Papillomavirus Research 7 (2019) 67–74

Contents lists available at ScienceDirect

Papillomavirus Research

journal homepage: www.elsevier.com/locate/pvr

Human papillomavirus 16 sub-lineage dispersal and cervical cancer risk worldwide: Whole viral genome sequences from 7116 HPV16-positive women☆

Gary M. Clifford[a,*], Vanessa Tenet[a], Damien Georges[a], Laia Alemany[b,c], Miquel Angel Pavón[b,d], Zigui Chen[e], Meredith Yeager[f,g], Michael Cullen[f,g], Joseph F. Boland[f,g], Sara Bass[f,g], Mia Steinberg[f,g], Tina Raine-Bennett[h], Thomas Lorey[i], Nicolas Wentzensen[f], Joan Walker[j], Rosemary Zuna[j], Mark Schiffman[f], Lisa Mirabello[f]

[a] International Agency for Research on Cancer, Lyon, France
[b] Infections and Cancer Unit, Cancer Epidemiology Research Program, Catalan Institute of Oncology, Bellvitge Institute of Biomedical Research, Barcelona, Spain
[c] CIBER en Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain
[d] CIBER en Oncología (CIBERONC), Barcelona, Spain
[e] Department of Microbiology, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China
[f] Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD, USA
[g] Cancer Genomics Research Laboratory, Leidos Biomedical Research, Inc., Frederick, MD, USA
[h] Women's Health Research Institute, Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA
[i] Regional Laboratory, Kaiser Permanente Northern California, Oakland, CA, USA
[j] University of Oklahoma Health Sciences Center, Oklahoma City, USA

ARTICLE INFO

Keywords:
HPV16
Cervical cancer
HPV carcinogenesis
HPV epidemiology
HPV genomics
Whole virus genome sequencing

ABSTRACT

Background: Human papillomavirus (HPV)16 can be separated into genetic sub-lineages (A1–4, B1–4, C1–4, D1–4) which may have differential cervical cancer risk.
Methods: A next-generation sequencing assay was used to whole-genome sequence 7116 HPV16-positive cervical samples from well-characterised international epidemiological studies, including 2076 controls, 1878 squamous cell carcinoma (SCC) and 186 adenocarcinoma/adenosquamous cell carcinoma (ADC), and to assign HPV16 sub-lineage. Logistic regression was used to estimate region-stratified country-adjusted odds ratios (OR) and 95%CI.
Results: A1 was the most globally widespread sub-lineage, with others showing stronger regional specificity (A3 and A4 for East Asia, B1–4 and C1–4 for Africa, D2 for the Americas, B4, C4 and D4 for North Africa). Increased cancer risks versus A1 were seen for A3, A4 and D (sub)lineages in regions where they were common: A3 in East Asia (OR = 2.2, 95%CI:1.0–4.7); A4 in East Asia (6.6, 3.1–14.1) and North America (3.8, 1.7–8.3); and D in North (6.2, 4.1–9.3) and South/Central America (2.2, 0.8–5.7), where D lineages were also more frequent in ADC than SCC (3.2, 1.5–6.5; 12.1, 5.7–25.6, respectively).
Conclusions: HPV16 genetic variation can strongly influence cervical cancer risk. However, burden of cervical cancer attributable to different sub-lineages worldwide is largely driven by historical HPV16 sub-lineage dispersal.

Abbreviations: ADC, adenocarcinoma/adenosquamous cell carcinoma; ASCUS, atypical squamous cells of undetermined significance; CI, confidence interval; CIN, cervical intraepithelial neoplasia; HPV, human papillomavirus; HC2, Hybrid Capture2; HSIL, high-grade squamous intraepithelial lesion; IARC, International Agency for Research on Cancer; ICO, Catalan Institute of Oncology; KPNC, Kaiser Permanente Northern California; LSIL, low-grade squamous intraepithelial lesion; NCI, National Cancer Institute; OR, odds ratio; SCC, squamous cell carcinoma

☆ The authors alone are responsible for the views expressed in this article and they do not necessarily represent the views, decisions or policies of the institutions with which they are affiliated.

* Correspondence to: International Agency for Research on Cancer, 150 cours Albert Thomas, 69372 Lyon Cedex 08, France.
E-mail addresses: CliffordG@iarc.fr (G.M. Clifford), TenetV@iarc.fr (V. Tenet), GeorgesD@iarc.fr (D. Georges), lalemany@iconcologia.net (L. Alemany), mpavon@iconcologia.net (M.A. Pavón), zigui.chen@cuhk.edu.hk (Z. Chen), yeagerm@mail.nih.gov (M. Yeager), michael.cullen@nih.gov (M. Cullen), bolandj2@mail.nih.gov (J.F. Boland), sara.bass2@nih.gov (S. Bass), mia.steinberg@nih.gov (M. Steinberg), Tina.R.Raine-Bennett@kp.org (T. Raine-Bennett), thomas.lorey@kp.org (T. Lorey), wentzenn@mail.nih.gov (N. Wentzensen), joan-walker@ouhsc.edu (J. Walker), rosemary-zuna@ouhsc.edu (R. Zuna), schiffmm@exchange.nih.gov (M. Schiffman), mirabellol@mail.nih.gov (L. Mirabello).

https://doi.org/10.1016/j.pvr.2019.02.001
Received 21 November 2018; Accepted 3 February 2019
Available online 06 February 2019
2405-8521/ © 2019 International Agency for Research on Cancer; licensee Elsevier B.V. This is an open access article under the CC BY-NC-ND 3.0 license (http://creativecommons.org/licenses/BY-NC-ND 3.0/IGO/).

1. Introduction

Of 13 high-risk types considered to cause cervical cancer [1], human papillomavirus (HPV)16 is by far the most carcinogenic [1,2], accounting for more than half of the estimated 620,000 cervical cancers diagnosed worldwide every year [3]. Although only a minority of HPV16 infections persist beyond 1 year [4], a large proportion of those that do persist will go on to develop (pre)cancer [5]. Nevertheless, the factors that determine which HPV infections will persist/progress are not well understood. Given that large known differences in carcinogenicity between HPV16 and closely related high-risk types (e.g. HPV31 and HPV35) must be explained by differences in their relatively small (8 kb) genomes, finer genetic variation may also play a significant role in determining which HPV16 infections progress to cancer.

HPV16 can be divided into four main lineages (A-D), which themselves can each be divided into four sub-lineages (A1–4, B1–4, C1–4 and D1–4). Targeted sequencing of HPV16 genomes has already linked genetic variation to differential cervical cancer risk [6–10], and with differential potential for adenocarcinoma/adenosquamous cell carcinoma (ADC) versus squamous cell carcinoma (SCC) [8,11–15].

More recently, a novel next-generation HPV16 sequencing method has been developed [16], allowing high-throughput sequencing of large numbers of whole HPV16 genomes. This technique has led to further refinement of differential HPV16 risks [17], as well as other novel discoveries in HPV genomics [18,19].

Here we report on the application of this next-generation sequencing assay to a large well epidemiologically-characterised collection of HPV16-positive cervical samples, with the objective to describe the worldwide distribution of HPV16 (sub)lineages, and their relative cervical cancer risks. Given that co-evolution of HPV16 and humans has been suggested to influence cervical cancer risk [17,20,21], we also had reason to investigate whether relative cancer risks by HPV16 (sub) lineage might be population-specific.

2. Material and methods

2.1. Origin of clinical specimens

2.1.1. International Agency for Research on Cancer (IARC)

IARC, Lyon, France, has coordinated cervical cancer case series, case-control studies, and population-based HPV prevalence surveys in many countries, spanning 25 years (see IARC HPV Variant Study Group in acknowledgements) [22–25]. Cervical samples (exfoliated cells or tissue/biopsy specimens) were genotyped for 37 HPV types using GP5+/6+-based PCR [26] in a centralized laboratory (Department of Pathology, VU University Medical Center, Amsterdam, The Netherlands). All 2848 HPV16-positive samples with DNA remaining in the IARC biobank were sent to the National Cancer Institute, National Institutes of Health, Rockville, MD (NCI) for whole HPV16 genome sequencing. A subset of these samples have previously contributed to an analysis of HPV16 genetic variation based on the targeted sequencing of the E6 gene and/or long control regions only [9,27].

2.1.2. Catalan Institute of Oncology (ICO)

ICO coordinated an international invasive cervical cancer case series (RIS HPV TT study) [28]. Formalin-fixed paraffin-embedded tumour tissues from over 20 countries were tested for 25 HPV types using the SPF10-LiPA25 protocol (Laboratory Biomedical Products, Rijswijk, The Netherlands). From this repository, 795 HPV16-positive invasive cervical cancers with available DNA were sent to NCI for whole HPV16 genome sequencing. A subset of these samples have contributed to analyses of HPV16 genetic variation based on the targeted sequencing of E6 gene only [14,29].

2.1.3. U.S. National Cancer Institute (NCI)

The Kaiser Permanente Northern California (KPNC) PaP cohort includes specimens from approximately 55,000 women that underwent routine cervical cancer screening at KPNC using Hybrid Capture2 (HC2) (QIAGEN, Germantown, MD) and cytology between January 2007 and January 2011. HPV typing of HC2-positive specimens was performed according to different methods [18], but all HPV16-positive samples were selected for whole HPV16 genome sequencing analysis at NCI. In the current work, we have included the 3215 of 3579 HPV16-positive specimens for which valid whole HPV16 genome sequence results were previously reported [17].

The SUCCEED study enrolled women undergoing colposcopy and/or diagnosed with cervical cancer at the University of Oklahoma Dysplasia Clinic between 2003 and 2009. In this analysis, we included 989 participants with HPV16 DNA detected in liquid-based cytology specimens by Linear Array® HPV Genotyping System (Roche Diagnostics) [30]. A subset of these samples has contributed to analyses of HPV16 genetic variation based on the targeted sequencing of L1 gene only [10].

The Proyecto Epidemiologico Guanacaste study recruited 10,049 women residing in a high-risk Costa Rican province in 1993–1994, among whom cervical HPV DNA was detected using a MY09/M11 L1 primer PCR system [31]. For the current analysis, 93 were available for whole HPV16 genome sequencing from among 503 HPV16-positive samples with previous targeted sequencing of E6 gene and LCR only [32].

Local ethical review committees in each of the participating countries approved the original studies, from which the samples for the current study are derived, and which were all conducted in accordance with the Declaration of Helsinki.

2.2. Whole HPV16 genome sequencing

All HPV16-positive DNA samples from the above described studies were processed according to the same protocol in the Cancer Genomics laboratory, Leidos Biomedical Research, Inc., Frederick, MD. A custom Thermo Fisher Ion Torrent AmpliSeq HPV16 panel approach was used to amplify the entire 7906 bp HPV16 genome, as previously reported in detail [16]. The next-generation sequencing assay used the Ion Torrent Proton and a custom HPV16 Ion Ampliseq panel of 47 multiplexed primer sets. Custom overlapping degenerate primers were designed to cover the entire viral genomes for all HPV16 variant lineages. Raw sequencing reads were quality and adaptor trimmed using the Torrent Suite Software and aligned to the HPV16 reference sequence (7906 bp) from GenBank (NC_001526.4) [16]. Variants were identified using the Torrent Variant Caller v.5.0.3 and annotated using snpEff v.3.6c [33]. The pipeline was executed using Snakemake [34], and settings and parameters can be found at https://github.com/cgrlab/cgrHPV16.

2.3. HPV16 sub-lineage classification and nomenclature

HPV16 variant lineage assignment was based on the maximum likelihood tree topology constructed using sixteen (A1–4, B1–4, C1–4, and D1–4) HPV16 variant sub-lineage reference sequences [19]. In the presence of multiple HPV16 lineages, a 'predominant' variant lineage was assigned if the more common variable sites were present in at least 60% of the sequence reads. Specimens with approximately equal levels of multiple isolates were excluded from the analysis. Samples with overall poor coverage across the genome (< 2000 total reads per sample or < 2000 nucleotide positions callable), and individual nucleotide sites per sample with low reads (< 5), were also excluded. A total of 9% of IARC, 68% of ICO, 10.2% of PaP cohort and 8% of SUCCEED/Guanacaste specimens were excluded due to poor read depth, poor coverage across the genome, HPV16 variant coinfection for which a predominant lineage could not be assigned and/or inability to assign sub-lineage. A large proportion of ICO samples failed sequencing due to the older FFPE-derived DNA being too degraded for amplification. We used these stringent quality control thresholds to minimize sequencing errors in the dataset for analysis.

Table 1
Geographic distribution of 7 116 HPV16-positive cervical samples.

Region and country	HPV16-positive samples (N)					
	Invasive cervical cancer			HSIL or CIN2/3	Control	Total
	SCC	ADC	Unknown histology			
North Africa	**209**	**10**	**0**	**0**	**12**	**231**
Algeria	98	3	0	0	10	111
Morocco	111	7	0	0	2	120
Sub-Saharan Africa	**378**	**13**	**7**	**19**	**130**	**547**
Benin[a,b]	3	0	0	0	0	3
Guinea	26	2	0	3	33	64
Kenya[a]	46	1	3	0	0	50
Mali[a,b]	40	0	0	0	0	40
Mozambique[a,b]	44	0	0	0	0	44
Nigeria	9	1	0	3	24	37
Rwanda	114	6	2	5	73	200
South Africa[a]	49	2	2	8	0	61
Tanzania[a]	21	1	0	0	0	22
Uganda[a,b]	26	0	0	0	0	26
East Asia	**361**	**26**	**42**	**39**	**126**	**594**
China/ Taiwan	75	6	1	13	33	128
Indonesia[a,b]	15	0	0	0	0	15
South Korea[b]	0	0	39	21	3	63
Mongolia[b]	45	0	2	5	56	108
Philippines	43	8	0	0	2	53
Thailand	183	12	0	0	20	215
Vietnam[b]	0	0	0	0	12	12
South Asia	**215**	**14**	**5**	**143**	**166**	**543**
Bangladesh[a,b]	2	0	1	0	0	3
Bhutan	51	2	3	132	94	282
India	96	9	0	6	58	169
Kuwait[a]	24	2	0	0	0	26
Nepal[b]	20	0	0	5	12	37
Pakistan	22	1	1	0	2	26
Europe	**172**	**18**	**0**	**2**	**33**	**225**
Bosnia-Herzegovina[a,b]	0	8	0	0	0	8
Croatia[a,b]	0	1	0	0	0	1
Czech Republic[a,b]	0	1	0	0	0	1
Georgia[a]	51	1	0	0	0	52
Germany[a,b]	13	0	0	0	0	13
Greece[a,b]	0	1	0	0	0	1
Israel[a]	6	1	0	0	0	7
Italy[a,b]	0	0	0	0	8	8
Poland	78	4	0	2	25	109
Spain[a]	24	1	0	0	0	25
North America	**214**	**59**	**0**	**2550**	**1415**	**4238**
Canada[a,b]	27	0	0	0	0	27
United States	187	59	0	2550	1415	4211
South/Central America	**429**	**46**	**4**	**26**	**121**	**626**
Argentina	33	2	0	2	30	67
Bolivia[a,b]	15	0	1	0	0	16
Brazil	64	13	0	0	5	82
Chile	61	4	0	0	15	80
Colombia[a]	20	4	0	0	0	24
Costa Rica[b]	0	0	2	24	66	92
Cuba[a]	23	3	0	0	0	26
Honduras[a,b]	8	0	0	0	0	8
Panama[a,b]	33	0	1	0	0	34
Paraguay	67	3	0	0	3	73
Peru	86	17	0	0	2	105
Venezuela[a,b]	19	0	0	0	0	19
Oceania[b]	**0**	**0**	**5**	**14**	**93**	**112**
Fiji[b]	0	0	1	1	34	36
Vanuatu[b]	0	0	4	13	59	76
Total	**1978**	**186**	**63**	**2793**	**2096**	**7116**

ADC, adenocarcinoma/adenosquamous cell carcinoma; CIN, cervical intraepithelial neoplasia; HPV, human papillomavirus; HSIL, high-grade squamous intraepithelial lesion; SCC, squamous cell carcinoma.

[a] Not included in case: control analysis (Table 2).

[b] Not included in analysis by histological type of cancer (Table 3).

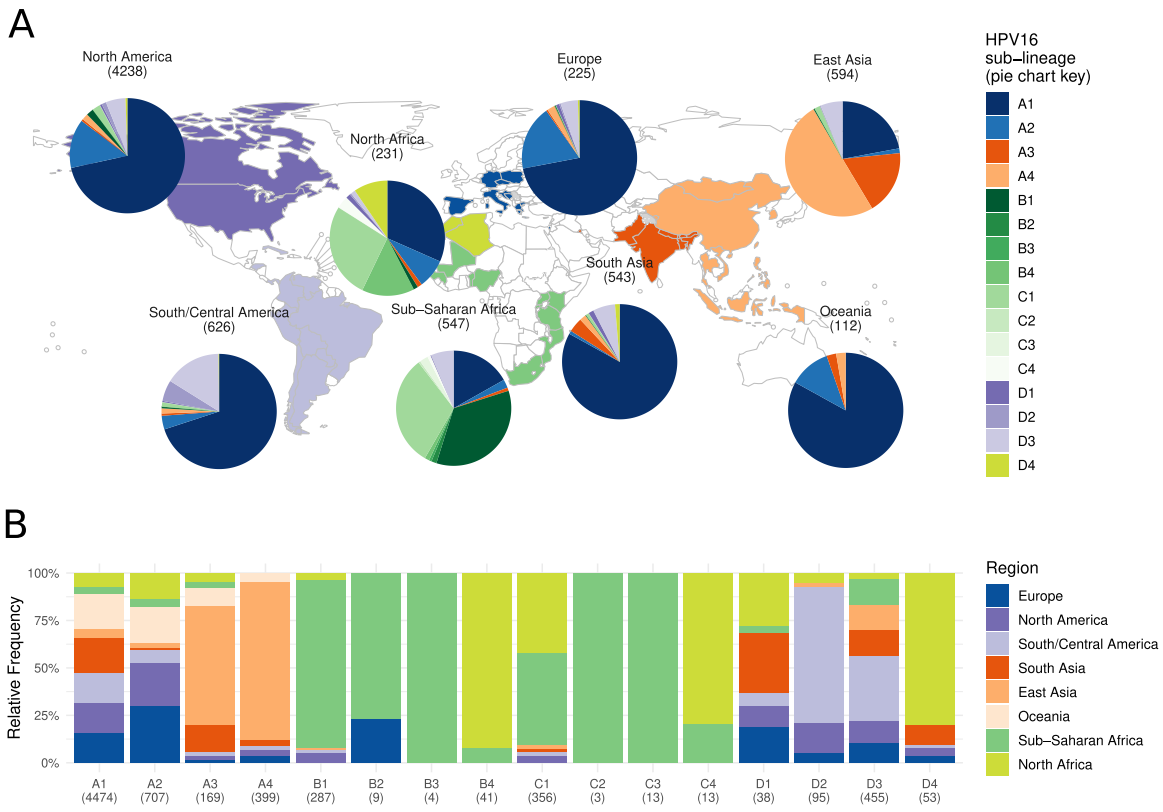

Fig. 1. Distribution of sub-lineages in 7116 HPV16-positive samples, by geographic region. The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement. Data source: IARC Map production: IARC World Health Organization © WHO 2018. All rights reserved.

2.4. Statistical analysis

2.4.1. Case-control analysis

Subjects were classified as either controls (including normal cells, atypical squamous cells of undetermined significance [ASCUS], low-grade squamous intraepithelial lesion [LSIL], or cervical intraepithelial neoplasia grade 1 [CIN1]), or as cancer cases (including SCC, ADC, or unspecified invasive cervical cancer). Samples from IARC population-based HPV prevalence studies for which histology and cytology were unavailable were also considered as controls. Samples reported as CIN grade 2 or 3 or high-grade squamous intraepithelial lesion (HSIL) were excluded from case-control analyses. Region-specific and worldwide odds ratios (OR), adjusted for country, were calculated using unconditional logistic regression to assess associations between HPV16 (sub)lineages and case: control status, predominantly using the A1 sub-lineage as the reference group (subjects from countries with only cases or only controls are thus dropped from analyses).

Table 2
Comparison of HPV16 (sub)lineages between cervical cancer cases and controls, by region.

Region	Cervical cancer cases								Controls								P value[a]	OR (95% CI)[b,c] versus A1		
	A				B	C	D	*Total*	A				B	C	D	*Total*		A3	A4	D
	A1	A2	A3	A4					A1	A2	A3	A4								
North Africa	68	19	2	0	37	67	26	***219***	5	0	1	0	0	3	3	***12***	0.093	0.1 (0.0–1.5)	–	0.6 (0.1–2.7)
Sub-Saharan Africa	18	0	0	0	76	53	13	***160***	16	1	3	0	54	39	17	***130***	0.180	0 (0 – 1.3)[d]	–	0.8 (0.2–2.5)
East Asia	63	4	48	255	2	9	33	***414***	47	3	33	28	0	0	3	***114***	**< 0.001**	**2.2 (1.0–4.7)**	**6.6 (3.1–14.1)**	0.5 (0.1–5.8)
South Asia	177	0	6	4	0	1	17	***205***	143	1	6	1	0	1	14	***166***	0.842	1.3 (0.4–4.4)	1.7 (0.2–16.4)	0.8 (0.4–1.8)
Europe	60	19	1	1	0	0	1	***82***	19	5	0	1	0	0	0	***25***	0.756	ND (0 - ∞)[d]	0.3 (0.0–5.3)	ND (0 - ∞)[d]
South/Central America	232	11	3	10	0	6	91	***352***	96	11	0	0	0	0	14	***121***	**0.006**	ND (0.3 - ∞)[d]	ND (1.0 - ∞)[d]	2.2 (0.8–5.7)
North America	141	35	3	10	3	2	52	***246***	1077	185	8	20	39	22	64	***1415***	**< 0.001**	2.9 (0.7–10.9)	**3.8 (1.7–8.3)**	**6.2 (4.1–9.3)**
Oceania	4	1	0	0	0	0	0	***5***	77	11	2	3	0	0	0	***93***	0.623	0 (0 – 46.3)[d]	0 (0 – 30.5)[d]	–
Total	**763**	**89**	**63**	**279**	**118**	**138**	**233**	***1683***	**1480**	**217**	**53**	**53**	**93**	**65**	**115**	***2076***		1.6 (0.9–2.8)	**4.5 (2.7–7.5)**	**3.1 (2.2–4.4)**

ADC, adenocarcinoma/adenosquamous cell carcinoma; CI, confidence interval; HPV, human papillomavirus; OR, odds ratio; SCC, squamous cell carcinoma; ND, not defined.

[a] Fischers exact test.

[b] Adjusted for country.

[c] All regional ORs for A2, B and C versus A1 were non-significant and are not shown.

[d] Crude OR and 95%CI (i.e. not adjusted for country).

2.4.2. HPV16 sub-lineages and histological type of cervical cancer

ADC cases included samples diagnosed as adenocarcinoma or adenosquamous cell carcinoma. Region-specific and worldwide ORs, adjusted for country, were calculated using unconditional logistic regression to assess associations between difference HPV16 (sub)lineages and histological type of cancer (ADC versus SCC), predominantly using the A1 sub-lineage as the reference group. As a sensitivity analysis, region-specific ORs were also calculated using conditional logistic regression, in which ADC and SCC were individually matched by country (in both analyses, cancer cases from countries with only ADC or only SCC are dropped).

3. Results

3.1. Geographical distribution of HPV16 sub-lineages

Valid whole HPV16 genome sequence data was available for 7116 HPV16-positive women, derived from 52 countries (Table 1). Subjects were categorized into eight geographic regions, namely North Africa (n = 231 from 2 countries), sub-Saharan Africa (n = 547, 10 countries), East Asia (n = 594, 8 countries), South Asia (n = 543, 6 countries), Europe (n = 225, 10 countries), North America (n = 4 238, 2 countries), South/Central America (n = 626, 12 countries) and Oceania (n = 112, 2 countries) (Table 1).

HPV16 isolates predominantly belonged to A lineage, which represented 78.7% of all subjects, followed by D (9.2%), C (6.4%) and B (5.8%) lineages. The distribution by HPV16 sub-lineages in the whole population were as follows: A1 (n = 3686, 61.1%), A2 (550, 9.1%), A3 (165, 2.7%), A4 (389, 6.5%), B1 (278, 4.6%), B2 (9, 0.2%), B3 (4, 0.1%), B4 (41, 0.7%), C1 (332, 5.5%), C2 (3, 0.05%), C3 (13, 0.2%), C4 (13, 0.2%), D1 (31, 0.5%), D2 (80, 1.3%),D3 (393, 6.5%), and D4 (48, 0.8%).

As illustrated in Fig. 1, HPV16 sub-lineage dispersal was population-specific. Fig. 1a describes the HPV sub-lineage distribution in each geographic region (pie charts), whereas Fig. 1b shows, for each sub-lineage, their relative geographical representation (after equal weighting of each of the eight regions). Globally, A1 was most widespread, being the predominant sublineage (> 70% of isolates) in all of Europe, South/Central America, North America, South Asia and Oceania, as well as being present in 15–30% of isolates from North Africa, sub-Saharan Africa and East Asia. A2, the most closely related sub-lineage to A1, was found in 10–20% of isolates from Europe, North America and Oceania. A3 and A4 accounted for the majority of isolates from East Asia (50% and 20%, respectively), but were rarely seen elsewhere. B and C sub-lineages were associated almost uniquely with Africa, accounting for 28.5% and 25.2% of sub-Saharan and 20.2% and 27.7% of North African isolates, respectively. B1 and C1 accounted for the large majority of B and C isolates, respectively, with the exception of North Africa, the only region where B4 was common. D lineages were most common in South/Central America, where they accounted for 23.8% of all isolates, but also accounted for 10–20% of isolates from all other regions except Oceania. In most regions, D3 was the predominant D sub-lineage, with the exception of North Africa, where D4 predominated. D2 was relatively specific to the Americas.

3.2. Case-control analysis

HPV16 (sub)lineages were compared between invasive cervical cancer cases (n = 1683, irrespective of histological type) and controls (n = 2076) within strata of geographical region, after exclusion of 2218 precancers (CIN2, CIN3, HSIL) and countries that were not represented with both cases and controls (544 cases and 20 controls, in order to allow adjustment by country) (Table 2).

Patterns of HPV16 (sub)lineages differed strongly between cases and controls in East Asia, South/Central America and North America (Fishers's exact test P < 0.001), but not in sub-Saharan Africa, North Africa, South Asia, Europe or Oceania (P > 0.09)(Table 2). In East Asia, this difference was driven by a significant association of A3 (OR = 2.2, 95% confidence interval (CI) 1.0–4.7) and A4 (OR = 6.6, 95%CI 3.1–14.1) sub-lineages with cancer; in North America, by a significant association of A4 (OR = 3.8, 95%CI 1.7–8.3) and D (6.2, 95%CI 4.1–9.3) (sub)lineages with cancer; and in South/Central America by a borderline significant association of A4 (OR not defined, 95%CI = 1.0 - ∞) and D (OR = 2.2, 95%CI 0.8–4.7) lineages with cancer (Table 2). A4 (OR = 4.5, 95%CI 2.7–7.5) and D (OR = 3.1, 95%CI 2.2–4.3) sub-lineages were significantly associated with cervical cancer in a pooled worldwide analysis (adjusted by country), but the association with A3 did not meet statistical significance (OR = 1.6, 95%CI 0.9–2.8). Using the A1 sublineage as a reference group, there were no significant differences between cases and controls, neither overall nor in any region, for A2 sub-lineage, nor for B or C lineages (data not shown).

3.3. Analyses by histological type of cancer

HPV16 (sub)lineages were compared between ADC (n = 175) and SCC (n = 1668) within strata of geographical region, after exclusion of cancers of unknown histological type (n = 63), as well as countries that were not represented with both ADC and SCC (11 ADC and 310 SCC, in order to allow adjustment by country) (Table 3).

Table 3
Comparison of HPV16 (sub)lineages by cervical cancer histological type.

Region	ADC								SCC								P value[a]	OR (95% CI)[b,c] versus A1	
	A				B	C	D	*Total*	A				B	C	D	*Total*		A2	D
	A1	A2	A3	A4					A1	A2	A3	A4							
North Africa	3	1	0	0	1	3	2	***10***	65	18	2	0	36	64	24	***209***	0.906	1.6 (0.1–17.3)	1.8 (0.3–11.4)
Sub-Saharan Africa	4	0	0	0	5	3	1	***13***	39	9	0	0	96	107	14	***265***	0.385	0 (0–4.7)[d]	0.5- (0.0–7.6)
East Asia	4	0	0	17	0	1	4	***26***	33	2	15	213	2	8	28	***301***	0.632	0 (0–20.1)[d]	2.4 (0.4–14.7)
South Asia	9	0	2	1	0	0	2	***14***	155	2	9	5	1	4	17	***193***	0.285	0 (0–36.7)[d]	2.0 (0.4–11.4)
Europe	6	0	0	0	0	0	1	***7***	113	31	1	2	1	0	11	***159***	0.405	0 (0–2.4)[d]	1.4 (0.1–18.3)
South/Central America	17	3	0	2	0	1	23	***46***	241	10	3	7	1	5	87	***354***	**0.001**	**5.0 (1.2–21.2)**	**3.2 (1.5–6.5)**
North America	19	3	0	2	1	0	34	***59***	122	32	3	8	2	2	18	***187***	**< 0.001**	0.6 (0.2–2.2)	**12.1 (5.7–25.6)**
Total	**62**	**7**	**2**	**22**	**7**	**8**	**67**	***175***	**768**	**104**	**33**	**235**	**139**	**190**	**199**	***1668***		0.8 (0.3–1.9)	**4.5 (2.9–7.0)**

ADC, adenocarcinoma/adenosquamous cell carcinoma; CI, confidence interval; HPV, human papillomavirus; OR, odds ratio; SCC, squamous cell carcinoma.
[a] Fischers exact test.
[b] Adjusted for country.
[c] All regional ORs for A3, A4, B and C versus A1 were non-significant and are not shown.
[d] Crude OR (i.e. not adjusted for country).

Patterns of HPV16 (sub)lineages differed significantly between ADC and SCC in South/Central America (P = 0.001) and North America (P < 0.001) only (Table 3), driven by a significant association of D lineages with ADC in both regions (OR = 3.2, 95%CI 1.5–6.5 and 12.1, 95%CI 5.7–25.6, respectively). The pooled worldwide country-adjusted OR for D versus A1 was also significant (4.5, 95%CI 2.9–7.0).

There was also a significant association of A2 sub-lineages with ADC in South/Central America (OR = 5.0, 95%CI 1.2–21.2), but no significant differences between cases and controls in any region for A3 sublineage, nor for B or C lineages (data not shown).

4. Discussion

By sequencing whole HPV16 genomes in the largest and most geographically and ethnically widespread set of samples reported to date, it was confirmed that HPV16 genetic variation can strongly influence risk and histologic type of cervical cancer. However, given the strong population-specific dispersal of HPV16 sub-lineages, their relative carcinogenicity and contribution to cancer burden needs to be considered by worldwide region.

It has long been clear that the distribution of major HPV16 lineages is population-dependent [9,14,27,35–37], considered to be driven by ancient co-evolution of HPV and humans, combined with more recent migration patterns (most notably from Europe and Africa to the Americas) [37,38]. However, our present global mapping of all 16 known HPV16 sub-lineages is the first of its kind. It confirms some previously established knowledge (e.g. specificity of A4 to East Asia, and of B and C lineages to Africa), as well as providing further refinement. For example, worldwide dispersal of A3 sub-lineages closely resembles that of A4 (and less so that of A1 and A2, with which it has often been combined in previous descriptions [37,38]). Furthermore, whereas D3 is the predominant D sub-lineage in most world regions, D4 predominates in, and is specific to North Africa, whereas D2 is highly specific to the Americas. There was additional specificity of B4, C4 and D4 sub-lineages for North Africa. Lastly, the A1 sub-lineage is well represented in all world regions, a phenomena speculated to be driven by interbreeding of modern humans with Neanderthals whilst migrating throughout Eurasia [37,38], making it a natural reference group for studies of HPV16 variant carcinogenicity around the world. These HPV16 variant dispersal patterns highlight how old-style geographical nomenclature (e.g. “European” for A1–3 sub-lineages, or “Asian American” for D lineages) are misleading, as the match between HPV16 sub-lineage distribution and geography is not as close as these historical names suggest. We recommend adherence to the modern nomenclature (A1–4, B1–4, C1–4 and D1–4) [19].

Using old nomenclature, often combining all “European” (A) and all “non-European” (BCD) lineages (because of small numbers and difficulty for further stratification based on targeted sequencing only), early studies already identified that even with these broad groups there was evidence that HPV16 variants could influence risk of persistence and progression [7,8,39,40]. However, such grouping masks important heterogeneity in HPV16 sub-lineage-specific cancer risk. Indeed, we were able to show that, among the A lineage, the A3 and A4 sub-lineages have increased cancer risk versus A1, and among the BCD lineages, only D (and perhaps only D2/D3) is associated with higher cancer risk. These findings corroborate those reporting higher cervical cancer risk for the A4 lineage in Asian [41], and for D2/D3 lineages in American [17] studies.

Indeed, cancer associations were specific for certain world regions (higher A3 risk in East Asia only, higher A4 risk in East Asia, South/Central America and North America only; higher D risk in South/Central and North America only). Lack of association for A3/A4 in other regions could be explained by their relative rarity and/or small sample sizes, as 95%CI of OR for A3 and A4 were not inconsistent with those regions where the (sub)lineages were more common. For D lineages, however, there was more evidence for regional heterogeneity, as no associations were seen in African or Asian regions, despite D lineages being common enough that an association could be expected given the strength of that seen in the Americas. Unfortunately, we did not have power to estimate region-specific country-adjusted ORs by D sub-lineages, but we do note that D2 accounted for a substantial fraction of D lineages in the two regions of South/Central (34 of 129) and North (43 of 256) America (where elevated cancer risk with D lineages was seen), and that D2 was virtually absent elsewhere. Indeed, D2 showed the highest risks for cervical (pre)cancer in a U.S. study contributing to this analysis for North America [17]. Alternatively, regional heterogeneity may represent an interaction between race/ethnicity and HPV16 variants [20]. Indeed, Asian and Hispanic women have showed increased CIN3+ risks for A4 and D variants, respectively, compared with other races [17]. Whilst the mechanism for such an interaction is unclear, it is possible that an HPV16 lineage that has co-evolved with a particular human race has an advantage in evading immune surveillance and persisting, thus increasing cancer risk [17,20].

We found no evidence of differential cervical cancer risks among B and C (sub)lineages, including in sub-Saharan Africa, where they were common enough to be compared robustly. Significant associations have previously been reported for CIN3 risk (lower for B and higher for C, versus A1), although not with cervical cancer [17].

Important heterogeneity in HPV16 (sub)lineages existed also according to cervical cancer histology, with a marked association of D lineages with ADC. Interestingly, the ADC association was only seen in the Americas (where D2/D3 sub-lineages predominate), and is consistent with findings from previous studies from this region [8,11–13,17], from where an association has also been reported with adenocarcinoma in-situ [12,17]. However, no differences in sub-lineage patterns by histological type were seen in any other region, which is consistent with a similar negative finding in the only previous study from outside the Americas [42]. Although we also observed a significant association between A2 variants with ADC in South/Central America which, if true, must be explained by the only 0.3% (~27 bp) difference with A1. However, given the rarity of A2, and the absence of an effect at the global level, this finding should be taken with precaution.

We did not have the statistical power to compare ADC separately versus controls. However, increased cancer risk for D lineages was not driven entirely by the association with ADC, as D lineages were also associated with SCC in a sensitivity analysis (data not shown). A4 has been associated specifically with ADC risk in a large study contributing North American samples also to this analysis [17]. However, we saw no evidence of differential association of A4 by histological outcome, including in East Asia, where A4 was the predominant sub-lineage.

In a recent report of E6 sequencing of HPV16-positive cervical cancer cases that partially overlap with the current study [14], significant associations of D lineages with ADC were observed, not only for South/Central America, in agreement with current findings, but also for Europe and Asia, a finding that we did not replicate. This inconsistency is likely to be due to the controlling for country. Indeed, we believe an important strength of the current work is the attention paid to adjustment of region-specific ORs by country, even if this required excluding countries where ADC and SCC were not both represented. In analyses unadjusted for country (data not shown), erroneous significant associations arose simply as an artefact of country imbalance between the groups being compared, combined with residual differences in HPV16 variant dispersal between countries (see Supplementary Fig. 1). It is clear that crude pooling of worldwide data on HPV16 variants across regions is inappropriate for meaningful comparisons, however, crude pooling of countries across vast world regions is also perilous.

Another important strength of this work is the comparison of control HPV16 infections with those in invasive cervical cancer, whereas many previous studies have focussed on CIN2+ or CIN3+, which may not be entirely representative of invasive disease. In fact, in most regions the number of controls is actually the limiting factor in the statistical power

of our analysis, given that these need to be derived from large population-based samples, often in the absence of cervical cancer screening programmes in low and middle income countries. Nevertheless, the number of HPV16-positive controls from outside Europe and North America represents the largest studied to date. Of further note, HPV16-positive controls are likely to over-represent persistent HPV16 infection, of which a proportion may be destined to progress to cervical cancer in the future. Hence, the true size of the differential risks associated with HPV16 variants may be even bigger than those estimated here.

In the current analysis that had a worldwide scope, we focussed on assessing cancer risks down to the level of HPV16 sub-lineage only, and the differential cancer risks identified should thus be contained in the set of polymorphisms that define (sub)lineages (e.g., all D and A sub-lineages differ by ~2.0% ~150 bp, A1 and A4 by ~0.7% ~60 bp). In addition, recent findings based on whole HPV16 genome sequencing of A1 isolates have shown also that rarer, non-lineage defining SNPs, can also offer significant differences in cancer risk [19], and that each HPV16 infection can actually be considered unique [19]. We were not able to investigate down to this level of genetic detail and indeed, for BCD lineages, numbers were insufficient even to allow comprehensive comparisons by sub-lineage.

5. Conclusion

In summary, HPV16 genetic variation can result in considerable risk differences for cancer. Most notably, D lineages (particularly D2 and D3), as well as A3 and A4 sub-lineages, show increased cervical cancer risk in comparison to the widespread A1, and for D2/D3, this risk seems to be particularly elevated for ADC. These epidemiological findings should help direct studies to elucidate HPV carcinogenesis at the biochemical and mechanistic level. However, all HPV16 sub-lineages are found in cervical cancer somewhere in the world, and are each expected, by default, to be more carcinogenic than any other high-risk HPV type. In fact, the greatest driver of the fraction of cervical cancer attributable to different HPV16 variant lineages around the world remains the dispersal of HPV16 variants through ancient human co-evolution and migration, highlighting how epidemiological studies of HPV genomics and carcinogenicity need to take into account global variability.

Acknowledgements

The members of the IARC HPV Variant Study Group include current/previous IARC staff (N. Munoz, R. Herrero, X. Bosch, S. Franceschi, M. Tommasino, T. Gheit) and local study coordinators in the following countries: Algeria (D. Hammouda), Argentina (D. Loria, E. Matos), Benin (E. Alihonou), Bolivia (J. L. Rios-Dalenz), Brazil (J. Eluf-Neto), Canada (P. Ghadirian), China (Qiao YL), Chile (C. Ferreccio, A. Luzoro, J. M. Ojeda, R. Prado), Colombia (N. Aristizabal, L. A. Tafur, M. Molano, H. Posso), Cuba (M. Torroella), Georgia (T. Alibegashvili, D. Kordzaia), Guinea (N. Keita, M. Koulibaly), India (T. Rajkumar, R. Rajkumar), South Korea (D.-H. Lee, H. R. Shin), Mali (S. Bayo), Morocco (N. Chaouki), Nigeria (J. O. Thomas, C. Okolo, I. Adewole), The Netherlands (C. J. L. M. Meijer, D.A.M. Heideman, P. J. F. Snijders), Panama (E. de los Rios), Paraguay (P. A. Rolon), Peru (E. Caceres, C. Santos), the Philippines (C. Ngelangel), Poland (W. Zatonski), South Africa (D. Moodley), Kenya (P. Gichangi, H. de Vuyst), Spain (S. de Sanjosé, X. Castellsagué), Tanzania (J. N. Kitinya), Thailand (S. Chichareon, S. Sukvirach, S. Tunsakul), and Uganda (H. R. Wabinga).

Cervical cancer cases were also contributed by members of the RIS HPV TT study group from the following countries/regions: Bangladesh (A. Nessa, A.J.E Nahar Rahman, F. Ahmed, M. Kamal), Bosnia Herzegovina (E. Iljazovic), Brazil (M. Aurelho de Lima, S. Henrique), Chile (R. Prado, X. Rodriguez, M. Guerrero, V. Leiva, E. Olave, C. Ramis, V. Toro), Colombia (H.J. Posso, L.E. Bravo, T. Collazos, L.S. García, R. Murillo, G.A.H. Suárez, C.E. Pinzón, G.I. Sánchez), Croatia (M. Grce, I. Sabol, S. Dzebro, M. Dominis), Czech Republic (I. Steiner, V. Vonka), Greece (M. Tzardi, T. Agorastos), Honduras (A. Ferrera, R. Bulnes), India (A. Jain, P.S.A. Sarma, S.K. Giri, M. Bhalla, B. Patel), Israel (J. Bornstein, A. Livoff, H.I. Cohen), Kuwait (W.F. Al-Jassar, Ra Al-Safi), Mozambique (C. Menéndez, C. Romagosa, C. Carrillo), Nigeria (A. Banjo, R. Anorlu, F.B. Abdulkareem, A.O. Daramola, C.C. Anunobi), The Philippines (E.J. Domingo, M.J.V. Germar, J.T.P. Luna, C. Zalameda-Castro, A.M. Fernandez, R. Balacuit, C.L. Llave, J.A. Toral), Poland (A.M. Nowakowski), Taiwan (C. Cheng-Yang, T.Y. Chu, C. Wen-Fang, C.M. Ho), Uganda (M. Odida, E. Weiderpass), Venezuela (E.L. Loyo, V.G. Barriola, M.N. de Gómez, A. Figueredo, J. Navarro).

Conflict of interests

The authors have no conflict of interest to disclose.

Financial support

The work at IARC was supported by the Institut National Du Cancer, France (HPV Genomics; grant number 2016-127). The funder had no role in study design, data collection, analysis and interpretation, in the writing of the report or in the decision to submit the work for publication.

Authors' contributions

GC, LM and MSchiffman conceived the study design. GC drafted the manuscript and supervised data analysis and interpretation. LM supervised all aspects of producing whole HPV16 genome sequencing data. MY, MC, JFB, SB and MSteinberg developed the HPV16 whole genome sequencing platform and/or were responsible for generation of the sequence data. ZC undertook phylogenetic classification of whole genome sequencing data into HPV16 sub-lineages. GC was responsible for acquisition of biological samples and epidemiological data for the IARC studies. LA was responsible for acquisition of data for the ICO RIT HPV VTT study. MSchiffman, TR-B and TL were responsible acquisition of biological samples and epidemiological data for KPNC Pap Cohort. NW, JW and RZ were responsible acquisition of biological samples and epidemiological data for SUCCEED study. VT and DG analysed the epidemiological data and performed statistical analysis for the present analysis.

All authors made substantial contributions to interpretation of data, were involved in critical revision of the manuscript, approved the final version for submission, and agree to be accountable for all aspects of the work.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.pvr.2019.02.001.

References

[1] IARC. Biological agents, IARC monogr eval carcinog risks hum 100B, 2012, 1–441. Available from: ⟨http://monographs.iarc.fr/ENG/Monographs/vol100B/index.php⟩. Accessed 31 Jul 2018.
[2] P. Guan, R. Howell-Jones, N. Li, L. Bruni, S. de Sanjosé, S. Franceschi, G.M. Clifford, Human papillomavirus types in 115,789 HPV-positive women: a meta-analysis from cervical infection to cancer, Int. J. Cancer 131 (2012) 2349–2359.
[3] C. de Martel, M. Plummer, J. Vignat, S. Franceschi, Worldwide burden of cancer attributable to HPV by site, country and HPV type, Int. J. Cancer 141 (2017) 664–670, https://doi.org/10.1002/ijc.30716.
[4] M. Plummer, M. Schiffman, P.E. Castle, D. Maucort-Boulch, C.M. Wheeler, A 2-year prospective study of human papillomavirus persistence among women with a cytological diagnosis of atypical squamous cells of undetermined significance or low-grade squamous intraepithelial lesion, J. Infect. Dis. 195 (2007) 1582–1589.
[5] S.K. Kjaer, K. Frederiksen, C. Munk, T. Iftner, Long-term absolute risk of cervical intraepithelial neoplasia grade 3 or worse following human papillomavirus

infection: role of persistence, J. Natl. Cancer Inst. 102 (2010) 1478–1488.
[6] I. Zehbe, G. Voglino, H. Delius, E. Wilander, M. Tommasino, Risk of cervical cancer and geographical variations of human papillomavirus 16 E6 polymorphisms, Lancet 352 (1998) 1441–1442.
[7] A. Hildesheim, M. Schiffman, C. Bromley, S. Wacholder, R. Herrero, A. Rodriguez, M.C. Bratti, M.E. Sherman, U. Scarpidis, Q.Q. Lin, M. Terai, R.L. Bromley, K. Buetow, R.J. Apple, R.D. Burk, Human papillomavirus type 16 variants and risk of cervical cancer, J. Natl. Cancer Inst. 93 (2001) 315–318.
[8] J. Berumen, R.M. Ordonez, E. Lazcano, J. Salmeron, S.C. Galvan, R.A. Estrada, E. Yunes, A. Garcia-Carranca, G. Gonzalez-Lira, A. Madrigal-de la Campa, Asian-American variants of human papillomavirus 16 and risk for cervical cancer: a case-control study, J. Natl. Cancer Inst. 93 (2001) 1325–1330.
[9] I. Cornet, T. Gheit, M.R. Iannacone, J. Vignat, B.S. Sylla, M.A. Del, S. Franceschi, M. Tommasino, G.M. Clifford, HPV16 genetic variation and the development of cervical cancer worldwide, Br. J. Cancer 108 (2013) 240–244.
[10] R.E. Zuna, W.E. Moore, R.P. Shanesmith, S.T. Dunn, S.S. Wang, M. Schiffman, G.L. Blakey, T. Teel, Association of HPV16 E6 variants with diagnostic severity in cervical cytology samples of 354 women in a US population, Int. J. Cancer 125 (2009) 2609–2613.
[11] R.D. Burk, M. Terai, P.E. Gravitt, L.A. Brinton, R.J. Kurman, W.A. Barnes, M.D. Greenberg, O.C. Hadjimichael, L. Fu, L. McGowan, R. Mortel, P.E. Schwartz, A. Hildesheim, Distribution of human papillomavirus types 16 and 18 variants in squamous cell carcinomas and adenocarcinomas of the cervix, Cancer Res. 63 (2003) 7215–7220.
[12] K.D. Quint, M.N. de Koning, L.J. van Doorn, W.G. Quint, E.C. Pirog, HPV genotyping and HPV16 variant analysis in glandular and squamous neoplastic lesions of the uterine cervix, Gynecol. Oncol. 117 (2010) 297–301, https://doi.org/10.1016/j.ygyno.2010.02.003.
[13] S.H. Rabelo-Santos, L.L. Villa, S.F. Derchain, S. Ferreira, L.O. Sarian, L.A. Angelo-Andrade, M.C. do Amaral Westin, L.C. Zeferino, Variants of human papillomavirus types 16 and 18: histological findings in women referred for atypical glandular cells or adenocarcinoma in situ in cervical smear, Int. J. Gynecol. Pathol. 25 (2006) 393–397.
[14] S. Nicolas-Parraga, L. Alemany, S. de Sanjose, F.X. Bosch, I.G. Bravo, Differential HPV16 variant distribution in squamous cell carcinoma, adenocarcinoma and adenosquamous cell carcinoma, Int. J. Cancer 140 (2017) 2092–2100, https://doi.org/10.1002/ijc.30636.
[15] The Cancer Genome Atlas Research Network, Integrated genomic and molecular characterization of cervical cancer, Nature 543 (2017) 378–384, https://doi.org/10.1038/nature21386.
[16] M. Cullen, J.F. Boland, M. Schiffman, X. Zhang, N. Wentzensen, Q. Yang, Z. Chen, K. Yu, J. Mitchell, D. Roberson, S. Bass, L. Burdette, M. Machado, S. Ravichandran, B. Luke, M.J. Machiela, M. Andersen, M. Osentoski, M. Laptewicz, S. Wacholder, A. Feldman, T. Raine-Bennett, T. Lorey, P.E. Castle, M. Yeager, R.D. Burk, L. Mirabello, Deep sequencing of HPV16 genomes: a new high-throughput tool for exploring the carcinogenicity and natural history of HPV16 infection, Papillomavirus Res. 1 (2015) 3–11, https://doi.org/10.1016/j.pvr.2015.05.004.
[17] L. Mirabello, M. Yeager, M. Cullen, J.F. Boland, Z. Chen, N. Wentzensen, X. Zhang, K. Yu, Q. Yang, J. Mitchell, D. Roberson, S. Bass, Y. Xiao, L. Burdett, T. Raine-Bennett, T. Lorey, P.E. Castle, R.D. Burk, M. Schiffman, HPV16 sublineage associations with histology-specific cancer risk using HPV whole-genome sequences in 3200 women, J. Natl. Cancer Inst. 108 (2016) djw100, https://doi.org/10.1093/jnci/djw100.
[18] L. Mirabello, M. Yeager, K. Yu, G.M. Clifford, Y. Xiao, B. Zhu, M. Cullen, J.F. Boland, N. Wentzensen, C.W. Nelson, T. Raine-Bennett, Z. Chen, S. Bass, L. Song, Q. Yang, M. Steinberg, L. Burdett, M. Dean, D. Roberson, J. Mitchell, T. Lorey, S. Franceschi, P.E. Castle, J. Walker, R. Zuna, A.R. Kreimer, D.C. Beachler, A. Hildesheim, P. Gonzalez, C. Porras, R.D. Burk, M. Schiffman, HPV16 E7 genetic conservation is critical to carcinogenesis, Cell 170 (2017), https://doi.org/10.1016/j.cell.2017.08.001.
[19] L. Mirabello, M.A. Clarke, C.W. Nelson, M. Dean, N. Wentzensen, M. Yeager, M. Cullen, J.F. Boland, M. Schiffman, R.D. Burk, The intersection of HPV epidemiology, genomics and mechanistic studies of HPV-mediated carcinogenesis, Viruses (2018) 10, https://doi.org/10.3390/v10020080.
[20] L.F. Xi, L.A. Koutsky, A. Hildesheim, D.A. Galloway, C.M. Wheeler, R.L. Winer, J. Ho, N.B. Kiviat, Risk for high-grade cervical intraepithelial neoplasia associated with variants of human papillomavirus types 16 and 18, Cancer Epidemiol. Biomark. Prev. 16 (2007) 4–10.
[21] L.F. Xi, N.B. Kiviat, A. Hildesheim, D.A. Galloway, C.M. Wheeler, J. Ho, L.A. Koutsky, Human papillomavirus type 16 and 18 variants: race-related distribution and persistence, J. Natl. Cancer Inst. 98 (2006) 1045–1052.
[22] F.X. Bosch, M.M. Manos, N. Munoz, M. Sherman, A.M. Jansen, J. Peto, M.H. Schiffman, V. Moreno, R. Kurman, K.V. Shah, Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International biological study on cervical cancer (IBSCC) Study Group, J. Natl. Cancer Inst. 87 (1995) 796–802.
[23] G.M. Clifford, S. Gallus, R. Herrero, N. Munoz, P.J. Snijders, S. Vaccarella, P.T. Anh, C. Ferreccio, N.T. Hieu, E. Matos, M. Molano, R. Rajkumar, G. Ronco, S.S. de, H.R. Shin, S. Sukvirach, J.O. Thomas, S. Tunsakul, C.J. Meijer, S. Franceschi, Worldwide distribution of human papillomavirus types in cytologically normal women in the International Agency for Research on Cancer HPV prevalence surveys: a pooled analysis, Lancet 366 (2005) 991–998.
[24] N. Munoz, F.X. Bosch, S. de Sanjose, R. Herrero, X. Castellsague, K.V. Shah, P.J. Snijders, C.J. Meijer, Epidemiologic classification of human papillomavirus types associated with cervical cancer, N. Engl. J. Med. 348 (2003) 518–527.
[25] E.J. Crosbie, M.H. Einstein, S. Franceschi, H.C. Kitchener, Human papillomavirus and cervical cancer, Lancet 382 (2013) 889–899.
[26] M.V. Jacobs, J.M. Walboomers, P.J. Snijders, F.J. Voorhorst, R.H. Verheijen, N. Fransen-Daalmeijer, C.J. Meijer, Distribution of 37 mucosotropic HPV types in women with cytologically normal cervical smears: the age-related patterns for high-risk and low-risk types, Int. J. Cancer 87 (2000) 221–227.
[27] I. Cornet, T. Gheit, S. Franceschi, J. Vignat, R.D. Burk, B.S. Sylla, M. Tommasino, G.M. Clifford, Human papillomavirus type 16 genetic variants: phylogeny and classification based on E6 and LCR, J. Virol. 86 (2012) 6855–6861.
[28] S. de Sanjosé, W.G. Quint, L. Alemany, D.T. Geraets, J.E. Klaustermeier, B. Lloveras, S. Tous, A. Felix, L.E. Bravo, H.R. Shin, C.S. Vallejos, P.A. de Ruiz, M.A. Lima, N. Guimera, O. Clavero, M. Alejo, A. Llombart-Bosch, C. Cheng-Yang, S.A. Tatti, E. Kasamatsu, E. Iljazovic, M. Odida, R. Prado, M. Seoud, M. Grce, A. Usubutun, A. Jain, G.A. Suarez, L.E. Lombardi, A. Banjo, C. Menendez, E.J. Domingo, J. Velasco, A. Nessa, S.C. Chichareon, Y.L. Qiao, E. Lerma, S.M. Garland, T. Sasagawa, A. Ferrera, D. Hammouda, L. Mariani, A. Pelayo, I. Steiner, E. Oliva, C.J. Meijer, W.F. Al-Jassar, E. Cruz, T.C. Wright, A. Puras, et al., Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study, Lancet Oncol. 11 (2010) 1048–1056.
[29] S. Nicolas-Parraga, C. Gandini, V.N. Pimenoff, L. Alemany, S. de Sanjose, F. Xavier Bosch, I.G. Bravo, HPV16 variants distribution in invasive cancers of the cervix, vulva, vagina, penis, and anus, Cancer Med. 5 (2016) 2909–2919, https://doi.org/10.1002/cam4.870.
[30] N. Wentzensen, M. Schiffman, T. Dunn, R.E. Zuna, M.A. Gold, R.A. Allen, R. Zhang, M.E. Sherman, S. Wacholder, J. Walker, S.S. Wang, Multiple human papillomavirus genotype infections in cervical cancer progression in the study to understand cervical cancer early endpoints and determinants, Int. J. Cancer 125 (2009) 2151–2158.
[31] R. Herrero, A. Hildesheim, C. Bratti, M.E. Sherman, M. Hutchinson, J. Morales, I. Balmaceda, M.D. Greenberg, M. Alfaro, R.D. Burk, S. Wacholder, M. Plummer, M. Schiffman, Population-based study of human papillomavirus infection and cervical neoplasia in rural Costa Rica, J. Natl. Cancer Inst. 92 (2000) 464–474.
[32] M. Schiffman, A.C. Rodriguez, Z. Chen, S. Wacholder, R. Herrero, A. Hildesheim, R. Desalle, B. Befano, K. Yu, M. Safaeian, M.E. Sherman, J. Morales, D. Guillen, M. Alfaro, M. Hutchinson, D. Solomon, P.E. Castle, R.D. Burk, A population-based prospective study of carcinogenic human papillomavirus variant lineages, viral persistence, and cervical neoplasia, Cancer Res. 70 (2010) 3159–3169, https://doi.org/10.1158/0008-5472.CAN-09-4179.
[33] P. Cingolani, A. Platts, L. Wang le, M. Coon, T. Nguyen, L. Wang, S.J. Land, X. Lu, D.M. Ruden, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of drosophila melanogaster strainw1118; iso-2; iso-3, Fly (Austin) 6 (2012) 80–92, https://doi.org/10.4161/fly.19695.
[34] J. Koster, S. Rahmann, Snakemake–a scalable bioinformatics workflow engine, Bioinformatics 28 (2012) 2520–2522, https://doi.org/10.1093/bioinformatics/bts480.
[35] L. Ho, S.Y. Chan, V. Chow, T. Chong, S.K. Tay, L.L. Villa, H.U. Bernard, Sequence variants of human papillomavirus type 16 in clinical samples permit verification and extension of epidemiological studies and construction of a phylogenetic tree, J. Clin. Microbiol. 29 (1991) 1765–1772.
[36] T. Yamada, M.M. Manos, J. Peto, C.E. Greer, N. Munoz, F.X. Bosch, C.M. Wheeler, Human papillomavirus type 16 sequence variation in cervical cancers: a worldwide perspective, J. Virol. 71 (1997) 2463–2472.
[37] V.N. Pimenoff, C.M. de Oliveira, I.G. Bravo, Transmission between archaic and modern human ancestors during the evolution of the oncogenic human papillomavirus 16, Mol. Biol. Evol. 34 (2017) 4–19, https://doi.org/10.1093/molbev/msw214.
[38] Z. Chen, R. DeSalle, M. Schiffman, R. Herrero, C.E. Wood, J.C. Ruiz, G.M. Clifford, P.K.S. Chan, R.D. Burk, Niche adaptation and viral transmission of human papillomaviruses from archaic hominins to modern humans, PLoS Pathog. 14 (2018) e1007352, https://doi.org/10.1371/journal.ppat.1007352.
[39] L.L. Villa, L. Sichero, P. Rahal, O. Caballero, A. Ferenczy, T. Rohan, E.L. Franco, Molecular variants of human papillomavirus types 16 and 18 preferentially associated with cervical neoplasia, J. Gen. Virol. 81 (2000) 2959–2968.
[40] L.F. Xi, J.J. Carter, D.A. Galloway, J. Kuypers, J.P. Hughes, S.K. Lee, D.E. Adam, N.B. Kiviat, L.A. Koutsky, Acquisition and natural history of human papillomavirus type 16 variant infection among a cohort of female university students, Cancer Epidemiol. Biomark. Prev. 11 (2002) 343–351.
[41] D. Hang, Y. Yin, J. Han, J. Jiang, H. Ma, S. Xie, X. Feng, K. Zhang, Z. Hu, H. Shen, G.M. Clifford, M. Dai, N. Li, Analysis of human papillomavirus 16 variants and risk for cervical cancer in Chinese population, Virology 488 (2016) 156–161, https://doi.org/10.1016/j.virol.2015.11.016.
[42] M.A. De Boer, L.A. Peters, M.F. Aziz, B. Siregar, S. Cornain, M.A. Vrede, E.S. Jordanova, G.J. Fleuren, Human papillomavirus type 18 variants: histopathology and E6/E7 polymorphisms in three countries, Int. J. Cancer 114 (2005) 422–425.