

RESEARCH

Open Access



Genetic variation of *E6* and *E7* genes of human papillomavirus type 16 from central China

Ting Li¹, Zhiping Yang¹, Chunlin Zhang¹, Sutong Wang¹ and Bing Mei^{1*}

Abstract

Background Persistent high-risk human papillomavirus (HR-HPV) infection is an important factor in the development of cervical cancer, and human papillomavirus type 16 (HPV-16) is the most common HR-HPV type worldwide. The oncogenic potential of HPV-16 is closely related to viral sequence variation.

Methods In order to clarify the variant characteristics of HPV-16 *E6* and *E7* genes in central China, *E6* and *E7* sequences of 205 HPV-16 positive samples were amplified by polymerase chain reaction. PCR products of *E6* and *E7* genes were further sequenced and subjected to variation analysis, phylogenetic analysis, selective pressure analysis and B-cell epitope prediction.

Results Twenty-six single nucleotide variants were observed in *E6* sequence, including 21 non-synonymous and 5 synonymous variants. Twelve single nucleotide variants were identified in *E7* sequence, including 6 non-synonymous and 6 synonymous variants. Four new variants were found. Furthermore, nucleotide variation A647G (N29S) in *E7* was significantly related to the higher risk of HSIL and cervical cancer. Phylogenetic analysis showed that the *E6* and *E7* sequences were all distributed in A lineage. No positively selected site was found in HPV-16 *E6* and *E7* sequences. Non-conservative substitutions in *E6*, H31Y, D32N, D32E, I34M, L35V, E36Q, L45P, N65S and K75T, affected multiple B-cell epitopes. However, the variation of *E7* gene had little impact on the corresponding B-cell epitopes (score < 0.85).

Conclusion HPV-16 *E6* and *E7* sequences variation data may contribute to HR-HPV prevention and vaccine development in Jingzhou, central China.

Keywords Human papillomavirus type 16 (HPV-16), Sequencing, Variation, Epitope prediction, Selection pressure

Background

Cervical cancer is one of the most frequently cancer in women worldwide, with an estimated 604,000 new cases and 342,000 deaths worldwide in 2020 [1]. High-risk human papillomavirus (HR-HPV) infection is the main cause of cervical cancer, which including 15 common

types (16, 18, 52, 58, 68, 66, 56, 31, 33, 35, 45, 82, 39, 51,59) [2]. The pathogenicity of HPV types is associated with a variety of factors including geography and ethnicity. Among these HR-HPV types, HPV-16 is the one with the highest carcinogenic risk in cervical cancer cases [3].

HPV is a double-stranded circular DNA virus with a genome of roughly 8000 base pairs, having three functional regions: (1) long control region (LCR); (2) early gene region (*E*): *E1*, *E2*, *E4–E7*, which encode *E1*, *E2*, *E4–E7* early proteins, respectively; (3) late gene region (*L*): composed of the ORFs of *L1* and *L2*, encoding *L1* and *L2* late proteins [4]. The viral *E6* and *E7* oncogenes play central roles in both HPV-induced oncogenesis

*Correspondence:

Bing Mei
meibing008@163.com

¹ Department of Laboratory Medicine, Jingzhou Hospital Affiliated to Yangtze University, Jingzhou 434020, China



initiation and malignant growth of HPV-positive cancer cells through the encoded E6 and E7 protein, which bind to p53 tumor suppressor protein and retinoblastoma gene product (pRb) respectively and result in degradation of the suppressor protein and the following cell proliferation and tumor formation [5].

Sequence diversity exists in each HPV type, with intra-type variation spectra varying from 1 to 10% and sub-spectra varying from 0.5 to 1% [6]. There are 4 main lineages (A, B, C, D) and 16 sub-lineages of HPV-16, including A1–3 (European), A4 (Asian), B1–4 (African 1), C1–4 (African 2), D1 (North American), D2–3 (Asian-American) and D4 [7]. HPV-16 intra-type variants are preferentially associated with specific histological types of cancer, and non-European HPV-16 variants have a higher oncogenicity due to their association with high-grade lesions of the cervix and invasive tumors [8, 9]. In one study, the association of HPV variant lineages with precancer and cancer is analyzed, showing that the presence of HPV-16 A4 variants is significantly associated with an increased risk of cervical cancer, compared to A1/A2; while women with non-A sub-lineages (B, C, D) have a higher risk of CIN 3 and cervical cancer [10]. Furthermore, molecular and epidemiological data indicate that variants of the same HPV type are biologically distinct and may confer differential pathogenic risks. HPV-16 *E6* variant E-G350 is closely related to HPV-16 persistent infection and cervical lesions transformation from LSIL to HSIL [11]. High mutation rates of HPV-16 *E6* L83V are associated with viral persistence and cervical intraepithelial neoplasia progression, conferring a higher risk for high-grade cervical intraepithelial neoplasia development [11]. The HPV-16 *E7* variant A647G (N29S) has been observed previously in some cervical cancer cases, and shows an increased oncogenicity in vitro, though at present cervical cancer risk association is conflicting across the various studies [12–14].

HPV prevalence and type distribution vary across the countries and the different regions of a country [15]. HPV-16 is the most oncogenic HPV type, with relevance to above 50% of cervical cancer worldwide [16]. The aims of this study were to identify gene variations and corresponding amino acid variations in HPV-16 *E6* and *E7* gene in Jingzhou, whose longitude and latitude are 111° 15′–114° 05′ E and 29° 26′–31° 37′ N respectively, west of Wuhan city in central China, construct phylogenetic trees of *E6* and *E7* sequences, predict the secondary structure and B-cell epitopes of the proteins, and perform selective pressure analysis. The results of this study may provide useful data for the prevention and treatment of HPV in Jingzhou area, central China.

Methods

Sample collection

In this study, cervical exfoliated cell samples were collected from the female patients underwent cervical screening of possible HR-HPV infection at gynecology department of Jingzhou Hospital Affiliated to Yangtze University from September 2019 to October 2021. Informed consents were obtained from all patients and the study was approved by the Ethics Committee of Jingzhou Hospital Affiliated to Yangtze University.

DNA extraction and typing

Based on the method of magnetic beads, DNA was extracted according to the instruction of the nucleic acid extracting kit (Guangzhou Magen Biotechnology Co., Ltd.). The DNA extraction products were measured by real-time quantitative PCR according to the instructions of HR-HPV typing kit (Shanghai ZJ Bio-Tech Co., Ltd.), covering 15 HR-HPV types (16, 18, 52, 58, 68, 66, 56, 31, 33, 35, 45, 82, 39, 51, 59). 4 µL extraction products, as PCR template, were used for HPV typing. Amplification parameters were initially 94 °C for 2 min, 93 °C for 10 s, and 62 °C for 31 s for 40 cycles. In each DNA amplification reaction, the total amount of product after each PCR cycle was measured with a fluorescent chemical at 62 °C. Both positive and negative controls were used for PCR amplification. Finally, the remained extracted DNA products were stored in – 80 °C for subsequent experiments.

PCR amplification and sequencing

In this study, samples with single positive HPV-16 were selected to further analyze the variation of HPV-16 sequence. The HPV-16 *E6* and *E7* sequence-specific primers were designed using Primer Premier 6.0 according to the reference sequence of the HPV-16 prototype (GenBank: K02718). The primers were synthesized by Sangon Biotech (Shanghai). The PCR amplification system was ddH₂O 15.875 µL, 10× Buffer (MgCl₂) (Takara) 2.5 µL, dNTP (Takara) 2 µL, Taq polymerase (5 U/L) (Takara) 0.125 µL, forward primer (20 µM) 1.25 µL, reverse primer (20 µM) 1.25 µL and DNA extraction 2.0 µL. PCR amplification parameters included pre-denaturation 95 °C for 3 min, denaturation at 94 °C for 45 s, annealing at 59.3 °C (*E6*)/58.9 °C (*E7*) for 45 s, extension at 72 °C for 60 s for 35 cycles, and final extension at 72 °C for 10 min. PCR products were subjected to 2.5% agarose gel electrophoresis, stained with gel red nucleic acid dye (Biotium) and sent to Sangon Biotech for sequencing. The following primers were used:

HPV-16 E6-F: 5'-CTAAGGGCGTAACCGAAATCG-3';

HPV-16 E6-R: 5'-TGCTCATAACAGTAGAGATCA GTTG-3';

HPV-16 E7-F: 5'-CCACTGTGTCCTGAAGAA-3';

HPV-16 E7-R: 5'-TCACCTGTACTGTCATT-3'.

Variation analysis

The sequencing results were examined for peak plots using Chromas software. Variation analysis of *E6* and *E7* variants was performed in comparison with the reference sequence of HPV-16 on NCBI (GenBank: K02718) through MEGA11 software. The variant sequences were further transformed into amino acid sequences to analyze the variation of the encoded proteins and predict their secondary structures using GOR4.

Phylogenetic analysis

The Maximum Likelihood tree was constructed using MEGA11 software with repeat parameters set to 1000 times. The reference sequences of HPV-16 lineages downloaded from NCBI were used to construct phylogenetic branches, including K02718 (A1), HQ644284 (A1), AF536179 (A2), HQ644236 (A3), AF534061 (A4), HQ644235 (A4), KU053908 (B1), HQ644238 (B1), HQ644298 (B2), KU015 (B3), KU053914 (B4), KU053917 (C1), HQ644239 (C1), HQ644244 (C2), KU053920 (C3), KU053925 (C4), HQ644257 (D1), AY686579 (D2), HQ644279 (D2), AF402678 (D3), HQ644269 (D3) and KU053931 (D4) [17].

Selective pressure analysis

The codeML program in pamlX software based on Maximum likelihood method (<http://abacus.gene.ucl.ac.uk/software/paml.html>) was used to infer dn/ds and positively selected site. Bayesian empirical Bayesian method was also used to calculate the posterior probability of positively selected site. Seven codon-substitution models (M0, M1, M2, M3, M5, M7, M8) were used to determine whether positive selection impacts the evolution of *E6*–*E7*. These models treat codons as basic units of evolutionary change and take into account genealogical history when calculating parameters. The log-likelihood score evaluates the quality of fit of the input data to the model conditions. These seven codon models use different assumptions to estimate different sets of parameters. In these models, dn/ds, estimated as separate cryptographic subclasses, are assumed to evolve independently of each other [18]. To assess the effect of the positive selection on a specific coding region, a likelihood ratio test (LRT) was performed to compare the nested models [19].

B-cell epitope prediction

B-cell epitopes of HPV-16 *E6* and *E7* reference and variant sequences were predicted using ABCpred Server (<http://crdd.osdd.net/raghava/abcpred/>), an artificial neural network-based B-cell epitope prediction server, based on default parameters. The higher the prediction score, the better the affinity of the epitope.

Results

Epidemiological characteristics of HR-HPV in Jingzhou

Epidemiological and typing data of HR-HPV were investigated in samples collected from January 2018 to December 2021 in Jingzhou, central China. Out of 31,633 samples tested, 6014 were positive for HR-HPV (19.01%). HPV-52 was the most common HR-HPV type (2029, 6.41%), followed by HPV-58 (1048, 3.31%) and HPV-16 (843, 2.66%).

HPV-16 *E6* and *E7* variation

Two hundred and five HPV-16 single positive samples were amplified and sequenced, of which *E6* and *E7* sequences were successfully obtained for 176 samples (median age 43.5 years; range 18–68 years), and the other 29 samples were excluded due to PCR or sequencing failure. In 176 samples, 110 women were histological normal (62.5%), 9 were LSIL (5.1%), 42 were HSIL (23.9%) and 15 were cervical cancer (8.5%). Among the 176 *E6* and *E7* sequences, 168 sequences had nucleotide variants, and the remaining 8 sequences were completely homologous to the reference sequence. In this study, the *E6* and *E7* variant sequences were further divided into 45 different variant groups, noted as 16HB01–16HB45, and the same sequences represented a specific variant group. These sequence data were submitted to GenBank with accession numbers OQ659416–OQ659460. 16HB01 (58/168, 34.5%) and 16HB02 (24/168, 14.3%) were the sequences with the highest variation frequency. Thirty-eight single nucleotide variants were identified in the *E6* and *E7* sequence, including 27 nonsynonymous variants and 11 synonymous variants. The sequence variability of *E6* gene was higher than that of *E7* gene. The most common nucleotide variation in *E6* was T178G (D32E) (102/168, 60.7%), while the most common nucleotide variations in *E7* were A647G (N29S) (101/168, 60.1%) and T846C (102/168, 60.7%). Distribution of major nucleotide variation in HPV16 *E6* and *E7* based on cervical disease status was shown in Table 1. We combined the groups of histological normal and LSIL into one category, meanwhile, combined HSIL and cervical cancer into another category. The statistical results suggested that A647G (N29S) was significantly related to HSIL

Table 1 Distribution of major nucleotide variation in HPV16 E6 and E7 genes based on cervical disease status

Gene	Nucleotide variation	Women with cervical disease status					
		HN	LSIL	HSIL	CC	Total	
		n=110	n=9	n=42	n=15	n=176	
E6	G94A	8	1	2	0	11	
	A95G	0	0	1	0	1	
	T109C	1	0	0	0	1	
	A131C	2	0	0	0	2	
	T137G	1	0	0	0	1	
	C153T	0	0	0	1	1	
	C158A	0	0	1	0	1	
	C173T	1	0	0	0	1	
	G176A	1	1	0	0	2	
	T178G	60	3	30	9	102	
	T178A	3	0	1	0	4	
	A184G	0	0	1	0	1	
	T185G	5	0	1	2	8	
	G188C	1	1	1	0	3	
	T216C	1	0	0	0	1	
	T241G	2	0	1	0	3	
	A276G	2	0	1	2	5	
	T292C	1	0	0	0	1	
	A306C	1	0	0	0	1	
	C335T	2	0	2	0	4	
	T350G	8	0	1	1	10	
	C387T	1	0	0	0	1	
	G395T	0	0	0	1	1	
	A426C	0	0	0	1	1	
	A442C	7	2	4	0	13	
	G489A	0	0	0	1	1	
	G553T	0	0	1	0	1	
	E7	G597A	0	0	1	0	1
		A646C	6	1	0	1	8
		A647G	59	3	30	9	101
G663A		2	0	3	0	5	
G666A		13	1	4	5	23	
G753C		1	1	0	0	2	
T760C		10	2	1	0	13	
G781C		0	1	0	0	1	
C790T		4	0	4	0	8	
G823C		1	0	0	0	1	
T843C		15	1	6	4	26	
T846C		60	3	30	9	102	
Total		279	21	127	46	473	

HN, histological normal; LSIL, low-grade squamous intraepithelial lesion; HSIL, high-grade squamous intraepithelial lesion; CC, cervical cancer

and cervical cancer ($OR=1.99$, 95% $CI=1.03$ to 3.87 , $P=0.04$). T292C of E6 sequence and G597A, G753C and G781C of E7 sequence were novel variants. Five amino acid variants, including two non-conservative

substitutions, were found in the sequence affecting the alpha helix. Ten amino acid variants, all with non-conservative substitutions, were found in the sequence affecting the fold. The results were shown in Table 2.

Phylogenetic analysis of E6 and E7

Phylogenetic analysis was performed by comparing *E6* and *E7* nucleotide sequences, including 45 viral variant sequences and 16 viral reference sequences mentioned above. The phylogenetic tree of *E6* sequence of HPV-16 was shown in Fig. 1. According to the phylogenetic tree of *E6* sequence of HPV-16, all variants were distributed in lineage A, with the most in sub-lineage A4 (119/168, 70.8%), followed by A1 (24/168, 14.3%), A2 (20/168, 11.9%) and A3 (5/168, 3.0%). The phylogenetic tree of *E7* sequence of HPV-16 was shown in Fig. 2. According to the phylogenetic tree of *E7* sequence of HPV-16, all variants were distributed in lineage A, with the most in

sub-lineage A4 (102/168, 60.7%), followed by A3 (43/168, 25.6%) and A1 (23/168, 13.7%).

Selective pressure analysis

No positively selected sites were found in *E6* and *E7* sequences, and the results were shown in Tables 3 and 4.

B-cell epitope prediction

B-cell epitope prediction was based on the corresponding amino acid reference sequences and variant sequences of HPV-16 *E6* and *E7*. Only epitope prediction scores greater than 0.85 were listed, and the predicted results were shown in Table 5. For the E6 protein,

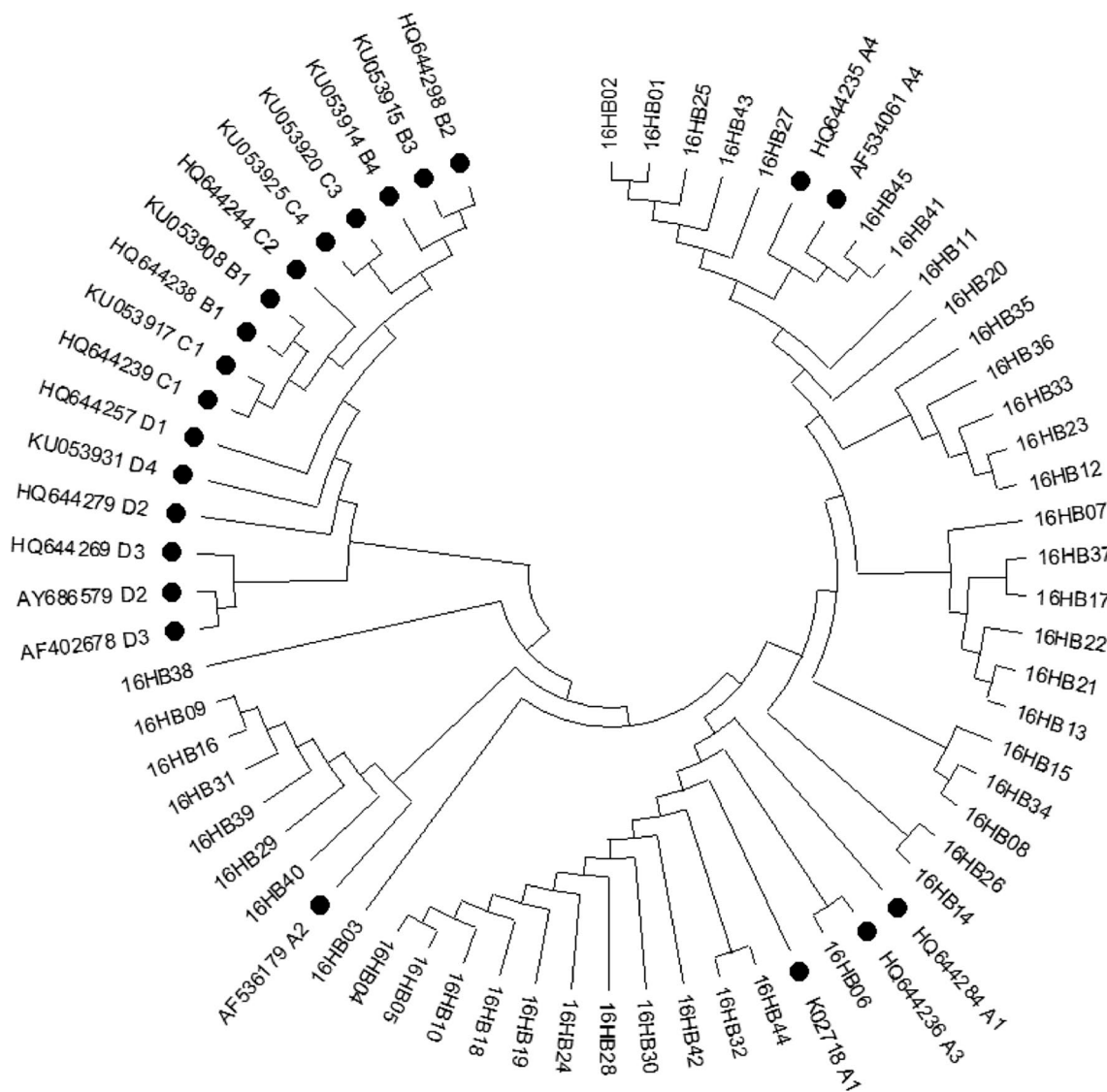


Fig. 1 Maximum Likelihood phylogenetic tree of HPV-16 *E6*. A1–4, B1–4, C1–4 and D1–4 represented the reference sequences of sub-lineages, and the others were variant sequences

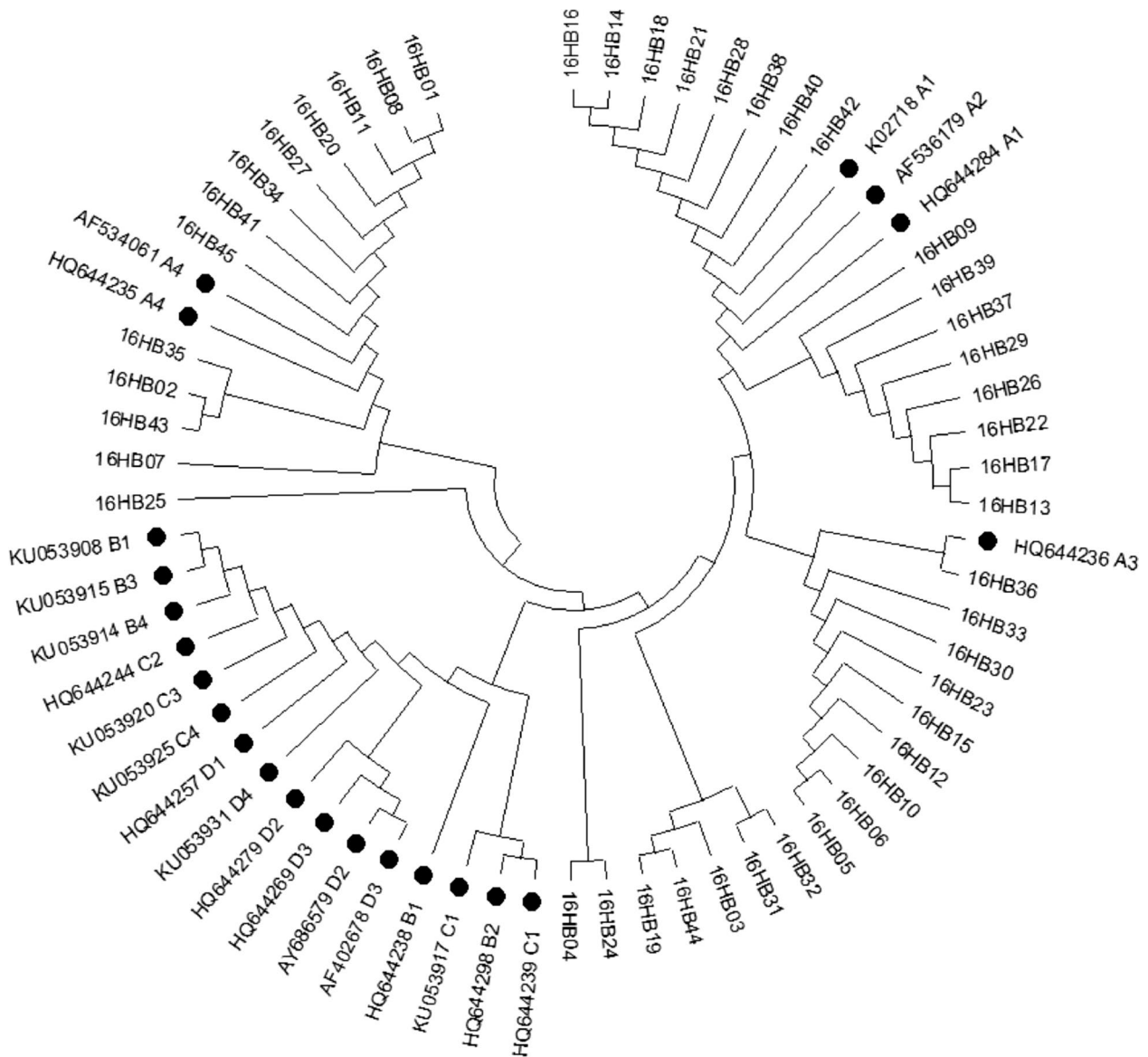


Fig. 2 Maximum Likelihood phylogenetic tree of HPV-16 E7. A1–4, B1–4, C1–4 and D1–4 represented the reference sequences of sub-lineages, and the others were variant sequences

Table 3 Results of PAML estimation for HPV-16 E6–E7 genes

Gene	Model	ln likelihood	Parameter estimates
E6–E7	M0 (one-ratio)	–1456.339336	$\omega = 0.53573$
	M1 (neutral)	–1450.807799	$p_0 = 0.63531$ $p_1 = 0.36469$
	M2 (selection)	–1448.843792	$p_0 = 0.77049$ $p_1 = 0.00000$ $p_2 = 0.22951$ $\omega_2 = 2.34961$
	M3 (discrete)	–1449.747371	$p_0 = 0.64437$ $p_1 = 0.12615$ $p_2 = 0.22948$ $\omega_0 = 0.00000$ $\omega_1 = 0.00000$ $\omega_2 = 2.35011$
	M5 (gamma)	–1449.081486	$\alpha = 0.16628$ $\beta = 0.26303$
	M7 (beta)	–1450.874346	$p = 0.00557$ $q = 0.00842$
	M8 (beta and ω)	–1448.843791	$p_0 = 0.77049$ $p = 0.00500$ $q = 52.87609$ ($p_1 = 0.22951$) $\omega = 2.34960$

Table 4 Results of likelihood ratio tests of positive selection

Gene	Model 1	Model 2	LRT	P value	$d_N/d_S > 1$ detected?	Positively selected codons
<i>E6-E7</i>	M1	M2	3.928014	0.140296113	Yes	None
	M1	M3	2.120856	0.547716257	Yes	None
	M7	M8	4.061110	0.131263306	Yes	None

Table 5 Predicted B-cell epitopes of the HPV-16 *E6* gene

Reference sequence		Start position	Score	Variants sequence		Start position	Score
Rank	Sequence			Rank	Sequence		
1	YSLYGTTLEQQYNKPL	88	0.87	1	IHEIILECVYCKQQLP	30	0.91
2	RDLCIVYRDGNPYAVC	55	0.86	2	YRDGNPYAVCDKCLTF	61	0.88
2	FHNIRGRWTGRMSSC	132	0.86	3	LCIVYRDGSPYAVCDK	57	0.87
3	TAMFQDPQERPRKLPQ	6	0.85	3	YAVCDKCLTFYSKISE	67	0.87
				4	YSVYGTTLEQQYNKPL	88	0.86
				4	QTTIYDIILECVYCKQ	27	0.86
				4	RDLCIVYRDGNPYAVC	55	0.86
				4	FHNIRGRWTGRMSSC	132	0.86
				5	TAMFQDPQERPRKLPQ	6	0.85
				5	QDPQERPRKVPQLCTE	10	0.85

there were 10 B-cell epitopes in the variant sequence, of which 6 B-cell epitopes were novel (30–45 IHEIILECVYCKQQLP, 61–76 YRDGNPYAVCDKCLTF, 57–72 LCIVYRDGSPYAVCDK, 67–82 YAVCDKCLTFYSKISE, 27–42 QTTIYDIILECVYCKQ, and 10–25 QDPQERPRKVPQLCTE) and 4 B-cell epitopes also existed in the reference sequence. According to the score ranking, the top four B-cell epitopes predicted by E6 protein were all novel epitopes, namely 30–45 IHEIILECVYCKQQLP (0.91), 61–76 YRDGNPYAVCDKCLTF (0.88), 57–72 LCIVYRDGSPYAVCDK (0.87), and 67–82 YAVCDKCLTFYSKISE (0.87). For the E7 protein, the reference and variant sequences shared a common epitope 39–54 DGPAGQAEPDRAHYNI (0.85).

Discussion

Cervical cancer is the leading cause of cancer deaths among women in developing countries [20]. Globally, the most common HR-HPV types in cervical cancer are HPV-16 (15.56–83.78%) and HPV-18 (3.4–41.1%) [21]. HPV-16 is the major cause of more than half of cervical cancer worldwide and the main HPV type causing invasive cervical cancer. There are significant geographic and ethnic differences in HR-HPV infection types [22, 23]. In China, HPV-52 and HPV-58 are the most common types, followed by HPV-16, which differed from the HPV type spectrum in other countries [24–26]. In Jingzhou, central

China, the three major HR-HPV types in women are HPV-52, HPV-58, and HPV-16, respectively [27].

The E6 and E7 proteins encoded by the *E6* and *E7* genes in HPV-16 are major oncogenic proteins and play important roles in viral replication, termination of cell differentiation and oncogenesis [28]. Previous studies have shown that HPV variants may affect viral infectivity, pathogenicity and host immune response [29]. Therefore, it is meaningful to analyze *E6* and *E7* gene variation and corresponding amino acid sequences. The sequencing results of *E6* and *E7* sequences indicated that the variation rate of *E6* sequences (5.5%) was greater than that of *E7* sequences (4.0%), implying *E7* sequences were more conserved than *E6* and more suitable as a target for therapeutic vaccine development. In this study, T178G (D32E) (102/168, 60.7%) of *E6* and A647G (N29S) (101/168, 60.1%) and T846C (102/168, 60.7%) of *E7* were the dominant variants, which were also found in other regions of the world. In Korea, the most prevalent variants are *E6* T178G (68%) and *E7* A647G (73%) in HPV-16 [30]. In India, the most frequent variants are T350G (100%) in *E6* and T789C (87.5%) in *E7* [31]. In Xinjiang, China, the most frequent variants in *E6* are T350G (36/75, 48%) and T178G (19/75, 25.3%); the most frequent variant in *E7* is A647G (18/75, 24%) [32]. In Northeast China, the most common variants are T178G (32.69%) in *E6* and A647G (34.62%), G666A (38.46%) and T846C (32.69%) in

E7 [33]. 16HB01, which simultaneously had the variants T178G (D32E), A647G (N29S) and T846C, was the most prevalent (58/168, 34.5%) in the 45 variant groups. In Taizhou, China, T178G (D32E) in *E6* and A647G (N29S) and T846C in *E7* occur in 96.4% of the A4 (Asian) variants [34]. In this study, twelve nonsynonymous variants in the sequences encoding *E6* or *E7* proteins were found, which may affect the folding of oncoproteins in the secondary structure and result in differences in their ability to interact with tumor suppressor proteins and the pathogenicity of HPV-16. T292C, G597A, G753C and G781 were new variants that had never been reported before. Specific mutations of HPV-16 are associated with an increased risk of high-grade squamous intraepithelial lesions and invasive cervical cancer development [35]. In Swedish, prototype HPV-16 and its *E6* variant L83V are both prevalent in preinvasive and invasive cervical lesions [36]. In Shanghai, China, T7220G (D32E) variation in *E6* and A7689G (N29S) in *E7* increase the incidence of HSIL [37]. In Kunming, China, the C749T (S63F) variation in *E7* is associated with cervical cancer [9]. In this study, A647G (N29S) variation in *E7* was significantly related to the higher risk of HSIL and cervical cancer, though the sample size of cervical cancer is relatively limited due to the fact that the patients in this study were mainly outpatients underwent HR-HPV screening.

HPV-16 is divided into genetic sub-lineages A1–4, B1–4, C1–4 and D1–4. The distribution of the lineages varies geographically and ethnically [38]. Globally, variants of HPV-16 sub-lineages A1–A3 are predominantly found in Europe, A4 sub-lineage in Asia, B and C lineages exclusively in Africa, and D lineage most common in South/Central America [39]. Different lineages exhibit disparity in carcinogenic potential. In comparison to other lineages, A4 sub-lineage and D lineage often show more strong association with cervical cancer [39, 40]. The worldwide burden of cervical cancer in different HPV lineages is largely driven by the spread of the historical HPV-16 sub-lineages, and HPV-16 gene variants can significantly affect the risk of cervical cancer development [39, 41]. In this study, we obtained 176 complete *E6* and *E7* gene sequences from HPV-16 isolates. All HPV-16 variants belonged to the A lineage, with the A4 sub-lineage predominating (70.8% in *E6* and 60.7% in *E7*). Our findings were consistent with the data from North China (64.70% of A4) and East China (62.50% of A4) [40, 42].

The main feature of positive selection is that it incurs an abnormal increase in allele frequency, which allows the virus to adapt rapidly to environmental changes [43]. In Greek population, the selection pressure analysis of amino acid residues of HPV-16 shows that codon 83 of *E6* protein and codon 85 of *E7* protein are positive selection sites [44]. However, in this study, positively selected

site of HPV-16 *E6* and *E7* sequences was never found in Jingzhou, central China, which was also in consistence with the results in southwest China [45]. This may be one of the possible reasons why HPV-16 is not the most prevalent in China, though needed to further verify in the future.

The HPV vaccines currently available on the market are preventive vaccines, and the development of therapeutic vaccines for people who have been infected with HPV is of great application prospects. *E6* and *E7* proteins are persistently expressed in HPV infection cells, most cervical cancer and precancerous lesions, but not in normal tissues [46]. Therefore, HPV *E6* and *E7* proteins are ideal targets for diagnostic detection and therapeutic vaccine design. In this study, the B-cell epitopes of HPV-16 *E6* and *E7* proteins were predicted using the ABCpred server, and the effects of amino acid changes caused by genetic variants on B-cell epitopes were analyzed based on changes in epitope fraction. The results predicted 10 B-cell epitopes in *E6* protein, of which six were new epitopes never encoded by the reference sequence. In addition, non-conservative substitutions of some amino acids improved B-cell epitope prediction scores, such as H31Y, D32N, D32E, I34M, L35V, E36Q, L45P, N65S, and K75T. Therefore, non-conservative substitutions of amino acids should be fully considered when developing therapeutic vaccines. The reference sequence and variant sequence of *E7* predicted the B-cell epitope at the same site (39–54 DGPAGQAEPDRAHYNI). The optimal epitope for therapeutic vaccines is often selected from the same region of the reference sequence and the variant sequence [47]. Therefore, *E7* protein may be more suitable as an ideal target for therapeutic vaccine design. These results are valuable for the development of HPV-16 therapeutic vaccines for the population in specific region, though these predicted epitopes still need to be further verified *in vivo*.

Conclusion

This study researched on the genetic variability of HPV-16 *E6* and *E7* genes in Jingzhou, central China. The sequencing results showed *E7* sequences were more conservative than *E6* sequences, meanwhile several novel variations (T292C, G597A, G753C and G781C) were detected. All HPV-16 variants belonged to the A lineage, with the A4 sub-lineage predominating. In addition, non-conservative substitutions, H31Y, D32N, D32E, I34M, L35V, E36Q, L45P, N65S and K75T in *E6*, affected multiple B-cell epitopes. This study provided useful data for the epidemiological characteristics, immunotherapy and vaccine development of HPV-16, in central China.

Abbreviations

HPV-16	Human papillomavirus type 16
HR-HPV	High-risk human papillomavirus
PCR	Polymerase chain reaction
LCR	Long control region
LSIL	Low-grade squamous intraepithelial lesion
HSIL	High-grade squamous intraepithelial lesion
The amino acid acronyms H, Y, D, N, E, I, M, L, V, Q, P, S, K and T	Histidine, tyrosine, aspartic acid, asparagine, glutamic acid, isoleucine, methionine, leucine, valine, glutamine, proline, serine, lysine and threonine

Acknowledgements

Not applicable.

Author contributions

TL and BM conceived and designed the study. TL, ZPY, CLZ and STW collected the samples. TL performed the experiments and drafted the manuscript. BM, ZPY and CLZ revised the paper. All authors read and approved the final manuscript.

Funding

This study was supported financially by Key research and development plan project supporting local special funds in the field of comprehensive health of Hubei Province (Grant Number: 2022BCE029), and Science and technology plan project of Jingzhou City (Grant Number: 2022HC62).

Availability of data and materials

The data generated during the current study are available in the NCBI repository (Home—Nucleotide—NCBI (nih.gov)). The sequence data were submitted to GenBank with accession numbers OQ659416–OQ659460.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of Jingzhou Hospital Affiliated to Yangtze University, and all the work followed the ethics guidelines of the hospital. Informed consent of patients had been obtained and the privacy of the study subjects were protected before sample collection.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 2 June 2023 Accepted: 19 September 2023

Published online: 27 September 2023

References

- Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. *Lancet*. 2007;370(9590):890–907.
- Cohen PA, Jhingran A, Oaknin A, Denny L. Cervical cancer. *Lancet*. 2019;393(10167):169–82.
- Chen Q, Yao L, Wu Q, Xu J, Yan C, Guo C, et al. Rapid and simultaneous visual typing of high-risk HPV-16/18 with use of integrated lateral flow strip platform. *Mikrochim Acta*. 2022;189(9):350.
- Piña-Sánchez P. Human papillomavirus: challenges and opportunities for the control of cervical cancer. *Arch Med Res*. 2022;53(8):753–69.
- 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.
- Muñoz-Bello JO, Carrillo-García A, Lizano M. Epidemiology and molecular biology of HPV variants in cervical cancer: the state of the art in Mexico. *Int J Mol Sci*. 2022;23(15):8566.
- Asensio-Puig L, Alemany L, Pavón MA. A straightforward HPV16 lineage classification based on machine learning. *Front Artif Intell*. 2022;5:851841.
- 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.
- 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.
- Gheit T, Cornet I, Clifford GM, Iftner T, Munk C, Tommasino M, et al. Risks for persistence and progression by human papillomavirus type 16 variant lineages among a population-based sample of Danish women. *Cancer Epidemiol Biomark Prev*. 2011;20(7):1315–21.
- 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.
- 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.
- Zine El Abidine A, Tomaić V, Bel Haj Rhouma R, Massimi P, Guizani I, Bou-baker S, et al. A naturally occurring variant of HPV-16 E7 exerts increased transforming activity through acquisition of an additional phospho-acceptor site. *Virology*. 2017;500:218–25.
- Eschle D, Dürst M, ter Meulen J, Luande J, Eberhardt HC, Pawlita M, et al. Geographical dependence of sequence variation in the E7 gene of human papillomavirus type 16. *J Gen Virol*. 1992;73(Pt 7):1829–32.
- Tang SY, Liao YQ, Hu Y, Shen HY, Wan YP, Wu YM. HPV Prevalence and genotype distribution among women from Hengyang district of Hunan province, China. *Front Public Health*. 2021;9:710209.
- Oyouni AAA. Human papillomavirus in cancer: infection, disease transmission, and progress in vaccines. *J Infect Public Health*. 2023;16(4):626–31.
- Mirabello L, 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.
- DeFilippis VR, Ayala FJ, Villarreal LP. Evidence of diversifying selection in human papillomavirus type 16 E6 but not E7 oncogenes. *J Mol Evol*. 2002;55(4):491–9.
- Yang Z. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. *J Mol Evol*. 2000;51(5):423–32.
- Bedell SL, Goldstein LS, Goldstein AR, Goldstein AT. Cervical cancer screening: past, present, and future. *Sex Med Rev*. 2020;8(1):28–37.
- Xia C, Li S, Long T, Chen Z, Chan PKS, Boon SS. Current updates on cancer-causing types of human papillomaviruses (HPVs) in east, southeast, and south Asia. *Cancers (Basel)*. 2021;13(11):2691.
- Wang Q, Song R, Zhao C, Liu H, Yang Y, Gu S, et al. HPV16 E6 promotes cervical cancer cell migration and invasion by downregulation of NHERF1. *Int J Cancer*. 2019;144(7):1619–32.
- Ai W, Wu C, Jia L, Xiao X, Xu X, Ren M, et al. Deep sequencing of HPV16 E6 region reveals unique mutation pattern of HPV16 and predicts cervical cancer. *Microbiol Spectr*. 2022;10(4):e0140122.
- Bao HL, Jin C, Wang S, Song Y, Xu ZY, Yan XJ, et al. Prevalence of cervicovaginal human papillomavirus infection and genotypes in the pre-vaccine era in China: a nationwide population-based study. *J Infect*. 2021;82(4):75–83.
- Li K, Li Q, Song L, Wang D, Yin R. The distribution and prevalence of human papillomavirus in women in mainland China. *Cancer*. 2019;125(7):1030–7.
- Zhu X, Wang Y, Lv Z, Su J. Prevalence and genotype distribution of high-risk HPV infection among women in Beijing, China. *J Med Virol*. 2021;93(8):5103–9.
- Yang Z, Zhang C, Luo P, Ye M, Gong Q, Mei B. Genetic variability of E6 and E7 genes of human papillomavirus type 58 in Jingzhou, Hubei Province of central China. *Virol J*. 2022;19(1):71.
- Zheng Y, Li X, Jiao Y, Wu C. High-risk human papillomavirus oncogenic E6/E7 mRNAs splicing regulation. *Front Cell Infect Microbiol*. 2022;12:929666.

29. Xi LF, Schiffman M, Koutsky LA, Hughes JP, Hulbert A, Shen Z, et al. Variant-specific persistence of infections with human papillomavirus types 31, 33, 45, 56 and 58 and risk of cervical intraepithelial neoplasia. *Int J Cancer*. 2016;139(5):1098–105.
30. 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.
31. 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.
32. 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;19:65.
33. Shang Q, Wang Y, Fang Y, Wei L, Chen S, Sun Y, et al. Human papillomavirus type 16 variant analysis of E6, E7, and L1 genes and long control region in identification of cervical carcinomas in patients in northeast China. *J Clin Microbiol*. 2011;49(7):2656–63.
34. Dai MZ, Qiu Y, Di XH, Shi WW, Xu HH. Association of cervical carcinogenesis risk with HPV16 E6 and E7 variants in the Taizhou area, China. *BMC Cancer*. 2021;21(1):769.
35. Bletsas G, Zagouri F, Amoutzias GD, Nikolaidis M, Zografos E, Markoulatos P, et al. Genetic variability of the HPV16 early genes and LCR. Present and future perspectives. *Expert Rev Mol Med*. 2021;23:e19.
36. Andersson S, Alemi M, Rylander E, Strand A, Larsson B, Sällström 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.
37. Zhao J, Zhan Q, Guo J, Liu M, Ruan Y, Zhu T, et al. Phylogeny and polymorphism in the E6 and E7 of human papillomavirus: alpha-9 (HPV16, 31, 33, 52, 58), alpha-5 (HPV51), alpha-6 (HPV53, 66), alpha-7 (HPV18, 39, 59, 68) and alpha-10 (HPV6, 44) in women from Shanghai. *Infect Agent Cancer*. 2019;14:38.
38. Mandal P, Bhattacharjee B, Sen S, Bhattacharya A, Saha SS, Chowdhury RR, et al. Predominance of genomically defined A lineage of HPV16 over D lineage in Indian patients from eastern India with squamous cell carcinoma of the cervix in association with distinct oncogenic phenotypes. *Transl Oncol*. 2022;15(1):101256.
39. 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.
40. 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.
41. 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.
42. 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.
43. Zhang Y, Cao M, Wang M, Ding X, Jing Y, Chen Z, et al. Genetic variability in E6, E7, and L1 genes of human papillomavirus genotype 52 from Southwest China. *Gene*. 2016;585(1):110–8.
44. Tsakogiannis D, Papadopoulou A, Kontostathi G, Ruether IGA, Kyriakopoulou Z, Dimitriou TG, et al. Molecular and evolutionary analysis of HPV16 E6 and E7 genes in Greek women. *J Med Microbiol*. 2013;62(Pt 11):1688–96.
45. Yang L, Yang H, Wu K, Shi X, Ma S, Sun Q. Prevalence of HPV and variation of HPV 16/HPV 18 E6/E7 genes in cervical cancer in women in South West China. *J Med Virol*. 2014;86(11):1926–36.
46. Wang R, Pan W, Jin L, Huang W, Li Y, Wu D, et al. Human papillomavirus vaccine against cervical cancer: opportunity and challenge. *Cancer Lett*. 2020;471:88–102.
47. He J, Yang Y, Chen Z, Liu Y, Bao S, Zhao Y, et al. Identification of variants and therapeutic epitopes in HPV-33/HPV-58 E6 and E7 in Southwest China. *Virology*. 2019;16(1):72.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

