Geographical distribution and risk association of human papillomavirus 52 variant lineages

Chuqing Zhang1, Jong-Sup Park2, Magdalena Grce3, Samantha Hibbitts4, Joel M. Palefsky5, Ryo Konno6, Karen K. Smith-McCune7,8, Lucia Giovannelli9, Tang-Yuan Chu10, María Alejandra Picconi11, Patricia Piña-Sánchez12, Wannapa Setheetham-Ishida13, Francois Coutlée14, Federico De Marco15, Yin-Ling Woo16, Wendy C. S. Ho1, Martin C. S. Wong17, Mike Z. Chirenje18, Tsitsi Magure18, Anna-Barbara Moscicki8,19, Ivan Sabol3, Alison N. Fiander4, Ziguí Chen20, Martin C. W. Chan1, Tak-Hong Cheung21, Robert D. Burk20, Paul K. S. Chan1,*

1Department of Microbiology, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong Special Administrative Region, China

2Department of Obstetrics and Gynecology, Seoul St. Mary's Hospital, The Catholic University of Korea, Seoul, Korea

3Division of Molecular Medicine, Rudjer Boskovic Institute, Zagreb, Croatia

4HPV Research Group, Cancer and Genetics Research Institute, School of Medicine, Cardiff University, United Kingdom

5Department of Medicine, University of California San Francisco, California, United States of America

6Department of Obstetrics and Gynecology, Jichi Medical University, Saitama Medical Center, Saitama, Japan

7Department of Obstetrics, Gynecology and Reproductive Sciences, University of California San Francisco, California, United States of America
8 Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, California, United States of America

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

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

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

12 Unidad de Investigación Médica en Enfermedades Oncológicas, Hospital de Oncología, CMN Siglo XXI, México D. F.

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

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

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

16 Department of Obstetrics and Gynaecology, Faculty of Medicine, Affiliated to UM Cancer Research Institute, University of Malaya, Malaysia

17 School of Public Health and Primary Care, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong Special Administrative Region, China

18 Department of Obstetrics and Gynecology, University of Zimbabwe, Harare, Zimbabwe

19 Department of Pediatrics, University of California San Francisco, California, United States of America

20 Albert Einstein College of Medicine, Bronx, New York, United States
Abstract

Human papillomavirus (HPV) 52 is commonly found in Asian cervical cancers, but rare elsewhere. Analysis of 611 isolates collected worldwide revealed a remarkable geographical distribution, with lineage B predominating in Asia (89.0% vs. 0-5.5%, $P_{\text{corrected}}<0.001$); whereas Africa, Americas and Europe were predominated by lineage A. We propose to name lineage B as “Asian” lineage to signify this feature. Preliminary analysis suggested a higher disease risk for lineage B, though ethnogeographical confounders could not be excluded. Further studies to verify whether the reported high disease attribution of HPV52 in Asia is due to the high prevalence of lineage B are warranted.
Introduction

Overall, human papillomavirus (HPV) 52 ranks the sixth or seventh among cervical cancers worldwide [1, 2]. However, studies from East Asia have reported a much higher ranking of HPV52. For instance, HPV52 was the third in squamous cell carcinoma, and the second among cervical intraepithelial neoplasia (CIN) 2 and CIN3 in Hong Kong [3]. Furthermore, HPV52 was the most common type in cervical cancers from Shanghai [4], and second in Taiwan [5] and Japan [6]. The underlying reason for such geographical predilection in disease attribution remains obscure. We investigated the geographical distribution and risk association of HPV52 variant lineages using a large series of samples collected worldwide to improve our knowledge on this non-vaccine-targeted HPV type.

Material and Methods

Study Samples

Cervical and vaginal samples from women or anal samples from men tested positive for HPV52 were transferred to a central laboratory for sequence analysis. DNA quality was assessed by amplifying a 932-bp fragment of LCR, and HPV type was ascertained by demonstrating a nucleotide sequence similarity of >90% compared with HPV52 prototype (GenBank accession no. X74481). The local institutional research ethics committee approved the collection of samples. Samples used in this study were sent without identifying patient information.
**Nucleotide Sequencing**

The E6, E7, L1 and LCR were amplified by long- or short-fragment polymerase chain reaction (PCR). Long-fragment PCR was applied to good-quality samples using primers 5'-ATG TCC ATT GAG TCA GGT CC-3' and 5'-TGC ATT TTC ATC CTC GTC C-3', and a second PCR using inner primers 5'-GGT CCT GCT GAC ATT CCA TTA CC-3' and 5'-CCT CTA CTT CAA ACC AGC CT-3' when necessary. Short-fragment PCR was used when the long-fragment approach failed. E6, E7, L1 and LCR were amplified using primer pairs E6E7 (5'-TGC ACT ACA CGA CCG GTT A-3' and 5'-CAT CCT CTA CCTG AAA TG-3'), L1A (5'-ATG TCC ATT GAG TCA GGT CC-3' and 5'-GCA CAG GGT CAC CTA AGG TA-3'), L1B (5'-AGG ATG GGG ACA TGG TAG AT-3' and 5'-CAC AGA CAA TTA CCC AAC AGA C-3') and LCR (5'-GTC TGC ATC TTT GGA GGA CA-3' and 5'-TGC GTT AGC TAC ACT GTG TTC-3'), respectively. When necessary, a second PCR using inner primers E6E7 (5'-TTC CCG TAC CCA CAA CCA CT-3' and 5'-CCT CTA CTT CAA ACC AGC CT-3'), L1A (5'-GGT CCT GAC ATT CCA TTA CC-3' and 5'-GGG CAC ATC ACT TTT ACT AGC-3'), L1B (5'-ACA GGA TTT GGT TGC ATG G-3' and 5'-TCT TTT GTG GAG GTA CGT GG-3') and LCR (5'-TTT GTT ACA GGC AGG GCT AC-3' and 5'-CGT TTT CGG TTA CAC CCT A-3') was performed. The PCR products were sequenced from both directions, and analyzed with SeqScape software (version 2.5, Applied Biosystems). Mutations that occurred only once were confirmed by repeat sequencing from the original sample.

**Phylogenetic Tree Construction**

The concatenated nucleotide sequences assembled from the E6, E7, L1 and LCR regions were used for phylogenetic tree construction. Representative variants identified previously were incorporated for lineage identification [7]. Maximum-likelihood trees were constructed...
using the Subtree-Pruning-Regrafting (SPR) search approach by the Molecular Evolutionary Genetic Analysis (MEGA) Software program (version 5.10, http://www.megasoftware.net/) [8]. The data were bootstrap resampled 1,000 times for tree topology evaluation.

Geographical Distribution of Variant Lineages

The detection rate of each variant lineage was compared among regions by Chi-squared test or Fisher’s exact test as appropriate with correction for multiple comparisons using the Bonferroni method. Epi Info (version 7.0.8.3, Centers for Disease Control and Prevention, Atlanta) was used to calculate P values. Multivariate analyses were performed to investigate the association between each lineage and disease, controlling for age. Subjects with normal cervical cytology were used as controls, whereas subjects with histologically confirmed CIN3 or invasive cervical cancer were categorized as cases. Two-tailed P values of < .05 were regarded as significant. The software package for statistical analysis (SPSS version 20, IBM) was used for multivariate analysis.

Results

Altogether, 611 specimens collected from 14 sites had DNA quality sufficient for sequencing (Supplementary Table S1). Of these, 73.2% were from Asia, 15.5% from Europe, 9% from Americas and 2.3% from Africa. Most samples were from women with normal cervical cytology (31.3%) and high-grade lesions (30.1%) including high-grade squamous intraepithelial lesions (HGSIL), CIN2 and CIN3. Altogether, 25 (4.1%) cervical samples had no associated
cytological or histological information, and another 14 (2.3%) were vaginal samples. The mean age of study subjects was 41.1 years (standard deviation: 14.0, range: 13-88).

**Lineage Identification**

The concatenated E6-E7-L1-LCR sequences derived from referent strains of each sublineage formed distinct branches in the phylogenetic tree, suggesting that these concatenated sequences comprising 40.6% (3226 nt) of the whole HPV52 genome can be used for lineage identification (Supplementary Figure S1). The tree topology of E6-E7-L1-LCR sequences derived from the 324 unique strains collected in this study revealed three closely related but distinct branches representing lineages A, B and C; and one distantly related branch representing lineage D.

The phylogenetic trees constructed from L1 or LCR sequences alone showed a topology similar to that of E6-E7-L1-LCR, and were able to identify variants up to the lineage level, but could not differentiate sublineages. Signature sequences within the L1 and LCR regions useful for lineage/sublineage identification are shown in Supplementary Figure S2. The phylogenetic trees constructed from E6 or E7 sequences alone showed topologies different from that of E6-E7-L1-LCR, and were not useful for lineage identification.

**Geographical Distribution of Variant Lineages**

Variation in geographical distribution of HPV52 lineages was observed (Figure 1). Lineage B was significantly more prevalent in Asia compared to elsewhere (89.0% in Asia vs. 0-5.5% elsewhere, $P_{\text{corrected}} < 0.001$ for each comparison). In contrast, Africa, the Americas and
Europe were all predominated by lineage A that accounted for 78.6-96.8% of the isolates compared to 5.5% in Asia (P_{corrected} < 0.001 for each comparison). Lineage C was uncommon across all regions (0% to 9.1%) and without significant variation. Lineage D was rarely detected in the Americas, Asia and Europe (0-1.8%); but was found in 3 of 14 samples from Africa giving a wide 95% confidence interval (CI) of 0-42.9%.

The majority (93.7%) of lineage A variants belonged to sublineage A1, which was consistently observed across regions. All lineage B variants identified in this study were sublineage B2, and all lineage C variants belonged to sublineage C2.

**Risk Association of Variant Lineages**

The distribution of variant lineages and sublineages according to cervical pathology status is shown in Table 1. Multivariate analyses adjusting for age were performed to compare subjects with normal cervical cytology as controls against subjects with histologically confirmed CIN3 or invasive cervical cancer as cases. Lineage B was found to associate with a significantly higher risk than lineage A (age-adjusted OR [95%CI] = 5.46 [2.28-13.07]). Lineage C was also associated with a significantly higher risk than lineage A (age-adjusted OR [95%CI] = 7.78 [2.26-26.75]). Lineage B appeared to associate with a higher risk than lineage C, but the difference was not statistically significant (age-adjusted OR [95% CI] = 1.42 [0.56-3.56]. The number of isolates belonging to lineage D was not enough for analysis.
Discussion

Intratypic variants of HPV are divided into lineages based on the topology of phylogenetic tree and a difference of >1% in their full genome sequences [7]. Such classification of variants is important not only for understanding the evolution of HPV, but also because it carries biological implications. HPV52 has evolved into four lineages, for which the geographical distribution and risk implication have been uncertain [7], but are addressed in this study. The main strengths of our study are the large number of samples collected around the world, and the ability to restrict risk association analysis to cases with histologically confirmed diagnoses. Nevertheless, this study had limitations in not being able to account for coinfection with other high-risk HPV types, the number of samples available from some regions such as Africa was small, and some samples did not have associated cytological/histological diagnoses. Furthermore, due to co-linearity between lineage and geographical distribution, the geographical source of sample could not be one of the covariants in the regression equation. Therefore, we cannot exclude the possibility of ethnogeographical effects on risk association of lineage variants.

To date, only a few studies have investigated the distribution of HPV52 lineages. Chang et al. found that among Taiwanese women, lineage B was the most prevalent (88.2%), followed by lineage C (11.1%), while lineage A was rare (0.7%) [9]. In contrast, lineage A was the most frequently found in Canada, especially among Caucasian [10, 11]. Another study examined samples collected from Japan, the Philippines and Vietnam, and reported that lineage B was the most prevalent followed by lineage A [12]. However, that study used E6 and E7 sequences to identify variant lineages, which is suboptimal for such purposes.
Our study assessed the distribution of HPV52 variants based on 611 samples collected from 14 cities across 4 continents providing the largest data set for assessing geographical distribution. The most remarkable finding was the dominance of lineage B, but rare occurrence of lineage A, in Asia. The exact opposite was true in non-Asian regions. Therefore, we propose to name lineage B of HPV52 as “Asian (As)”, and lineage A as “non-Asia (nAs)” to signify their characteristic geographical distribution.

Studies on risk association of HPV52 variants are limited and inconclusive. Ding et al. examined the E6 and E7 sequences of 121 samples from Zhejiang, Eastern China, but could not identify any variant with increased or decreased oncogenic risk [13]. Sun et al. analyzed the L1, E6, E7 and LCR sequences of 72 samples from Shengjing, Northeast China [14]. In that study, the variants were not grouped according to the lineage classification system proposed by Chen et al. [7], and no significant risk association was observed. Ishizaki et al. studied 109 samples from Japan, the Philippines and Vietnam. Again, no significant association between E6 and E7 sequence variation and abnormal cytology was found [12].

Although examination of E6 and E7 sequence variation did not reveal any significant risk association, some interesting findings were observed when lineage classification was taken into account. Chang et al. used LCR-E6-E7 sequences to identify the lineage of 280 samples from Taiwan, and reported a higher risk of CIN for lineage C compared to lineage B variants [9]. Unfortunately, lineage A was found in two samples only and therefore precluded from risk association comparison. Two studies on risk association of HPV52 variants were available from Canada. Aho et al. showed that non-prototypic LCR variant was an independent predictor for viral persistence [11]. The observations from Formentin et al. suggested that variant MTL-52-
LCR-21 that belongs to sublineage A1 and variant MTL-52-LCR-02 that belongs to sublineage A2 conferred a higher risk. However, most of the isolates available in these Canadian studies were of lineage A, precluding the comparison among different lineages. Schiffman et al. examined HPV52 samples derived from the Guanacaste Cohort Study, and observed that all CIN2+ cases were infected with lineages A/B/C suggesting a lower risk for lineage D [15]. However, the observation was highly unstable and not statistically significant.

The current study has generated the most comprehensive data for analyzing risk association of HPV52 variant lineages with cervical disease. Based on our observations, we propose to classify lineage A as a “low-risk” lineage of HPV52, whereas lineages B as a “high-risk” lineage. Lineage C is probably “high-risk” as well. Lineage D is rare and cannot be assigned to a risk category at this stage. Nevertheless, this risk classification should be further evaluated preferably with assessment on the transforming ability of these variants using in-vitro or in-vivo models.

In conclusion, we found that classifying HPV52 variants into lineages carries epidemiological and pathological implications. Lineage B can be regarded as “Asian” and “high-risk” based on its geographical distribution and risk for cervical neoplasia. The reported higher disease attribution of HPV52 in Asia is likely to be a result of the higher prevalence of lineage B in that region. The unique epidemiological feature of HPV52 in Asia should be considered in the design and evaluation of diagnostic assays and vaccines intended for Asia.
Funding: This work was supported by the Research Fund for the Control of Infectious Diseases, Food and Health Bureau, Hong Kong Special Administrative Region (reference no.: 10090482). Magdalena Grce and Ivan Sabol received support from the Croatian Ministry of Research, Education and Sport on HPV research. Karen K. Smith-McCune and Anna-Barbara Moscicki received support from Bill and Melinda Gates Foundation. Francois Coutlée received support from the Cancer Society of Canada for research on HPV variants. Federico De Marco received support from the Ministry of Foreign Affairs, and the Italian Ministry of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of interest: Paul KS Chan is participating in a clinical trial supported by, and has received honoraria as advisory board member and support for attending academic conferences from GlaxoSmithKline. Joel Palefsky has received travel support, research grant funding and served as board member for Merck & Co., received stock options from Aura Biosciences, served as board member for Pharmajet, and received consultation fee (paid to institution) from Qiagen. Ryo Konno has received grants from GlaxoSmithKline, and fees from GlaxoSmithKline, Merck Sharp Dome and Qiagen. Karen K Smith-McCune is in the Scientific and Clinical Advisory Board of, and has received stock options from OncoHealth Inc. Francois Coutlée has received honoraria from Merck Sharp Dome for lectures on HPV. Anna-Barbara Moscicki has received honoraria from Merck Sharp Dome for being advisory board member, and from Becton Dickinson as speaker. Alison Fiander has served advisory boards for GlaxoSmithKline Biologicals SA and Sanofi
Pasteur MSD, and received research grant funding and support to attend HPV-related conferences from both companies.

An abstract of this manuscript has been accepted for presentation in the Third Workshop on Emerging Oncogenic Viruses in Mandura, Italy, 4-8 June 2014.

*Correspondence to: Paul KS Chan MD FRCPath
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
References


5. Ho CM, Chien TY, Huang SH, Lee BH, Chang SF. Integrated human papillomavirus types 52 and 58 are infrequently found in cervical cancer, and high viral loads predict risk of cervical cancer. Gynecol Oncol 2006; 102: 54-60.


**Figure legends**

**Figure 1. Distribution of HPV52 variant lineages and sublineages according to geographical regions.**
<table>
<thead>
<tr>
<th>Lineage / sublineage (N)</th>
<th>Normal cervical cytology (N=191)</th>
<th>Cervical intraepithelial neoplasia 3 (N=111)</th>
<th>Invasive cervical cancer (N=41)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (36)</td>
<td>30 (15.7%)</td>
<td>4 (3.6%)</td>
<td>2 (4.9%)</td>
</tr>
<tr>
<td></td>
<td>A1 (34)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>28 (14.6%)</td>
<td>4 (3.6%)</td>
<td>2 (4.9%)</td>
</tr>
<tr>
<td></td>
<td>A2 (2)</td>
<td>2 (1.0%)</td>
<td>0</td>
</tr>
<tr>
<td>B (289)</td>
<td>155 (81.2%)</td>
<td>98 (88.3%)</td>
<td>36 (87.8%)</td>
</tr>
<tr>
<td></td>
<td>B1 (0)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>B2 (289)</td>
<td>155 (81.2%)</td>
<td>98 (88.3%)</td>
</tr>
<tr>
<td>C (14)</td>
<td>5 (2.6%)</td>
<td>6 (5.4%)</td>
<td>3 (7.3%)</td>
</tr>
<tr>
<td></td>
<td>C1 (0)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>C2 (14)</td>
<td>5 (2.6%)</td>
<td>6 (5.4%)</td>
</tr>
<tr>
<td>D (4)</td>
<td>1 (0.5%)</td>
<td>3 (2.7%)</td>
<td>0</td>
</tr>
</tbody>
</table>

1 All cervical intraepithelial neoplasia 3 and invasive cervical cancer cases were diagnosed by histology.