JID MS#54593-R1

Title page

Geographical distribution and risk association of human papillomavirus 52 variant lineages

Chuqing Zhang[1], Jong-Sup Park[2], Magdalena Grce[3], Samantha Hibbitts[4], Joel M. Palefsky[5], Ryo Konno[6], Karen K. Smith-McCune[7,8], Lucia Giovannelli[9], Tang-Yuan Chu[10], María Alejandra Picconi[11], Patricia Piña-Sánchez[12], Wannapa Settheetham-Ishida[13], Francois Coutlée[14], Federico De Marco[15], Yin-Ling Woo[16], Wendy C. S. Ho[1], Martin C. S. Wong[17], Mike Z. Chirenje[18], Tsitsi Magure[18], Anna-Barbara Moscicki[8,19], Ivan Sabol[3], Alison N. Fiander[4], Zigui Chen[20], Martin C. W. Chan[1], Tak-Hong Cheung[21], Robert D. Burk[20], Paul K. S. Chan[1*]

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

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

Laboratory of Molecular Virology and Bacteriology, Rudjer Boskovic Institute, Zagreb, Croatia[3]

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

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

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

Department of Obstetrics, Gynecology and Reproductive Sciences, University of California San Francisco, California, United States of America[7]

Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, California, United States of America[8]

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

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

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

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

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

Département de Microbiologie et Infectiologie, Centre Hospitalier de l'Université de Montréal, Montréal, Québec, Canada[14]

Laboratory of Virology – The "Regina Elena" National Cancer Institute, Rome, Italy[15]

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

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

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

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

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

Department of Obstetrics and Gynaecology, Prince of Wales Hospital, Hong Kong Special Administrative Region, China[21]

Short running head: Distribution and risk of HPV52 lineages

Word count: Abstract (100); main content (1992); table (1); figure (1); references (15); online supplementary materials (1 table, 2 figures)

***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

Abstract

Human papillomavirus (HPV) 52 is commonly found in Asian cervical cancers, but rare elsewhere. Analysis of 611 isolates collected worldwide revealed geographical variation in lineage distribution, with lineage B predominating in Asia (89.0% vs. 0-5.5%, $P_{corrected}$<0.001); whereas Africa, Americas and Europe were predominated by lineage A. Lineage B conferred a higher risk for cervical intraepithelial neoplasia 3 and invasive cancers than lineage A (OR [95%CI] = 5.46 [2.28-13.07]). The reported high disease attribution of HPV52 in Asia is likely due to the high prevalence of lineage B. We propose to name lineage B as “Asian” lineage to signify this feature.

Keywords

HPV, cervical cancer, phylogeny, oncogenic risk, Asia

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 being the second in Taiwan [5] and Japan [6]. While the geographical predilection in disease attribution is obvious, its underlying reason 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-fragment 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 then a second PCR using inner primers 5'-GGT CCT 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 separately using primer pairs E6E7 (5'-TGC ACT ACA CGA CCG GTT A-3' and 5'-CAT CCT CGT CCT CTG 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'). When necessary, a second-round of PCR using inner primer pairs E6E7 (5'-TTA 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'-TTC 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 of each lineage and sub-lineage identified from a previous study 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 the 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 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 of HPV52 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 a topology quite different from that of E6-E7-L1-LCR, and were not useful for lineage identification.

Geographical Distribution of Variant Lineages

Variation in distribution of HPV52 lineages according to geographical location was observed (Figure 1). Lineage B was significantly more prevalent in Asia compared to elsewhere (89.0% in Asia vs. 0-5.5% elsewhere, $P_{corrected} < 0.001$ 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 too few 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 the 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 relatively small, and some samples did not have associated cytological/histological diagnoses.

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 from a worldwide perspective. 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 differences in risk association among HPV52 variants, 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). Dr. Magdalena Grce's research on HPV is supported by the Croatian Ministry of Research, Education and Sport. Dr. Francois Coutlée's research on HPV variants is supported by the Cancer Society of Canada. Dr. Federico De Marco is partly supported by the Italian Ministry of Foreign Affairs, DGPC Uff V and by 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: Karen K. Smith-McCune is in the Scientific and Clinical Advisory Board of and has received stock options for Oncohealth Inc. Francois Coutlée has received grants through his institution from Merck and F. Hoffman-La Roche, as well as honoraria from Merck and F. Hoffman-La Roche for lectures on HPV. Paul KS Chan is participating in a clinical trial supported by, and has received honorarium as advisory board member and support for attending academic conferences from GlaxoSmithKline.

An abstract of this manuscript has been submitted 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

1. Muñoz N, Bosch FX, Castellsagué X, et al. Against which human papillomavirus types shall we vaccinate and screen? The international perspective. Int J Cancer **2004**; 111: 278-285.

2. Smith JS, Lindsay L, Hoots B, et al. Human papillomavirus type distribution in invasive cervical cancer and high-grade cervical lesions: A meta-analysis update. Int J Cancer **2007**; 121: 621-632.

3. Chan PK, Cheung TH, Li WH, et al. Attribution of human papillomavirus types to cervical intraepithelial neoplasia and invasive cancers in southern china. Int J Cancer **2012**; 131: 692-705.

4. Huang S, Afonina I, Miller BA, Beckmann AM. Human papillomavirus types 52 and 58 are prevalent in cervical cancers from Chinese women. Int J Cancer **1997**; 70: 408-411.

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.

6. Sasagawa T, Basha W, Yamazaki H, Inoue M. High-risk and multiple human papillomavirus infections associated with cervical abnormalities in Japanese women. Cancer Epidemiol Biomarkers Prev **2001**; 10: 45-52.

7. 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.

8. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol **2011**; 28: 2731-2739.

9. Chang YJ, Chen HC, Lee BH, et al. Unique variants of human papillomavirus genotypes 52 and 58 and risk of cervical neoplasia. Int J Cancer **2011**; 129: 965-973.

10. Formentin A, Archambault J, Koushik A, et al. Human papillomavirus type 52 polymorphism and high-grade lesions of the uterine cervix. Int J Cancer **2013**; 132: 1821-1830.

11. Aho J, Hankins C, Tremblay C, et al. Genomic polymorphism of human papillomavirus type 52 predisposes toward persistent infection in sexually active women. J Infect Dis **2004**; 190: 46-52.

12. Ishizaki A, Matsushita K, Hoang HT, et al. E6 and E7 variants of human papillomavirus-16 and -52 in Japan, the Philippines, and Vietnam. J Med Virol **2013**; 85: 1069-1076.

13. 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: 436-443.

14. Sun Z, Lu Z, Liu J, et al. Genomic polymorphism of human papillomavirus type 52 in women from northeast China. Int J Mol Sci **2012**; 13: 14962-14972.

15. 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: 3159-3169.

Figure legends

Figure 1. Distribution of HPV52 variant lineages and sublineages according to geographical regions.

Online supplementary materials

Supplementary Figure S1. Phylogenetic trees showing HPV52 variant lineages and sublineages.

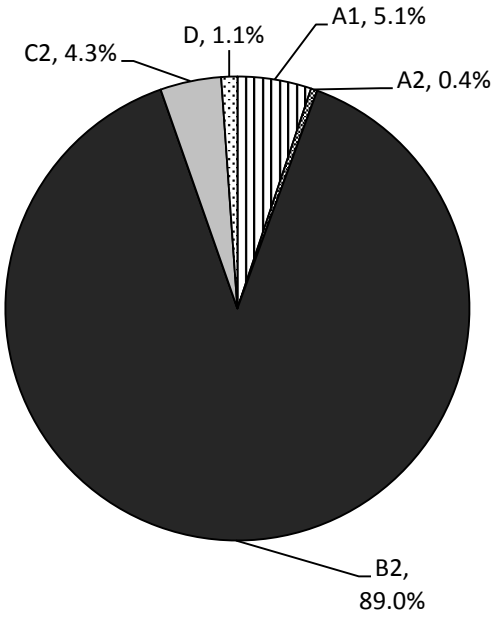
Maximum-likelihood trees were constructed using the program MEGA 5 based on concatenated E6-E7-L1-LCR nucleotide sequences of 324 unique HPV52 strains collected in this study and published referent strains of each sublineage (A1: X74481, A2: HQ537739, B1: HQ537740, B2: HQ537743, C1: HQ537744, C2: HQ537746, D: HQ537748). The unrooted tree showing sublineage clustering of each sequence. Samples examined in this study are labeled with different colours and symbols according to the region and country, respectively. Solid black triangles denote sublineage referent strains. The box in right lower quadrant shows the rooted tree displaying overall topology. Bootstrap values of key nodes generated by 1,000 resampling are shown. The length of the scale bar represents 0.003 substitutions per nucleotide position. HPV67 prototype sequence (GenBank accession no. NC_004710) was set as the outgroup and represented by a grey broken line.

Supplementary Figure S2. L1 and LCR sequence signatures for HPV52 lineage/sublineage identification.

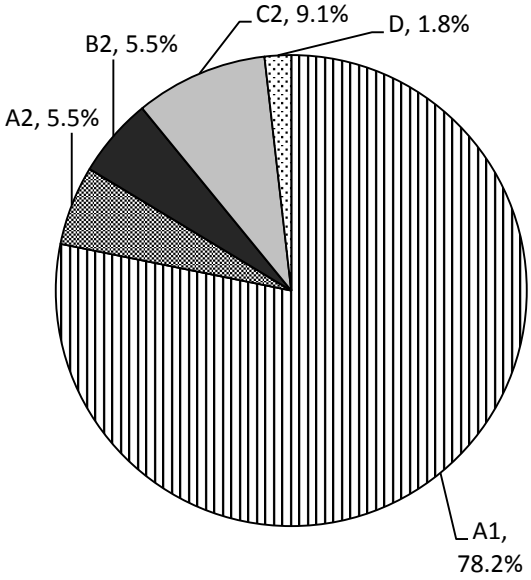
Vertical numbers represent nucleotide positions according to the HPV52 prototype (GenBank accession no. X74481). Sites with no changes are marked with dashes. C6483A, A6711G, C6983G are nonsynonymous substitutions, others are synonymous. Y = C or T, S = G or C, M = A or C, del = deletion, ins = insertion. [a] Deletion of TG at nucleotide position 7287 and 7288. [b] Deletion in 1 sample. [c] Insertion of GT between nucleotide positions 7287 and 7288. [d] T to G substitution in 2 samples, deletion in 7 samples. GenBank accession no.

for E6 variants: KJ675743 - KJ675789; E7 variants: KJ675790 - KJ675805; LCR variants: KJ675806 - KJ675962; L1 variants: KJ675963 - KJ676095.

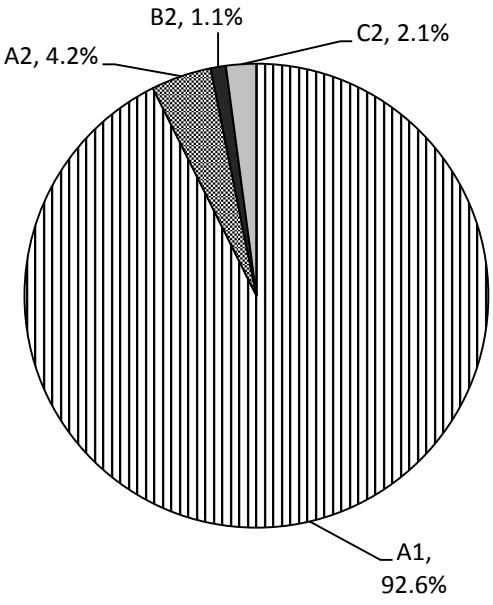
(i) Asia, N=447



(ii) Americas, N=55



(iii) Europe, N=95



(iv) Africa, N=14

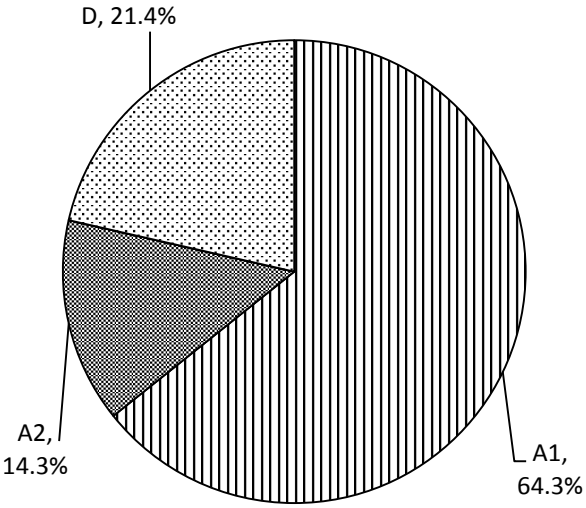
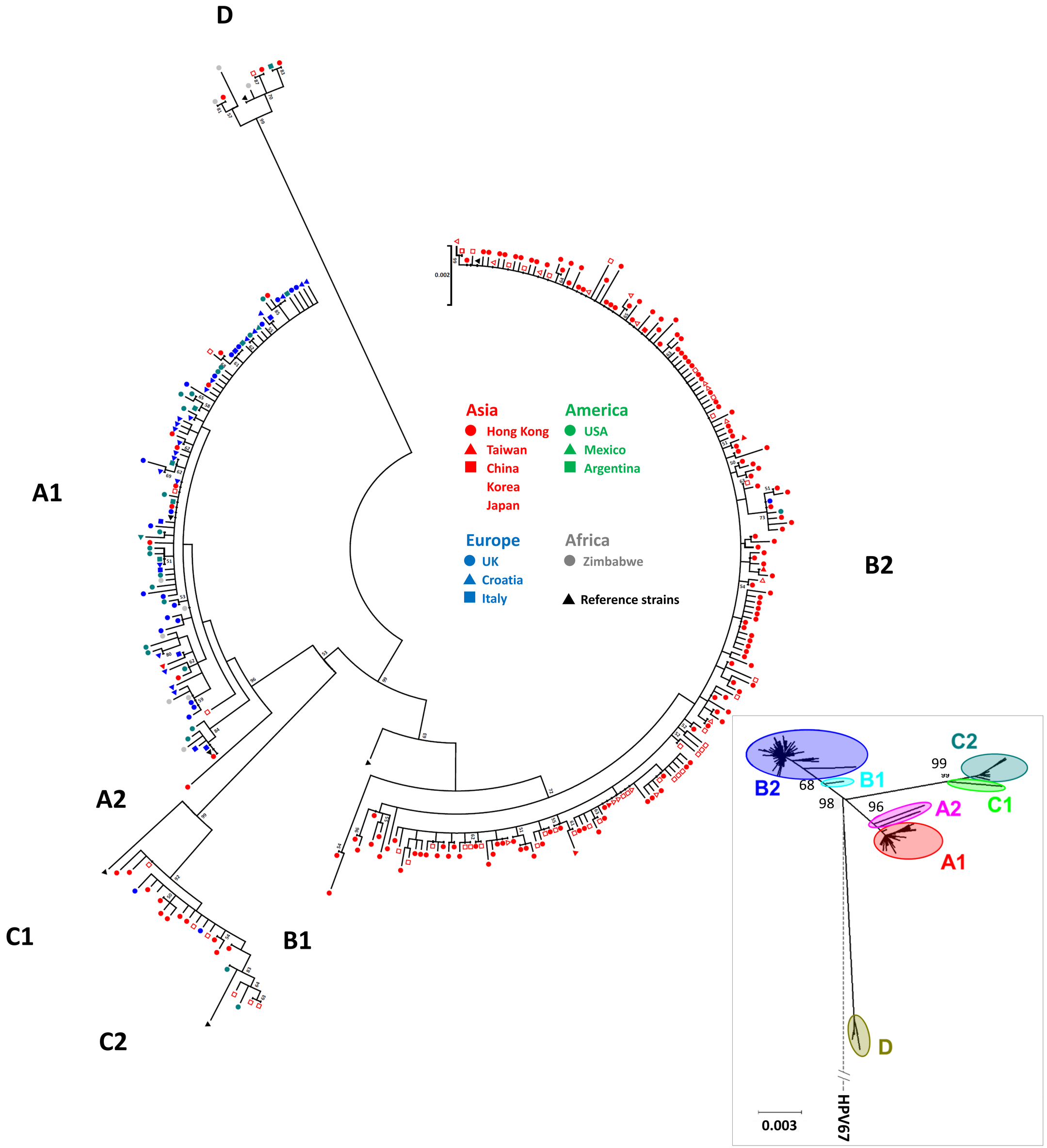


Table 1. Distribution of HPV52 variant lineages according to cervical pathology status

Lineage / sublineage (N)	No. (%) of subjects[1]		
	Normal cervical cytology (N=191)	**Cervical intraepithelial neoplasia 3 (N=111)**	**Invasive cervical cancer (N=41)**
A (36)	30 (15.7%)	4 (3.6%)	2 (4.9%)
A1 (34)	28 (14.6%)	4 (3.6%)	2 (4.9%)
A2 (2)	2 (1.0%)	0	0
B (289)	155 (81.2%)	98 (88.3%)	36 (87.8%)
B1 (0)	0	0	0
B2 (289)	155 (81.2%)	98 (88.3%)	36 (87.8%)
C (14)	5 (2.6%)	6 (5.4%)	3 (7.3%)
C1 (0)	0	0	0
C2 (14)	5 (2.6%)	6 (5.4%)	3 (7.3%)
D (4)	1 (0.5%)	3 (2.7%)	0

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


D
A1
A2
C1
C2
B1
B2
0.002
Asia
Hong Kong
Taiwan
China
Korea
Japan
America
USA
Mexico
Argentina
Europe
UK
Croatia
Italy
Africa
Zimbabwe
Reference strains
C2
B2
B1
A2
C1
A1
D
99
68
98
96
HPV67
0.003

Supplementary Figure S2. L1 and LCR sequence signatures for HPV52 lineage/sublineage identification

Lineage/ Sublineage	No.	L1																LCR																									
		5	5	6	6	6	6	6	6	6	6	6	6	6	7	7		7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7					
		7	9	1	2	4	4	5	5	6	7	7	7	9	9	0	0	1	2	2	2	3	3	3	3	4	4	4	5	5	6	6	6	6	7	7	8	9				
		6	0	9	1	4	8	3	6	9	0	1	1	8	9	4	7	6	4	8	8	7	8	8	9	1	2	2	7	7	2	5	7	9	1	8	6	1	2	8	9	
		3	9	7	8	3	3	3	9	8	1	1	2	3	2	9	9	8	9	7	8	1	1	2	3	4	1	4	1	9	4	9	1	6	2	8	1	7	8	6	7	
Prototype		**C**	**A**	**T**	**G**	**C**	**C**	**T**	**C**	**G**	**T**	**A**	**G**	**C**	**T**	**T**	**G**	**G**	**G**	**T**	**G**	**G**	**G**	**T**	**T**	**T**	**G**	**C**	**C**	**A**	**T**	**T**	**C**	**T**	**G**	**A**	**G**	**C**	**A**	**T**	**C**	
A																																											
A1	163	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	
A2	11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	
B																																											
B1	0																																										
B2	402	-	-	-	A	-	-	-	-	-	G	-	-	-	-	-	-	S	-	del[a]	T	-	-	-	-	-	-	-	-	-	G	C	-	-	C	-	A	M	-	-	-	
C																																											
C1	0																																										
C2	26	-	G	-	A	T	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	C	A	-	-	-	G	C	T	C	C	G	A	A	A[b]	-	-	
D	9	A	-	C	A	-	A	C	A	-	-	G	A	G	C	C	A	T	C	Ins[c]	-	T	G	-	-	-	T	G	C	G	-	-	S[d]	C	-	A	G	G	C	T		

Vertical numbers represent nucleotide positions according to the HPV52 prototype (GenBank accession no. X74481). Sites with no changes are marked with dashes. C6483A, A6711G, C6983G are nonsynonymous substitutions, others are synonymous. Y = C or T, S = G or C, M = A or C, del = deletion, ins = insertion.

[a] Deletion of TG at nucleotide position 7287 and 7288.

[b] Deletion in 1 sample.

[c] Insertion of GT between nucleotide positions 7287 and 7288.

[d] T to G substitution in 2 samples, deletion in 7 samples.

Supplementary Table S1. Geographical source and pathological status of specimens

Region, country or city	All lesion grades	No. of specimens according to cervical pathology status					
		Normal	ASCUS	Low-grade	High-grade	Carcinoma	Unknown
Africa	**14**	**0**	**0**	**0**	**0**	**0**	**14**
Zimbabwe	14	0	0	0	0	0	14
Americas	**55**	**1**	**0**	**1**	**2**	**1**	**9**
Argentina	9	0	0	0	0	0	9
Canada	1	0	0	0	1	0	0
Mexico	4	1	0	1	1	1	0
United States	41[a]	0	0	0	0	0	0[b]
Asia	**447**	**168**	**5**	**69**	**162**	**40**	**3**
Hong Kong	321	127	3	36	128	27	0
Japan	21	0	0	11	9	1	0
Korea	91	40	2	20	23	5	1
Shanghai	2	0	0	0	0	0	2
Taiwan	10	0	0	2	2	6	0
Thailand	2	1	0	0	0	1	0
Europe	**95**	**22**	**0**	**40**	**20**	**0**	**13**
Croatia	40	0	0	21	19	0	0
Italy	14	4	0	4	1	0	5
United Kingdom	41	18	0	15	0	0	8
Total	**611**	**191**	**5**	**110**	**184**	**41**	**39[b]**

ASCUS, atypical squamous cells of undetermined significance; CIN, cervical intraepithelial neoplasia; High-grade includes high-grade squamous intraepithelial lesions (HGSIL), CIN2 and CIN3; Low-grade includes low-grade squamous intraepithelial lesions (LGSIL) and CIN1.

[a] anal swab samples.

[b] not include 41 anal samples.