Li *et al. Virology Journal* (2025) 22:121
https://doi.org/10.1186/s12985-025-02752-4

Virology Journal

RESEARCH Open Access

Study on the detection rate, genetic polymorphism, viral load, persistent infection capacity, and pathogenicity of human papillomavirus type 33

Qiongyao Li[1,2†], Qichen Cheng[1,2†], Di Tian[1,2], Zhengyuan An[3], Lei Li[2], Feng Yang[2], Mingjing Zhang[4], Ganglin Liu[1], A. Peixin[1], Yan Yang[1*] and Zuyi Chen[1,2*]

Abstract

Background There is a lack of research on the relations among genetic polymorphisms, viral load, adaptability, persistent infection ability, and pathogenicity of human papillomavirus (HPV) type 33. Understanding these relations is crucial for revealing its pathogenic mechanisms and formulating prevention strategies.

Methods Exfoliated cervical cells were harvested from female participants in three hospitals located in the southwestern region of China (Guizhou, Sichuan, and Chongqing). Real-time fluorescence PCR technology was used for HPV genotyping and genomic quantification, and Sanger sequencing was used to obtain the gene sequence. then, changing trends in HPV33 detection rates and *E6/E7* allele frequencies were compared. Positive selection, viral load, pathogenicity, and persistent infection capacity of different *E6/E7* variants/mutations were analyzed.

Results Among 239,743 samples, HPV detection number was 56,681, the HPV33 detection rate was 3.72% (2,110/56,681) among all detected HPV genotypes. Between 2009 and 2023, a downward trend in the HPV33 detection rate was observed. The *E6* + *E7* prototype (*E6* + *E7* on the same variant is consistent with the reference sequence) was the dominant variant, with a significantly increased allele frequency. This dominant variant showed a significantly higher relative risk in causing persistent infection and cervical diseases (cervical intraepithelial neoplasia and cervical cancer). The viral load in the cervical disease group was significantly higher than that in the lesion-free group, and the viral load in the persistent infection group was significantly higher than that in the viral clearance group. There was no correlation between viral load and major genetic variants/mutations.

Conclusions The *E6* + *E7* prototype has a significant impact on the pathogenicity and persistent infection capacity of HPV33. Viral load is positively correlated with pathogenicity and persistent infection capacity. It may serve as

[†]Qiongyao Li and Qichen Cheng contributed equally to this work.

*Correspondence:
Yan Yang
yanyangzmu@163.com
Zuyi Chen
514373742@qq.com

Full list of author information is available at the end of the article

a biomarker for predicting disease progression during HPV33 screening. Other mechanisms underlying allele replacement require further investigation.

Keywords Human papillomavirus, Cervical cancer, Cervical intraepithelial neoplasia, Viral persistence, Viral load, Single nucleotide polymorphisms

Introduction

Cervical cancer is the fourth most common cancer among women, and infection with human papillomavirus (HPV) is a key factor in its development [1]. HPVs are classified into high- and low-risk types based on their carcinogenicity [2]. Low-risk HPVs usually cause benign lesions, primarily including HPV 6 and 11 [3]. High-risk HPVs are closely associated with cervical cancer, and common high-risk HPV types include HPV 16, 18, 31, 33, 45, 52, and 58 [4].

HPV33 is an important high-risk HPV type, with a global detection rate of approximately 3.5%, ranking tenth among high-risk HPV types [5]. However, its proportion of cervical cancer cases is 5%, ranking fourth among all high-risk HPV types [5]. The fact that HPV33's carcinogenic ranking is much higher than its prevalence ranking indicates its strong pathogenic potential and high research value.

HPV33 has six early genes (*E1*, *E2*, *E4*, *E5*, *E6*, and *E7*) and two late genes (*L1* and *L2*) [6]. *E6* and *E7* are the most important oncogenic HPV genes. E6 can inhibit the function of P53, leading to an uncontrolled cell cycle and continuous division of infected cells [7]. E7 inhibits the function of retinoblastoma tumor suppressor proteins, promoting infected cell division [8]. Under the combined action of E6 and E7, infected cells initiate and maintain cell division, accumulate mutations (permanent alterations in the DNA sequence of HPV33 genes), and increase the risk of cancer development [9]. Mutations in *E6* and *E7* may alter the biological activity of HPV, leading to changes in its pathogenicity. Studying genetic polymorphisms (the occurrence of multiple forms of a gene within the HPV33 intratype variant) aids in understanding the evolution and pathogenic mechanisms of microorganisms [10]. In recent years, some studies have examined the genetic polymorphisms of HPV33 [11–14], but only a few have analyzed the pathogenicity of *E6*/*E7*, and these analyses have been carried out at the sublineage level [14].

HPV itself does not encode an enzyme for DNA replication and relies on host cell division enzymes to replicate its genome [15]. The combined action of *E6* and *E7* promotes and maintains infected cell division and drives the replication of the HPV genome. Viral load is a core indicator of the infectivity of a virus and may be related to persistent infection capacity and pathogenicity [16]. Therefore, exploring the relations among HPV33 viral load, *E6*/*E7* genetic polymorphisms, persistent infection capacity, and pathogenicity can deepen our understanding of its adaptation and pathogenic mechanisms.

The relation between HPV33 viral load and pathogenicity remains controversial, and few studies have examined its correlation with persistent infections. Research on the association between HPV33 polymorphisms and pathogenicity is limited, particularly regarding specific variants (different forms of the virus that result from genetic changes) and mutations. Moreover, the correlation between genetic polymorphisms and viral load has rarely been studied, and there is a lack of research on long-term gene polymorphism changes. Thus, our study aimed to explore the changing trends in HPV33 detection rates and *E6*/*E7* genetic polymorphisms and further analyze the relations among genetic polymorphisms, viral load, pathogenicity, and persistent infection ability.

Methods

A schematic overview of the study design is provided in Supplementary Material 1. Below, we detail each component.

Study population

From January 1, 2009, to December 31, 2023, women who underwent gynecological examinations or related consultations at the Affiliated Hospital of Zunyi Medical University (Guizhou Province, China), Chengdu Huada Hospital (Sichuan Province, China), and Chongqing Tongnan Maternal and Child Health Hospital (Chongqing municipality, China) were recruited, and their exfoliated cervical cells were collected for HPV genotyping. Women who visited the hospital for gynecological examinations or consultations were included in the study. The exclusion criteria were as follows: no history of sexual activity, total hysterectomy, use of uterine or vaginal medications/surgery within the past 3 d, and menstrual period.

Sample collection

Gynecologists collected exfoliated cervical cells using cervical swabs (Chaozhou Kaipu Co., Ltd.) and stored the samples in a cell preservation solution (Chaozhou Kaipu Co., Ltd.) for HPV genotyping. If cytological examination was required, samples were collected using cervical swabs (Hologic, Inc) and stored in ThinPrep™ Pap Test PreservCyt (Hologic, Inc). All samples were temporarily stored at 4 °C and tested or examined within 24 h of collection.

HPV genotyping

We used the HPV Nucleic Acid Genotyping Detection Kit [17] (National Medical Device Registration Certificate No. 20143402188, Chaozhou Kaipu Biotech Co., Ltd.; detection limit: 20 copies/reaction; specificity: 100%) and its supporting equipment and reagents to detect HPV nucleic acids and determine genotypes, according to the manufacturer's instructions. This kit can genotype HPV types 16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 6, 11, 42, 43, 44, 81, 73, and 82. Quality control was carried out for each batch of products, and positive samples were selected and stored at -80 °C.

Quantification and sequencing

First stage (January 1, 2009–December 31, 2015): Since the disease symptoms of patients were not collected and there was no follow-up for the samples in this stage, the samples in this stage were not quantified. Only HPV33 samples were randomly selected for sequencing to compare *E6* and *E7* allele frequencies (the relative abundances of different versions of *E6* and *E7* genes within HPV33).

Second stage (January 1, 2017–December 31, 2023): Patients were followed up after their information had been collected. HPV33 single-infection samples were randomly selected for *E6* and *E7* gene sequencing and DNA quantification.

We used the HPV Nucleic Acid Genotyping Detection Kit [18, 19] (fluorescent polymerase chain reaction (PCR) method, National Medical Device Registration Certificate No. 20153400364, Jiangsu BioPerfectus Co., Ltd.; detection limit: 20 copies/reaction; specificity: 99.6%) to quantify HPV33. This kit can genotype and quantify HPV types 6, 11, 16, 18, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 26, 82, 73, and 81. To avoid concentration differences caused by the sampling method, the kit evaluates the viral load based on the concentration of human *TOP3* gene per 10,000 cells [18].

We used the primers 5'-AAAAAAGTAGGGTGTAA CCGA-3' and 5'-TGCCACTGTCATCTGCTGT-3' to amplify the complete *E6*/*E7* genes of HPV33 by PCR in a thermal cycler (Suzhou Tianlong, China). The 50 μL reaction mixture for amplification was as follows: 5 μL of DNA sample (10–100 ng), 25 μL of 2× San Taq PCR Mix ($MgCl_2$, dNTPs, DNA polymerase), 4 μL of each forward and reverse primer (0.25 μmol), and 12 μL of ddH_2O. The PCR reagents were purchased from Sangon Biotech Co., Ltd. (Shanghai, China). The cycling conditions were as follows: pre-denaturation: 95 °C for 10 min; 35 cycles, each cycle consisting of denaturation: 94 °C for 50 s, annealing: 54 °C for 60 s, and extension: 72 °C for 60 s; final extension: 72 °C for 7 min; cooling: 10 °C for 5 min.

The 989-bp amplified products were bidirectionally sequenced using the same primer pair by a Sanger sequencing assay (Sangon Biotech, Shanghai, China).

Comparison of persistent infection capacity and pathogenicity

Based on previous studies, we designed a cross-sectional study with a short-term follow-up of 6–24 months in the second stage (January 1, 2017–December 31, 2023) [20–22]. Patients with a single HPV33 infection were recruited for this cross-sectional study and short-term follow-up. The exclusion criteria for the cross-sectional study were being immunocompromised patients (including those with HIV infection or those taking immunosuppressive drugs) or refusing histopathological biopsy. After all patients underwent gynecological colposcopy or pathological Liquid-Based Cytology Test examination, a histopathological examination was performed for any detected abnormalities to distinguish between the lesion-free and diseased groups. The disease group included only cervical intraepithelial neoplasia (CIN) (grades 1–3) and cervical cancer.

Cervical cancer, CIN3, and CIN2 usually require treatment and are regarded as disease endpoints; thus, these patients were not followed up. Patients with CIN1 and lesion-free women were included in the follow-up period. Inclusion criteria for follow-up: women who tested positive for a single HPV33 infection were advised to undergo re-examinations every six months, including HPV retesting starting from the first examination point. The follow-up period lasted 6–24 months. Exclusion criteria: surgical treatment, loss to follow-up, and coinfection with other high-risk HPV types. During follow-up, persistent HPV33 positivity was defined as persistent infection, whereas negative conversion was defined as non-persistent infection.

Finally, the data analyzed included only samples with successfully obtained cross-sectional study results and complete *E6*/*E7* sequences.

Data analysis

The obtained sequences were compared with the GenBank database sequences and with the HPV33 *E6*/*E7* sequence with GenBank reference ID: M12732 (obtained from PaVE: pave.niaid.nih.gov). MEGA 7.0 (MEGA Limited, New Zealand) was used to determine and record the nucleotide substitution sites in the sequenced HPV33 *E6*/*E7* genes.

The PAML software was used for the likelihood ratio test to infer the non-synonymous/synonymous nucleotide differences, calculate the positive selection sites in the HPV33 *E6* and *E7* gene sequences, and determine the non-synonymous/synonymous mutation ratio to measure the selection pressure of the environment on genes and mutation sites [23].

The first stage (January 1, 2009–December 31, 2015) was named Time 1, and the second stage (January 1, 2017–December 31, 2023) was named Time 2 to enable

the comparison of allele frequencies over a longer period. Time 2 was evenly divided into Time 2.1 (January 1, 2017–June 30, 2020) and Time 2.2 (July 1, 2020–December 31, 2023) to compare allele frequencies in the short term.

EXCEL 2019 was used to draw the trend line for HPV33 detection rates. The chi-square test was used to evaluate the differences between the two groups of data, and the independent samples t-test was used to compare the means of the two groups. Statistical significance was set at $p < 0.05$. SPSS 28 (IBM, USA) was used to calculate the relative risk, significance of differences and 95% confidence intervals (CI).

Results

Detection rate and its trend

Among the 239,743 samples tested in this study, 42,332 were positive. And due to the presence of multiple infections, HPV detection number was 56,681. The overall detection rate of HPV33 was 3.72% (2,110/56,681; 95% CI: 3.57-3.88%), ranking tenth among high-risk HPVs. From 2009 to 2023, the detection rates of HPV33 were 4.34% (18/415; 95% CI: 2.38-6.30%), 5.15% (33/641; 95% CI: 3.44-6.86%), 5.88% (50/850; 95% CI: 4.30-7.46%), 6.51% (35/538; 95% CI: 4.42-8.59%), 6.07% (73/1,203; 95% CI: 4.72-7.42%), 4.07% (60/1,474; 95% CI: 3.06-5.08%), 3.93% (103/2,623; 95% CI: 3.18-4.67%), 4.13% (103/2,623; 95% CI: 3.50-4.75%), 4.37% (258/5,900; 95% CI: 3.85-4.89%), 3.70% (183/4,951; 95% CI: 3.17-4.22%), 3.29% (185/5,625; 95% CI: 2.82-3.75%), 3.24% (199/6,142; 95% CI: 2.80-3.68%), 3.12% (228/7,309; 95% CI: 2.72-3.52%), 3.12% (239/7,656; 95% CI: 2.73-3.51%), and 3.48% (285/8,181; 95% CI: 3.09-3.88%) ($p < 0.001$, $\chi^2 = 67.99$). Overall, the detection rate exhibited a downward trend, as illustrated in Fig. 1.

Gene polymorphism and positive selection

In this study, 458 samples met the requirements for gene polymorphism analysis, including 196 samples at Time 1 and 262 samples at Time 2. The mutation status of the HPV33 *E6*/*E7* gene sequence is shown in Tables 1 and 2. The number of samples for each variant is shown in Supplementary Material 2.

Based on the analyzed sequences, we identified multiple positive selection sites for both *E6* and *E7*, as shown in Table 3.

Comparison of allele frequencies

Since rare mutations lacked statistical significance, we only analyzed variants/mutations with an allele frequency > 2% ($n \geq 5$) in Time 2. The variants/mutations included in the analysis were *E6* prototype ($n = 194$), A213C (K35N) ($n = 57$), A364C (N86H) ($n = 39$), A387C (K93N) ($n = 51$), A446G (Q113R) ($n = 41$), G542A (R145K) ($n = 6$), *E7* prototype ($n = 200$), C706A (A45E) ($n = 10$), C706T (A45V) ($n = 11$), and A862T (Q97L) ($n = 38$).

When comparing allele frequencies between Times 1 and 2, the allele frequencies of the *E6* and *E7* prototypes increased significantly. Given their potential future relevance, we conducted an additional analysis of the *E6* prototype + *E7* prototype (both the *E6* and *E7* genes of the same viral strain are consistent with the reference sequence, subsequently named *E6* + *E7* prototype) and found that its frequency increased significantly. Conversely, the allele frequencies of G542T (R145K), C706T (A45V), and C706A (A45E) significantly decreased.

When comparing allele frequencies between Times 2.1 and 2.2, only the frequency of C706A (A45E) changed significantly (decreasing significantly). The allele frequencies are listed in Table 4.

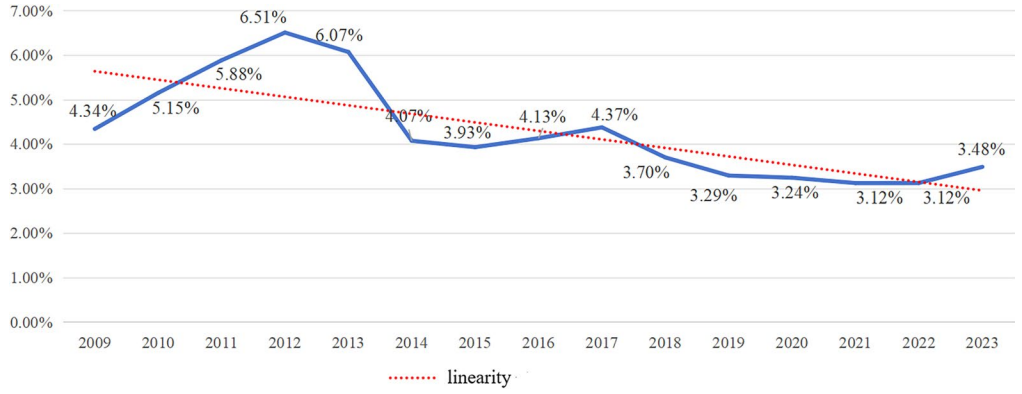

Fig. 1 Change trend of HPV33 detection rates. *Note*: The solid line represents the detection rate. The dotted line represents the linear trend of the detection rate, and this linear trend was created using Microsoft Excel 2019

Table 1 Mutations in HPV33 *E6* sequences

nt position	213	214	253	273	329	364	373	387	446	479	542	549
Reference	A	C	G	A	G	A	G	A	A	G	G	T
33E6M01	C	-	-	G	-	C	-	C	G	-	-	-
33E6M02	C	-	-	G	-	C	-	C	G	-	A	-
33E6M03	-	-	-	-	-	-	-	-	-	-	A	-
33E6M04	C	-	-	-	-	-	-	-	-	-	-	-
33E6M05	C	-	-	-	-	-	-	C	-	-	-	C
33E6M06	-	A	-	-	-	-	-	-	-	-	-	-
33E6M07	C	-	T	G	A	C	A	C	G	-	-	-
33E6M08	C	-	-	G	-	-	-	C	G	-	-	-
33E6M09	-	-	-	-	-	-	A	-	-	-	-	-
33E6M10	-	-	-	-	-	-	-	-	-	A	-	-
33E6M11	-	-	-	-	-	-	-	-	-	-	-	-
Reference	K	P	D	R	S	N	E	K	Q	R	R	T
AA position	35	36	49	55	74	86	89	93	113	124	145	147
mutation	N	T	Y	-	N	H	K	N	R	Q	K	-

Note The nucleotides conserved with respect to the reference sequence were marked with a horizontal line (-), whereas a variation position was indicated by a letter. The "nt" represents nucleotide. The "AA" means amino acid

Pathogenicity, persistent infection capacity, variant/mutation, and viral load

In the cross-sectional study, the mean logarithmic value of viral load in the cervical disease group (CIN + cervical cancer) was 4.86 ± 0.95 (units: $\log_{10}$ copies per 10,000 cells), which was significantly higher than that in the lesion-free group (4.47 ± 0.88) ($p = 0.001$, t = 3.41). In the follow-up study, the mean logarithmic value of viral load in the persistent infection group was 4.91 ± 0.87, which was significantly higher than that in the viral clearance group (4.10 ± 0.91) ($p < 0.001$, t = 5.36).

In the analysis of the relation between variant/mutation and viral load, only G542A (R145K) showed a significantly high viral load. Further details are presented in Table 5. Viral load, cross-sectional results, and follow-up results for each sample are shown in Supplementary Material 3.

Persistent infection capacity, pathogenicity, and variant

Compared with the control groups, the *E6* prototype and *E6* + *E7* prototype led to a significantly higher relative risk of CIN1 and CIN3+ (CIN3 and cervical cancer); A213C (K35N), A364C (N86H), A387C (K93N), A446G (Q113R), and A862T (Q87L) resulted in a significantly lower relative risk of CIN+ (CIN and cervical cancer). Further details are presented in Table 6.

Discussion

E6 and *E7* play crucial roles in the life cycle and pathogenesis of HPV, and polymorphisms in *E6* and *E7* have been reported in various HPV types [24–26]. HPV33 *E6* and *E7* gene polymorphisms have also been reported in several papers [14, 21]. In contrast to studies on the genetic polymorphism of HPV33, this study, from a unique perspective, combined the detection rate of HPV33 with gene frequency changes and used multicenter samples to study the *E6* and *E7* gene polymorphisms, viral load, persistent infection capacity, pathogenicity, and adaptation ability of HPV33. The results deepen our understanding of the evolution and pathogenic mechanisms of HPV and provide a scientific basis for the prevention and treatment of HPV33.

According to reports from multiple regions in China, the prevalence of HPV33 ranks seventh to tenth among common high-risk types (5.32% in Xianning, ranking seventh; 6.3% in Nanjing, ranking seventh; 5.0% in Zhejiang, ranking ninth; and 4.22% in Sichuan, ranking tenth) [27–30]. In our study, the detection rate of HPV33 among women in gynecological outpatient departments and those undergoing physical examinations in the southwest region of China was approximately 3.72%, ranking tenth among high-risk types, similar to other regions in China. However, our trend study on HPV33 suggested that its detection rate slowly decreased, indicating that the

Table 2 Mutations in HPV33 *E7* sequences

nt position	669	706	706	737	758	780	798	845	850	850	862
Reference	G	C	C	A	C	A	C	C	C	C	A
33E7M01	-	-	-	-	-	-	-	-	-	-	T
33E7M02	-	A	-	-	-	-	-	-	-	-	-
33E7M03	-	-	T	-	-	-	-	-	-	-	-
33E7M04	-	-	-	-	-	-	-	-	A	-	T
33E7M05	C	-	-	-	-	-	-	A	-	-	-
33E7M06	-	-	-	G	-	-	-	-	-	-	T
33E7M07	-	-	-	-	-	-	G	-	-	-	-
33E7M08	-	-	-	-	G	G	G	-	-	G	-
33E7M09	-	-	-	-	-	-	-	-	-	-	-
Reference	D	A	A	V	N	N	L	C	T	T	Q
AA position	33	45	45	55	63	70	76	91	93	93	97
mutation	H	E	V	-	K	D	V	-	S	S	L

Note The nucleotides conserved with respect to the reference sequence were marked with a horizontal line (-), whereas a variation position was indicated by a letter. The "nt" represents nucleotide. The "AA" means amino acid

Table 3 HPV33 *E6* and *E7* positive selection

Genes	Models	LnL	Estimates of parameters	2ΔI and *P*	Positively selected sites
E6	M7	-797.193306	$p = 0.00500$, q = 0.02000		NA
	M8	-776.099507	$p_0 = 0.95623$, $p = 43.03699$, q = 99.00000, $p1 = 0.04377$, ω = 36.18486	42.2, $p < 0.001$	35 K**, 86 N**, 89E*, 93 K**, 113Q**
E7	M7	-479.935234	$p = 0.00749$, q = 0.00500		NA
	M8	-473.988826	$p_0 = 0.92441$, $p = 0.03104$, q = 0.02469, $p1 = 0.07559$, ω = 36.49653	11.9, $p < 0.001$	45 A*, 93T*, 97Q**

Note M7 means specifies the model of nucleotide substitution (REV); M8 means specifies the model of nucleotide substitution (UNREST). LnL: the log-likelihood difference between the two models. 2ΔI: twice the log-likelihood difference between the two models ($P < 0.05$ was considered statistically significant). NA means not applicable. NS means no sites reached the significance level of Bayes empirical Bayes 0.90. An asterisk indicates posterior probability ≥ 0.95, and two asterisks indicate posterior probability ≥ 0.99

Table 4 Comparison of HPV33 *E6/E7* allele frequencies

Prototype/mutation	Time 1	Time 2	p_1	Time 2.1	Time 2.2	p_2
***E6*prototype**	**126 (64.29%)**	**194 (74.05%)**	**0.024**	92 (70.77%)	102 (77.27%)	0.230
A213C (K35N)	44 (22.45%)	57 (21.76%)	0.859	32 (24.62%)	25 (18.94%)	0.266
A364C (N86H)	21 (10.71%)	39 (14.89%)	0.191	17 (13.08%)	22 (16.67%)	0.414
A387C (K93N)	43 (21.94%)	51 (19.47%)	0.517	29 (22.31%)	22 (16.67%)	0.249
A446G (Q113R)	24 (12.24%)	41 (15.65%)	0.302	19 (14.62%)	22 (16.67%)	0.648
G542T (R145K)	**22 (11.22%)**	**6 (2.29%)**	**< 0.001**	4 (3.08%)	2 (1.52%)	0.398
***E7*prototype**	**123 (62.76%)**	**200 (76.34%)**	**0.002**	96 (73.85%)	104 (78.79%)	0.347
C706T (A45V)	**27 (13.78%)**	**11 (4.20%)**	**< 0.001**	7 (5.38%)	4(3.03%)	0.342
C706A (A45E)	**20 (10.20%)**	**10 (3.82%)**	**0.006**	**10 (7.69%)**	**0**	**0.001**
A862T (Q87L)	24 (12.24%)	38 (14.50%)	0.484	16 (12.31%)	22 (16.67%)	0.316
***E6* + *E7*prototype**	**98 (50.00%)**	**182 (69.47%)**	**< 0.001**	86 (66.15%)	96 (72.73%)	0.248

Note The data with significant differences are in bold. *E6* + *E7* prototype indicates that both *E6* and *E7* on the same sequence are identical to the reference sequence. Time 1: January 1, 2009 to December 31, 2015. Time 2: January 1, 2017 to December 31, 2023. Time 2.1: January 1, 2017 to June 30, 2020. Time 2.2: July 1, 2020 to December 31, 2023. The "number (percentage)" represents number and percentage of HPV33 *E6* or *E7* prototype/mutation allele detected within the corresponding time period. p_1: The p-value of the comparison between time1 and time2. p_2: The p-value of the comparison between time 2.1 and time 2.2

overall adaptability of HPV33 might be slower than that of other major types. Notably, in other studies, the number of cervical cancer cases caused by HPV33 was higher than its prevalence. For example, in a study in Xianning, HPV33 caused 7.78% of CIN2/3 (ranking fourth) and 6.55% of cervical cancer (ranking fourth), which was higher than its prevalence ranking of 5.32% (seventh) [27]. A worldwide meta-analysis also showed that HPV33 causes approximately 5% of cervical cancers, ranking fourth, which is significantly higher than its prevalence (3.5%, ranking tenth) [5]. A carcinogenic rate higher than the prevalence indicates the strong pathogenicity of HPV33 after infection, highlighting the importance of studying HPV33 and suggesting the possible existence of variants with strong pathogenicity.

The detection rate of HPV81 has increased rapidly, and the replacement of its internal variants has also been rapid [31]. The external detection rate of HPV33 did not change rapidly, and its internal genetic polymorphisms showed only a slow change (no significant

Table 5 Viral loads comparison of HPV33 *E6/E7* variant/mutation

Variant/mutation	with	without	*p*	t
E6 prototype	4.62 ± 0.90	4.65 ± 1.02	0.786	0.27
A213C (K35N)	4.55 ± 1.01	4.65 ± 0.91	0.494	0.69
A364C (N86H)	4.59 ± 1.07	4.63 ± 0.91	0.790	0.27
A387C (K93N)	4.52 ± 1.05	4.65 ± 0.90	0.380	0.88
A446G (Q113R)	4.51 ± 1.10	4.65 ± 0.90	0.388	0.87
G542A (R145K)	**5.72 ± 0.84**	**4.60 ± 0.92**	**0.003**	**2.96**
E7 prototype	4.64 ± 0.90	4.57 ± 1.03	0.611	0.51
C706A (A45E)	4.58 ± 0.90	4.63 ± 0.93	0.859	0.18
C706T (A45V)	4.61 ± 0.89	4.63 ± 0.93	0.950	0.60
A862T (Q97L)	4.47 ± 1.09	4.65 ± 0.90	0.265	1.12
E6 + *E7* prototype	4.60 ± 0.90	4.68 ± 0.99	0.543	0.61

Note The data with significant differences are in bold. "With" means carrying/belonging to the variant/mutation, while "Without" means not carrying/not belonging to the variant/mutation. The numbers below "With/Without" represent the mean logarithmic value of viral load (units: $\log_{10}$ copies per 10,000 cells)

change in dominant variants/mutations from 2017 to 2023). However, after a long period of accumulation, significant changes occurred (from 2009 to 2023), which could reflect the current environmental pressures faced by HPV33. One of the main evolutionary paths of pathogenic microorganisms is to reduce disease symptoms to achieve symbiosis with the host [32], because pronounced external symptoms are more likely to be detected and treated, thus blocking transmission [33]. However, for HPV, persistent infection of a single type is a necessary condition for its pathogenicity; the longer the persistent infection period, the longer the lesion development time and the greater the severity of the lesion [34]. Therefore, whether persistent infection is beneficial or harmful to HPV's survival warrants in-depth study. Our results showed that the HPV33 *E6* + *E7* prototype significantly increased the relative risk for causing CIN1, CIN3 and persistent infection. While the HPV33 *E6* + *E7* prototype reduced the genetic diversity of HPV33 by outcompeting other variants, it also increased pathogenicity, leading to external symptoms that might weaken the overall adaptability of HPV33. This could be one of the factors contributing to the decline in HPV33 detection rates. HPV is an ancient virus that emerged during the early evolution of human ancestors and has co-evolved with them [35]. Our analysis showed that there were multiple positive selection sites on *E6* and *E7* that could enhance survival capacity [36]. However, in our study, the frequencies of positive selection sites accumulated during the long-term evolution of HPV33 *E6*/*E7* did not increase significantly. Instead, the detection rate of the HPV33 *E6*/*E7* prototype, which lacks these positive selection mutations, rose remarkably. This suggests that the survival ability of the variants carrying *E6*/*E7* positive selection mutations is inferior to that of the *E6*/*E7* prototype variants, indicating that with environmental changes, different sites may have different adaptation effects [37].

The variants of HPV33 are divided into three main phylogenetic lineages (A, B, and C) and five sublineages (A1, A2, A3, B1, and C1) [21]. These lineages have different distribution patterns and exhibit distinct biological activities worldwide [21]. The A1 sublineage, which is closest to the reference sequence, significantly increases the risk of CIN2/3 or cervical cancer and is also the predominant sublineage in Asia and Europe [14, 21]. However, the A1 sublineage comprises more than 10 variants, and the variant responsible for increasing the pathogenicity and adaptability of A1 has not been reported. Our results not only confirm that A1 is the most dominant and pathogenic sublineage in HPV33 but also further identify the *E6* + *E7* prototype variant as the key determinant of the strong pathogenicity and survival ability of A1. The mutations A213C (K35N), A364C (N86H), A387C (K93N), A446G (Q113R), and A862T (Q87L) significantly weakened persistent infection capacity and pathogenicity of HPV33, possibly because the control group predominantly consisted of *E6* + *E7* prototypes, which have stronger pathogenicity and greater persistent infection capacity.

Viral load is a core indicator of the infectivity, and it is also related to persistent infection capacity and pathogenicity [16, 38, 39]. However, the association between high-risk HPV viral load and lesion severity remains unclear. Most cross-sectional studies on HPV viral load and cervical diseases have reported that the viral load of HPV16 is positively correlated with more severe cervical lesions [40–43]. However, for other HPV types, the relation between viral load and disease is less well established. Among the four high-risk types—HPV18, 33, 52, and 58—which rank second in importance to HPV16, a few studies have suggested that a high viral load is positively correlated with severe lesions, whereas most studies have found no such significant association [40–43]. Regarding other relatively less prevalent high-risk types, very few studies have demonstrated a correlation between viral load and cervical lesions [42, 43]. This inconsistency may arise because the correlation between viral load and pathogenicity varies among HPV types. Alternatively, insufficient sample sizes or interference from mixed infections may have led to discrepancies in research findings. Therefore, whether HPV viral load can serve as a predictor of disease progression or function as an auxiliary marker for assessing cervical disease severity requires further investigation. In this cross-sectional study, the viral load was significantly higher in the disease group than in the lesion-free group. Follow-up analysis further revealed that a high viral load was positively correlated with persistent infection. Given that persistent

Table 6 Pathogenicity and persistent infection capacity comparison of *E6*/*E7* variants/mutations

Variant/mutation		With	Without	Relative risk (95% CI)	*p*
E6 prototype	**CIN3+**	**40**	**6**	**1.30 (1.11–1.52)**	**0.008**
	CIN2	15	6	1.07 (0.80–1.43)	0.676
	CIN1	**34**	**4**	**1.34 (1.15–1.56)**	**0.006**
	CIN + total	**89**	**16**	**1.27 (1.11–1.45)**	**0.001**
	Persistent	**48**	**14**	**1.30 (1.04–1.62)**	**0.025**
A213C (K35N)	**CIN3+**	**5**	**41**	**0.37 (0.16–0.88)**	**0.011**
	CIN2	3	18	0.49 (0.17–1.43)	0.148
	CIN1	**3**	**35**	**0.27 (0.09–0.82)**	**0.006**
	CIN + total	**11**	**94**	**0.36 (0.19–0.66)**	**< 0.001**
	Persistent	**11**	**51**	**0.49 (0.26–0.89)**	**0.013**
A364C (N86H)	**CIN3+**	**3**	**43**	**0.033 (0.11–1.03)**	**0.035**
	CIN2	3	18	0.72 (0.24–2.16)	0.55
	CIN1	**2**	**36**	**0.27 (0.07–1.07)**	**0.033**
	CIN + total	**8**	**97**	**0.38 (0.19–0.81)**	**0.007**
	Persistent	**7**	**55**	**0.46 (0.21–1.03)**	**0.046**
A387C (K93N)	**CIN3+**	**5**	**41**	**0.43 (0.18–1.02)**	**0.036**
	CIN2	3	18	0.56 (0.19–1.65)	0.26
	CIN1	**3**	**35**	**0.31 (0.10–0.95)**	**0.019**
	CIN + total	**11**	**94**	**0.41 (0.22–0.76)**	**0.003**
	Persistent	**10**	**52**	**0.49(0.26–0.94)**	**0.022**
A446G (Q113R)	**CIN3+**	**3**	**43**	**0.31 (0.1–0.97)**	**0.024**
	CIN2	3	18	0.68 (0.23–2.02)	0.471
	CIN1	**2**	**36**	**0.25 (0.06-1.00)**	**0.023**
	CIN + total	**8**	**97**	**0.36 (0.17–0.75)**	**0.003**
	Persistent	**7**	**55**	**0.42 (0.19–0.92)**	**0.021**
E7 prototype	CIN3+	39	7	1.20 (1.02–1.41)	0.056
	CIN2	17	4	1.15 (0.91–1.44)	0.326
	CIN1	**33**	**5**	**1.23 (1.05–1.44)**	**0.042**
	CIN + total	**89**	**16**	**1.20 (1.05–1.36)**	**0.009**
	Persistent	**50**	**12**	**1.32 (1.07–1.64)**	**0.011**
A862T (Q87L)	**CIN3+**	**3**	**43**	**0.34 (0.11–1.07)**	**0.042**
	CIN2	3	18	0.75 (0.25–2.24)	0.593
	CIN1	**2**	**36**	**0.28 (0.07–1.10)**	**0.039**
	CIN + total	**8**	**97**	**0.40 (0.19–0.84)**	**0.01**
	Persistent	**5**	**57**	**0.32 (0.13–0.79)**	**0.007**
E6 + *E7* prototype	**CIN3+**	**38**	**8**	**1.32 (1.11–1.58)**	**0.01**
	CIN2	14	7	1.07 (0.77–1.48)	0.705
	CIN1	**32**	**6**	**1.35 (1.12–1.62)**	**0.011**
	CIN + total	**84**	**21**	**1.28 (1.10–1.50)**	**0.002**
	Persistent	**45**	**17**	**1.35 (1.05–1.74)**	**0.021**

Note This table only shows the variants/mutations with significant statistical significance. For more detailed data, please refer to Supplementary Material 4. The results of significant differences are shown in bold. "Lesion-free" is the control "1" for the disease group, and "Non-Persistent" is the control "1" for the persistent infection ability group. "With" means carrying/belonging to the variant/mutation, while "Without" means not carrying/not belonging to the variant/mutation. The numbers under "With/Without" represent the corresponding sample numbers. "CI" stands for "Confidence Interval". "CIN": cervical intraepithelial neoplasia. "CIN3+" represents CIN3 and cervical cancer, and "CIN+" represents CIN and cervical cancer

infection is a prerequisite for HPV pathogenicity, this finding provides supporting evidence for the positive correlation between high HPV33 viral load and severe disease. Thus, it can be inferred that detecting a high HPV33 viral load not only suggests that current cervical disease severity may be greater but also indicates an increased risk of future persistent infection and a higher likelihood of disease progression.

In the HPV life cycle, *E6* and *E7* play a critical role by promoting infected cell division and driving viral replication. Mutations in *E6* and *E7* are likely to affect viral load. The *E6* prototype of HPV81 is particularly notable, as it significantly increases viral load, leading to the rapid replacement of allele frequencies [31]. However, the situation differs in HPV33. Although we identified a G542T (R145K) mutation in HPV33 *E6*/*E7* that

significantly influences viral load—and is also a positive selection site—the frequency of this mutation declined significantly. This suggests that HPV33 strains carrying this mutation have recently struggled to adapt to environmental pressures, indicating that viral load has not been a decisive factor in HPV33 variant replacement. From a viral transmission perspective, during the spread of severe acute respiratory syndrome coronavirus 2 α and δ variants, a high viral load was a key driver of increased transmission rates [44]. This phenomenon is analogous to the rapid replacement of mutant strains driven by the high viral load of the HPV81 *E6* prototype. However, during the transition from the severe acute respiratory syndrome coronavirus 2 δ variant to the omicron variant, omicron did not exhibit a substantial advantage in viral load over delta [45]. This scenario resembles HPV33, where variant replacement has not been driven primarily by changes in viral load [45]. Based on these findings, we speculate that alternative, unexplored mechanisms may be involved in HPV33 variant replacement [45].

Conclusion

The detection rate of HPV33 declined gradually between 2009 and 2023. Among its variants, the *E6*+*E7* prototype—the dominant variant with the greatest survival advantage—exhibited a slight increase in persistent infection capacity but a significant increase in pathogenicity. A high viral load was positively correlated with CIN + disease and persistent infection. Therefore, viral load may serve as a biomarker for predicting disease progression in HPV33 screening. HPV33 *E6* and *E7* alleles were gradually replaced and exhibited significant changes following long-term accumulation. However, no correlation was observed between viral load and major genetic variants/mutations, suggesting that alternative mechanisms may contribute to allele replacement.

Abbreviations

HPV Human papillomavirus
CIN Cervical intraepithelial neoplasia

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12985-025-02752-4.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Acknowledgements

We sincerely appreciate the contributions of medical staff from the gynecology, physical examination, and pathology departments to this study. Also, our gratitude goes to the patients who cooperated.

Author contributions

Zuyi Chen, and Qiongyao Li designed the study. Zuyi Chen, Yan Yang, Mingjing Zhang and Qiongyao Li provided the samples. Zuyi Chen, Qichen Cheng, Di Tian, Mingjing Zhang, Ganglin Liu, Zhengyuan An, Peixin A, Qiongyao Li conducted the experiments. Zuyi Chen, Qiongyao Li, Ganglin Liu, Peixin A, Lei Li and Feng Yang analyzed the data. Zuyi Chen, Qichen Cheng, and Qiongyao Li wrote the paper. The authors read and approved the final manuscript.

Funding

This work was supported by grants from the National Natural Science Foundation of China (No.82302527), Guizhou Province "Thousand Talents" Plan (No. xmrc120240204), Affiliated Hospital of Zunyi Medical University Excellent Talents Plan (No. rc220220916), Zunyi Science and Big Data Bureau (Zunyi Science and Technology Cooperation HZ (2023) 239), College Students' Innovative Entrepreneurial Training Plan Program (National Level No. 2024106610893, Provincial level No. S2024106612334).

Data availability

The data underpinning the results of this research are publicly accessible in GENEBANK at the website https://www.ncbi.nlm.nih.gov/genbank/. For HPV33 E7, the reference numbers are PV091834 - PV091842, while for HPV33 E6, the reference numbers are PV091866 - PV091876.

Declarations

Ethics approval and consent to participate

The research project received approval from the Ethics Committee of the Affiliated Hospital of Zunyi Medical University. The approval number was ZYFYLS2018(81). Written informed consents were obtained from all the patients or their guardians. The entire study was carried out in strict accordance with the standards of medical ethics.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Laboratory Medicine, Affiliated Hospital of Zunyi Medical University, Zunyi, Guizhou 563000, China
[2]Information Division, Affiliated Hospital of Zunyi Medical University, Zunyi, Guizhou, China
[3]Department of Medical Laboratory, People's Hospital of Dejang, Dejang, Guizhou, China
[4]Laboratory of Family Planning Service Center, Tongnan Maternal and Child Health Care Hospital, Tongnan, Chongqing, China

Received: 24 February 2025 / Accepted: 21 April 2025
Published online: 26 April 2025

References

1. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Cancer J Clin. 2024;74(3):229–63.
2. Rizzo A, Moschese D, Salari F, et al. High prevalence of high-risk HPV genotypes in individuals attending an infectious diseases clinic from 2018 to 2022 in Milan, Italy. J Med Virol. 2024;96(1):e29370.
3. Gholamzad A, Khakpour N, Hashemi M, et al. Prevalence of high and low risk HPV genotypes among vaccinated and non-vaccinated people in Tehran. Virol J. 2024;21(1):9.
4. Zhang J, Zha T, Wang X, et al. Prevalence and genotype distribution of HPV infections among women in Chengdu, China. Virol J. 2024;21(1):52.
5. Wei F, Georges D, Man I, et al. Causal attribution of human papillomavirus genotypes to invasive cervical cancer worldwide: a systematic analysis of the global literature. Lancet. 2024;404(10451):435–44.
6. Taguchi A, Yoshimoto D, Kusakabe M, et al. Impact of human papillomavirus types on uterine cervical neoplasia. J Obstet Gynecol Res. 2024;50(8):1283–8.

7. Niiya A, Hamaguchi Y, Mishima H, et al. Four conserved amino acids on human papillomavirus E6 predict clinical high-risk types. J Med Virol. 2023;95(8):e29049.
8. Zhao X, Zhang R, Song Z, et al. Curcumin suppressed the proliferation and apoptosis of HPV-positive cervical cancer cells by directly targeting the E6 protein. Phytother Res. 2024;38(10):4967–81.
9. Chen Z, Chan ABW, Kam LS, et al. Changes in the incidence and human Papillomavirus-Positive portion of oropharyngeal squamous cell carcinoma in Hong Kong. Cancers. 2024;16(1):226.
10. Yadav C, Yadav R, Chabbra R, et al. Overview of genetic and epigenetic regulation of human papillomavirus and apoptosis in cervical cancer. Apoptosis. 2023;28(5):683–701.
11. Mobini Kesheh M, Shavandi S, Azami J, et al. Genetic diversity and bioinformatic analysis in the L1 gene of HPV genotypes 31, 33, and 58 Circulating in women with normal cervical cytology. Infect Agents Cancer. 2023;18(1):19.
12. Gyöngyösi E, László B, Szalmás A, et al. Transcriptional activity of the long control region in human papillomavirus type 33 intratype variants. Virol J. 2023;20(1):152.
13. Yu M, Wu S, Wang S, et al. Polymorphism of E6 and E7 genes in human papillomavirus types 31 and 33 in Northeast China. Can J Infect Dis Med Microbiol. 2023;2023(1):9338294.
14. Yan ZY, Di XH, Qiu Y, et al. Cervical carcinogenesis risk association of HPV33 E6 and E7 genetic variations in Taizhou, Southeast China. Virol J. 2023;20(1):156.
15. Bappy SS, Haque Asim MM, Ahasan MM, et al. Virus-induced host cell metabolic alteration. Rev Med Virol. 2024;34(1):e2505.
16. Jones TC, Biele G, Mühlemann B, et al. Estimating infectiousness throughout SARS-CoV-2 infection course. Science. 2021;373(6551):eabi5273.
17. Mo B, Ye Y, Yu M, et al. Prevalence and genotype distribution of HPV combined with cervical pathological results in women from Sichuan, China: A cross-sectional study based on post-vaccination period 2019 to 2023. Cancer Med. 2024;13(16):e70148.
18. Sun Z, Zhang R, Liu Z, et al. Development of a fluorescence–based multiplex genotyping method for simultaneous determination of human papillomavirus infections and viral loads. BMC Cancer. 2015;15(1):1–11.
19. Li M, Zhao C, Zhang X, et al. PAX1/JAM3 methylation and HPV viral load in women with persistent HPV Infection. Cancers. 2024;16(7):1430.
20. Bonde J, Bottari F, Iacobone AD, et al. Human papillomavirus same genotype persistence and risk: a systematic review. J Lower Genit Tract Dis. 2021;25(1):27–37.
21. Chen AA, Heideman DAM, Boon D, et al. Human papillomavirus 33 worldwide genetic variation and associated risk of cervical cancer. Virology. 2014;448:356–62.
22. Seong J, Ryou S, Lee JG, et al. Enhanced disease progression due to persistent HPV-16/58 infections in Korean women: a systematic review and the Korea HPV cohort study. Virol J. 2021;18:1–13.
23. Zapata L, Caravagna G, Williams MJ, et al. Immune selection determines tumor antigenicity and influences response to checkpoint inhibitors. Nat Genet. 2023;55(3):451–60.
24. Li T, Yang Z, Luo P, et al. Genetic variability of human papillomavirus type 18 based on E6, E7 and L1 genes in central China. Virol J. 2024;21(1):152.
25. Totaro ME, Gili JA, Liotta DJ, et al. Genetic variation in the E6 and E7 genes of human papillomavirus type 16 in Northeastern Argentina. J Med Virol. 2022;94(2):745–51.
26. Shabanpour M, Jalali-Alhosseini P, Shoja Z, et al. Lineage and sublineage analysis of human papillomavirus type 58 in Iranian women. Virol J. 2024;21(1):244.
27. Qiu B, Jiang N, Jiang J, et al. The prevalence and genotype distribution of high-risk human papillomaviruses among women in Xianning, China. Virol J. 2024;21(1):140.
28. Yin X, Zhang C, Wu X, et al. HPV prevalence and distribution characteristics in postmenopausal women from Nanjing, China. BMC Womens Health. 2024;24(1):68.
29. Yan X, Shen L, Xiao Y, et al. Prevalence, characteristics, and distribution of HPV genotypes in women from Zhejiang Province, 2016–2020. Virol J. 2021;18:1–12.
30. Wang Q, Xu M, Zhou H, et al. Prevalence characteristics of cervical human papillomavirus infection in Chengdu and Aba district, Sichuan Province, China. PLoS ONE. 2024;19(6):e0304760.
31. Chen Z, Li Q, Xu Y, et al. Study on the detection rate, genetic polymorphism, viral load, persistent infection capacity, and pathogenicity of human papillomavirus type 81. J Med Virol. 2024;96(6):e29753.
32. Drew GC, Stevens EJ, King KC. Microbial evolution and transitions along the parasite–mutualist continuum. Nat Rev Microbiol. 2021;19(10):623–38.
33. Johansson MA, Quandelacy TM, Kada S, et al. SARS-CoV-2 transmission from people without COVID-19 symptoms. JAMA Netw Open. 2021;4(1):e2035057–2035057.
34. Perkins RB, Wentzensen N, Guido RS, et al. Cervical cancer screening: a review. JAMA. 2023;330(6):547–58.
35. Tanaka K, Kogure G, Onuki M, et al. Ancient evolutionary history of human papillomavirus type 16, 18 and 58 variants prevalent exclusively in Japan. Viruses. 2022;14(3):464.
36. Viana R, Moyo S, Amoako DG, et al. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in Southern Africa. Nature. 2022;603(7902):679–86.
37. Pereira L, Mutesa L, Tindana P, et al. African genetic diversity and adaptation inform a precision medicine agenda. Nat Rev Genet. 2021;22(5):284–306.
38. Bekker LG, Smith P, Ntusi N. A B. HIV is sexually untransmittable when viral load is undetectable. Lancet. 2023;402(10400):428–9.
39. McBride AA. Human papillomaviruses: diversity, infection and host interactions. Nat Rev Microbiol. 2022;20(2):95–108.
40. Oyervides-Muñoz MA, Pérez-Maya AA, Sánchez-Domínguez CN, et al. Multiple HPV infections and viral load association in persistent cervical lesions in Mexican women. Viruses. 2020;12(4):380.
41. Long W, Yang Z, Li X, et al. HPV-16, HPV-58, and HPV-33 are the most carcinogenic HPV genotypes in Southwestern China and their viral loads are associated with severity of premalignant lesions in the cervix. Virol J. 2018;15:1–8.
42. Zhou Y, Shi X, Liu J, et al. Correlation between human papillomavirus viral load and cervical lesions classification: A review of current research. Front Med. 2023;10:1111269.
43. Fobian SF, Mei X, Crezee J, et al. Increased human papillomavirus viral load is correlated to higher severity of cervical disease and poorer clinical outcome: A systematic review. J Med Virol. 2024;96(6):e29741.
44. Kidd M, Richter A, Best A, et al. S-variant SARS-CoV-2 lineage B1. 1.7 is associated with significantly higher viral load in samples tested by TaqPath polymerase chain reaction. J Infect Dis. 2021;223(10):1666–70.
45. Puhach O, Adea K, Hulo N, et al. Infectious viral load in unvaccinated and vaccinated individuals infected with ancestral, Delta or Omicron SARS-CoV-2. Nat Med. 2022;28(7):1491–500.

Publisher's note