</td>
<td>3</td>
<td>0</td>
<td>0.253</td>
</tr>
<tr>
<td>V46I</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
</tr>
<tr>
<td>V46T</td>
<td>1</td>
<td>0</td>
<td>1.000</td>
</tr>
<tr>
<td>D86E</td>
<td>4</td>
<td>1</td>
<td>0.380</td>
</tr>
<tr>
<td>D86N</td>
<td>1</td>
<td>0</td>
<td>1.000</td>
</tr>
<tr>
<td>K93N</td>
<td>6</td>
<td>4</td>
<td>0.987</td>
</tr>
<tr>
<td>C95R</td>
<td>1</td>
<td>0</td>
<td>1.000</td>
</tr>
<tr>
<td>N97D</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
</tr>
<tr>
<td>N97E</td>
<td>1</td>
<td>0</td>
<td>1.000</td>
</tr>
<tr>
<td>E7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R9K</td>
<td>0</td>
<td>2</td>
<td>0.199</td>
</tr>
<tr>
<td>T20I</td>
<td>15</td>
<td>27</td>
<td>0.004</td>
</tr>
<tr>
<td>G41R</td>
<td>46</td>
<td>25</td>
<td>0.035</td>
</tr>
<tr>
<td>T50I</td>
<td>1</td>
<td>0</td>
<td>1.000</td>
</tr>
<tr>
<td>G63D</td>
<td>47</td>
<td>25</td>
<td>0.024</td>
</tr>
<tr>
<td>G63H</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
</tr>
<tr>
<td>G63S</td>
<td>18</td>
<td>27</td>
<td>0.021</td>
</tr>
<tr>
<td>T74A</td>
<td>7</td>
<td>2</td>
<td>0.188</td>
</tr>
<tr>
<td>D76E</td>
<td>7</td>
<td>2</td>
<td>0.188</td>
</tr>
<tr>
<td>V77A</td>
<td>1</td>
<td>6</td>
<td>0.045</td>
</tr>
</tbody>
</table>

¹ Histology-confirmed cervical intraepithelial neoplasia grade III (CIN3) or invasive cervical cancer (ICC). Normal, no abnormalities detected by cytology or colposcopy or biopsy.

² 2x2 comparison between samples with and without the indicated amino acid substitution using chi-squared test or Fisher’s exact test. Significant associations are bolded.

³ --, odds ratio undefined; CI, confidence interval
Figure 1. Nucleotide sequence variations at E6 of 19 HPV58 isolates

<table>
<thead>
<tr>
<th>E6 sequence pattern</th>
<th>No. (%)</th>
<th>Nucleotide sequence</th>
<th>Amino acid sequence</th>
<th>Lineage</th>
<th>Sub-lineage</th>
</tr>
</thead>
<tbody>
<tr>
<td>E6-A1</td>
<td>296 (73.8%)</td>
<td>G C G G T C G C A T A T T T A</td>
<td>A A1 / A2</td>
<td>A1</td>
<td>A1 / A2</td>
</tr>
<tr>
<td>E6-A2</td>
<td>27 (6.7%)</td>
<td>- - - - - - t - - - - - - - -</td>
<td>A A1</td>
<td>A1</td>
<td>A1</td>
</tr>
<tr>
<td>E6-A3</td>
<td>15 (3.7%)</td>
<td>- - - - - - C - - - - - - - -</td>
<td>K93N</td>
<td>A</td>
<td>A2</td>
</tr>
<tr>
<td>E6-A4</td>
<td>2 (0.5%)</td>
<td>- t - - - - t - - - - - - - -</td>
<td>A A2 / A3</td>
<td>A2 / A3</td>
<td>A2 / A3</td>
</tr>
<tr>
<td>E6-A5</td>
<td>2 (0.5%)</td>
<td>- - - - - - t - - - - - - - -</td>
<td>c A</td>
<td>A2 / A3</td>
<td>A2 / A3</td>
</tr>
<tr>
<td>E6-A6</td>
<td>1 (0.2%)</td>
<td>- - - - - - t - - - - - - - -</td>
<td>c A</td>
<td>A2 / A3</td>
<td>A2 / A3</td>
</tr>
<tr>
<td>E6-A7</td>
<td>1 (0.2%)</td>
<td>- - - - - - t - - - - - - - -</td>
<td>g A</td>
<td>A2 / A3</td>
<td>A2 / A3</td>
</tr>
<tr>
<td>E6-A8</td>
<td>1 (0.2%)</td>
<td>- - - - - - t - - - - - - - -</td>
<td>G I105M</td>
<td>A</td>
<td>A2 / A3</td>
</tr>
<tr>
<td>E6-A9</td>
<td>1 (0.2%)</td>
<td>A - - - - - - t - - - - - - -</td>
<td>E25K</td>
<td>A</td>
<td>A1 / A2 / A3</td>
</tr>
<tr>
<td>E6-B1</td>
<td>2 (0.5%)</td>
<td>- - C - - - t - A - C - - - -</td>
<td>E32Q / D86E / K93N</td>
<td>B</td>
<td>B1</td>
</tr>
<tr>
<td>E6-B2</td>
<td>1 (0.2%)</td>
<td>- - C - - - t - C - - - -</td>
<td>E32Q / K93N</td>
<td>B</td>
<td>B1</td>
</tr>
<tr>
<td>E6-B3</td>
<td>1 (0.2%)</td>
<td>- - C - - - t - A - - - -</td>
<td>E32Q / D86E</td>
<td>B</td>
<td>B1</td>
</tr>
<tr>
<td>E6-C1</td>
<td>33 (8.2%)</td>
<td>- - - - - - t - - - - - - -</td>
<td>D86E</td>
<td>C</td>
<td>--</td>
</tr>
<tr>
<td>E6-C2</td>
<td>6 (1.5%)</td>
<td>- - - - - - t - A - - - -</td>
<td>D86E</td>
<td>B</td>
<td>B2</td>
</tr>
<tr>
<td>E6-C3</td>
<td>3 (0.7%)</td>
<td>- - C - - - t - A - - - G</td>
<td>V46L / D86E / N97D</td>
<td>C</td>
<td>--</td>
</tr>
<tr>
<td>E6-D1</td>
<td>5 (1.2%)</td>
<td>- - A C t - - - C G G - -</td>
<td>V46T / C95R / N97E</td>
<td>D</td>
<td>D2</td>
</tr>
<tr>
<td>E6-D2</td>
<td>2 (0.5%)</td>
<td>- - A t - - - G - -</td>
<td>V46I / N97D</td>
<td>D</td>
<td>D1</td>
</tr>
<tr>
<td>E6-D3</td>
<td>1 (0.2%)</td>
<td>- - A C t - - - G - -</td>
<td>V46T / N97K</td>
<td>D</td>
<td>D2</td>
</tr>
<tr>
<td>E6-D4</td>
<td>1 (0.2%)</td>
<td>- - A C t A - - - C G G - -</td>
<td>V46T / D86N / C95R / N97E</td>
<td>D</td>
<td>D2</td>
</tr>
</tbody>
</table>
Figure 2. Phylogenetic trees of E6 and E7 of HPV58.
The maximum-likelihood trees were constructed by MEGA5 using the close-neighbor-interchange (CNI) search approach. Bootstrap values of ≥ 50% generated by 1,000 resamplings are shown.

A. Constructed from 450-bp E6 nucleotide sequences of 19 HPV58 isolates. The length of scale bar represents 0.004 substitutions per nucleotide position. The sample HK969 with E6 nucleotide sequence identical to the prototype (GenBank accession no. NC_001443) is indicated by an arrow. The group containing the prototype sequence was assigned as E6-A. Sample CAN301 was grouped into sequence pattern E6-A.

1159x813mm (96 x 96 DPI)
B. Constructed from 297-bp E7 nucleotide sequences of 21 HPV58 isolates. The length of scale bar represents 0.004 substitutions per nucleotide position. The sample HK523 with nucleotide sequence identical to the prototype (GenBank accession no. NC_001443) is indicated by an arrow. The group containing the prototype sequence was assigned as E7-A.
Figure 3. Nucleotide sequence variations at E7 of 21 HPV58 isolates

<table>
<thead>
<tr>
<th>E7 sequence pattern^1</th>
<th>No. (%)^2</th>
<th>Nucleotide sequence^3</th>
<th>Amino acid sequence^4</th>
<th>Lineage^5</th>
<th>Sub-lineage^5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5 6 6 7 7 7 7 7 8 8 8 8 9 2 2 4 5 6 6 7 9 9 0 0 4 4 5 9 2 4 0 6 6 0 1 3 5 3 8 1 3 0 9 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>G C G C T T T G A T A C C T C C</td>
<td>A</td>
<td>A1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E7-A1</td>
<td>24 (6.0%)</td>
<td>- - - - - - - - - -</td>
<td>G41R / G63D</td>
<td>A</td>
<td>A1/A3</td>
</tr>
<tr>
<td>E7-A2</td>
<td>12 (3.0%)</td>
<td>- - - - - g - - - - - C</td>
<td>G41R / G63D</td>
<td>A</td>
<td>A2</td>
</tr>
<tr>
<td>E7-A3</td>
<td>10 (2.5%)</td>
<td>- - - - - g - - - - -</td>
<td>G41R / G63D</td>
<td>A</td>
<td>A1</td>
</tr>
<tr>
<td>E7-A4</td>
<td>3 (0.7%)</td>
<td>- - - - c - - - - -</td>
<td>G41R / G63D</td>
<td>A</td>
<td>A1</td>
</tr>
<tr>
<td>E7-A5</td>
<td>1 (0.2%)</td>
<td>- - - - - c - - - - -</td>
<td>G41R / G63D</td>
<td>A</td>
<td>A2</td>
</tr>
<tr>
<td>E7-B1</td>
<td>196 (48.9%)</td>
<td>- - A - - - g - - A - - - - -</td>
<td>G41R / G63D</td>
<td>A</td>
<td>A2</td>
</tr>
<tr>
<td>E7-B2</td>
<td>5 (1.2%)</td>
<td>- - A - - - g c - - A - - - - -</td>
<td>G41R / G63D</td>
<td>A</td>
<td>A2</td>
</tr>
<tr>
<td>E7-B3</td>
<td>3 (0.7%)</td>
<td>- - A - - - g - - A - - - - t</td>
<td>G41R / G63D</td>
<td>A</td>
<td>A2</td>
</tr>
<tr>
<td>E7-B4</td>
<td>1 (0.2%)</td>
<td>- - A - - - g - - A - - c - -</td>
<td>G41R / G63D</td>
<td>A</td>
<td>A2</td>
</tr>
<tr>
<td>E7-B5</td>
<td>1 (0.2%)</td>
<td>- - A - - - g - - A A - - - -</td>
<td>G41R / G63N</td>
<td>A</td>
<td>A2</td>
</tr>
<tr>
<td>E7-C1</td>
<td>86 (21.4%)</td>
<td>T - - - - - g A - - - - -</td>
<td>T20I / G63S</td>
<td>A</td>
<td>A3</td>
</tr>
<tr>
<td>E7-C2</td>
<td>3 (0.7%)</td>
<td>A T - - - - - g A - - - - -</td>
<td>R9K / T20I / G63S</td>
<td>A</td>
<td>A3</td>
</tr>
<tr>
<td>E7-D1</td>
<td>33 (8.2%)</td>
<td>- - - - - - g - - - - - - G T A - t - c</td>
<td>T74A / D76E</td>
<td>C</td>
<td>--</td>
</tr>
<tr>
<td>E7-D2</td>
<td>6 (1.5%)</td>
<td>- - - - - - g - - A G - - - - - - G T A - t - -</td>
<td>G63D / T64A / T74A / D76E</td>
<td>D</td>
<td>D2</td>
</tr>
<tr>
<td>E7-D3</td>
<td>5 (1.2%)</td>
<td>- - - - - - c g - - - - - - G T A - t - -</td>
<td>T74A / D76E</td>
<td>B</td>
<td>B2</td>
</tr>
<tr>
<td>E7-D4</td>
<td>4 (1.0%)</td>
<td>- - - - - - g - - A - - - - - - G T A - - - -</td>
<td>G63S / T74A / D76E</td>
<td>B</td>
<td>B1</td>
</tr>
<tr>
<td>E7-D5</td>
<td>3 (0.7%)</td>
<td>A - - - - - - g - - A - - - - - - G T A - t - c</td>
<td>R9K / T74A / D76E</td>
<td>C</td>
<td>--</td>
</tr>
<tr>
<td>E7-D6</td>
<td>2 (0.5%)</td>
<td>- - - - - - g - - C A - - - - - - G T A - t - -</td>
<td>G63H / T74A / D76E</td>
<td>D</td>
<td>D1</td>
</tr>
<tr>
<td>E7-D7</td>
<td>1 (0.2%)</td>
<td>- - - - - - g - - - - - - G - - - - - -</td>
<td>T74A</td>
<td>A</td>
<td>A2</td>
</tr>
<tr>
<td>E7-D8</td>
<td>1 (0.2%)</td>
<td>- - - - - - g - - A T - - - - - - G T A - t - -</td>
<td>G63D / T74A / D76E</td>
<td>D</td>
<td>D2</td>
</tr>
<tr>
<td>E7-D9</td>
<td>1 (0.2%)</td>
<td>- - - - - - T c g - - - - - - G T A - t - -</td>
<td>T50I / T74A / D76E</td>
<td>B/C/D</td>
<td>NA</td>
</tr>
</tbody>
</table>
Figure 4. Geographical distribution of HPV58 isolates according to E6 and E7 sequence variation patterns.

A. E6

824x1119mm (96 x 96 DPI)
B. E7

753x1050mm (96 x 96 DPI)