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

Paul K.S. Chan[1], **Chuqing Zhang**[1], **Jong-Sup Park**[2], **Karen K. Smith-McCune**[3,4], **Joel M. Palefsky**[5], **Lucia Giovannelli**[6], **Francois Coutlée**[7], **Samantha Hibbitts**[8], **Ryo Konno**[9], **Wannapa Settheetham-Ishida**[10], **Tang-Yuan Chu**[11], **Annabelle Ferrera**[12], **María Alejandra Picconi**[13], **Federico De Marco**[14], **Yin-Ling Woo**[15], **Tainá Raiol**[16], **Patricia Piña-Sánchez**[17], **Jeong-Hoon Bae**[2], **Martin C.S. Wong**[18], **Mike Z. Chirenje**[19], **Tsitsi Magure**[19], **Anna-Barbara Moscicki**[4,20], **Alison N. Fiander**[8], **Giuseppina Capra**[6], **Eun Young Ki**[2], **Yi Tan**[21], **Zigui Chen**[22], **Robert D. Burk**[22], **Martin C.W. Chan**[1], **Tak-Hong Cheung**[23], **David Pim**[24], and **Lawrence Banks**[24]

[1]Department of Microbiology, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong Special Administrative Region, People's Republic of China [2]Department of Obstetrics and Gynecology, Seoul St. Mary's Hospital, The Catholic University of Korea, Seoul, Korea [3]Department of Obstetrics, Gynecology and Reproductive Sciences, University of California San Francisco, San Francisco, CA [4]Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA [5]Department of Medicine, University of California San Francisco, San Francisco, CA [6]Sezione di Microbiologia, Dipartimento di Scienze per la Promozione della Salute, Azienda Ospedaliera Universitaria Policlinico P. Giaccone, Palermo, Italy [7]Département de Microbiologie et Infectiologie, Centre Hospitalier de l'Université de Montréal, Montréal, Québec, Canada [8]HPV Research Group, Cancer and Genetics Research Institute, School of Medicine, Cardiff University, Cardiff, Wales, United Kingdom [9]Department of Obstetrics and Gynecology, Jichi Medical University, Saitama Medical Center, Saitama, Japan [10]Department of Physiology, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand [11]Department of Obstetrics and Gynaecology, Buddhist Tzu Chi General Hospital, Institute of Medical Sciences, Tzu Chi University, Hualien, Taiwan [12]School of Microbiology, Science Faculty, Universidad Nacional Autónoma de Honduras, Tegucigalpa, Honduras [13]Servicio Virus Oncogénicos, Departamento Virología, Instituto Nacional de Enfermedades Infecciosas, ANLIS "Carlos G. Malbrán," Buenos Aires, Argentina [14]Laboratory of Virology, The "Regina Elena" National Cancer Institute, Rome, Italy [15]Department of Obstetrics and Gynaecology, University of Malaya, Kuala Lumpur, Malaysia [16]Department of Cellular Biology, Institute of Biology, University of Brasilia, Brasilia, Brazil [17]Unidad de Investigación Médica en Enfermedades Oncológicas, Hospital de Oncología, CMN Siglo XXI, México D.F [18]School of Public Health and Primary Care, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong Special Administrative Region, People's Republic of China [19]Department of Obstetrics and Gynecology, University of Zimbabwe, Harare, Zimbabwe [20]Department of Pediatrics, University of California San Francisco, San Francisco, CA [21]Stanley Ho Centre for Emerging Infectious Diseases, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong Special Administrative Region, People's Republic of China [22]Albert Einstein College of Medicine, Bronx, NY [23]Department of Obstetrics and Gynaecology, Prince of Wales Hospital, Hong Kong Special Administrative Region, People's Republic of China [24]International Centre for Genetic Engineering and Biotechnology, Trieste, Italy

Correspondence to: Paul K.S. 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, paulkschan@cuhk.edu.hk.

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 our 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.

Keywords

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.[1] 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.[2,3] 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),[4–9] whereas it only accounts for about 2% of cervical cancers in other regions.[2,3] 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,[10] 21% in Southeastern Mexico,[11] 9.5% in Northeastern Brazil[12] and was the second most common type (about 13%) found in precancerous lesions in Central and Southern Brazil.[13,14]

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.[15,16] Studies comparing the risk association of these variant lineages are very limited and a conclusion has not yet been reached.[17,18]

The E6 and E7 proteins encoded by high-risk HPV play a key role in cellular transformation. Previous studies from Eastern[19] and Southern Chinese populations[20] 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.[18–20] Our 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.

Material 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 our study (Table 1). These samples have been used to establish a lineage classification system for HPV58 variants.[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 GACCGA 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, forward and reverse primers at 0.25 μM each, 1.25 unit of HotStar*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 Terminator v.3.1 sequencing reaction mix (Applied Biosystems, Foster City, USA), 3 μL of 5× 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) and run on an ABI 3130 automated sequence analyzer (Applied Biosystems). Sequence data were obtained from both directions and analyzed with SeqScape v2.5 (Applied Biosystems). 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.[21] 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.[22,23]

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 χ^2 test and using the Bonferroni method to account for multiple comparisons.[24] 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 nonsynonymous substitutions was first assessed by univariate analysis using χ^2 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 USA. 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.[16] In our study also, there were 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 our study.

Results

E6 sequence variation

Altogether, 19 variants showing nucleotide sequence variations in the E6 ORF were identified (Fig. 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 nonsynonymous 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 intergroup nucleotide sequence variability range from 0.9% (E6-A and -B) to 2.0% (E6-B and -D), whereas the maximum intragroup variability range 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 (Fig. 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 nonsynonymous substitutions (Fig. 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 (Fig. 4a). In Asia, Americas and Europe, E6-A predominated and accounted for 86–96% of isolates. In Africa, E6-A (51%) cocirculated 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 three regions).

The geographical distribution of E7 sequence patterns was more diverse (Fig. 4b). In Asia, both E7-B and E7-C patterns were commonly found, whereas E7-B and E7-D patterns codominated 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 (T20I and G63S) 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

Our study has delineated the sequence variations of two onco-protein-encoding genes of HPV58. This dataset 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 our 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.[15,25–27] Phylogenetically related E6 and E7 sequence patterns could be identified by distinct signatures of nonsynonymous 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.[28,29] 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.[30] This strain is regarded as the prototype and commonly used for primer design. The E6-A2 and E7-A1 sequence pattern, as assigned by our 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 codominated with E6-C that was uncommon elsewhere. Two characteristics in the distribution of E7 sequence variation patterns were observed. First, 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. Our 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 harbored 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.[20] Another study from eastern China (Zhejiang) also reported a significant positive trend of association between these substitutions and the severity of cervical lesions.[19]

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).[15,16] Although, naturally occurring LCR sequence variations have been shown to carry different promoter activity *in vitro*,[31] 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 sublineages.[16,18] This is special for HPV58 as an association between LCR polymorphisms and lesion has been reported for other types. Our 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.

Acknowledgments

References

1. Ferlay J, Shin HR, Bray F, et al. Estimates of worldwide burden of cancer in 2008 : GLOBOCAN 2008. Int J Cancer. 2010; 127:2893–917. [PubMed: 21351269]

2. de Sanjose S, Quint WG, Alemany L, et al. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. Lancet Oncol. 2010; 11:1048–56. [PubMed: 20952254]

3. Li N, Franceschi S, Howell-Jones R, et al. Human papillomavirus type distribution in 30,848 invasive cervical cancers worldwide: variation by geographical region, histological type and year of publication. Int J Cancer. 2011; 128:927–35. [PubMed: 20473886]

4. Shah W, Hongwei C, Jin Z, et al. The prevalence of human papillomavirus type 58 in Chinese patients with cervical carcinoma and its influence on survival. Clin Oncol (R Coll Radiol). 2009; 21:768–74. [PubMed: 19875274]

5. Huang S, Afonina I, Miller BA, et al. Human papillomavirus types 52 and 58 are prevalent in cervical cancers from Chinese women. Int J Cancer. 1997; 70:408–11. [PubMed: 9033647]

6. Hwang T. Detection and typing of human papillomavirus DNA by PCR using consensus primers in various cervical lesions of Korean women. J Korean Med Sci. 1999; 14:593–9. [PubMed: 10642935]

7. Chan PK, Ho WC, Yu MY, et al. Distribution of human papillomavirus types in cervical cancers in Hong Kong: current situation and changes over the last decades. Int J Cancer. 2009; 125:1671–7. [PubMed: 19507252]

8. Ding DC, Hsu HC, Huang RL, et al. Type-specific distribution of HPV along the full spectrum of cervical carcinogenesis in Taiwan: an indication of viral oncogenic potential. Eur J Obstet Gynecol Reprod Biol. 2008; 140:245–51. [PubMed: 18502563]

9. Asato T, Maehama T, Nagai Y, et al. A large case-control study of cervical cancer risk associated with human papillomavirus infection in Japan, by nucleotide sequencing-based genotyping. J Infect Dis. 2004; 189:1829–32. [PubMed: 15122519]

10. Herrero R, Hildesheim A, Bratti C, et al. Population-based study of human papillomavirus infection and cervical neoplasia in rural Costa Rica. J Natl Cancer Inst. 2000; 92:464–74. [PubMed: 10716964]

11. Canche JC, López IR, Suárez NG, et al. High prevalence and low E6 genetic variability of human papillomavirus 58 in women with cervical cancer and precursor lesions in Southeast Mexico. Mem Inst Oswaldo Cruz. 2010; 105:144–8. [PubMed: 20428672]

12. Fernandes JV, Meissner RV, Carvalho MG, et al. Prevalence of human papillomavirus in archival samples obtained from patients with cervical pre-malignant and malignant lesions from Northeast Brazil. BMC Res Notes. 2010; 3:96. [PubMed: 20377903]

13. Camara GN, Cerqueira DM, Oliveira AP, et al. Prevalence of human papillomavirus types in women with pre-neoplastic and neoplastic cervical lesions in the Federal District of Brazil. Mem Inst Oswaldo Cruz. 2003; 98:879–83. [PubMed: 14762511]

14. Paesi S, Serafini EP, Barea F, et al. High prevalence of human papillomavirus type 58 in patients with cervical pre-malignant lesions in southern Brazil. J Med Virol. 2009; 81:1270–5. [PubMed: 19475616]

15. Chen Z, Schiffman M, Herrero R, et al. Evolution and taxonomic classification of human papillomavirus 16 (HPV16)-related variant genomes: HPV31, HPV33, HPV35, HPV52, HPV58 and HPV67. PLoS One. 2011; 6:e20183. [PubMed: 21673791]

16. Chan PK, Luk AC, Park JS, et al. Identification of human papillomavirus type 58 lineages and the distribution worldwide. J Infect Dis. 2011; 203:1565–73. [PubMed: 21592985]

17. Schiffman M, Rodriguez AC, Chen Z, et al. A population-based prospective study of carcinogenic human papillomavirus variant lineages, viral persistence, and cervical neoplasia. Cancer Res. 2010; 70:3159–69. [PubMed: 20354192]

18. Chang YJ, Chen HC, Lee BH, et al. CBCSP-HPV Study Group. Unique variants of human papillomavirus genotypes 52 and 58 and risk of cervical neoplasia. Int J Cancer. 2011; 129:965–73. [PubMed: 20949622]

19. Ding T, Wang X, Ye F, et al. Distribution of human papillomavirus 58 and 52 E6/E7 variants in cervical neoplasia in Chinese women. Gynecol Oncol. 2010; 119:436–43. [PubMed: 20870281]

20. Chan PK, Lam CW, Cheung TH, et al. Association of human papillomavirus type 58 variant with the risk of cervical cancer. J Natl Cancer Inst. 2002; 94:1249–53. [PubMed: 12189229]

21. Tamura K, Peterson D, Peterson N, et al. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011; 28:2731–9. [PubMed: 21546353]

22. Delport W, Poon AF, Frost SD, et al. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics. 2010; 26:2455–7. [PubMed: 20671151]

23. Pond SL, Frost SD. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics. 2005; 21:2531–3. [PubMed: 15713735]

24. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ. 1995; 310:170. [PubMed: 7833759]

25. Stewart AC, Eriksson AM, Manos MM, et al. Intratype variation in 12 human papillomavirus types: a worldwide perspective. J Virol. 1996; 70:3127–36. [PubMed: 8627792]

26. Prado JC, Calleja-Macias IE, Bernard HU, et al. Worldwide genomic diversity of the human papillomaviruses-53, 56, and 66, a group of high-risk HPVs unrelated to HPV-16 and HPV-18. Virology. 2005; 340:95–104. [PubMed: 16039686]

27. Calleja-Macias IE, Villa LL, Prado JC, et al. Worldwide genomic diversity of the high-risk human papillomavirus types 31, 35, 52, and 58, four close relatives of human papillomavirus type 16. J Virol. 2005; 79:13630–40. [PubMed: 16227283]

28. Chen Z, Terai M, Fu L, et al. Diversifying selection in human papillomavirus type 16 lineages based on complete genome analyses. J Virol. 2005; 79:7014–23. [PubMed: 15890941]

29. Chan PK, Lam CW, Cheung TH, et al. Human papillomavirus type 16 intratypic variant infection and risk for cervical neoplasia in southern China. J Infect Dis. 2002; 186:696–700. [PubMed: 12195358]

30. Matsukura T, Sugase M. Molecular cloning of a novel human papillomavirus (type 58) from an invasive cervical carcinoma. Virology. 1990; 177:833–6. [PubMed: 2164740]

31. Raiol T, de Amorim RM, Galante P, et al. HPV-58 molecular variants exhibit different transcriptional activity. Intervirology. 2011; 54:146–50. [PubMed: 20948225]

What's new?

For unknown reasons, human papillomavirus (HPV) 58 accounts for a significant proportion of cervical cancers in East Asia and parts of Latin America, but it is uncommon elsewhere. In this study, the authors analyzed the E6 and E7 genes from HPV58 samples from around the world. They found that patients with an HPV58 variant containing T20I and G63S substitutions in E7 had a greater risk of developing cancer. This variant was also more prevalent in Asia than in other regions, and may therefore help to explain the higher disease burden observed in this region.

E6 sequence pattern[1]	No. (%)[2]	Nucleotide sequence[3]																Amino acid sequence[4]	Lineage[5]	Sub-lineage[5]
		1	1	2	2	2	3	3	3	3	3	3	3	4	4	4	4			
		8	8	0	4	4	0	6	6	7	8	9	9	0	1	2	7			
		2	7	3	5	6	7	5	7	3	8	2	8	0	8	4	9			
Reference		G	C	G	G	T	C	G	C	A	A	T	A	T	T	T	A		A	A1
E6-A1	296 (73.8%)	-	-	-	-	-	t	-	-	-	-	-	-	-	-	-	-		A	A1 / A2
E6-A2	27 (6.7%)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		A	A1
E6-A3	15 (3.7%)	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-	-	K93N	A	A2
E6-A4	2 (0.5%)	-	t	-	-	-	t	-	-	-	-	-	-	-	-	-	-		A	A2 / A3
E6-A5	2 (0.5%)	-	-	-	-	-	t	-	-	-	-	-	-	-	-	-	c		A	A2 / A3
E6-A6	1 (0.2%)	-	-	-	-	-	t	-	-	-	-	-	-	c	-	-			A	A2 / A3
E6-A7	1 (0.2%)	-	-	-	-	-	t	-	-	g	-	-	-	-	-	-	-		A	A2 / A3
E6-A8	1 (0.2%)	-	-	-	-	-	t	-	-	-	-	-	-	-	-	G	-	I105M	A	A2 / A3
E6-A9	1 (0.2%)	A	-	-	-	-	t	-	-	-	-	-	-	-	-	-	-	E25K	A	A1 / A2 / A3
E6-B1	2 (0.5%)	-	-	C	-	-	t	-	A	-	C	-	-	-	-	-	-	E32Q / D86E / K93N	B	B1
E6-B2	1 (0.2%)	-	-	C	-	-	t	-	-	-	C	-	-	-	-	-	-	E32Q / K93N	B	B1
E6-B3	1 (0.2%)	-	-	C	-	-	t	-	A	-	-	-	-	-	-	-	-	E32Q / D86E	B	B1
E6-C1	33 (8.2%)	-	t	-	-	-	t	-	A	-	-	-	-	-	-	-	-	D86E	C	--
E6-C2	6 (1.5%)	-	-	-	-	-	t	-	A	-	-	-	-	-	-	-	-	D86E	B	B2
E6-C3	3 (0.7%)	-	t	-	C	-	t	-	A	-	-	-	G	-	-	-	-	V46L / D86E / N97D	C	--
E6-D1	5 (1.2%)	-	-	-	A	C	t	-	-	-	-	C	G	G	-	-	-	V46T / C95R / N97E	D	D2
E6-D2	2 (0.5%)	-	-	-	A	-	t	-	-	-	-	-	G	-	-	-	-	V46I / N97D	D	D1
E6-D3	1 (0.2%)	-	-	-	A	C	t	-	-	-	-	-	-	G	-	-	-	V46T / N97K	D	D2
E6-D4	1 (0.2%)	-	-	-	A	C	t	A	-	-	-	C	G	G	-	-	-	V46T / D86N / C95R / N97E	D	D2

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 sublineage classification based on full genome or concatenated E6-E7-E2-E5-L1-PCR sequences described in previous studies.[15,16] Some E6 sequence patterns are not sublineage-specific. –, no sublineage assigned. Single nucleotide substitutions diagnostic for E6 variants are boxed. GenBank accession numbers for E6 sequences are KC190273 - KC190291.

Figure 2.
Phylogenetic trees of E6 and E7 of HPV58. The maximum-likelihood trees were constructed by MEGA5 using the 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.

E7 sequence pattern[1]	No. (%)[2]	Nucleotide sequence[3]																		Amino acid sequence[4]	Lineage[5]	Sub-lineage[5]
		5	6	6	7	7	7	7	7	7	7	7	7	7	8	8	8	8	8			
		9	3	9	2	2	4	5	6	6	6	7	9	9	0	0	4	4	5			
		9	2	4	2	6	4	6	0	1	3	5	3	8	1	3	0	9	2			
Reference		G	C	G	C	T	T	T	G	G	A	T	A	C	C	T	C	C	T		A	A1
E7-A1	24 (6.0%)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		A	A1/A3
E7-A2	12 (3.0%)	-	-	-	-	-	g	-	-	-	-	-	-	-	-	C	-	-	-	V77A	A	A2
E7-A3	10 (2.5%)	-	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	-	-		A	A1
E7-A4	3 (0.7%)	-	-	-	-	c	-	-	-	-	-	-	-	-	-	-	-	-	-		A	A1
E7-A5	1 (0.2%)	-	-	-	-	-	-	c	-	-	-	-	-	-	-	-	-	-	-		A	A2
E7-B1	196 (48.9%)	-	-	A	-	-	g	-	-	A	-	-	-	-	-	-	-	-	-	G41R / G63D	A	A2
E7-B2	5 (1.2%)	-	-	A	-	-	g	c	-	A	-	-	-	-	-	-	-	-	-	G41R / G63D	A	A2
E7-B3	3 (0.7%)	-	-	A	-	-	g	-	-	A	-	-	-	-	-	-	t	-	G41R / G63D	A	A2
E7-B4	1 (0.2%)	-	-	A	-	-	g	-	-	A	-	c	-	-	-	-	-	-	-	G41R / G63D	A	A2
E7-B5	1 (0.2%)	-	-	A	-	-	g	-	A	A	-	-	-	-	-	-	-	-	-	G41R / G63N	A	A2
E7-C1	86 (21.4%)	-	T	-	-	-	g	-	A	-	-	-	-	-	-	-	-	-	-	T20I / G63S	A	A3
E7-C2	3 (0.7%)	A	T	-	-	-	g	-	A	-	-	-	-	-	-	-	-	-	-	R9K / T20I / G63S	A	A3
E7-D1	33 (8.2%)	-	-	-	-	-	g	-	-	-	-	-	G	T	A	-	t	-	c	T74A / D76E	C	--
E7-D2	6 (1.5%)	-	-	-	-	-	g	-	-	A	G	-	G	T	A	-	t	-	-	G63D / T64A / T74A / D76E	D	D2
E7-D3	5 (1.2%)	-	-	-	-	c	g	-	-	-	-	-	G	T	A	-	t	-	-	T74A / D76E	B	B2
E7-D4	4 (1.0%)	-	-	-	-	-	g	-	A	-	-	-	G	T	A	-	-	-	-	G63S / T74A / D76E	B	B1
E7-D5	3 (0.7%)	A	-	-	-	-	g	-	-	-	-	-	G	T	A	-	t	-	c	R9K / T74A / D76E	C	--
E7-D6	2 (0.5%)	-	-	-	-	-	g	-	C	A	-	-	G	T	A	-	t	-	-	G63H / T74A / D76E	D	D1
E7-D7	1 (0.2%)	-	-	-	-	-	g	-	-	-	-	-	G	-	-	-	-	-	-	T74A	A	A2
E7-D8	1 (0.2%)	-	-	-	-	-	g	-	-	A	T	-	G	T	A	-	t	-	-	G63D / T74A / D76E	D	D2
E7-D9	1 (0.2%)	-	-	-	T	c	g	-	-	-	-	-	G	T	A	-	t	-	-	T50I / T74A / D76E	B/C/D	NA

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 sublineage classification based on full genomes or concatenated E6-E7-E2-E5-L1-PCR sequences described in previous studies.[15,16] Some E7 sequence patterns are not lineage/sublineage-specific. –, no sublineage assigned. NA, not applicable. Single nucleotide substitutions diagnostic for E7 variants are boxed. GenBank accession numbers for E7 sequences are JX896403 - JX896423.
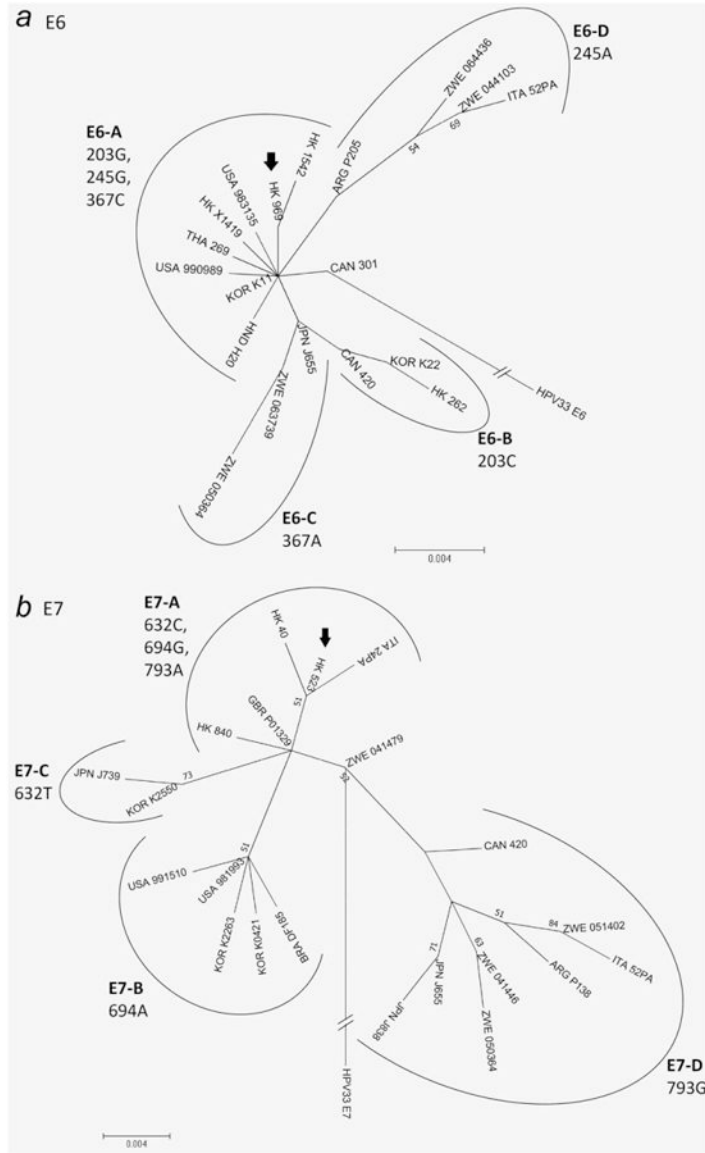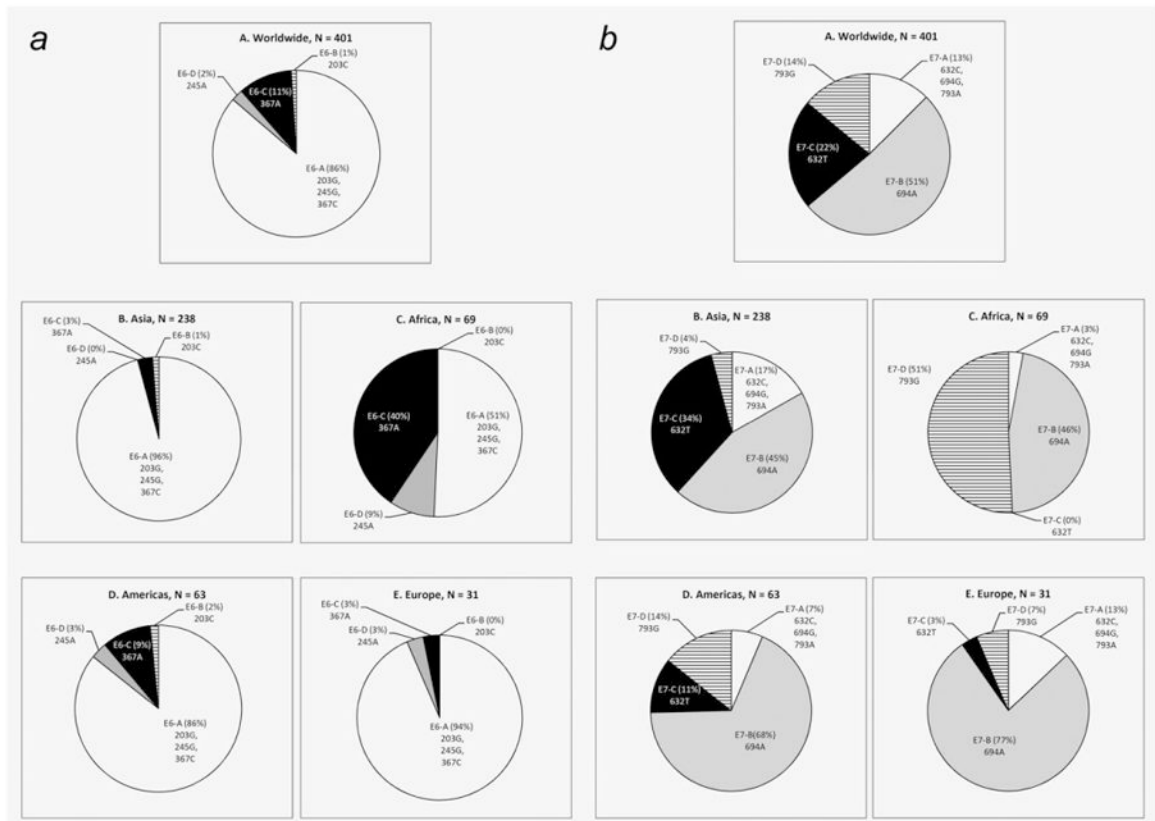
Figure 4.
Geographical distribution of HPV58 isolates according to E6 and E7 sequence variation
patterns. (*a*) *E6*; (*b*) *E7*.

Table 1

Source and cervical pathology status of study samples

Region	Country/city	Total no. of specimen	No. of specimen per cervical pathology status					
			Normal	ASCUS[1]	LGSIL[2]	HGSIL[3]	Carcinoma	Unknown
Africa	Zimbabwe[4]	69	0	0	0	0	0	69
Americas		63	6	1	6	10	2	1
	Canada	10	2	0	4	4	0	0
	USA[5]	37						
	Mexico	2	0	0	0	1	1	0
	Argentina	6	3	1	0	0	1	1
	Brazil	3	1	0	1	1	0	0
	Honduras	5	0	0	1	4	0	0
Asia		238	57	23	57	66	31	4
	Mainland China	3	0	0	0	0	0	3
	Hong Kong	90	17	0	15	41	17	0
	Taiwan	6	0	0	0	0	6	0
	Korea	119	30	23	40	22	4	0
	Japan	14	9	0	2	1	1	1
	Thailand	6	1	0	0	2	3	0
Europe		31	16	1	9	3	0	2
	Great Britain	14	6	0	6	2	0	0
	Italy	17	10	1	3	1	0	2

[1] ASCUS, atypical squamous cells of undetermined significance.

[2] LGSIL, low-grade squamous intraepithelial lesions, including 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

Amino acid substitution	Cervical pathology[1]		p-value[2]	Odds ratio (95% CI)[3]
	Normal (N = 79)	CIN3/ICC (N = 64)		
E6				
E25K	0	1	0.448	–
E32Q	3	0	0.253	–
V46I	1	1	1.000	1.24 (0.00–46.34)
V46T	1	0	1.000	–
D86E	4	1	0.380	0.30 (0.01–2.94)
D86N	1	0	1.000	–
K93N	6	4	0.987	0.81 (0.18–3.45)
C95R	1	0	1.000	–
N97D	1	1	1.000	1.24 (0.00–46.34)
N97E	1	0	1.000	–
E7				
R9K	0	2	0.199	–
T20I	**15**	**27**	**0.004**	**3.11 (1.38–7.07)**
G41R	**46**	**25**	**0.035**	**0.46 (0.22–0.95)**
T50I	1	0	1.000	–
G63D	**47**	**25**	**0.024**	**0.44 (0.21–0.90)**
G63H	1	1	1.000	1.24 (0.00–46.34)

Amino acid substitution	Cervical pathology[1]		p-value[2]	Odds ratio (95% CI)[3]
	Normal (*N* = 79)	CIN3/ICC (*N* = 64)		
G63S	**18**	**27**	**0.021**	**2.47 (1.13–5.44)**
T74A	7	2	0.188	0.33 (0.05–1.84)
D76E	7	2	0.188	0.33 (0.05–1.84)
V77A	1	6	0.045	8.07 (0.92–182.76)

[1] Histology-confirmed CIN3 or invasive cervical cancer. Normal, no abnormalities detected by cytology or colposcopy or biopsy.

[2] 2×2 comparison between samples with and without the indicated amino acid substitution using χ^2 test or Fisher's exact test. Significant associations are bolded.

[3] –, odds ratio undefined; CI, confidence interval.