

Identification of Human Papillomavirus Type 58 Lineages and the Distribution Worldwide

Paul K. S. Chan,¹ Alfred C.S. Luk,¹ Jong-Sup Park,² Karen K. Smith-McCune,^{3,4} Joel M. Palefsky,⁵ Ryo Konno,⁶ Lucia Giovannelli,⁷ Francois Coutlée,⁸ Samantha Hibbitts,⁹ Tang-Yuan Chu,¹⁰ Wannapa Settheetham-Ishida,¹¹ María Alejandra Picconi,¹² Annabelle Ferrera,¹³ Federico De Marco,¹⁴ Yin-Ling Woo,¹⁵ Tainá Raiol,¹⁶ Patricia Piña-Sánchez,¹⁷ Jo L. K. Cheung,¹ Jeong-Hoon Bae,² Mike Z. Chirenje,¹⁸ Tsitsi Magure,¹⁸ Anna-Barbara Moscicki,^{4,19} Alison N. Fiander,⁹ Rosa Di Stefano,⁷ Tak-Hong Cheung,²⁰ May M. Y. Yu,²¹ Stephen K. W. Tsui,²² David Pim,²³ and Lawrence Banks²³

¹Department of Microbiology, Faculty of Medicine, Chinese University of Hong Kong, Hong Kong Special Administrative Region, China; ²Department of Obstetrics and Gynecology, Seoul St Mary's Hospital, Catholic University of Korea, South Korea; ³Department of Obstetrics, Gynecology and Reproductive Sciences; ⁴Helen Diller Family Comprehensive Cancer Center and; ⁵Department of Medicine, University of California, San Francisco; ⁶Department of Obstetrics and Gynecology, Jichi Medical University, Saitama Medical Center, Japan; ⁷Sezione di Microbiologia, Dipartimento di Scienze per la Promozione della Salute, Azienda Ospedaliera Universitaria Policlinico P. Giaccone, Palermo, Italy; ⁸Département de Microbiologie et Infectiologie, Centre Hospitalier de l'Université de Montréal, Québec, Canada; ⁹Section of Obstetrics and Gynaecology, School of Medicine, Cardiff University, United Kingdom; ¹⁰Department of Obstetrics and Gynaecology, Buddhist Tzu Chi General Hospital, Graduate Institute of Clinical Medicine, Tzu Chi University, Hualien, Taiwan; ¹¹Department of Physiology, Faculty of Medicine, Khon Kaen University, Thailand; ¹²Servicio Virus Oncogénicos, Departamento Virología, Instituto Nacional de Enfermedades Infecciosas, ANLIS Carlos G. Malbrán, Buenos Aires, Argentina; ¹³School of Microbiology, Science Faculty, Universidad Nacional de Honduras, Tegucigalpa, Honduras; ¹⁴Laboratory of Virology—Regina Elena Cancer Institute, Rome, Italy; ¹⁵Department of Pathology, University of Cambridge, United Kingdom; ¹⁶Department of Cellular Biology, Institute of Biology, University of Brasilia, Brazil; ¹⁷Unidad de Investigación Médica en Enfermedades Oncológicas, Hospital de Oncología, CMN Siglo XXI, Mexico City, Mexico; ¹⁸Department of Obstetrics and Gynecology, University of Zimbabwe, Harare, Zimbabwe; ¹⁹Department of Pediatrics, University of California, San Francisco; ²⁰Department of Obstetrics and Gynaecology and; ²¹Department of Anatomical and Cellular Pathology, Prince of Wales Hospital, Hong Kong Special Administrative Region, China; ²²School of Biomedical Sciences, Chinese University of Hong Kong, Hong Kong Special Administrative Region, China; and ²³International Centre for Genetic Engineering and Biotechnology, Trieste, Italy

Background. Human papillomavirus type 58 (HPV-58) accounts for a much higher proportion of cervical cancers in East Asia than other types. A classification system of HPV-58, which is essential for molecular epidemiological study, is lacking.

Methods and results. This study analyzed the sequences of 401 isolates collected from 15 countries and cities. The 268 unique concatenated E6-E7-E2-E5-L1-LCR sequences that comprised 57% of the whole HPV-58 genome showed 4 distinct clusters. L1 and LCR produced tree topologies that best resembled the concatenated sequences and thus are the most appropriate surrogate regions for lineage classification. Moreover, short fragments from L1 (nucleotides 6014–6539) and LCR (nucleotides 7257–7429 and 7540–52) were found to contain sequence signatures informative for lineage identification. Lineage A was the most prevalent lineage across all regions. Lineage C was more frequent in Africa than elsewhere, whereas lineage D was more prevalent in Africa than in Asia. Among lineage A variants, sublineage A2 dominated in Africa, the Americas, and Europe, but not in Asia. Sublineage A1, which represents the prototype that originated from a patient with cancer, was rare worldwide except in Asia.

Conclusions. HPV-58 can be classified into 4 lineages that show some degree of ethnogeographic predilection in distribution. The evolutionary, epidemiological, and pathological characteristics of these lineages warrant further study.

Received 8 December 2010; accepted 24 January 2011.

Potential conflicts of interest: none reported.

Presented in part: 26th International Papillomavirus Conference, Montreal, Canada, 3–8 July 2010. Abstract 345.

Correspondence: Paul K.S. 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, China (paulkschan@cuhk.edu.hk).

The Journal of Infectious Diseases 2011;203:1565–73

© The Author 2011. Published by Oxford University Press on behalf of the Infectious Diseases Society of America. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com

0022-1899 (print)/1537-6613 (online)/2011/20311-0010\$14.00

DOI: 10.1093/infdis/jir157

Human papillomavirus (HPV) infection is a necessary, although insufficient, cause of cervical cancer, which is still the second leading cancer to affect women worldwide [1]. Two prophylactic vaccines (Cervarix and Gardasil) targeting the 2 most prevalent high-risk HPV types (HPV types 16 and 18) found in cervical cancer are available [2, 3]. These vaccines are expected to prevent ~70% of cervical cancers worldwide [4]. There remain at least 13 high-risk HPV types not targeted by the

current vaccines that account for the remaining cervical cancers [5]. Globally, HPV types 31, 33, and 45 form the second group and HPV types 35, 52, and 58 form the third group in the ranking of cancer association [6]. However, the disease impact associated with these other HPV types shows considerable geographical variation. As shown in a meta-analysis, HPV type 58 (HPV-58) was found in 3.3% of cervical cancers globally and 5.6% of cervical cancers in Asia, whereas the prevalence in high-grade cervical intraepithelial lesions was 7.0% globally and 12.2% in Asia [4]. Studies from East Asian populations have reported an even higher rate. For instance, HPV-58 was detected in 26% of cervical squamous cell carcinoma in Shanghai [7], 16% in South Korea [8], 10% in Hong Kong [9] and Taiwan [10], and 8% in Japan [11]; and HPV-58 ranked third in cervical cancer cases from Asia overall [12]. Although HPV-58 may not play an etiological role in all these cases, as some of them are found in coinfections with other high-risk types, the disease impact conferred by HPV-58, especially in East Asian populations, cannot be neglected. The reason for a geographical or ethnical predilection of HPV-58-associated cervical neoplasia is not fully understood. Previous studies have suggested that host genetic factors and the circulation of variants with higher oncogenicity could play a role [13, 14]. To date, information on sequence variability of HPV-58 is very limited. This study was conducted to elucidate the phylogenetic relationship between HPV-58 variants collected worldwide to establish a classification system that will facilitate further study on the oncogenic potential of HPV-58.

MATERIALS AND METHODS

Study Samples

Cervical, vaginal, or anal samples were collected by study collaborators. Those samples that had tested positive for HPV-58 were transferred to a central laboratory for sequence analysis. The quality of DNA was assessed by amplifying a 1039-bp fragment of the L1 region, and the identity of HPV-58 was ascertained by demonstrating a nucleotide sequence similarity of >90% compared with the corresponding L1 fragment of the prototype (GenBank accession no. NC_001443). Altogether, 401 samples collected from 15 geographical locations had sufficient DNA quality for sequencing of the whole length of the E6, E7, E2 (containing E4), E5, L1, and LCR regions (Table 1). All samples except 37 anal swab specimens from men in the United States were cervical, vaginal scrape, swab, or tissue specimens from women. The distribution of cervical pathology status is shown in Table 1.

Nucleotide Sequencing

The whole lengths of E6, E7, E2 (containing E4), E5, L1, and LCR were amplified separately by polymerase chain reaction (PCR) using primers designed on the basis of the prototype

(<http://www.ncbi.nlm.nih.gov/genbank> accession no. NC_001443; primer sequences are shown in the Supplemental Data). Briefly, the PCR was conducted in a 50- μ L reaction mix containing 4 μ L of extracted DNA, 200 μ mol/L deoxynucleotide triphosphates, forward and reverse primers at .25 μ mol/L each, and 1.25 U of HotStarTaq Plus polymerase (Qiagen). The cycling conditions were as follows: activation of polymerase at 95°C for 5 min, 40 cycles of denaturation at 94°C for 50 s, annealing at 56°C–60°C for 50 s, and extension at 72°C for 50–70 s, followed by a final extension at 72°C for 8 min. Amplification was visualized by agarose gel electrophoresis. Samples with insufficient amplification product for sequencing were subjected to another round of PCR using the nested primers.

PCR products were purified by Microspin S-400 columns (GE Healthcare). Ten microliters of the purified PCR products were mixed with 2 μ L of BigDye Terminator sequencing reaction mix (version 3.1; Applied Biosystems), 3 μ L of 5 \times sequencing buffer, and 3.2 pmol of the sequencing primer; and made up to a final volume of 20 μ L according to the manufacturer's instructions. The cycling conditions for the labeling PCR were 25 cycles at 95°C for 15 s, 50°C for 15 s, and 65°C for 75 s. 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 software (version 2.5; Applied Biosystems). Mutations that occurred only once were confirmed by repeating from the original sample.

Naming of Variants and Phylogenetic Tree Construction

Variant sequences were named WW for "worldwide," followed by a number according to its prevalence as found in this study. The concatenated nucleotide sequences from the 5 complete open reading frames (ORFs; E6, E7, E2, E5, and L1) and the LCR region of individual variants were used for phylogenetic tree construction. The tree construction processes were repeated for each genomic region to identify the most informative surrogate region for lineage classification of HPV-58 variants.

Maximum-likelihood trees were constructed using the program PAUP* (version 4.0b10) [15]. Modeltest (version 3.7) [16] was used to identify the best evolutionary model. A neighbor-joining tree was constructed as a starting tree, followed by a maximum-likelihood tree using the subtree pruning and regrafting (SPR) search approach. The data were bootstrap resampled 1,000 times. To verify the tree topologies observed from maximum-likelihood trees, the program MrBayes (version 3.1.2) [17] was used for Bayesian tree construction, with the nucleotide substitution model set according to the Modeltest results. The Markov chain Monte Carlo analysis was run for 5,000,000 generations with trees sampled at every 1,000 generations. A burn-in rate of 25% was used in summarizing the data. The trees were displayed with Figtree (version 1.1.2; <http://tree.bio.ed.ac.uk/software>).

Table 1. Distribution of Study Samples According to the Source of Collection and Cervical Pathology Status

Region, country or city	Total no. of specimens	No. of specimens of each cervical pathology status					
		Normal	ASCUS	LGSIL	HGSIL	Carcinoma	Unknown
Africa
Zimbabwe ^a	69	0	0	0	0	0	69
Americas	63	6	1	6	10	2	1 ^b
Canada	10	2	0	4	4	0	0
United States	37	... ^c	... ^c	... ^c	... ^c	... ^c	... ^c
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
South 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
United Kingdom	14	6	0	6	2	0	0
Italy	17	10	1	3	1	0	2
Total	401	83	26	68	72	34	81 ^b

NOTE. ASCUS, atypical squamous cells of undetermined significance; CIN, cervical intraepithelial neoplasia; HGSIL, high-grade squamous intraepithelial lesions (including CIN2, CIN3, and severe dysplasia); LGSIL, low-grade squamous intraepithelial lesions (including CIN1 and mild dysplasia);

^a Cervical or vaginal swabs.

^b Not including 37 anal swab samples.

^c Anal swab samples.

Geographical Distribution of Variant Lineages

The rate of detection of each variant lineage was compared among the 4 regions by use of a Pearson χ^2 test. When a significant difference was obtained between regions (defined as $P < .05$), detection rates between regions were compared 2 by 2 with a Fisher exact test, for a total of 6 comparisons. The level of significance for 6 possible comparisons was then set at .008 according to the Bonferroni correction.

The distribution of variant lineages among anal swabs that were collected from men at a single center in the United States was compared with that among cervical samples collected from other parts of the Americas by use of a Pearson χ^2 test or a Fisher exact test as appropriate. The association between variant lineage and cervical pathology status was assessed by a Fisher exact test. Two-tailed P values of $< .05$ were regarded as significant.

RESULTS

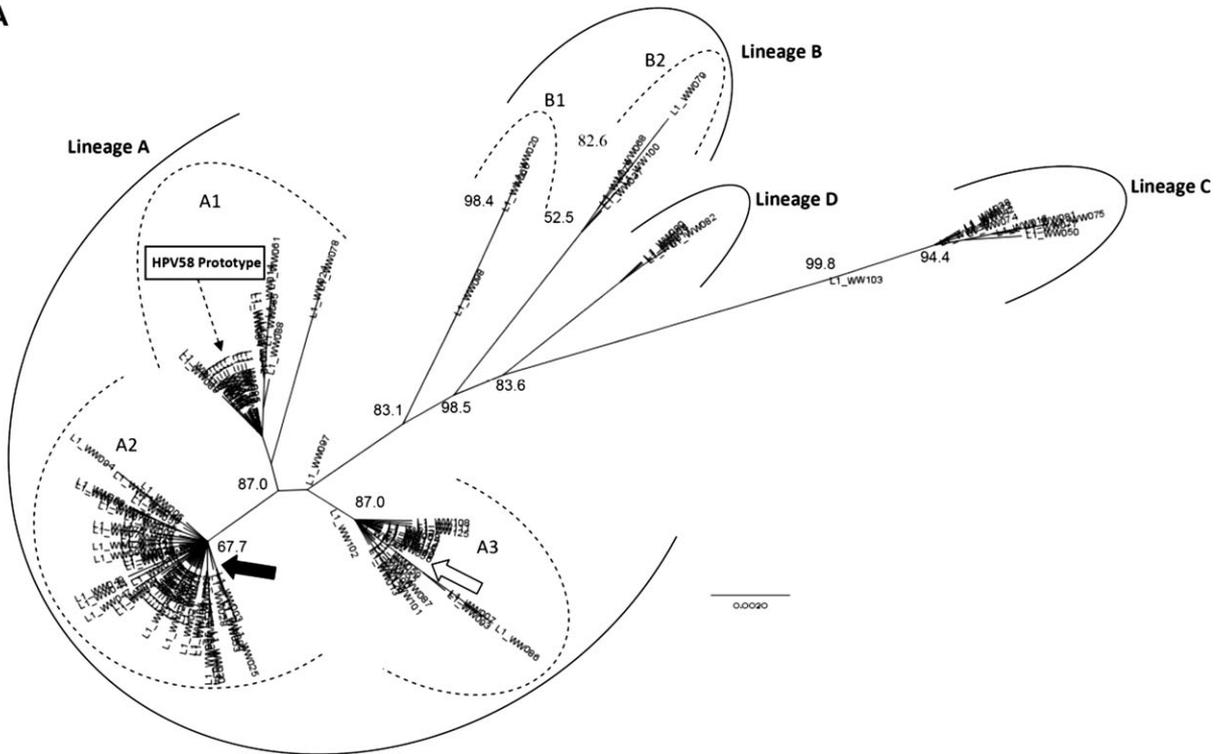
Lineage Classification

Altogether, 268 unique concatenated E6-E7-E2-E5-L1-LCR nucleotide sequences of HPV-58 variants were assembled. The lengths ranged from 4416 bp to 4462 bp, accounting for ~57% of the whole viral genome. Since none of the assembled

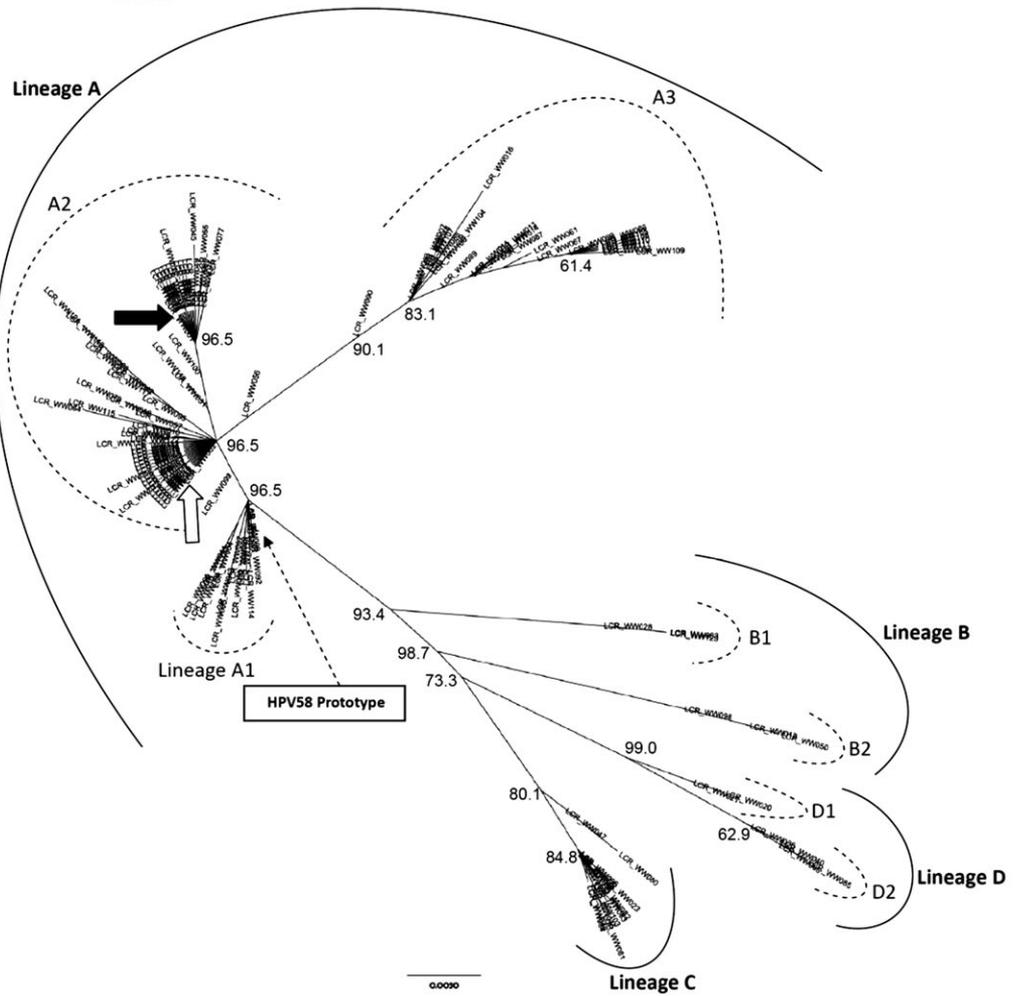
sequences was identical to the prototype, the sequence available at GenBank (GenBank accession no. NC_001443) was used to assemble a concatenated prototype sequence to serve as a reference. The maximum-likelihood tree revealed 4 clusters (Figure 1), and the topology was same as that observed from the Bayesian tree.

Among all the genomic regions examined, L1 and LCR displayed tree topologies that most closely resembled that of the concatenated E6-E7-E2-E5-L1-LCR sequences. The L1 and LCR regions were considered to be most informative surrogate regions for phylogenetic grouping of HPV-58 variants when the full genome sequence is not available. The lineage containing the prototype was assigned as lineage A, and the others were arbitrarily designated as lineages B, C, and D. Figures 2A and 2B show the maximum-likelihood trees of the L1 and LCR sequences, which revealed topologies that were same as those observed from the Bayesian trees. The 2 most prevalent L1 variants, L1_WW001 and L1_WW002, were found in 26.7% and 11.7% of samples, respectively; whereas the L1_WW054 variant, found in .2% of samples, is the prototype. The 2 most prevalent LCR variants, LCR_WW001 and LCR_WW002, were found in 20.8% and 11.5% of isolates, respectively. The variant LCR_WW009, which was found in 3.6% of isolates, is

A



B



Lineage	L1- 526 bp												LCR- 173 bp												LCR- 337 bp														
	6	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				
	0	0	0	0	2	4	4	4	4	4	4	5	5	2	2	2	2	3	3	3	3	3	3	4	4	5	6	6	7	7	7	7	7	7	7	7	7		
	1	3	3	5	2	1	3	4	5	5	5	9	0	3	5	6	7	8	0	1	3	4	6	9	2	2	4	1	8	1	3	4	5	7	7	9	3	5	
	4	8	9	1	2	6	4	0	0	8	9	6	0	9	7	6	7	4	4	3	2	5	9	5	1	9	0	9	6	4	0	5	5	8	9	2	0	2	
Prototype	A	C	A	C	A	A	T	A	G	G	G	T	C	A	T	C	C	C	A	A	G	T	T	G	G	T	A	G	G	C	C	G	A	T	A	C	C	C	
A1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	R ⁷	-	-	-	-	-	-	-	-	-	-	-	-
A2	C	-	-	-	-	G	C	-	-	-	-	-	G	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	
A3	M ¹	-	-	-	-	-	-	-	-	-	-	-	R ⁴	-	-	IN	-	G	-	-	-	-	-	-	-	-	-	-	C	-	R ³	-	-	-	-	-	T		
B1	-	-	-	-	-	-	G	-	-	-	-	G	G	-	S ⁵	-	C	-	-	-	-	-	-	-	G	A	-	-	-	-	-	-	-	-	-	G	T		
B2	-	-	-	-	G	-	-	-	T	-	C	T	G	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	A	C	-	-	G	-	G	K ¹⁰	-		
C	-	T	G	A	-	-	C	C	R ²	R ³	G	-	-	-	-	G	-	-	C	K ⁶	-	-	-	-	-	-	-	A	-	-	R ⁹	-	G	T	-	-			
D1	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	G	-	-	A	-	-	-	G	-	-	-	-	A	-	-	C	G	T	G	T	-			
D2	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	G	-	-	A	-	-	A	A	G	-	-	-	A	-	-	C	G	-	G	T	-	-		

Figure 3. Signature sequences for human papillomavirus type 58 (HPV-58) lineage identification. Numbers refer to nucleotide position of the HPV-58 prototype (<http://www.ncbi.nlm.nih.gov/genbank> accession no. NC_001443). A dash represents the same nucleotide as in the prototype. IN, insertion of 12 bp. For M¹, 91% of isolates have C and 9% have A; for R², 54.5% have G and 45.5% have A; for R³, 91% have A and 9% have G; for R⁴, 96% have G and 4% A; for S⁵, 67% have C and 33% have G; for K⁶, 86% have G and 14% have T; for R⁷, 78.6% have A and 21.4% have G; for R⁸, 96% have G and 4% have A; for R⁹, 86% have G and 14% have A; and for K¹⁰, 67% have T and 33% have G.

few for statistical analysis. Lineage C was found in 9.2% of isolates collected worldwide and was found significantly more frequently in Africa than in Asia, the Americas, or Europe ($P \leq .001$ for each comparison). Lineage D has a worldwide prevalence of 2.2% and seemed to be more frequent in Africa (8.7%). However, the number of lineage D isolates was too few for statistical analysis.

Since lineage A was the most prevalent lineage identified, a subgroup analysis was performed for the distribution of sub-lineages A1, A2, and A3. Sublineage A2 was the most frequently detected sublineage, accounting for 62.3% of lineage A isolates collected worldwide and dominating in Africa (94.1%), the Americas (87.0%), and Europe (93.1%). In contrast, sublineages A1, A2, and A3 were more evenly distributed in Asia, accounting for 16.2%, 47.8%, and 36.0%, respectively, of lineage A isolates found in this region. As a result, the proportion of lineage A isolates belonging to sublineage A2 was significantly lower in Asia than in other regions ($P < .001$), whereas the proportion belonging to sublineage A3 was significantly higher in Asia than in other regions ($P < .001$).

Of the 37 anal swabs collected from men in the United States, 30 (81.1%) were lineage A, 1 (2.7%) was lineage B, 5 (16.7%) were lineage C, and 1 (.3%) was lineage D. The proportion of each lineage among these samples was not

significantly different ($P = .06-.4$) from that of samples collected from women in the Americas.

Lineage Distribution and Cervical Pathology

In Asia, lineage A accounted for 94.8% of high-grade squamous intraepithelial lesion (HGSIL) and carcinoma samples and 96.4% of samples of normal cytology, low-grade squamous intraepithelial lesion (LGSIL), and atypical squamous cells of undetermined significance (ASCUS); no significant difference in the distribution of lineage A compared with non-A lineages was found ($P = .745$). Similarly, lineage A accounted for the majority of HGSIL/carcinoma samples in Hong Kong (94.8%) and South Korea (92.3%), and no significant association between lineage and lesion severity was found ($P = 1.0$ for Hong Kong; $P = .120$ for South Korea). The number of samples available from other regions was too few for a similar statistical analysis.

DISCUSSION

Analysis of intratypic sequence variation of HPV can provide important information for the design of diagnostic tools, development of vaccines, identification of molecular markers for epidemiological studies, elucidation of implications of sequence variation on biological and pathological properties, and

Figure 2. Phylogenetic trees of L1 and LCR sequences of human papillomavirus type 58 (HPV-58) variants. The maximum-likelihood tree was constructed with the PAUP* program (version 4.0) using the GTR+I+G model for nucleotide substitution. Bootstrap values of key positions generated with resampling 1,000 times are shown. The length of the scale bar represents the number of substitutions per nucleotide position. The position of the prototype (<http://www.ncbi.nlm.nih.gov/genbank> accession no. NC_001443) is indicated. The 2 most prevalent variant sequences, WW_001 and WW_002, are marked with black and white arrows, respectively. A, L1 open reading frame of 121 variants (HM639317–HM639717). B, LCR open reading frame of 123 variants (HQ338950–HQ339350).

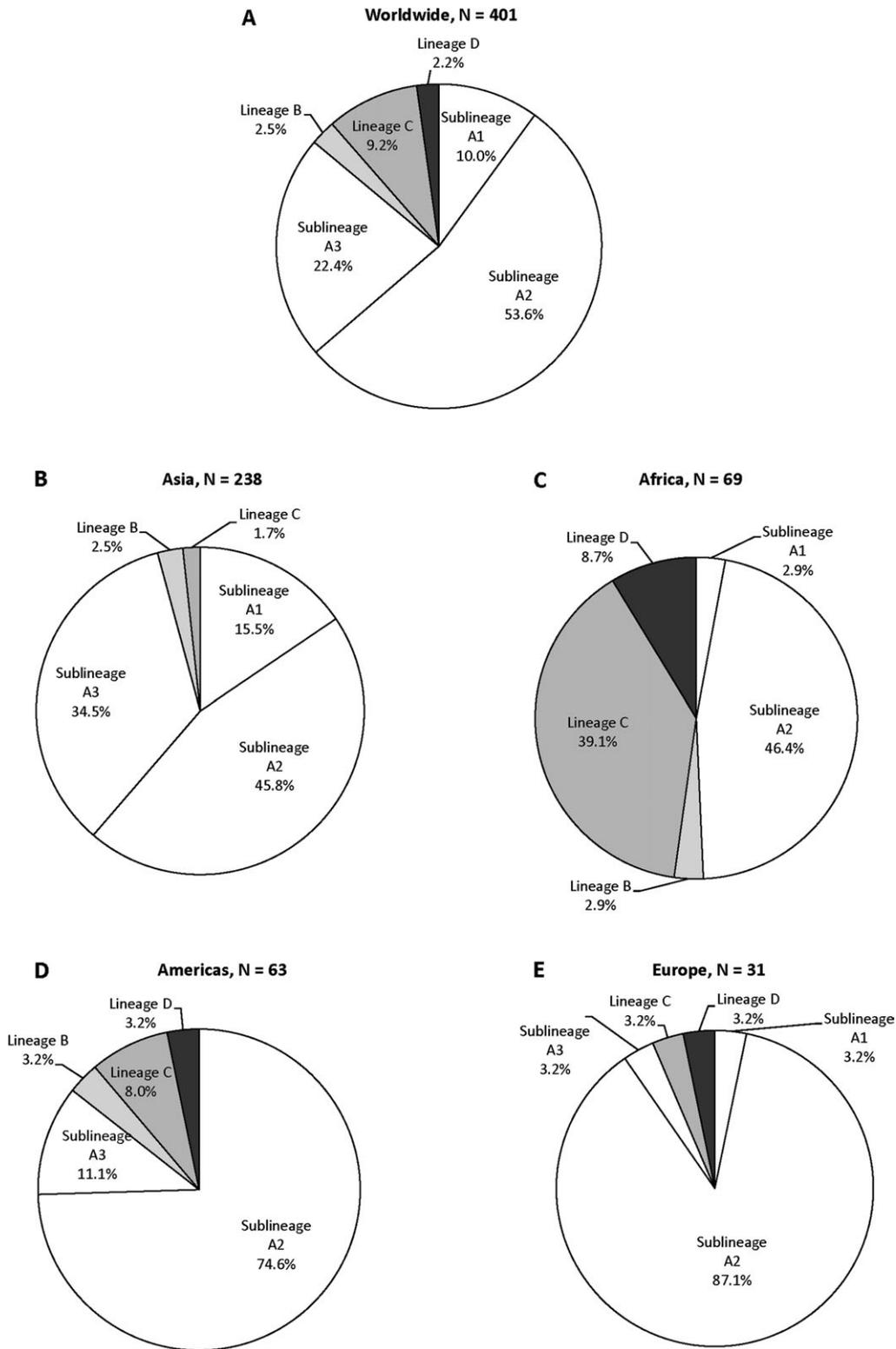


Figure 4. Geographical distribution of human papillomavirus type 58 (HPV-58) lineages. A, Worldwide ($N = 401$); B, Asia ($N = 238$); C, Africa ($N = 69$); D, the Americas ($N = 63$); E, Europe ($N = 31$).

understanding of the evolution and taxonomy of the virus [18, 19]. The currently available data are mainly derived from the 2 HPV types, HPV-16 and HPV-18, most commonly found in

cervical cancers. Data on sequence variation of HPV-58 isolates collected worldwide are scarcely available [20, 21]. In this study, ~57% of the whole viral genome was sequenced. The selected

regions included L1, which is the most important region for defining HPV type and variant; LCR, which is the most variable region; E6, which contains informative signatures for HPV-16 variant lineage classification; E7, which has been reported to be more variable than E6 for HPV-58 [14]; and E2 and E5, which are important in oncogenesis. To our knowledge, the number of HPV-58 isolates examined in this study represents the largest reported collection sampled from multiple countries around the world. Nevertheless, one should be aware of the fact that the number of samples available from Africa and Europe for this study was relatively small, and thus the distribution of variants in these regions might not be fully elucidated.

Our analysis on the E6-E7-E2-E5-L1-LCR concatenated sequences of HPV-58 variants showed 4 phylogenetically distinct clusters, suggesting that HPV-58 variants had evolved into 4 lineages. We then attempted to identify genomic regions that could best reproduce the 4 clusters. Among the 7 genomic regions examined, 5 of them (E6, E7, E2, E4, and E5) were relatively conserved, as expected for these proteins. The tree topologies generated from these 5 regions were quite different from that of the concatenated sequences. On the other hand, the LCR and L1 regions displayed a tree topology that most closely resembled that of the concatenated sequences and were therefore regarded as the most informative surrogate regions for HPV-58 variant lineage classification. A similar topology was also observed by Calleja-Macias et al [20], who used a 461-bp fragment of LCR of 21 HPV-58 variants for tree construction.

The error frequency estimated for a standard Taq polymerase-based PCR ranges from 2×10^{-4} to 30×10^{-4} [21]. To minimize the chance of recording artificial sequence variations, we performed sequencing from both directions in independent PCRs. In addition, sequence variations observed only once were repeated. It is unlikely that the sequence variations presented are due to errors produced during the amplification process. The observed maximum nucleotide sequence divergence of the L1 ORF within each lineage ranged from .4% to 1.7%, and was 2.2% for all variants together. This limited sequence divergence indicates the absence of subtypes or intermediary genomes within the HPV-58 variants. This observation concurs with previous studies on other HPV types [20, 22]. HPV-58, as with other HPV types, probably has gone through genetic drifts that became amplified by founder effects and bottlenecks of evolution.

A clear association between phylogenetic clustering and the ethnogeographic origin of HPV-16 variants has been observed previously, and thus HPV-16 lineages were named as follows: E (European), As (Asian), AA (Asian American), and Af-1 and Af-2 (African 1 and 2) [23, 24]. The largest available series of analyses on HPV-58 variants was reported by Calleja-Macias et al [20], which included 101 samples from different parts of the world. Their analysis on a 461-bp fragment of LCR revealed 21 variants, showing a limited amount of diversification in unique geographical locations and no clear geographical association with any

variants was observed. The present study allowed a more in-depth analysis based on a larger sample size. Although the ethnogeographic correlation for HPV-58 lineages was not as prominent as that for HPV-16, a predilection in distribution of HPV-58 lineages was observed in this study. Lineage A predominated in all regions except in Africa, where lineages A and C existed in comparable proportions. Although Asia comprised the largest number of samples in this study, none of them belonged to lineage D. The distribution of sublineages A1, A2, and A3 also displayed geographical variation. Although sublineage A2 predominated in Africa, the Americas and, Europe, a relatively higher frequency of sublineages A1 and A3 was found in Asia.

We hypothesize that lineage A (probably sublineage A2) was the oldest lineage, which disseminated with early human evolution and migration and had seeded into different parts of the world before other lineages emerged [24–25]. Host or environmental factors might have favored the emergence and spread of lineage C in Africa, whereas lineage D was difficult to establish in Asia.

We assigned A1 to the sublineage that contained the prototype, which was cloned from a patient with cervical cancer in Japan. In this study, sublineage A1 was rarely detected except in Asia. It is worthwhile to further investigate whether the reported higher contribution of HPV-58 to invasive cancers in East Asia is associated with a higher level of oncogenicity of sublineage A1 [7–12].

Since all the anal samples from men available for this study were collected from a single center in the United States, we compared their lineage distribution with samples from women collected from the rest of the Americas. The results showed that there were no significant differences between samples from men and those from women, and therefore pooling these samples together for the analysis of the geographical distribution of lineages was justified.

A potential limitation of the present study is the lack of sufficient samples to allow further analyses of the geographical distribution of variant lineages stratified according to cervical pathology status. Nevertheless, at least for Asia and the Americas, the proportion of samples with normal cytology or LGSIL was similar to that of samples with ASCUS, HGSIL, or carcinoma (48.7% and 48.0% of samples from Asia and the Americas, respectively, were normal/LGSIL), although Europe had a higher proportion of normal/LGSIL samples (86.2%), and information on cervical status for the samples from Africa was not known. We attempted to analyze the association between oncogenic risk and variant lineage on the basis of samples collected from Asia, Hong Kong, and South Korea, where a substantial number of samples in this study were collected, but no significant association was observed. However, such a result should not be regarded as final. Further studies are required to examine the oncogenic association of these variant lineages.

This study provides a detailed analysis on HPV-58 variant lineages and indicates that the distribution may be linked

ethnogeographically. Whether this reflects the survival fitness of these variants under different host genetic and environmental pressures or that some of these lineages are still slowly evolving and extending their ecological territories remains to be established. Further study on the evolution of HPV-58 and close monitoring of the possibility of type replacement by this virus following the widespread administration of HPV vaccines are warranted. It is worthwhile to further study the biological and pathological implications of this lineage classification system.

Supplementary Data

Supplementary tables are available online at <http://jid.oxfordjournals.org>.

Funding

This study was supported by the International Centre for Genetic Engineering and Biotechnology (project no. CRP/CHN08-03) and the Chinese University of Hong Kong Focused Investment Scheme Funding to Centre for Microbial Genomics and Proteomics. Specimens from Canada were obtained with the support of the Cancer Society of Canada. F. D. M. 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.

Acknowledgments

We are grateful for the contributions of Sergio Andres Tonon and Cláudia Renata F. Martins, who unfortunately have passed away before the completion of this study. We thank Robert D. Burk for his advice on lineage classification and naming.

References

- Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. *Lancet* **2007**; 370:890–907.
- FUTURE II Study Group. Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions. *N Engl J Med* **2007**; 356:1915–27.
- Paavonen J, Naud P, Salmerón J, et al. Efficacy of human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by oncogenic HPV types (PATRICIA): final analysis of a double-blind, randomised study in young women. *Lancet* **2009**; 374:301–14.
- 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**; 121:621–32.
- Muñoz N, Bosch FX, de Sanjosé S, et al. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med* **2003**; 348:518–27.
- 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**; 111:278–85.
- 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**; 70:408–11.
- 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.
- 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.
- 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.
- Asato T, Maehama T, Nagai Y, Kanazawa K, Uezato H, Kariya K. 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.
- Bao YP, Li N, Smith JS, Qiao YL; ACCPAB members. Human papillomavirus type distribution in women from Asia: a meta-analysis. *Int J Gynecol Cancer* **2008**; 18:71–9.
- Chan DP, Cheung TH, Tam AO, et al. Risk association of HLA-A allele and high-risk HPV infection for cervical neoplasia in Chinese women. *J Infect Dis* **2005**; 192:1749–56.
- 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.
- Swofford DL. PAUP: phylogenetic analysis using parsimony (*and other methods), version 4.0 b10. Sunderland, Massachusetts: Sinauer, 2002.
- Posada D, Crandall KA. Modeltest: testing the model of DNA substitution. *Bioinformatics* **1998**; 14:817–8.
- Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **2001**; 17:754–5.
- Bernard HU, Calleja-Macias IE, Dunn ST. Genome variation of human papillomavirus types: phylogenetic and medical implications. *Int J Cancer* **2006**; 118:1071–6.
- Sichero L, Villa LL. Epidemiological and functional implications of molecular variants of human papillomavirus. *Braz J Med Biol Res* **2006**; 39:707–17.
- 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.
- Eckert KA, Kunkel TA. DNA polymerase fidelity and the polymerase chain reaction. *PCR Methods Applic* **1991**; 1:17–24.
- Stewart AC, Eriksson AM, Manos MM, et al. Intratype variation in 12 human papillomavirus types: a worldwide perspective. *J Virol* **1996**; 70:3127–36.
- Chan SY, Ho L, Ong CK, et al. Molecular variants of human papillomavirus type 16 from four continents suggest ancient pandemic spread of the virus and its coevolution with humankind. *J Virol* **1992**; 66:2057–66.
- Ho L, Chan SY, Burk RD, et al. The genetic drift of human papillomavirus type 16 is a means of reconstructing prehistoric viral spread and the movement of ancient human populations. *J Virol* **1993**; 67:6413–23.
- Li Y, Li Z, He Y, et al. Phylogeographic analysis of human papillomavirus 58. *Sci China Ser C-Life Sci* **2009**; 52:1164–72.