JOURNAL OF CLINICAL MEDICINE OF KAZAKHSTAN (E-ISSN 2313-1519)
Original Article DOI: https://doi.org/10.23950/jcmk/14274

High-risk human papillomaviruses L1 gene isolates identified in Western Kazakhstan

Saule K. Balmagambetova[1], Elena V. Zholdybayeva[2], Oxana V. Zavalennaya[3], Ainur Amanzholkyzy[4], Victoria I. Kononets[5], Gulmira M. Zharmakhanova[5], Nadiar M. Mussin[6], Lazzat M. Zhamaliyeva[7], Nurgul M. Kereyeva[1]

[1]*Department of Oncology, West Kazakhstan Marat Ospanov Medical University, Aktobe, Kazakhstan*
[2]*Shared laboratory, National Center for Biotechnology, Astana, Kazakhstan*
[3]*PCR laboratory, Scientific-Production Center, West Kazakhstan Marat Ospanov Medical University, Aktobe, Kazakhstan*
[4]*Department of Physiology, West Kazakhstan Marat Ospanov Medical University, Aktobe, Kazakhstan*
[5]*Department of Natural Science Disciplines, West Kazakhstan Marat Ospanov Medical University, Aktobe, Kazakhstan*
[6]*Department of Transplantology, West Kazakhstan Marat Ospanov Medical University, Aktobe, Kazakhstan*
[7]*Department of Family Medicine, West Kazakhstan Marat Ospanov Medical University, Aktobe, Kazakhstan*

Received: 2023-04-01.
Accepted: 2024-01-04.

J Clin Med Kaz 2024; 21(1):26-34

Corresponding author:
Saule K. Balmagambetova.
E-mail: sau3567@gmail.com;
ORCID: 0000-0003-4080-5383.

Abstract

Kazakhstani researchers reported a significant prevalence of highly carcinogenic human papillomavirus types in the country.

The article **aimed** to present HPV L1 gene sequencing developments in women affected with cervical cancer throughout the western part of Kazakhstan with provided findings on the geographic pathways of obtained isolates.

Methods. The HPV L1 gene was amplified using the consensus primers MY09HPV 5'-CGTCCMARRGGAWACTGATC-3' and MY11HPV 5' – GCMCAGGGWCATAAYAATGG-3'. The purified DNA was used as the target for direct nucleotide sequencing. Phylogenetic analyses were conducted using the MegAlign program from the LASERGENE software package (version 6.0; DNA star, Madison, WI) and MEGA version 5.0 software. A multiple alignment was created through Clustal W software, and the neighbor-joining method was used to construct the phylogenetic tree.

Results. Of 70 HPV samples transported to the Astana shared laboratory for gene L1 sequencing, only ten appeared fit to obtain isolates (14.3%). The viral load of the samples ranged from 3.3 to 8.2, and the range of DNA concentration was from 8.16 to 69.6 ng/uL. Not yet registered in the world genebank, a unique HPV16 Kazakhstani isolate with its own branch was revealed in Aktobe. An isolate of potentially carcinogenic HPV53 forming a remote cluster with KF436822/1, KU951264.1 – Southwest China, and 97% identity with EU056643.1 – Ireland, and acted as a single agent for invasive cervical cancer was identified.

In general, the sequencing findings indicate the variety of ways for HPV pervasion into the western region of Kazakhstan: North and South America, Europe, and Asia.

The study was recorded in the ISRCTN registry, No. 7154910, 02/01/2018.

Keywords: Human papillomavirus, gene sequencing, isolates, gene L1, cervical cancer, western Kazakhstan.

Introduction

Cervical cancer remains a substantial health problem for women globally, requiring more effective prevention and control strategies. Despite plenty of various screening modalities, the incidence of cervical cancer in the world has demonstrated an increase of 68.5% since 1990 [1]. Human papillomaviruses (HPVs) are a causative factor for cervical cancer onset, as high-risk HPV DNA presents in 99.7% of cervical cancer specimens [2, 3]. HPVs constitute a large, diverse family of about 200 fully characterized types. In total, four groups are currently classified, including group 1, containing types of highly carcinogenic risk (HR-HPV); probably carcinogenic, such as type 68 (group 2a); possibly carcinogenic, such as 26, 53, 66, 67, 70, 73 and 82 types, considered potential carcinogens, with an arguable and still not thoroughly defined role in carcinogenesis (group 2b); not classified (group 3); and, possibly not carcinogenic types (group 4)

[4]. A growing number of proofs of HPV's continuous evolution resulted in clear molecular-biological evidence of the uniquely high carcinogenicity of types currently classified as probably / possibly carcinogenic [5, 6].

Some sources mentioned particular areas of Asia as countries with the lowest cervical cancer rates [7], but subsequent studies have proven this is not true. The first Kazakh researcher to study the HPV prevalence was M. Buleshov (2011). Buleshov et al. found that out of 17,000 women tested in South Kazakhstan, 1,870 were HPV-positive (11%). Their findings were then widely cited in international HPV bulletins [8-10]. However, more recent authors reported higher rates of HPV infection in various regions of the country, within 25% [11] and even higher. Among Kazakhstani women attending the gynecologic clinics, high-risk HPV was found in 39%, and 13% were infected with multiple HR-HPV types [12]. All homeland researchers report the predominance of the most carcinogenic HPV type 16 in the swabs of Kazakhstani women [11-14]. In the five Central Asian countries that were formerly part of the Soviet Union (Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan), cervical cancer incidence and mortality rates are far higher than those in most Western and high-income nations and are increasing [15]. Experts from the International Agency for Research on Cancer (IARC) Information Centre on HPV reported the prevalence of HPV16 and/or HPV18 among Kazakhstani women with normal cytology as 3.4%, with low-grade cervical lesions (LSIL/CIN-1) 21.2%, with high-grade cervical lesions (HSIL/CIN-2/CIN-3/CIS) 42.1%, and in cervical cancer cases 68.9%, respectively. Meanwhile, the crude incidence rate of cervical cancer per 100,000 women is 18.4. As to the HPV vaccine, it is unavailable and has not been introduced for now [16]. As known, the pilot HPV vaccination program started in Kazakhstan in 2013 but then was discontinued. Recent studies have shown that contrary attitudes towards HPV vaccination exist among Kazakhstani women, with approximately half having positive and almost half having negative or neutral attitudes towards the vaccine [17]. Nonetheless, the Kazakh Institute of Oncology and Radiology reported that the program relaunch is scheduled for 2024.

According to the nomenclature of HPV established by the International Committee on Taxonomy of Virus (ICTV), each HPV type can be differentiated into phylogenetic lineages in terms of geographic distribution, pathogenicity, regulation of transcription, and immunological response [18, 19, 20]. The HPV16 type has accordingly been divided into four phylogenetic lineages: A, B, C, and D. The lineage D consists of the three sub-lineages: D1, D2, and D3, that include Asian-American and North American sequences. HPV intratypic molecular variants can be distinguished based on oncogenic potentials despite their phylogenetic relatedness. Reportedly, the HPV16 lineage D appears more carcinogenic than other lineages [21].

The HPV genome is an 8,000 base pair (bp), double-stranded, circular DNA packaged within a protein capsid. The prototypical genome encodes 6 early genes (E1, E2, E4, E5, E6, and E7) and 2 late genes (L1 and L2) [22]. Specifically, the L1 gene encodes the major capsid protein, which forms a pentameric capsomer that self-arranges into a 72-subunit icosahedral capsid. The capsid is essential for viral binding and entry into host-specific tissues [23]. Furthermore, the L1 coding sequences of the immunogenic surface loops are distinctively poorly conserved due to selective pressures for mutagenesis and immune evasion [24].

Research on HPV gene sequencing that focuses on the geographic locations of the found isolates has not yet been performed in Kazakhstan. Making a nationwide map of HPV responsible for cervical cancer onset in local women and identifying HPV lineages and the isolates' countries of origin appears to be one of the foreground tasks for Kazakhstani researchers.

The purpose of HPV L1 gene sequencing in the present work was to confirm the HPVs identification established by PCR typing and to clarify the geographical distribution of identified HPV types circulating in the western region of Kazakhstan.

Thus, the article **aimed** to present HPV L1 gene sequencing developments in women affected with cervical cancer throughout the western part of Kazakhstan and provide findings on the geographic pathways of the obtained isolates.

Methods

This research constituted a part of a large multipurpose project on HPV infection in western Kazakhstan, with the published protocol [25]. The study protocol (version_2) was approved by the University Local Ethical Committee (Ref. 1 dated 09.01.2016). The work was carried out using the checklist of mandatory items for observational studies (STROBE) and STREGA for studies with genetic material. The informed consent form was developed based on WHO recommendations, and all participants who signed the form were informed of the research objectives. The patients' rights were protected.

Test systems "Quantum-21" (Russian manufacturing) were used for qualitative and quantitative detection of total HPV 21 types (6, 11, 16, 18, 26, 31, 33, 35, 39, 44, 45, 51, 52, 53, 56, 58, 59, 66, 68, 73, and 82). Of them, thirteen are highly carcinogenic, five are possibly/probably carcinogenic (26, 53, 66, 73, 82), and three are low carcinogenic types (6, 11, 44). DNA extraction for further L1 gene sequencing was performed with QIAamp DNA Mini Kit (Qiagen Ltd, Crawley, UK). All details on HPV detection in samples collected in women with newly diagnosed cervical cancer through the PCR Real-time method were published [25].

Description of HPV L1 gene sequencing

Amplification of PCR products:

Human papillomavirus fragment amplification was performed using universal primers:

MY09HPV – 5' – CGTCCMARRGGAWACTGATC-3'

and MY11HPV – 5' – GCMCAGGGWCATAAYAATGG.

The reaction mixture contained 50 pmol of each primer (MY09/11) 1x PCR buffer with KCl (Fermentas), six mM MgCl2, 200 mmol of each dNTP, and 2 U of Taq polymerase (Taq DNA Polymerase (recombinant, Fermentas). The PCR program was run on a GeneAmp amplifier 9700 (Applied Biosystems). It included long-term denaturation at 95°C for 5 minutes, 40 cycles at 95°C for 1 minute, 55°C for 1 minute, and 72°C for 1 minute, and final elongation at 72°C for 7 minutes.

Electrophoretic analysis of amplification products:

The amplified target DNA fragments were analyzed by separating DNA fragments in an agarose gel (agarose concentration from 2%) in the presence of an intercalating agent, ethidium bromide, which was used for further DNA visualization. Electrophoresis was performed in a PowerPac horizontal electrophoresis chamber and a BioRad Electrophoretic bath current source. 1x TAE buffer was used as electrode buffer.

The results were documented using the Gel Doc (Bio-Rad) gel documentation system with Quantity One software (Bio-Rad). The molecular sizes of the analyzed DNA samples were determined by comparing their electrophoretic mobility in the gel with the mobility of markers – a DNA fragment of a known molecular weight. "DNA Ladder 1kb" (Fermentas) was used as a molecular weight marker.

Determination and analysis of nucleotide sequences:

PCR products were purified from unbound primers by an enzymatic method using Exonuclease I (Fermentas) and alkaline phosphatase (Shrimp Alkaline Phosphatase, Fermentas).

To eliminate non-specific fragments, an additional method was used to purify PCR products by precipitation in the presence of polyethylene glycol (PEG 6000), based on early papers showing an inverse relationship between the percentage concentration of PEG and the size of the purified double-stranded DNA molecules.

Purification included:

– adding an equal volume of the prepared PEG mixture (52.4 PEG 6000, 40 ml 3M NaOAc pH 5.2, 1.32 ml 1M MgCl2, and up to 200 ml deionized water) to the PCR product;

– intensive mixing on a vortex for 20 seconds;

– keeping at room temperature for 30 minutes;

– centrifugation at 13 thousand rpm for 10 minutes;

– removal of the supernatant;

– washing the precipitate with 80% ethanol and eluting the DNA in water.

The sequencing reaction was performed using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) according to the manufacturer's instructions, followed by separating the fragments on a 3730xl DNA Analyzer (Applied Biosystems).

Nucleotide sequences obtained using forward and reverse primers were analyzed and combined into a joint sequence using SeqScape 2.6.0 software (Applied Biosystems).

The resulting nucleotide sequences were identified using the BLAST function in the international Gene Bank database (http://www.ncbi.nlm.nih.gov/). The genotype was determined by the maximum percentage of identity of the analyzed sequence to the reference samples in the Gene Bank. Phylogenetic analyses were conducted using the MegAlign program from the LASERGENE software package (version 6.0; DNA star, Madison, WI) and with MEGA version 5.0 software, http://www.megasoftware.net). A multiple alignment was created through Clustal W software, and the neighbor-joining method was used to construct the phylogenetic tree.

Results

For research, 70 samples were transported to the shared laboratory of the National Scientific Center for Biotechnology (Astana), but successful sequencing was performed only in 10 cases (14.3%). The overall effectiveness of L1 gene sequencing: of seven samples isolated in Aktau, two were sequenced successfully; of eight samples from Uralsk, one was sequenced; of three samples from Atyrau, sequencing failed in all cases. Fifty-two samples from Aktobe resulted in seven successful developments.

Table 1 Results of successful sequencing of the HPV L1 gene isolated in cervical cancer patients in western Kazakhstan.

№	Sample code	Viral load/ per sample	Region	Identified types	DNA concentration, ng/uL	Performing a sequence of the HPV L1 gene
1	2a/2/1	3,3*103	Aktau	18	69.6	An isolate similar to KC470221.1, country of origin the USA
2	3/2/189	8,2*103	Aktau	33	53.0	Isolate KC706450.1 from Saudi Arabia
3	2a/4/12	5,5 * 103	Uralsk	53	42.22	An isolate forming a remote cluster with KF436822/1, KU951264.1 – Southwest China, 97% identity with EU056643.1 – Ireland
4	4/1/27	6,7*103	Aktobe	16	24.66	A sequence similar to the EU918764 isolate from China is obtained
5	2a/1/30	7,1*103	Khromtau, Aktobe	16	25.84	An isolate similar to KU951264/1 was obtained, forming a separate cluster with AJ617545.1, the country of origin for both isolates is Cyprus
6	2a/1/34	7,4*103	Khromtau, Aktobe	16	28.0	Common cluster with isolates KU707481.1 (country – Netherlands), GQ465900 (country – Canada)
7	5/1/2	3,5*103 7,2*103	Aktobe, Shalkar	6 16	11.5	An isolate of genotype 16 similar to isolate AJ617545.1 from Cyprus
8	5/1/3	6,2*103	Aktobe, Temir	16	8.16	An isolate of genotype 16 similar to isolate EF133498.1, country – Portugal
9	4/1/314	6,5*103	Aktobe, Kayndy	31	22.0	An isolate of genotype 31, similar to isolate KX514424.1, country – Brazil
10	2a/1/2	5,2*103	Aktobe, Alga	16	19.3	An authentic isolate from the Aktobe region, not represented yet in the world Genbank

Table 1 displays the summarized results of the HPV L1 gene sequences obtained. The viral load of the samples, measured in 103 genomic equivalents (GE) per sample, ranged from 3.3 to 8.2, and the range of DNA concentration was broad, from 8.16 to 69.6 ng/uL (Table 1).

Mangystau samples

Results of the primary nucleotide sequence of sample 2a/2/1 (Aktau), identified with existing nucleotide sequences using the BLAST program, are shown in Figure 1.

Nucleotide sequence of HPV 2a/2/1 (Aktau, HPV18):

TTTAACAATATGTGCTTCTACACAGTCTATCCTGTA
CCTGGGCAATATGATGCTACCAAATTTAAGCAGTATAG
CAGACATGTTGAAGAATATGATTTGCAGTTTATTTTTCA
GTTATGTACTATTACTTTAACTGCAGATGTTATGTCCTAT
ATTCATAGTATGAATAGCAGTATTTTAGAGGATTGGAAC
TTTGGTGTTCCCCCCCCGCCAACTACTAGTTTGGTGGA
TACATATCGTTTTGTACAATCTGTTGCTATTACCTGTCA
AAAGGATGCTGCACCAGCTGAAAATAAGGATCCCTAT
GATACGTTAAAGTTTTGGAATGTGGATTTAAAGGAAAA
GTTTTCTTTAGACTTAGA

The identity with the isolate KC470221.1 (country of origin – the USA) is 99%. A phylogenetic tree was built for sample 2a/2/1 (Aktau) (Figure 2).

Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Ident	Accession
Human papillomavirus type 18 isolate Qv28775, complete genome	656	656	100%	0.0	99%	KC470221.1
Human papillomavirus type 18, isolate IS768, major capsid protein L1 (L1) gene, partial cds	656	656	100%	0.0	99%	U45893.1
Human papillomavirus type 18 isolate Z100, complete genome	651	651	100%	0.0	99%	KC470222.1
Human papillomavirus type 18 isolate Rw57, complete genome	651	651	100%	0.0	99%	KC470219.1
Human papillomavirus type 18 isolate Rw687, complete genome	651	651	100%	0.0	99%	KC470218.1
Human papillomavirus type 18 isolate Rw750, complete genome	651	651	100%	0.0	99%	KC470217.1
Human papillomavirus type 18 isolate Z53, complete genome	651	651	100%	0.0	99%	KC470216.1
Human papillomavirus type 18 isolate Z52, complete genome	651	651	100%	0.0	99%	KC470214.1
Human papillomavirus type 18 isolate MI_142_09 L1 protein gene, partial cds	651	651	100%	0.0	99%	JF728188.1
Human papillomavirus type 18 isolate Qv03814, complete genome	651	651	100%	0.0	99%	EF202154.1
Human papillomavirus type 18, isolate IS168, major capsid protein L1 (L1) gene, partial cds	651	651	100%	0.0	99%	U45892.1
Human papillomavirus type 18 isolate BF380, complete genome	645	645	100%	0.0	99%	KC470228.1
Human papillomavirus type 18 isolate Qv12693, complete genome	645	645	100%	0.0	99%	KC470227.1
Human papillomavirus type 18 isolate Z125, complete genome	645	645	100%	0.0	99%	KC470226.1
Human papillomavirus type 18 isolate BF172, complete genome	645	645	100%	0.0	99%	KC470225.1
Human papillomavirus type 18 isolate BF288, complete genome	645	645	100%	0.0	99%	KC470224.1
Human papillomavirus type 18 isolate BF309, complete genome	645	645	100%	0.0	99%	KC470223.1

Figure 1 - BLAST data for HPV18 (Mangystau, sample 2a/2/1)

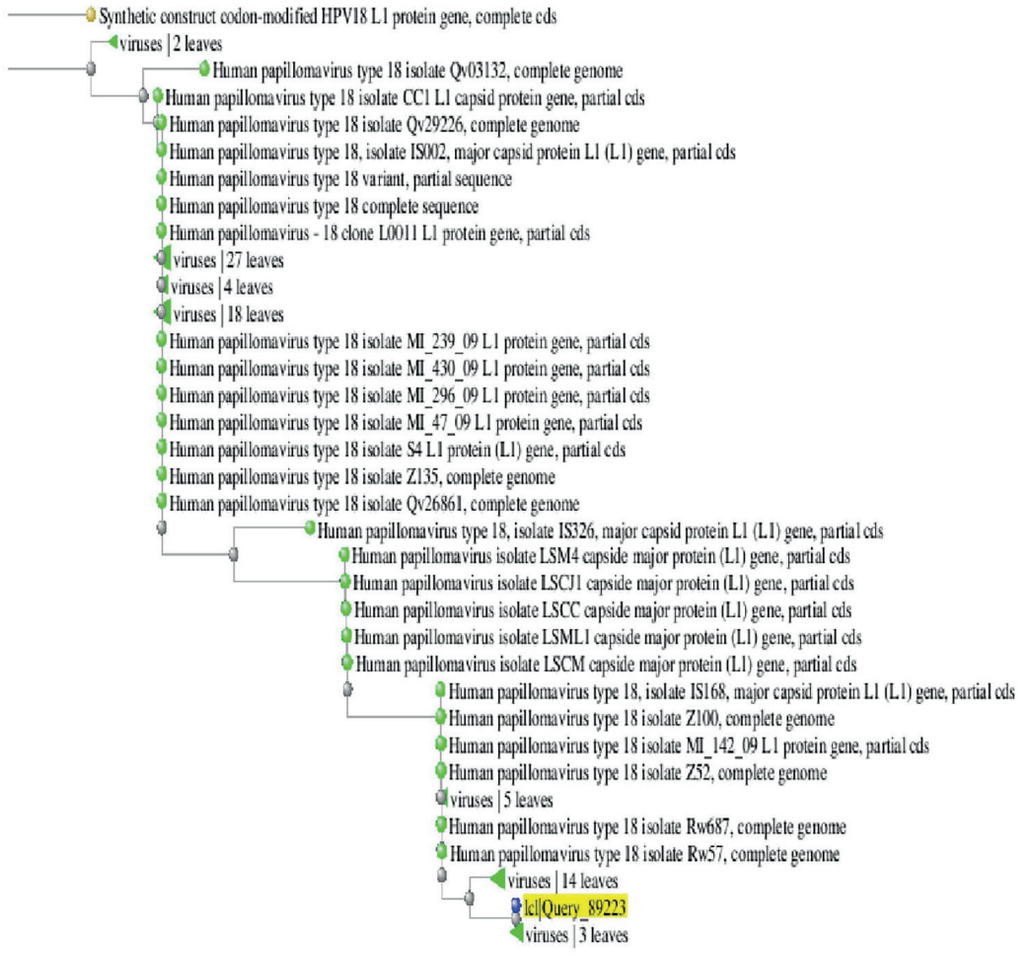

Figure 2 - HPV18 (Mangystau, sample 2a/2/1), phylogenetic tree

Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Ident	Accession
Human papillomavirus type 33 isolate 3 major capsid protein L1 (l1) gene, partial cds	592	592	100%	2e-165	99%	KC706450.1
Human papillomavirus type 33 isolate S5 L1 protein (L1) gene, partial cds	592	592	100%	2e-165	99%	JQ902114.1
Human papillomavirus type 33 isolate RW926, complete genome	592	592	100%	2e-165	99%	HQ537694.1
Human papillomavirus type 33 isolate Z84, complete genome	592	592	100%	2e-165	99%	HQ537693.1
Human papillomavirus type 33 isolate RW702, complete genome	592	592	100%	2e-165	99%	HQ537692.1
Human papillomavirus type 33 isolate BF266, complete genome	592	592	100%	2e-165	99%	HQ537691.1
Human papillomavirus type 33 isolate Qv35834, complete genome	592	592	100%	2e-165	99%	HQ537690.1
Human papillomavirus type 33 isolate Qv32494, complete genome	592	592	100%	2e-165	99%	HQ537688.1
Human papillomavirus isolate 1514 L1 protein gene, partial cds	592	592	100%	2e-165	99%	GU797244.1
Human papillomavirus type 33 isolate 3 major capsid protein L1 (L1) gene, complete cds	592	592	100%	2e-165	99%	GQ479014.1
Human papillomavirus type 33 isolate 2 major capsid protein L1 (L1) gene, complete cds	592	592	100%	2e-165	99%	GQ479013.1
Human papillomavirus isolate 06JAN_PHL_MY061_02 L1 capsid protein (L1) gene, partial cds	592	592	100%	2e-165	99%	EU911184.1
Human papillomavirus isolate 06JAN_PHL_MY035_06 L1 capsid protein (L1) gene, partial cds	592	592	100%	2e-165	99%	EU911114.1
Human papillomavirus type 33 strain SRB170AK L1 protein (L1) gene, partial cds	592	592	100%	2e-165	99%	EU779744.1
Human papillomavirus type 33 isolate Bsb-98 major capsid protein L1 gene, partial cds	592	592	100%	2e-165	99%	DQ486473.1
Human papillomavirus isolate PT170-04 L1 protein gene, partial cds	592	592	100%	2e-165	99%	DQ111016.1
Human papillomavirus type 33, isolate IS267, major capsid protein L1 (L1) gene, partial cds	592	592	100%	2e-165	99%	U45895.1
Human papillomavirus type 33, complete genome	592	592	100%	2e-165	99%	M12732.1
Human papillomavirus type 33 isolate 2207 major capsid protein L1 gene, partial cds	588	588	99%	2e-164	99%	EU056640.1

Figure 3 - The BLAST program data for HPV33 (Mangystau, sample 3/2/189)

The following sequence in Table 1 is also isolated in Aktau (Mangystau region). Figure 3 presents BLAST data for this sample (HPV33).

As can be seen from the results, the identity is 99%. The first in the GenBank alignments is isolate KC706450.1, corresponding to the HPV33 genotype detected in Saudi Arabia.

Nucleotide sequence of HPV33 sample 3/2/189 (Mangystau region):

ACCACTCGCAGTACTAATATGACTTTATGCACACA AGTAACTAGAaGACAGTACATATAAAAATGAAAATTTT AAAGAATATATAAGACATGTTGAAGAATATGATCTACA GTTTGTTTtTCAACTATGCAAAGTTACCTTAACTGCAGA-AGTTATGACATATATTCATGCTATGAATCCAGATATTTTA GAAGATTGGCAATTTGGTTTAACACCTCCTCCATCTGC TAGTTTACAGGATACCTAAAGGTTTGTTACCTCTCAGG CTATTACGTGTCAAAAAACAAGTACCTCCAAAGGAAA AGGAAGACCCCTTAGGTAAATATACATTTT

The phylogenetic tree of sample 3/2/189 with isolates presented in Genbank (sample marked in yellow) is shown in Figure 4.

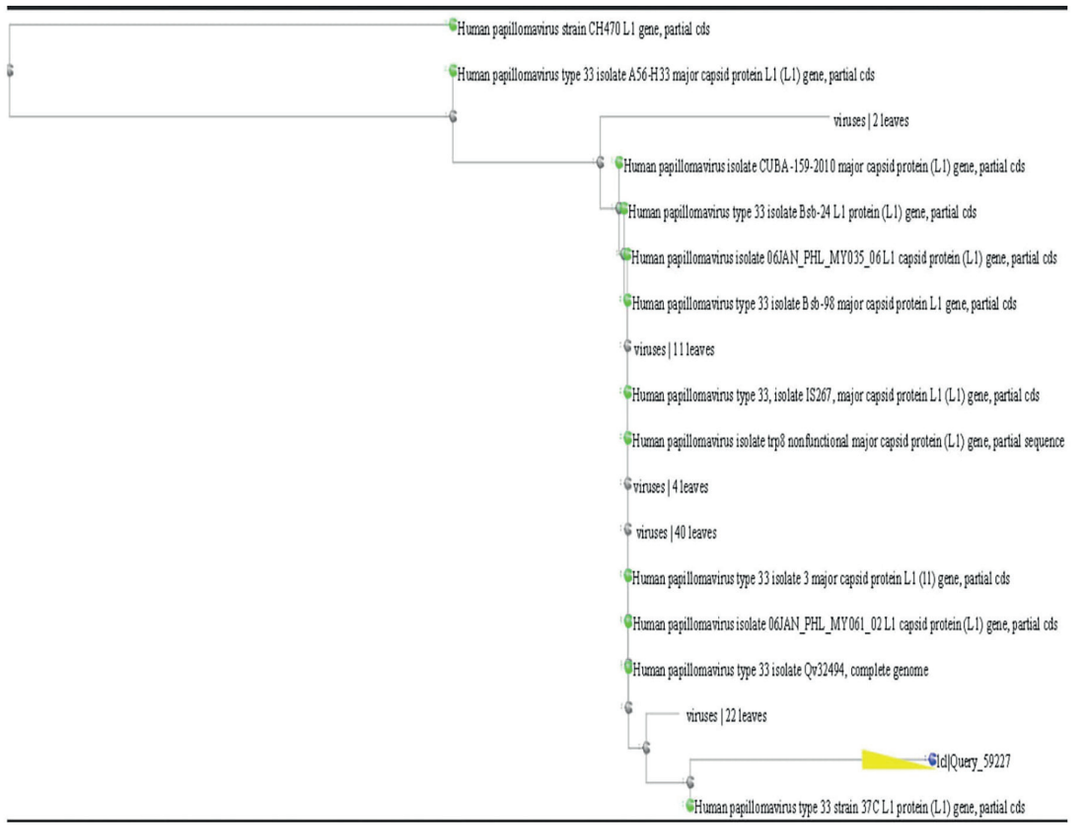

Figure 3 - HPV33 (Mangystau, sample 3/2/189), phylogenetic tree

Uralsk samples

Only one sample from Uralsk was fit for sequencing. An isolate genotype 53 (potentially / possibly carcinogenic) was obtained as a separate cluster from those from Southwest China and Ireland (identity 97%).

Aktobe samples

Sample 4/1/27 (HPV16), primary nucleotide sequence:

CAGTAAAGGATCCTAATTGTTACTGTTGTAGATA
CTACACGCAGTACACAATATGTCATGTATGGTGGCTGC
CATATGCTACCTCTGATACTACATATTAAAGTACTAAC
TTTAAAGAGTATCTACGACATGGGGAGGAATATGATTT
ACAGTTTATTTTTCAACTGTGCAAAATAACCTTATCTGC
AGACGTTATGACATACATTCATTCTATGAATTCCGCTAT
TTTGGAGGACTGGAATTTTGGACTACCCCCTCCCCCCT
GAGGCTCATTAGAAGATACTTATAGGTTTGTAACCTCC
CAGGCCATTGCTTGTCAAAAACATGCCCCTCCAACAC
CTAAAGAAGATCCCCTTAAAAAATACGCTTTTTGGGAA
GTAAATTTAAAGGAAAAGTTTTCTGCAGACCTAGATCA
GTTTCCCCCTTGGACGAAA

Figure 5 shows a print screen of the BLAST program results and the percent identity of the existing nucleotide sequences with the tested sample.

Description	Max score	Total score	Query cover	E value	Ident	Accession
Human papillomavirus type 16 isolate LZcc11-16, complete genome	590	590	94%	9e-165	93%	EU918764.1
Human papillomavirus type 16 isolate PT 62-05 L1 protein gene, partial cds	588	588	96%	3e-164	92%	EF133498.1
Human papillomavirus type 16 isolate PT 59-05 L1 protein gene, partial cds	588	588	96%	3e-164	92%	EF133497.1
Human papillomavirus proviral partial L1 gene for major capsid protein, isolate CY07-118	586	586	96%	1e-163	92%	AJ617545.1
Human papillomavirus type 16 isolate CNA138, complete genome	584	584	94%	4e-163	93%	KP212157.1
Human papillomavirus type 16 isolate CNA34, complete genome	584	584	94%	4e-163	93%	KP212153.1
Human papillomavirus type 16 isolate CNA33, complete genome	584	584	94%	4e-163	93%	KP212152.1
Human papillomavirus type 16 isolate CNA20, complete genome	584	584	94%	4e-163	93%	KP212151.1
Human papillomavirus type 16 isolate CNA15, complete genome	584	584	94%	4e-163	93%	KP212150.1
Human papillomavirus type 16 strain IR isolate IR-59 L1 protein (L1) gene, complete cds	584	584	94%	4e-163	93%	KP161014.1
Human papillomavirus type 16 strain IR isolate IR-44 L1 protein (L1) gene, complete cds	584	584	94%	4e-163	93%	KP160999.1
Human papillomavirus type 16, complete genome	584	584	94%	4e-163	93%	KF880690.1
Human papillomavirus type 16 isolate IR-32 L1 (L1) gene, complete cds	584	584	94%	4e-163	93%	KM058666.1
Human papillomavirus type 16 isolate IR-31 L1 (L1) gene, complete cds	584	584	94%	4e-163	93%	KM058665.1
Human papillomavirus type 16 isolate IR-19 L1 (L1) gene, complete cds	584	584	94%	4e-163	93%	KM058654.1
Human papillomavirus type 16 isolate IR-9 L1 (L1) gene, complete cds	584	584	94%	4e-163	93%	KM058645.1
Human papillomavirus type 16 isolate IR-1 L1 (L1) gene, complete cds	584	584	94%	4e-163	93%	KM058637.1
Human papillomavirus type 16 isolate IR-0 L1 (L1) gene, complete cds	584	584	94%	4e-163	93%	KM058636.1
Human papillomavirus type 16 isolate 16-Anhui12 from China, complete genome	584	584	94%	4e-163	93%	KC935953.1
Human papillomavirus type 16 isolate 78SE major capsid protein L1 (L1) gene, partial cds	584	584	94%	4e-163	93%	KJ467234.1
Human papillomavirus type 16 isolate 32SE major capsid protein L1 (L1) gene, partial cds	584	584	94%	4e-163	93%	KJ467230.1

Figure 5 - BLAST data for HPV16 (Aktobe, sample 4/1/27)

Figure 6 - HPV16 (Aktobe, sample 4/1/27), phylogenetic tree

The first in the table is the accession EU918764, which was isolated in China in 2007, but the percentage of identity with the tested sample is 93%. The phylogenetic tree of sample 4/1/27 with isolates presented in Genbank (sample marked in yellow) is displayed in Figure 6.

The explored sample forms a somewhat isolated branch, having a common node with the CNA33 isolate.

Identification of the genotype 16 Kazakhstani isolate from Aktobe, having its own branch, appears to be the study's most important finding. Unique sample 2a/1/2 (Aktobe, Alga district) and its nucleotide sequence:

CAGTAAAGGATCCTAATTGTTACTGTTGTAGATA CTACACGCAGTACACAATATGTCATGTATGGTGGCTGC CATATGCTACCTCTGATACTACATATTAAAGTACTAACT TTAAAGAGTATCTACGACATGGGGAGGAATATGATHA CAGTTTATTTTTCAACTGTGCAAAATAACCTTATCTGCA GACGTTATGACATACATTCATTCTATGAATTCCGCTATT TTGGAGGACTGGAATTTTGGACTACCCCCTCCCCCCTG AGGCTCATTAGAAGATACTTATAGGTTTGTAACCTCCC AGGCCATTGCTTGTCAAAAACATGCCCCTCCAACACCT AAAGAAGATCCCCTTAAAAAATACGCTTTTTGGGAAGT AAATTTAAAGGAAAAGTTTTCTGCAGACCTAGATCAGT TTCCCCCTTGGACGAAA

The results of identifying the sample with the BLAST program are shown in Figure 7.

The figure shows that the identity of the studied isolate with those deposited in Genbank reached only 93% percent.

Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Ident	Accession
Human papillomavirus type 16 isolate LZcc11-16, complete genome	590	590	94%	9e-165	93%	EU918764.1
Human papillomavirus type 16 isolate PT 62-05 L1 protein gene, partial cds	588	588	96%	3e-164	92%	EF133498.1
Human papillomavirus type 16 isolate PT 59-05 L1 protein gene, partial cds	588	588	96%	3e-164	92%	EF133497.1
Human papillomavirus proviral partial L1 gene for major capsid protein, isolate CY07-118	586	586	96%	1e-163	92%	AJ617545.1
Human papillomavirus type 16 strain IR isolate IR-59 L1 protein (L1) gene, complete cds	584	584	94%	4e-163	93%	KP161014.1
Human papillomavirus type 16 strain IR isolate IR-44 L1 protein (L1) gene, complete cds	584	584	94%	4e-163	93%	KP160999.1
Human papillomavirus type 16, complete genome	584	584	94%	4e-163	93%	KF880690.1
Human papillomavirus type 16 isolate IR-32 L1 (L1) gene, complete cds	584	584	94%	4e-163	93%	KM058666.1
Human papillomavirus type 16 isolate IR-31 L1 (L1) gene, complete cds	584	584	94%	4e-163	93%	KM058665.1
Human papillomavirus type 16 isolate IR-19 L1 (L1) gene, complete cds	584	584	94%	4e-163	93%	KM058654.1
Human papillomavirus type 16 isolate IR-9 L1 (L1) gene, complete cds	584	584	94%	4e-163	93%	KM058645.1
Human papillomavirus type 16 isolate IR-1 L1 (L1) gene, complete cds	584	584	94%	4e-163	93%	KM058637.1
Human papillomavirus type 16 isolate IR-0 L1 (L1) gene, complete cds	584	584	94%	4e-163	93%	KM058636.1
Human papillomavirus type 16 isolate 16-Anhui12 from China, complete genome	584	584	94%	4e-163	93%	KC935953.1
Human papillomavirus type 16 isolate 78SE major capsid protein L1 (L1) gene, partial cds	584	584	94%	4e-163	93%	KJ467234.1
Human papillomavirus type 16 isolate 32SE major capsid protein L1 (L1) gene, partial cds	584	584	94%	4e-163	93%	KJ467230.1
Human papillomavirus type 16 isolate 2SE major capsid protein L1 (L1) gene, partial cds	584	584	94%	4e-163	93%	KJ467225.1

Figure 7 - BLAST data for HPV16 (Aktobe, Alga district, sample 2a/1/2)

The program Mega 6.0 was used to construct the phylogenetic tree (Figure 8).

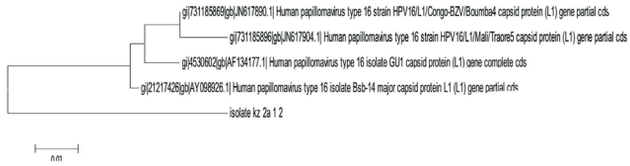

Figure 8 - HPV16 (Aktobe, sample 2a/1/2), phylogenetic tree

The presence of the Kazakh isolate's separate branch is seen in the figure. Further data collection is needed to build a phylogenetic tree and clarify the data obtained on the Kazakh HPV16 unique isolate.

The data from the presented study on the detected HPV types are summarized on the map of Western Kazakhstan (Figure 9). Regrettably, the findings are incomplete – there is a lack of data from the Atyrau region and insufficient data from Oral terrains. The research towards detecting all available information on HPV distribution to compile a complete map of the prevalence of HPV types in the Western region and throughout the country needs to be continued.

Figure 9 - HPV types detected in the western region of Kazakhstan

Discussion

As stated previously, the most convenient object for HPV gene sequencing is the L1 gene, often sequenced for additional genotype identification. Comparison of HPV 6b, 11, 16, 18, 31, and 33 sequences showed that the most conserved regions are within the open reading frames (ORFs) of the E1 and L1 genes (E, early; L, late). These HPV genes are responsible for virus replication. Two highly conserved 20 bp sequences were found in the L1 ORF, common to all sequenced HPV types [26, 27]. As known, HPV16 variants, classified based on less than 10% nucleotide variations in the major capsid (L1 ORF), contribute to persistent infection leading to cancer development. L1 protein forms the cornerstone of HPV structure and antigenicity [23]. Nowadays, all licensed HPV recombinant vaccines are designed based on HPV major capsid L1 protein [27]. But to explore the genetic variability of isolates, genomic regions E6/E7 are used, i.e., genes responsible for the expression of oncoproteins. Researchers observed thousands of unique HPV16 genomes exploring the E7 region. They reported that very few women shared the identical HPV16 sequence, which should stimulate a careful re-evaluation of the clinical implications of HPV mutation rates, transmission, clearance, and persistence [28].

In our study, the results of HPV L1 gene sequencing evidence that the geographical introduction of obtained isolates in western Kazakhstan is extensive and various. In fact, we established pervasion of HPV infection from almost all continents, excluding Africa and Australia. For types 18 and 33 circulating in the Mangistau region, we found an introduction from North America and the Arabian Peninsula. At the same time, researchers from Saudi Arabia established that HPVs circulating in their territory formed a closed cluster with African, Asian, East Asian, and American HPVs distributed into multiple lineages from various geographical locations [29].

We found that the isolate HPV53 from Uralsk was presented as a separate cluster from those from Southwest China and Ireland. Canada and the Netherlands have been identified as the geographic locations for the isolate obtained in the Khromtau sight of Aktobe. Other countries of origin for the Aktobe samples were the following: Brazil, South America (sample isolated in the Kayndy site), Cyprus (isolates obtained in the districts of Khromtau and Shalkar), Portugal (the isolate was obtained in the Temir district), and China.

We believe that the work done has resulted in two indisputable achievements. In the Alga area, an original isolate of genotype 16 has been detected as a separate branch for which it is necessary to complete the phylogenetic tree. The isolate has not yet been entered into the world Genbank, and the registration process should be commenced as the tree is built. Considering that HPV16 is the predominant type among HPV-infected women in western Kazakhstan (26.4% prevalence), besides it is the causative factor for 54.1% of women affected with cervical cancer, any research into the nature and behavior of this genotype seems very important [11].

In Uralsk, we detected a case of cervical cancer caused by HPV53 (sample 2a/4/12). The value of this sample's successful sequencing is that the apparent carcinogenicity of one of the so-called "potentially carcinogenic" HPV types was confirmed. By definition, highly carcinogenic types are those capable of causing invasive cancer by acting as a single agent. Moreover, we established the geographic path for this isolate, which forms a remote cluster with KF436822/1, KU951264.1 – Southwest China, and 97% identity with EU056643.1 – Ireland.

However, the work was accompanied by significant disadvantages. The first and the most prominent is unsatisfactory productivity, only 14.3% of successful cases. As known, collection, storage, and transport significantly affect the quality of the samples. In our study, some examples showed a very low viral load, and some lacked DNA, thus failing to obtain the L1 gene sequences in 60 cases. Specialists from the shared laboratory practiced repeated attempts to optimize the PCR conditions and the sequencing procedure, as well as additional experiments (reamplification, re-sequencing), but without success. Meanwhile, the sampling and transportation of the biomaterial were carried out according to the Instructions of the transport medium manufacturer – the Russian company "DNA-technologies". However, it cannot be denied that there were certain deviations from Instruction during the transportation of samples by train or plane due to some lack of control by the project staff.

One of the limitations is some incompleteness of work. Identifying the lineages for HPV16 isolates would be beneficial, given the different carcinogenicity of the four HPV16 lineages. The subsequent research on the genetic features of HPV types circulating in Kazakhstan should include these issues. Moreover, there is a need to generate complete genome sequence information to provide a clearer picture of the genetic diversity and evolution of HPVs in the country.

In general, the sequencing findings indicate the variety of ways for HPV pervasion into the western region of Kazakhstan. The presence of HPV53 as a single agent for invasive cervical cancer (sample 2a/4/12) confirms the current viewpoint on the particularities of HPV behavior – high genetic variability and a tendency to constant evolution. It is necessary to continue the research cycle and complete the Kazakhstani isolate of type HPV16 registration process in the world Genbank.

Author Contributions: Conceptualization, S. B. and E. Z.; methodology, S. B., E. Z. and O. Z.; validation, A. A. and G. Z.; formal analysis, V. K., G. Z., and N. K.; investigation, S. B., E. Z., and O. Z.; resources, A. A. and L. Z.; data curation, O. Z., V. K. and G. Z.; writing – original draft preparation, S. B.; writing – review and editing, S. B., E. Z. and N. M.; visualization, E. Z.; supervision, S. B.; project administration, N. M.; funding acquisition, L. Z. and N. K. The authors have read and agreed to the published version of the manuscript

Disclosures: There is no conflict of interest for all authors.

Acknowledgments: A shared laboratory of the National Center for Biotechnology of the Committee of Science carried out HPV L1 gene sequencing based on the scientific cooperation agreement.

Funding: This research was performed in frames of the scientific project "Epidemiological analysis of Human Papillomavirus in western Kazakhstan in relation to HPV-attributable cervical pathology – social, clinical and genetic aspects," funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant 2230/GF4, state registration No. 0115PK01224. The study is recorded in the ISRCTN registry, # 7154910, 01/02/2018.).

References

1. Zhang X, Zeng Q, Cai W, Ruan W. Trends of cervical cancer at global, regional, and national level: data from the Global Burden of Disease study 2019. *BMC Public Health.* 2021; 21(1): 894. https://doi.org/10.1186/s12889-021-10907-5.
2. zur Hausen H. Papillomaviruses in the causation of human cancers – a brief historical account. Virology. 2009; 384(2): 260-5. https://doi.org/10.1016/j.virol.2008.11.046.
3. Walboomers JM, Jacobs MV, Manos MM, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol*. 1999; 189(1): 12–19. https://doi.org/10.1002/(SICI)1096-9896(199909)189:1<12::AID-PATH431>3.0.CO;2-F.
4. Schiffman M, Clifford G, Buonaguro FM. Classification of weakly carcinogenic human papillomavirus types: addressing the limits of epidemiology at the borderline. *Infect Agent Cancer.* 2009; 4: 17-29. https://doi.org/10.1186/1750-9378-4-8.
5. Halec G, Alemany L, Lloveras B, Schmitt M, et al. Retrospective International Survey and HPV Time Trends Study Group. Pathogenic role of the eight probably/possibly carcinogenic HPV types 26, 53, 66, 67, 68, 70, 73 and 82 in cervical cancer. *J Pathol* 2014; 234: 441-51. ttps://doi.org/10.1002/path.4405.
6. McBride AA. Human papillomaviruses: diversity, infection and host interactions. *Nat Rev Microbiol* 2022; 20(2): 95–108. https://doi.org/10.1038/s41579-021-00617-5.
7. Shrestha AD, Neupane D, Vedsted P, Kallestrup P. Cervical Cancer Prevalence, Incidence and Mortality in Low and Middle Income Countries: A Systematic Review. *Asian Pac J Cancer* 2018; 19(2): 319-24. https://doi.org/10.1111/ecc.13207.
8. Rogovskaya SI, Shabalova IP, Mikheeva IV, Minkina GN, Podzolkova NM, Shipulina OY, Sultanov SN, Kosenko IA, Brotons M, Buttmann N, Dartell M, Arbyn M, Syrjänen S, & Poljak M. Human papillomavirus prevalence and type-distribution, cervical cancer screening practices and current status of vaccination implementation in Russian Federation, the Western countries of the former Soviet Union, Caucasus region and Central Asia. *Vaccine* 2013; 31 Suppl 7: H46–H58. https://doi.org/10.1016/j.vaccine.2013.06.043.
9. De Vuyst H, Alemany L, Lacey C, Chibwesha CJ, Sahasrabuddhe V, Banura C, Denny L, Parham GP. The burden of human papillomavirus infections and related diseases in sub-saharan Africa. *Vaccine* 2013; 31 Suppl 5: F32-46. https://doi.org/10.1016/j.vaccine.2012.07.092.
10. European Cancer Organisation. Putting HPV on the Map: The State of HPV Prevention Programmes in the WHO European Region. Available at: [https://www.europeancancer.org/resources/256:hpv-prevention-programmes.html] (Accessed 27 Apr 2022.)
11. Balmagambetova SK, Tinelli A, Urazayev ON, Sakieva KZ, Koyshybaev AK, Zholmukhamedova DA, et al. HPV Types Distribution in General Female Population and in Women Diagnosed with Cervical Cancer Across Western Kazakhstan. *Asian Pac J Cancer* 2019; 20(4): 1089-96. https://doi.org/10.31557/APJCP.2019.20.4.1089.
12. Babi A, Issa T, Issanov A, Akilzhanova A, Nurgaliyeva K, Abugalieva Z, et al. Prevalence of high-risk human papillomavirus infection among Kazakhstani women attending gynecological outpatient clinics. *Int J Infect Dis* 2021; 109: 8-16. https://doi.org/10.1016/j.ijid.2021.06.006.
13. Aimagambetova G, Azizan A. Epidemiology of HPV Infection and HPV-Related Cancers in Kazakhstan: a Review. *Asian Pac J Cancer Prev* 2018; 19(5): 1175-80. https://doi.org/10.22034/APJCP.2018.19.5.1175.
14. Niyazmetova L, Aimagambetova G, Stambekova N, et al. Application of molecular genotyping to determine prevalence of HPV strains in pap smears of Kazakhstan women. Int J Infect Dis 2016; 54: 85-8. https://doi.org/10.1016/j.ijid.2016.11.410.
15. Aimagambetova G, Chan CK, Ukybassova T, Imankulova B, Balykov A, Kongrtay K, et al. Cervical cancer screening and prevention in Kazakhstan and Central Asia. J Med Screen. 2021; 28(1): 48-50. https://doi.org/10.1177/0969141320902482.
16. Bruni L, Albero G, Serrano B, Mena M, Collado JJ, Gómez D, Muñoz J, Bosch FX, de Sanjosé S. ICO/IARC Information Centre on HPV and Cancer (HPV Information Centre). Human Papillomavirus and Related Diseases in Kazakhstan. Summary Report 22 October 2021. [Date Accessed]
17. Aimagambetova G, Babi A, Issa T, Issanov A. What Factors Are Associated with Attitudes towards HPV Vaccination among Kazakhstani Women? Exploratory Analysis of Cross-Sectional Survey Data. *Vaccines (Basel)*. 2022; 10(5): 824. https://doi.org/10.3390/vaccines10050824.
18. Pal A, & Kundu R. Human Papillomavirus E6 and E7: The Cervical Cancer Hallmarks and Targets for Therapy. *Front Microbiol* 2020; 10: 3116. https://doi.org/10.3389/fmicb.2019.03116.
19. Van Doorslaer K. Revisiting Papillomavirus Taxonomy: A Proposal for Updating the Current Classification in Line with Evolutionary Evidence. *Viruses* 2022; 14(10): 2308. https://doi.org/10.3390/v14102308.
20. Bhattacharjee R, Das SS, Biswal SS, Nath A, Das D, Basu A, Malik S, Kumar L, et al. Mechanistic role of HPV-associated early proteins in cervical cancer: Molecular pathways and targeted therapeutic strategies. *Crit Rev Oncol Hematol* 2022; 174: 103675. https://doi.org/10.1016/j.critrevonc.2022.103675.
21. Vonsky M, Shabaeva M, Runov A, Lebedeva N, Chowdhury S, Palefsky JM, & Isaguliants M. Carcinogenesis Associated with Human Papillomavirus Infection. Mechanisms and Potential for Immunotherapy. *Biochemistry. Biokhimiia* 2019; 84(7): 782–799. https://doi.org/10.1134/S0006297919070095.
22. Sayers EW, Cavanaugh M, Clark K, Ostell J, Pruitt KD, & Karsch-Mizrachi I. GenBank. *Nucleic Acids Res* 2020; 48(D1): D84–D86. https://doi.org/10.1093/nar/gkz956.
23. Mane A, Patil L, Limaye S, Nirmalkar A, Kulkarni-Kale U. Characterization of major capsid protein (L1) variants of Human papillomavirus type 16 by cervical neoplastic status in Indian women: Phylogenetic and functional analysis. *J Med Virol* 2020; 92(8): 1303-08. https://doi.org/10.1002/jmv.25675.
24. Shen-Gunther J, Cai H, Zhang H and Wang Y. Abundance of HPV L1 Intra-Genotype Variants With Capsid Epitopic Modifications Found Within Low- and High-Grade Pap Smears With Potential Implications for Vaccinology. *Front Genet* 2019; 10: 489. https://doi.org/10.3389/fgene.2019.00489.
25. Bekmukhambetov YZ, Mynbaev OA, Tinelli A, Balmagambetova SK, Koyshybaev AK, Urazayev ON, Ismagulova EK, et al. Human Papillomavirus related issues in western Kazakhstan: Protocol for a comprehensive study. *Rus Open Med J.* 2018; 7(4): e0408. https://doi.org/10.15275/rusomj.2018.0408.
26. Yu L, Majerciak V, Zheng ZM. HPV16 and HPV18 Genome Structure, Expression, and Post-Transcriptional Regulation. Int J Mol Sci. 2022; 23(9): 4943. doi: 10.3390/ijms23094943. Erratum in: *Int J Mol Sci.* 2022; 23(14). https://doi.org/10.3390/ijms23094943.
27. Oumeslakht L, Ababou M, Badaoui B, Qmichou Z. Worldwide genetic variations in high-risk human papillomaviruses capsid L1 gene and their impact on vaccine efficiency. *Gene* 2021; 782:145533. https://doi.org/10.1016/j.gene.2021.145533.
28. Nelson CW, & Mirabello L. Human papillomavirus genomics: Understanding carcinogenicity. *Tumour Virus Res* 2023; 15: 200258. https://doi.org/10.1016/j.tvr.2023.200258.
29. Sait K, Turki R, Abuzenadah AM, Jiffiri OH, Bohmaidah A, Sohrab SS. Genetic diversity and phylogenetic analysis of HPV 16 & 18 variants isolated from cervical specimens of women in Saudi Arabia. *Saudi J Biol Sci.* 2019; 26(2): 317-24. https://doi.org/10.1016/j.sjbs.2018.05.005.