BMC Cancer

RESEARCH ARTICLE

Open Access

Association of cervical carcinogenesis risk with HPV16 *E6* and *E7* variants in the Taizhou area, China

Mei-Zhen Dai[†], Yi Qiu[†], Xing-Hong Di, Wei-Wu Shi and Hui-Hui Xu[*]

Abstract

Background: Human papillomavirus (HPV) type 16 accounts for a larger share of cervical cancer and has been a major health problem worldwide for decades. The progression of initial infection to cervical cancer has been linked to viral sequence properties; however, the role of HPV16 variants in the risk of cervical carcinogenesis, especially with longitudinal follow-up, is not fully understood in China.

Methods: We aimed to investigate the genetic variability of HPV16 *E6* and *E7* oncogenes in isolates from cervical exfoliated cells. Between December 2012 and December 2014, a total of 310 single HPV16-positive samples were selected from women living in the Taizhou area, China. Sequences of all *E6* and *E7* oncogenes were analysed by PCR-sequencing assay. Detailed sequence comparison, genetic heterogeneity analyses and maximum-likelihood phylogenetic tree construction were performed with BioEdit Sequence Alignment Editor and MEGA X software. Data for cytology tests and histological diagnoses were obtained from our Taizhou Area Study with longitudinal follow-up for at least 5 years. The relationship between HPV16 variants and cervical carcinogenesis risk was analysed by the chi-square test or Fisher's exact test.

Results: In this study, we obtained 64 distinct variation patterns with the accession GenBank numbers MT681266-MT681329. Phylogenetic analysis revealed that 98.3% of HPV16 variants belong to lineage A, in which the A4 (Asian) sublineage was dominant (64.8%), followed by A2 (12.1%), A1 (11.4%), and A3 (10.0%). The A4 (Asian) sublineage had a higher risk of CIN2+ than the A1–3 (European) sublineages (OR = 2.69, 95% CI = 1.04–6.97, *P* < 0.05). Furthermore, nucleotide variation in HPV16 *E6* T178G is associated with the development of cervical cancer.

Conclusion: These data could provide novel insights into the role of HPV16 variants in cervical carcinogenesis risk in China.

Keywords: HPV16, Genetic variants, *E6* gene, *E7* gene, Carcinogenesis, Cervical cancer

Background

Cervical cancer ranks as the fourth most common cancer in women worldwide, and approximately 90% of cervical cancer deaths occur in less developed countries [1]. Persistent human papillomavirus (HPV) infections are the major risk factor for cervical cancer, where nearly 100% of cervical cancer tissues contain HPV DNA [2]. Currently, more than 170 HPV types have been well identified, and approximately 40 of them can be easily transmitted in human mucosa or epithelium [3, 4]. Of the known oncogenic (or high-risk [HR]) HPV types, HPV16 is the most frequently (65%) involved in cervical cancer worldwide [5, 6]. Our previous epidemiological studies showed that the HPV16 type, with an infection

* Correspondence: hui739@163.com
[†]Mei-Zhen Dai and Yi Qiu contributed equally to this work.
Medical Research Center, Taizhou Hospital of Zhejiang Province, Wenzhou Medical University, Linhai, Zhejiang 317000, China

Dai *et al. BMC Cancer* (2021) 21:769

Page 2 of 9

rate of 3.1%, was less common in the Chinese general population but was found most frequently in patients with cervical intraepithelial neoplasia 2 (CIN2), CIN3 and cervical cancer (31.6, 47.9 and 60.6%, respectively) [7, 8]. Many lines of evidence have indicated that the genetic variability of the HPV16 genome contributes to HPV-driven cervical carcinogenesis [9, 10].

HPV16 genomes are approximately 8000 bp and composed of eight to nine open reading frames (ORFs) (*E1, E2, E4, E5, E6, E7, L1,* and *L2*) and a noncoding long control region (*LCR*) [9, 11]. E6 and E7 are the major oncoproteins that can rapidly degrade the tumour suppressor protein p53 and downregulate Rb products, resulting in cellular transformation and cervical carcinogenesis [12]. Based on genomic analysis, HPV16 genetic variants can be divided into four main phylogenetic lineages: A (sublineages A1–4), B (sublineages B1–4), C (sublineages C1–4) and D (sublineages D1–4) [13, 14]. The ethnicity of the sublineages is as follows: A1–3 (traditionally classified as European), A4 (Asian), B1–4 (African-1), C1–4 (African-2), D1 (North American), D2 and D3 (Asian-American), and D4 [15, 16]. The variants of HPV differ in carcinogenicity, geographical distribution, and ethnic group [16, 17]. In the present study, we investigated the genetic variants in HPV16 *E6* and *E7* oncoprotein-encoding genes and their involvement in cervical carcinogenesis in the Taizhou area, Southeast China.

Methods

Subject recruitment and HPV genotyping

Cervical exfoliated cell specimens were collected from Chinese patients who underwent cervical cancer screening at our gynaecological clinic from December 2012 to December 2014. HPV genotyping was performed using the GP5+/bioGP6 + –PCR/MPG assay for 27 genotypes using standard procedures (CFDA Certified No. (2017): 3404697) as previously reported [8]. The inclusion criteria were single HPV16-infected patients with no pregnancy and no history of total uterus or cervix resection. Total cellular DNA was extracted using the DNeasy Blood & Tissue Kit (QIAGEN) and stored at – 20 °C.

PCR amplification and sequencing

For analysis of HPV16 variants, single HPV16-positive specimens were selected for this study. The HPV16 *E6* and *E7* genes were amplified by using specific primers for E6 and E7, 16E6E7_F 5′-ACTAAGGGCGTA ACCGAAAT-3′ and 16E6E7_R 5′-TGCAGTAAAC AACGCAT-3′, which were designed using Primer Premier 5 according to the prototype HPV16 reference sequence (GenBank accession number K02718). PCR conditions consisted of 40 repeated cycles of 30 s denaturation at 94 °C, 45 s annealing at 57 °C, 45 s

elongation at 72 °C, and a 7 min final incubation step at 72 °C (Thermo Hybaid, USA). PCR products had an amplicon size of 1061 bp (nucleotide sites [nt] 23–1083, including *E6* gene nt83–559 and *E7* gene nt562–858).

Subsequently, PCR products were purified and sequenced on the ABI 3730xl DNA Genetic Analyzer at BGI (Shanghai, China). To avoid PCR or sequencing errors, all the data were confirmed at least twice by repeating the PCR and sequencing reactions.

Phylogenetic tree analysis

K02718 was used as the standard for alignment and nucleotide site numbering in this study. The *E6* and *E7* variants in the HPV16 sequences were aligned using the ClustalW tool applied by BioEdit Sequence Alignment Editor. Phylogenetic construction was carried out using MEGA X software using the maximum-likelihood statistical method [18]. To construct the phylogenetic branches, *E6* and *E7* in the HPV16 sequences were downloaded from GenBank NCBI and included K02718 (A1), AF536179 (A2), HQ644236 (A3), AF534061 (A4), KU053908 (B1), HQ644298 (B2), KU053915 (B3), KU053914 (B4), KU053917 (C1), HQ644244 (C2), KU053920 (C3), KU053925 (C4), HQ644257 (D1), AY686579 (D2), AF402678 (D3), and KU053931 (D4) [13, 15, 19].

Treatment and follow-up

Patients with HPV16 infection were recommended for colposcopy examination immediately, according to the 2012 guidelines of cervical cancer screening [20]. Directed biopsy of suspected lesions was performed using cervical biopsy forceps. The biopsy tissues were placed in a 10% formalin solution for histological examination. Histological diagnoses were performed by pathologists and classified as normal, CIN1, CIN2, CIN3 or invasive cervical cancer, according to the WHO histological criteria.

Loop electrosurgical excision procedure (LEEP) was performed for treatment of CIN2 or worse (CIN2+). After the LEEP operation, re-examinations of HPV genotyping and cytology co-testing were performed every 3 months of follow-up, and colposcopy examination was performed for patients with abnormal follow-up results. Data for HPV genotyping, cytology tests and histological diagnoses were obtained from our Taizhou Area HPV Study [8] with longitudinal follow-up for at least 5 years.

Statistical analysis

SPSS 16.0 statistical software (SPSS Inc., Chicago, IL) was used for this study. The association of cervical carcinogenesis risk with HPV16 variants was analysed using the chi-square test or Fisher's exact test. Odds ratios

Dai *et al. BMC Cancer* (2021) 21:769

Page 3 of 9

(ORs) and relative 95% confidence intervals (95% CI) were calculated. All statistical tests were two-sided. *P* values < 0.05 were accepted as statistically significant.

Results

Characteristics of the study population

Between December 2012 and December 2014, single HPV16 positivity was detected in 310 women (median age 41.6 years; range 19–77) who were selected for this study. The flow diagram of the present study is shown Fig. 1. A total of 298 (96.1%) sequences of the *E6* and *E7* genes from HPV16 isolates were obtained. Twelve (3.9%) sequences were excluded due to the small number of HPV copies. Among 298 HPV16-infected women, 63 (21.1%) had normal and adequate cervix by colposcopy examination, so no further biopsy diagnosis was performed. 41 (13.8%) women refused further colposcopy

examination. The remaining 194 (65.1%) underwent colposcopy biopsy for diagnosis which were further used for risk association analysis, and 65 were diagnosed with normal cervices after biopsy, 22 with CIN1, 31 with CIN2, 61 with CIN3, and 15 with squamous cell carcinoma (SCC). Characteristics of the study population categorized by HPV16 (sub)lineage are shown in Table 1.

Variations in the E6 and E7 genes

Compared with the prototype HPV16 reference sequence K02718, 97.7% (291/298) of the HPV16 isolates showed nucleotide variation in our study. A summary of nucleotide and amino acid sequence variation throughout the E6 and E7 fragments is shown Fig. 2. We obtained 64 distinct variation patterns denoted as 16CNTZ01-16CNTZ64, which were published with the GenBank accession codes MT681266 to MT681329. In
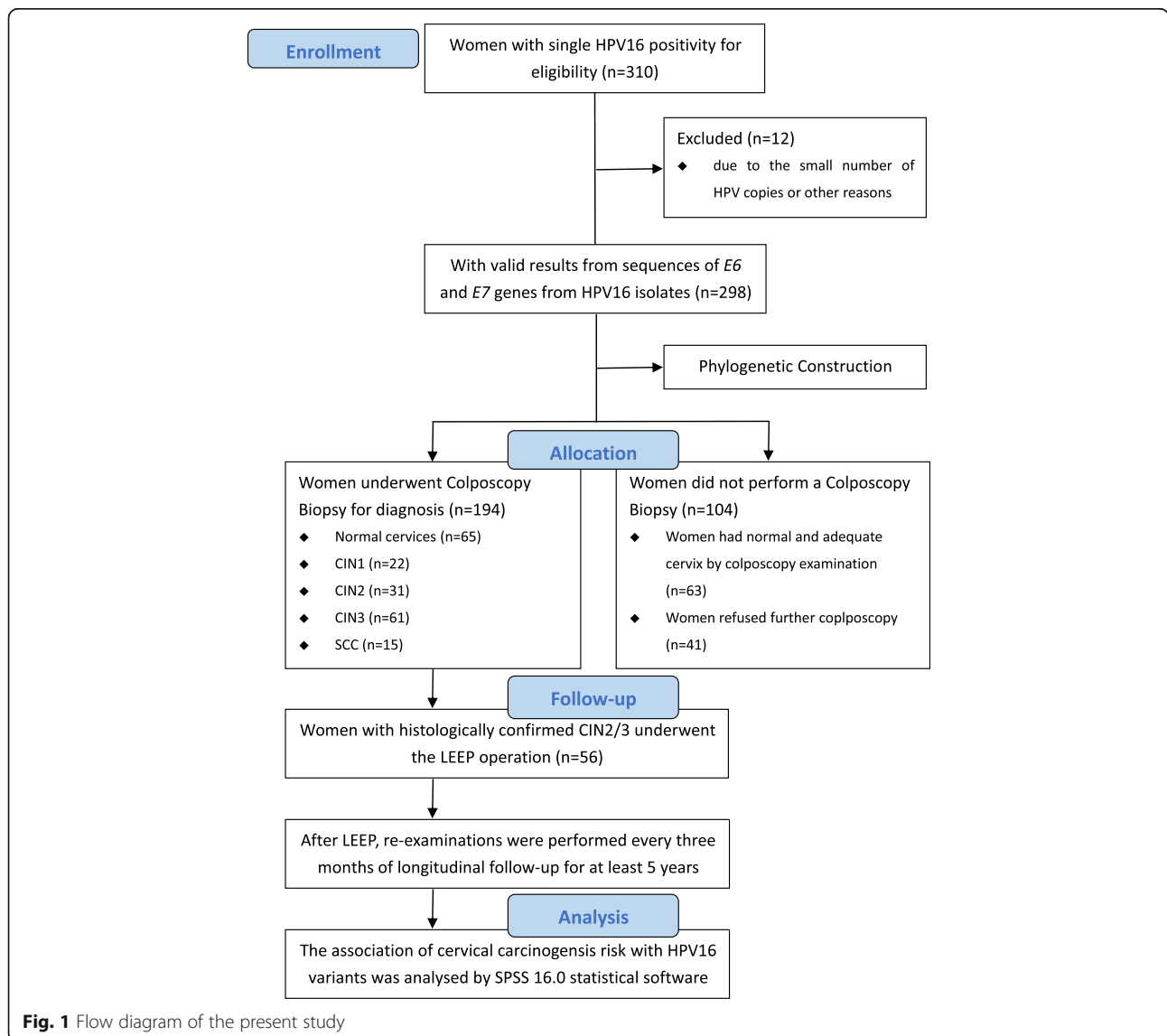

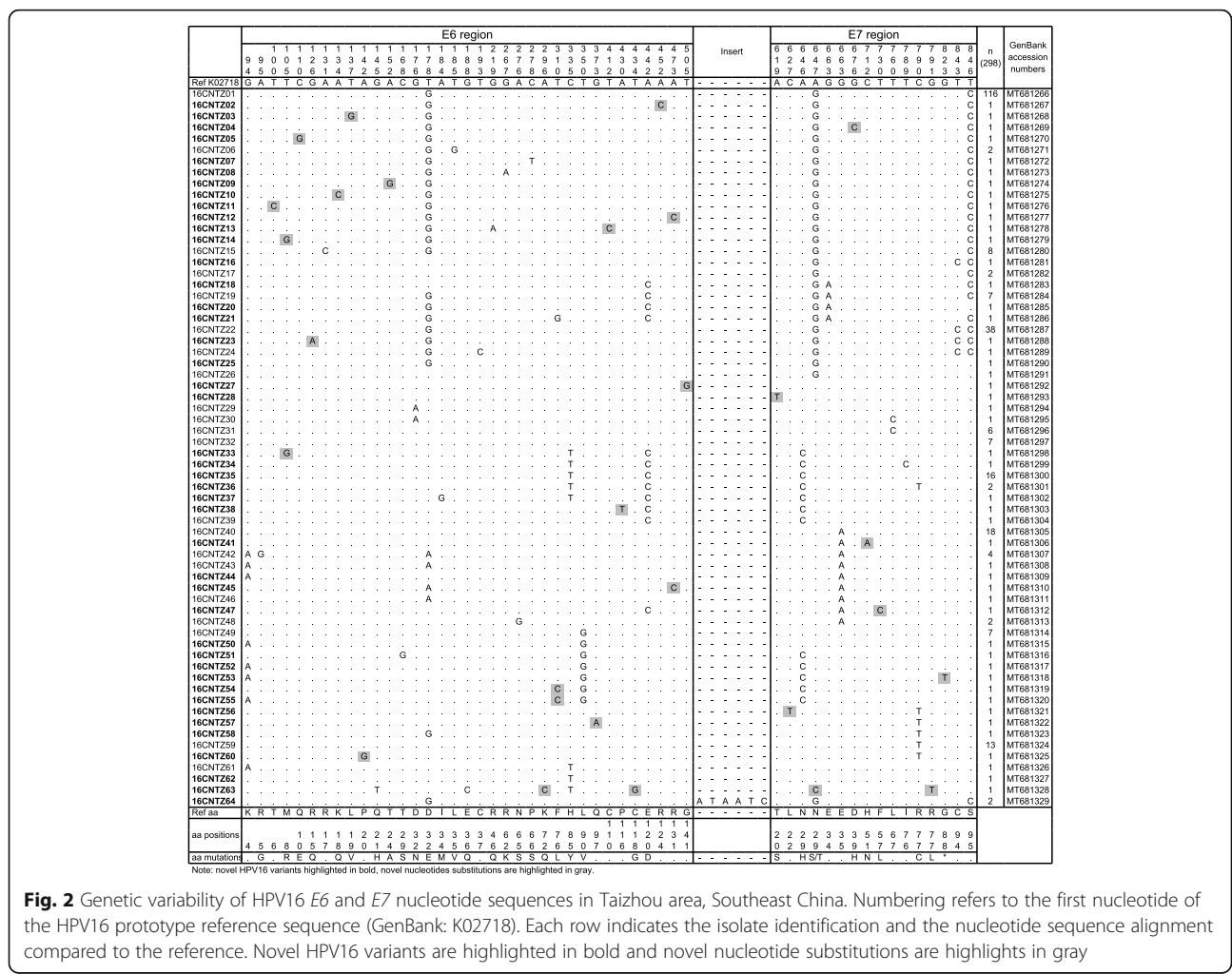**Fig. 1** Flow diagram of the present study

Dai *et al. BMC Cancer* (2021) 21:769

Page 4 of 9

Table 1 Distribution of HPV16 (sub)lineages according to cervical disease status ($n = 194$)

Ages	NILM	CIN1	CIN2	CIN3	SCC	Total
	42.2 ± 9.8	42.7 ± 8.5	39.6 ± 9.9	43.6 ± 6.7	55.3 ± 11.7	
A1	4	2	1	2	1	**10**
A2	8	3	5	5	2	**23**
A3	5	5	3	6	0	**19**
A4[*]	40	11	19	45	10	**125**
A5	7	0	1	2	2	**12**
non-A	1	1	2	1	0	**5**
Total	**65**	**22**	**31**	**61**	**15**	**194**

NILM negative for intraepithelial lesion or malignancy, *CIN* cervical intraepithelial neoplasia, *SCC* squamous cell carcinoma

[*]The A4 (Asian) variants had a higher risk of CIN2+ than the A1–3 (European) variants when compared to CIN1 (OR = 2.69, 95% CI = 1.04 to 6.97, $P < 0.05$)

this study, 43 (67.2%, 43/64) novel HPV16 variants were detected, which are highlighted in bold in Fig. 2. It is worth noting that the insertion of ATAATC between nt561 and nt562 was first detected in the 16CNTZ64 variant, which accounted for 0.7% (2/298) of HPV16 isolates.

In the *E6-E7* sequences, a total of 54 single nucleotide substitutions were identified, with 36 (66.7%) non-synonymous substitutions and 24 (44.4%) novel substitutions. The sequence variability in the *E6* gene was higher than that of the *E7* gene (Fig. 2). The three most prevalent nucleotide substitutions were T178G (D32E) (191/298,

Fig. 2 Genetic variability of HPV16 *E6* and *E7* nucleotide sequences in Taizhou area, Southeast China. Numbering refers to the first nucleotide of the HPV16 prototype reference sequence (GenBank: K02718). Each row indicates the isolate identification and the nucleotide sequence alignment compared to the reference. Novel HPV16 variants are highlighted in bold and novel nucleotide substitutions are highlights in gray

Dai *et al. BMC Cancer* (2021) 21:769

Page 5 of 9

64.1%) in the *E6* gene and A647G (N29S) (195/298, 65.4%) and T846C (192/298, 64.4%) in the *E7* gene, which are specific to the A4 (Asian) sublineage. Another nucleotide substitution was also found in nucleotide site 178, T178A (D32E), with the same amino acid change as T178G but belonging to the A3 (European) sublineage. Notably, C335T (H85Y) was previously found to be specific to B/C/D lineages, and A646C (N29H) belongs to the A2 sublineage. However, both C335T (H85Y) and A646C (N29H) non-synonymous substitutions appeared in 7% (21/298) of HPV16 isolates at the same time in this TZHPV study.

To the best of our knowledge, the base substitutions of T100C, T105G(M8R), C110G(Q10E), G126A(R15Q), A134C(K18Q), T137G(L19V), A142G, A152G(T24A), A296C(K72Q), T310G(F76L), G373A, T412C, A430T, T434G(C118G), A452C, A473C, and T505G in *E6* and A619T(T20S), C627T, A647C(N29T), G676C(D39H), T730C(F57L), G791T(R77L), and G823T(G88*) in *E7* have never been reported in previous studies.

Phylogenetic construction

The maximum-likelihood phylogenetic tree based on the HPV16 *E6-E7* sequences was inferred from 64 obtained HPV16 variants and 16 reference sequences (Fig. 3). According to the phylogenetic tree, 93.8% (60/64) of HPV16 variants belong to lineage A. In our study, the most predominant HPV16 variants belong to sublineage A4 (64.8%, 193/298), followed by sublineages A2 (12.1%, 36/298), A1 (11.4%, 34/298), and A3 (10.0%, 30/298). Notably, the 16CNTZ61, 16CNTZ62, 16CNTZ63 and 16CNTZ64 variants representing 5 samples (1.7%) belonging to non-A variant lineages were also identified in the Taizhou area.

16CNTZ01 is the most common variant (38.9%, 116/ 298) in our population, followed by 16CNTZ22 (12.8%, 38/298) and 16CNTZ40 (6.0%, 18/298). The novel variant 16CNTZ35 was the most common variant (44.4%, 16/36) in the A2 sublineage. A subset of the A2 sublineage appeared to be uniquely defined by the C335T (H85Y) and A442C (E120D) in *E6* and A646C (N29H) in *E7*, and a subset of the A1 sublineage was uniquely defined by the C790T (R77C) in the *E7* gene.
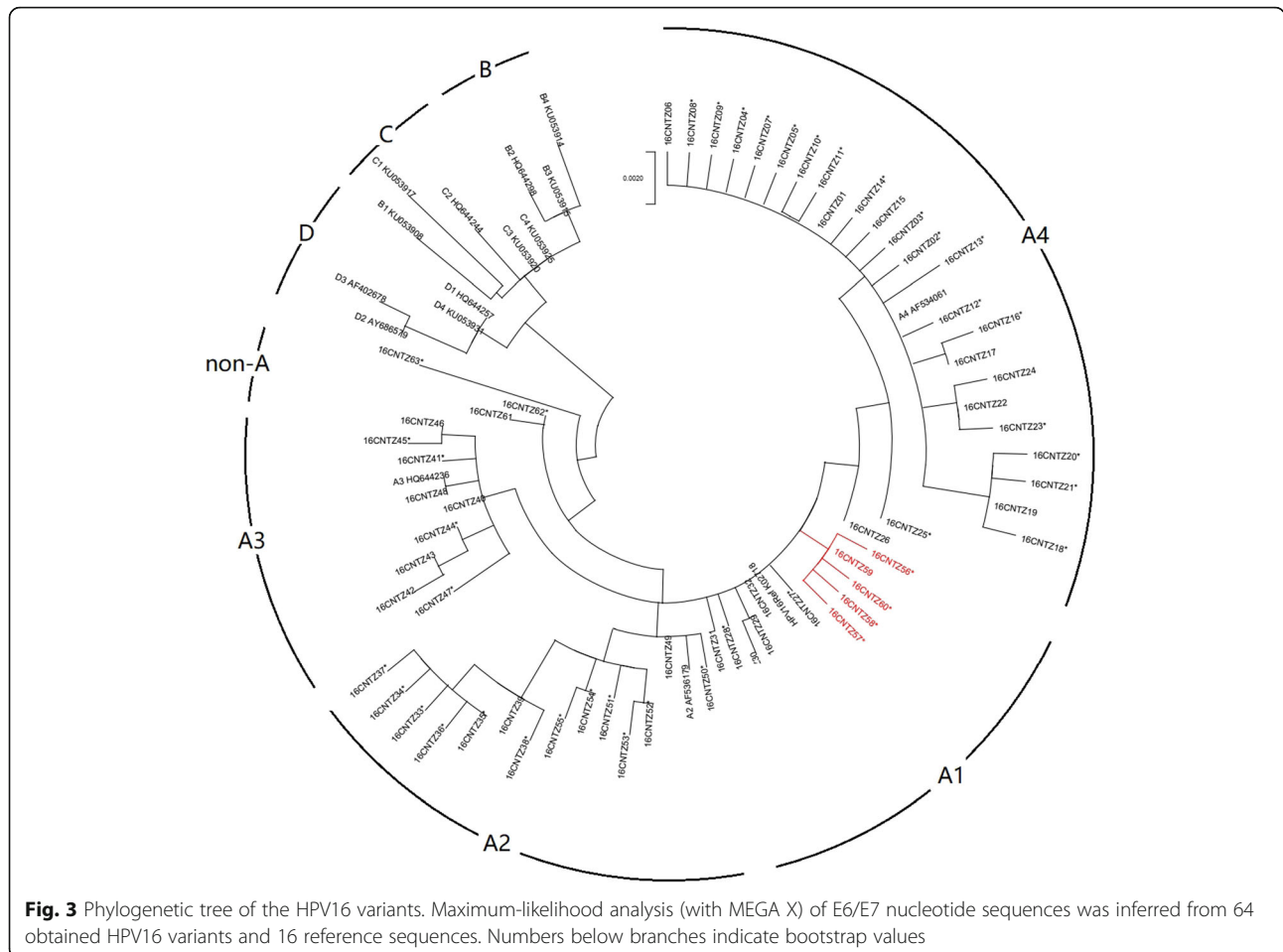

Fig. 3 Phylogenetic tree of the HPV16 variants. Maximum-likelihood analysis (with MEGA X) of E6/E7 nucleotide sequences was inferred from 64 obtained HPV16 variants and 16 reference sequences. Numbers below branches indicate bootstrap values

Dai *et al. BMC Cancer* (2021) 21:769

Page 6 of 9

Risk association with cervical carcinogenesis

Among women with histologically confirmed abnormal cervixes, the A4 (Asian) variants had a higher risk of CIN2+ than the A1–3 (European) variants (OR = 2.69, 95% CI = 1.04 to 6.97, $P < 0.05$) (Table 1). It was clear that A4 variants had stronger oncogenicity than A1–3 variants in the Taizhou population. Moreover, our data showed that the oncogenicity of HPV16 *E6* T178G

(D32E) and *E7* A647G (N29S) variation was associated with an increased risk of CIN2+ (OR = 2.24 ~ 2.45) (Table 2).

At follow-up, 56 women with histologically confirmed CIN2/3 underwent the LEEP operation, among whom, 10 women were diagnosed as the residual/ recurrent disease during the follow-up visit. Seven women with positive margin: 3 women with HPV16 positive and cytology

Table 2 Distribution of major nucleotide variation in HPV16 E6 and E7 genes based on cervical disease status

Gene	Nucleotide variation	Amino Acid variation	Women with cervical disease status					
			NILM *n* = 65	CIN1 *n* = 22	CIN2 *n* = 31	CIN3 *n* = 61	SCC *n* = 15	Total *n* = 194
synonymous substitutions								
E6	A95G	R5G	0	1	1	0	0	**2**
	T105G	M8R	0	0	1	1	0	**2**
	C110G	Q10E	1	0	0	0	0	**1**
	A134C	K18Q	0	0	0	0	1	**1**
	T137G	L19V	0	0	0	1	0	**1**
	G145T	Q21H	0	1	0	0	0	**1**
	A152G	T24A	0	0	0	1	0	**1**
	C168G	T29S	0	0	1	0	0	**1**
	T178G	D32E	38	11	18	46	10	**123**
	T178A	D32E	0	2	1	2	0	**5**
	T185G	L35V	0	0	0	1	0	**1**
	G188C	E36Q	0	1	0	0	0	**1**
	A296C	K72Q	0	1	0	0	0	**1**
	T310C	F76L	0	1	1	0	0	**2**
	C335T	H85Y	4	2	3	4	2	**15**
	T350G	L90V	3	2	3	1	0	**9**
	T434G	C118G	0	1	0	0	0	**1**
	A442C	E120D	7	2	2	8	2	**21**
E7	A646C	N29F	5	3	4	5	2	**19**
	A647G	N29S	40	11	20	46	10	**127**
	A647C	N29T	0	1	0	0	0	**1**
	C712A	H51N	0	0	0	1	0	**1**
	T730C	F57L	0	0	0	1	0	**1**
	C790T	R77C	7	0	2	3	2	**14**
	G791T	R77L	0	1	0	0	0	**1**
	G823T	G88*	0	1	0	0	0	**1**
non-synonymous substitutions								
E7	G663A	–	2	1	0	3	0	**6**
	G666A	–	5	5	3	6	0	**19**
	T843C	–	4	1	5	10	1	**21**
	T846C	–	40	11	19	44	10	**124**
	others	–	9	6	5	8	1	**29**
	Total		**165**	**66**	**89**	**192**	**41**	**553**

NILM negative for intraepithelial lesion or malignancy, *CIN* cervical intraepithelial neoplasia, *SCC* squamous cell carcinoma
The symbol *reflected terminal codon, the end of the protein sequence

Dai *et al. BMC Cancer* (2021) 21:769

Page 7 of 9

normal, 1 women with HPV16 positive and cytology abnormal, 2 women with HPV negative and cytology abnormal, and 1 women with histologically confirmed CIN3 at the 3rd month follow-up visit. Five women had recurrent CIN2/3 (rCIN2/3) within 5 years post-treatment, of which 3 women with clean margins: 1 had 16CNTZ01(A4), 1 had 16CNTZ25(A4), and 1 had 16CNTZ51(A2). At follow-up, 65 women with histologically confirmed normal cervix, among whom, 6 women were progressed to CIN (1CIN1, 3CIN2 and 2CIN3) within 5 years: 3 had 16CNTZ01(A4), 1 had 16CNTZ22(A4), 1 had 16CNTZ59(A5), and 1 had 16CNTZ61(non-A). Additionally, the substitution G823T at E7 leading to a premature stop codon occurred in an isolated CIN1 sample (16CNTZ53/A2), and follow-up for 3 years showed a normal cytology test.

Discussion

China is one of the top contributors to the global burden of cervical cancer, of which HPV16 causes 70% of all cervical cancer worldwide, and some HPV16 variants are more oncogenic than others [13, 21]. It has been clear that the distribution of HPV16 (sub)lineages depends on geographical origin and ethnicity, combined with more recent migration patterns (especially migration from Europe and Africa to the Americas) [22, 23]. Globally, the data showed that variants in the HPV16 sublineages A1–3 were the most widespread in Europe, lineage D in Central-South America, and sublineage A4 in Asia, and lineages B and C variants were mostly restricted to Africa [24]. Several epidemiological studies have suggested that HPV16 non-European variants are associated with the persistence of HPV infection and its progression to cervical cancer, especially sublineages A4, C, D2, and D3 [11, 14, 16, 17, 24, 25].

In this study, we aimed to shed light on the HPV16 variants currently circulating in the Taizhou area, China. We obtained 298 complete sequences of the *E6* and *E7* genes from HPV16 isolates. Most of the HPV16 variants belong to the A lineage (98.3%), of which the A4 (Asian) sublineage was dominant (64.8%). The results of our study are in agreement with previous data in Central (60.3%) and South China (65.5%) [26–29]. In China, HPV16 variants in the A1–3 (European) sublineages are common in Xinjiang, which lies in Northwest China [30, 31]. Our results indicated that the A4 (Asian) variants had an increased risk for CIN2+ compared to the A1–3 (European) variants ($P < 0.05$), which is in accordance with the previous findings reported in a pooled worldwide analysis (adjusted by country) [14]. Thus, the higher contribution of HPV16 to cervical cancer in China may be due to the higher oncogenicity of sublineage A4 (Asian) variants [25]. To our knowledge, the number of HPV16 isolates analysed in this study represents the largest reported collection sampled in Southeast China.

As the HPV16 oncoproteins play an important role in the development of cervical cancer, nucleotide variations in the HPV16 *E6* and *E7* genes could also be associated with the progression of cervical carcinogenesis. Here, we identified 54 single nucleotide substitutions in HPV16 isolates in the Taizhou population, including 37 in the *E6* gene and 17 in the *E7* gene (Fig. 2). In particular, we found that the three most common variations in HPV16 isolates were the T178G (D32E) in E6 (64.1%) and A647G (N29S) and T846C in E7 (65.4 and 64.4%, respectively), and these three nucleotide substitutions are apparently linked because of their simultaneous occurrence in 96.4% of A4 (Asian) variants in this study. It has been reported that the prevalence of both non-synonymous substitutions D32E and N29S is much higher in Asia (65.5% in China, 68.0% in Korea, 44.2% in Japan and 73.9% in Thailand) than in Europe (2%), North America (3%), and Africa (0%) [25, 32–35]. Our results showed that the oncogenicity of E6 D32E and E7 N29S variations was associated with the development of cervical cancer. It has been confirmed that the non-synonymous substitution A647G(N29S) may block the physiological function of Rb, thereby maintaining long-term infection of HPV and increasing the likelihood of persistent viral infection and cervical cancer progression [31, 36, 37]. Therefore, these data may help to explain the higher cervical cancer burden observed in China.

T350G (L90V) has been shown to be associated with the progression of cervical lesions [38–40]. Many studies have reported that 350 T/G polymorphism is a common variation located in the *E6* oncogene, but it has significant heterogeneity by the world region [41]. In Europe/Central Asia and East Asia, cervical cancer risk was significantly associated with the 350 T. However, the opposite was true in South Asia and South/Central America [40, 41]. In our study, we did not find that HPV16 *E6* T350G (L90V) was associated with cervical cancer risk, which is in good agreement with previous data in China [28]. Notably, 350 T/G polymorphism has been shown to influence the cervical cancer risk of European lineages, and also occurs in non-European lineages [40]. Additionally, HPV16 E6 H85Y and E120D belong to the A2 sublineage in this study but are not associated with cervical cancer. Therefore, the carcinogenicity of 350 T vs 350G, 85H vs 85 V, 120E vs 120D might be population-dependent. E6 variations might contribute to oncogenesis by disrupting p53 degradation by affecting the interaction between E6 and p53/E6AP.

Our follow-up data showed that 56 women with HPV16-positive CIN2/3 were treated by the LEEP operation, and 10 women (17.9%) were diagnosed as the residual/ recurrent disease during the follow-up visit. Among 5 women with rCIN2/3, 80% (4/5) of HPV16 isolates belong to the A4(Asian) sublineage. The reasons

Dai *et al. BMC Cancer* (2021) 21:769

Page 8 of 9

for the residual/ recurrent after the LEEP operation might be incomplete excision of baseline CIN2/3, high-risk HPV DNA persistence, or multifocal disease. Our follow-up data also showed that 65 women with histologically confirmed normal cervix, among whom, 6 women (9.2%) were progressed to CIN grade. 66.7%(4/6) of HPV16 isolates belong to the A4(Asian) sublineage. The persistence HPV16 DNA is the root cause of the recurrent disease for the women treated for the CIN2/3, and the same HPV16 variants were found in both lesions [42]. Therefore, our follow-up data showed that 24.7% (48/194) women with persistent infection of HPV16 genotype, which may involves a persistent infection with the same variant.

The largest limitation of our study was the time of follow-up. Since the time between HPV16 infections and CIN2/3 lesions was calculated to be 5–15 years post-infection, our longitudinal follow-up for HPV genotyping, cytology tests and histological diagnoses may affected the outcome in the present study.

Conclusions

In summary, the present study reported genetic variants in the HPV16 *E6* and *E7* genes in the Taizhou area, Southeast China, and provided evidence for their involvement in the increased risk of cervical carcinogenesis. Data about different HPV16 variants in the population of specific regions have significance in uncovering the carcinogenic mechanism of HPV16 and in developing preventive and therapeutic vaccines against HPV.

Abbreviations
CIN: Cervical intraepithelial neoplasia; HPV: Human papillomavirus; LCR: Long control region; LEEP: Loop electrosurgical excision procedure; ORF: Open reading frames; PCR: Polymerase chain reaction

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12885-021-08531-y.

Additional file 1. Clinical data for HPV16 study in Taizhou area, China.

Availability of data and materials
All data generated during this study are included in this published article. The supplementary materials included the nucleotide variations of the *E6* and *E7* genes from HPV16 isolates and the follow-up data of patients. In addition, these sequences have been released to GenBank database with the accession codes of MT681266 to MT681329. The links are https://www.ncbi.nlm.nih.gov/nuccore/MT681266 ~ https://www.ncbi.nlm.nih.gov/nuccore/MT681329.

Declarations

Ethics approval and consent to participate
This study was approved by the Institutional Medical Ethics Review Board of Taizhou Hospital of Zhejiang Province (approval # MERB-2017-020). All patients provided written informed consent before specimen collection.

Consent for publication
Written informed consent was obtained from all patients for the publication of their medical data.

Competing interests
The authors declare that they have no competing interests.

References
1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424. https://doi.org/10.3322/caac.21492.
2. Cohen PA, Jhingran A, Oaknin A, Denny L. Cervical cancer. Lancet. 2019; 393(10167):169–82. https://doi.org/10.1016/S0140-6736(18)32470-X.
3. de Sanjose S, Quint WG, Alemany L, Geraets DT, Klaustermeier JE, Lloveras B, et al. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. Lancet Oncol. 2010;11(11): 1048–56.
4. de Villiers EM. Cross-roads in the classification of papillomaviruses. Virology. 2013;445(1–2):2–10. https://doi.org/10.1016/j.virol.2013.04.023.
5. Alejo M, Alemany L, Clavero O, Quiros B, Vighi S, Seoud M, et al. Contribution of human papillomavirus in neuroendocrine tumors from a series of 10,575 invasive cervical cancer cases. Papillomavirus Res. 2018;5: 134–42. https://doi.org/10.1016/j.pvr.2018.03.005.
6. Xu HH, Wang K, Feng XJ, Dong SS, Lin A, Zheng LZ, et al. Prevalence of human papillomavirus genotypes and relative risk of cervical cancer in China: a systematic review and meta-analysis. Oncotarget. 2018;9(20):15386–97. https://doi.org/10.18632/oncotarget.24169.
7. Xu H, Lin A, Shao X, Shi W, Zhang Y, Yan W. Diagnostic accuracy of high-risk HPV genotyping in women with high-grade cervical lesions: evidence for improving the cervical cancer screening strategy in China. Oncotarget. 2016; 7(50):83775–83. https://doi.org/10.18632/oncotarget.11959.
8. Xu HH, Lin A, Chen YH, Dong SS, Shi WW, Yu JZ, et al. Prevalence characteristics of cervical human papillomavirus (HPV) genotypes in the Taizhou area, China: a cross-sectional study of 37967 women from the general population. BMJ Open. 2017;7(6):e014135. https://doi.org/10.1136/bmjopen-2016-014135.
9. Kukimoto I, Muramatsu M. Genetic variations of human papillomavirus type 16: implications for cervical carcinogenesis. Jpn J Infect Dis. 2015;68(3):169–75. https://doi.org/10.7883/yoken.JJID.2014.584.
10. Rader JS, Tsaih SW, Fullin D, Murray MW, Iden M, Zimmermann MT, et al. Genetic variations in human papillomavirus and cervical cancer outcomes. Int J Cancer. 2019;144(9):2206–14. https://doi.org/10.1002/ijc.32038.
11. van der Weele P, Meijer CJLM, King AJ. Whole-Genome Sequencing and Variant Analysis of Human Papillomavirus 16 Infections. J Virol. 2017; 91(19): pii: e00844–e00817.
12. Hoppe-Seyler K, Bossler F, Braun JA, Herrmann AL, Hoppe-Seyler F. The HPV E6/E7 oncogenes: key factors for viral carcinogenesis and therapeutic targets. Trends Microbiol. 2018;26(2):158–68. https://doi.org/10.1016/j.tim.2017.07.007.

Dai *et al. BMC Cancer*   (2021) 21:769

Page 9 of 9

13. Mirabello I, Clarke MA, Nelson CW, Dean M, Wentzensen N, Yeager M, et al. The Intersection of HPV Epidemiology, Genomics and Mechanistic Studies of HPV-Mediated Carcinogenesis. Viruses. 2018;10(2):80.

14. Clifford GM, Tenet V, Georges D, Alemany L, Pavón MA, Chen Z, et al. Human papillomavirus 16 sub-lineage dispersal and cervical cancer risk worldwide: whole viral genome sequences from 7116 HPV16-positive women. Papillomavirus Res. 2019;7:67–74. https://doi.org/10.1016/j.pvr.2019.02.001.

15. Burk RD, Harari A, Chen Z. Human papillomavirus genome variants. Virology. 2013;445(1–2):232–43. https://doi.org/10.1016/j.virol.2013.07.018.

16. Mirabello L, Yeager M, Cullen M, Boland JF, Chen Z, Wentzensen N, et al. HPV16 sublineage associations with histology-specific cancer risk using HPV whole-genome sequences in 3200 women. J Natl Cancer Inst. 2016;108(9): djw100.

17. Cornet I, Gheit T, Clifford GM, Combes JD, Dalstein V, Franceschi S, et al. Human papillomavirus type 16 E6 variants in France and risk of viral persistence. Infect Agent Cancer. 2013;8(1):4. https://doi.org/10.1186/1750-9378-8-4.

18. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35(6):1547–9. https://doi.org/10.1093/molbev/msy096.

19. Galati L, Equestre M, Bruni R, Accardi L, Torti C, Fiorillo MT, et al. Identification of human papillomavirus type 16 variants circulating in the Calabria region by sequencing and phylogenetic analysis of HPV16 from cervical smears. Infect Genet Evol. 2019;68:185–93. https://doi.org/10.1016/j.meegid.2018.12.024.

20. Saslow D, Solomon D, Lawson HW, Killackey M, Kulasingam SL, Cain JM, et al. American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer. J Low Genit Tract Dis. 2012;16(3):175–204.

21. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. CA Cancer J Clin. 2016;66(2):115–32. https://doi.org/10.3322/caac.21338.

22. Chen Z, DeSalle R, Schiffman M, Herrero R, Wood CE, Ruiz JC, et al. Niche adaptation and viral transmission of human papillomaviruses from archaic hominins to modern humans. PLoS Pathog. 2018;14(11):e1007352. https://doi.org/10.1371/journal.ppat.1007352.

23. Pimenoff VN, de Oliveira CM, Bravo IG. Transmission between archaic and modern human ancestors during the evolution of the oncogenic human papillomavirus 16. Mol Biol Evol. 2017;34(1):4–19. https://doi.org/10.1093/molbev/msw214.

24. Freitas LB, Chen Z, Muqui EF, Boldrini NA, Miranda AE, Spano LC, et al. Human papillomavirus 16 non-European variants are preferentially associated with high-grade cervical lesions. PLoS One. 2014;9(7):e100746. https://doi.org/10.1371/journal.pone.0100746.

25. Hang D, Yin Y, Han J, Jiang J, Ma H, Xie S, et al. Analysis of human papillomavirus 16 variants and risk for cervical cancer in Chinese population. Virology. 2016;488:156–61. https://doi.org/10.1016/j.virol.2015.11.016.

26. Liu Y, Pan Y, Gao W, Ke Y, Lu Z. Whole-genome analysis of human papillomavirus types 16, 18, and 58 isolated from cervical Precancer and Cancer samples in Chinese women. Sci Rep. 2017;7(1):263. https://doi.org/10.1038/s41598-017-00364-9.

27. Zhao J, Zhu J, Guo J, Zhu T, Zhong J, Liu M, et al. Genetic variability and functional implication of HPV16 from cervical intraepithelial neoplasia in Shanghai women. J Med Virol. 2020;92(3):372–81. https://doi.org/10.1002/jmv.25618.

28. Zhou Z, Yang H, Yang L, Yao Y, Dai S, Shi L, et al. Human papillomavirus type 16 E6 and E7 gene variations associated with cervical cancer in a Han Chinese population. Infect Genet Evol. 2019;73:13–20. https://doi.org/10.1016/j.meegid.2019.04.008.

29. Cao M, Chenzhang Y, Ding X, Zhang Y, Jing Y, Chen Z. Genetic variability and lineage phylogeny of human papillomavirus type-16 and -53 based on the E6, E7, and L1 genes in Southwest China. Gene. 2016;592(1):49–59. https://doi.org/10.1016/j.gene.2016.07.039.

30. He H, Li H, Fan P, Zhu J, Pan Z, Pan H, et al. Variants of human papillomaviruses 16 (HPV16) in Uigur women in Xinjiang. China Infect Agent Cancer. 2016;11(1):44. https://doi.org/10.1186/s13027-016-0089-2.

31. Zhe X, Xin H, Pan Z, Jin F, Zheng W, Li H, et al. Genetic variations in E6, E7 and the long control region of human papillomavirus type 16 among

patients with cervical lesions in Xinjiang. China Cancer Cell Int. 2019;19(1): 65. https://doi.org/10.1186/s12935-019-0774-5.

32. Choi BS, Kim SS, Yun H, Jang DH, Lee JS. Distinctive distribution of HPV16 E6 D25E and E7 N29S intratypic Asian variants in Korean commercial sex workers. J Med Virol. 2007;79(4):426–30. https://doi.org/10.1002/jmv.20826.

33. Chopjitt P, Ekalaksananan T, Pientong C, Kongyingyoes B, Kleebkaow P, Charoensri N. Prevalence of human papillomavirus type 16 and its variants in abnormal squamous cervical cells in Northeast Thailand. Int J Infect Dis. 2009;13(3):212–9. https://doi.org/10.1016/j.ijid.2008.06.017.

34. Ding T, Wang X, Ye F, Cheng X, Lu W, Xie X. Distribution of human papillomavirus 16 E6/E7 variants in cervical cancer and intraepithelial neoplasia in Chinese women. Int J Gynecol Cancer. 2010;20(8):1391–8. https://doi.org/10.1111/IGC.0b013e3181eced04.

35. Matsumoto K, Yoshikawa H, Nakagawa S, Tang X, Yasugi T, Kawana K, et al. Enhanced oncogenicity of human papillomavirus type 16 (HPV16) variants in Japanese population. Cancer Lett. 2000;156(2):159–65. https://doi.org/10.1016/S0304-3835(00)00457-2.

36. Chansaenroj J, Theamboonlers A, Junyangdikul P, Swangvaree S, Karalak A, Poovorawan Y. Whole genome analysis of human papillomavirus type 16 multiple infection in cervical cancer patients. Asian Pacific J Cancer Prev. 2012;13(2):599–606. https://doi.org/10.7314/APJCP.2012.13.2.599.

37. Jenison SA, Yu XP, Valentine JM, Galloway DA. Characterization of human antibody-reactive epitopes encoded by human papillomavirus types 16 and 18. J Virol. 1991;65(3):1208–18. https://doi.org/10.1128/jvi.65.3.1208-1218.1991.

38. Andersson S, Alemi M, Rylander E, Strand A, Larsson B, Sallstrom J, et al. Uneven distribution of HPV 16 E6 prototype and variant (L83V) oncoprotein in cervical neoplastic lesions. Br J Cancer. 2000;83(3):307–10. https://doi.org/10.1054/bjoc.2000.1247.

39. Grodzki M, Besson G, Clavel C, Arslan A, Franceschi S, Birembaut P, et al. Increased risk for cervical disease progression of French women infected with the human papillomavirus type 16 E6-350G variant. Cancer Epidemiol Biomark Prev. 2006;15(4):820–2. https://doi.org/10.1158/1055-9965.EPI-05-0864.

40. Pande S, Jain N, Prusty BK, Bhambhani S, Gupta S, Sharma R, et al. Human papillomavirus type 16 variant analysis of E6, E7, and L1 genes and long control region in biopsy samples from cervical cancer patients in North India. J Clin Microbiol. 2008;46(3):1060–6. https://doi.org/10.1128/JCM.02202-07.

41. Cornet I, Gheit T, Iannacone MR, Vignat J, Sylla BS, Del Mistro A, et al. HPV16 genetic variation and the development of cervical cancer worldwide. Br J Cancer. 2013;108(1):240–4. https://doi.org/10.1038/bjc.2012.508.

42. van der Weele P, King AJ, Meijer CJLM, Steenbergen RDM. HPV16 variant analysis in primary and recurrent CIN2/3 lesions demonstrates presence of the same consensus variant. Papillomavirus Res. 2019;7:168–72. https://doi.org/10.1016/j.pvr.2019.04.008.