Journal of Taibah University Medical Sciences (2025) 20(2), 242–250

Taibah University

Journal of Taibah University Medical Sciences

www.sciencedirect.com

Original Article

Exploring HPV-linked head and neck cancer in Southern Punjab, Pakistan: Insights from HPV-16 phylogenetic analysis

Muhammad Kashif, PhD [a,b,*], Sadia Minhas, PhD [c], Shah Jahan, PhD [d], Faheem Shahzad, PhD [d], Romeeza Tahir, PhD [e], Afia Abbas, MSc [e], Muhammad Idrees, PhD [f] and Nadeem Afzal, PhD [e]

[a] *Department of Oral Pathology, University of Health Sciences, Khyaban e Jamia Punjab, Lahore, Pakistan*
[b] *Department of Oral Pathology, Bakhtawar Amin Medical and Dental College, Northern Bypass Road, Multan, Pakistan*
[c] *Department of Oral Pathology, Akhtar Saeed Medical and Dental College, Bahria Town, Lahore, Pakistan*
[d] *Institute of Allied Health Sciences, University of Health Sciences, Khyaban e Jamia Punjab, Lahore, Pakistan*
[e] *Department of Immunology, University of Health Sciences, Khyaban e Jamia Punjab, Lahore, Pakistan*
[f] *Center of Excellence in Molecular Biology (CEMB), University of the Punjab, Lahore, Pakistan*

Received 11 May 2024; revised 12 February 2025; accepted 5 March 2025; Available online 1 April 2025

الملخص

أهداف البحث: سرطان الخلايا الحرشفية في منطقة الرأس والرقبة هو مرض غير متجانس ويمكن تصنيفه إلى نوعين فرعيين: إيجابي لفيروس الورم الحليمي البشري (20%) وسلبي لفيروس الورم الحليمي البشري (80%). ومع ذلك، فإن انتشار النمط الجيني لفيروس الورم الحليمي البشري في باكستان غير معروف جيدا. تهدف هذه الدراسة إلى بحث مدى شيوع النمط الجيني لفيروس الورم الحليمي البشري-16 في مرضى سرطان الخلايا الحرشفية في منطقة جنوب البنجاب في باكستان وما هي السمات الجزيئية المحددة التي يظهرها هذا النمط الجيني.

طرق البحث: في هذه الدراسة المقطعية، جمع الباحثون 85 عينة أنسجة من حالات تم تشخيصها بسرطان الخلايا الحرشفية في منطقة الرأس والرقبة. لاستخراج الحمض النووي الجينومي، تم استخدام مقاطع الأنسجة المضمنة بالبارافين المثبتة بالفورمالين. وللكشف عن الحمض النووي لفيروس الورم الحليمي البشري، تم تضخيم المنطقة L1 باستخدام بادئات GP5+ وGP6+. تم استخدام تفاعل البوليميراز المتسلسل في الوقت الحقيقي لتحديد النمط الجيني لفيروس الورم الحليمي البشري عالي الخطورة. تم استخدام تسلسل الجينوم الكامل للتحليل التطوري والطفري لفيروس الورم الحليمي البشري-16.

النتائج: من بين 85 عينة، كانت 7.1% إيجابية لفيروس الورم الحليمي البشري، مع 4.7% إيجابية لفيروس الورم الحليمي البشري-16 و2.4% إيجابية لفيروس الورم الحليمي البشري-18. كما كان هناك ارتباط كبير بين فيروس الورم الحليمي البشري عالي الخطورة الإيجابي والدرجة النسيجية. كان جينوم فيروس الورم الحليمي البشري-16 في الدراسة وثيق الصلة بتلك الموجودة في تايلاند والولايات المتحدة والهند والصين وأوروبا، وتم الكشف عن إجمالي 11 طفرة في الجينوم، منها أربع طفرات جديدة.

الاستنتاجات: تشير نتائج الدراسة الحالية إلى انتشار نادر لسرطان الخلايا الحرشفية للرأس والرقبة المرتبط بفيروس الورم الحليمي البشري عالي الخطورة في باكستان. من خلال التحليل التطوري، تم اكتشاف أن جينوم فيروس الورم الحليمي البشري-16 المعزول في هذه الدراسة له بنية وراثية مميزة وكان له أيضا أوجه تشابه مع الجينومات المبلغ عنها من تايلاند والولايات المتحدة والهند والصين وأوروبا.

Abstract

Introduction/objectives: Head and neck region squamous cell carcinoma (HNSCC) is a heterogeneous disease that can be categorized into human papillomavirus (HPV)-positive (20 %) and HPV-negative (80 %) subtypes. However, the prevalence of HPV genotypes is not clear in Pakistan. This study investigated how common the HPV-16 genotype is in patients with HNSCC in the Southern Punjab region of Pakistan, and the specific molecular features of this genotype.

Methods: For this cross-sectional study, 85 tissue samples were collected from diagnosed cases of HNSCC. Formalin-fixed paraffin-embedded tissue sections were

* Corresponding address: Department of Oral Pathology, Bakhtawar Amin Medical and Dental College, Northern Bypass Road, 60000, Multan, Pakistan.
E-mail: drkashifazam@gmail.com (M. Kashif)

Peer review under responsibility of Taibah University.

Production and hosting by Elsevier

 https://doi.org/10.1016/j.jtumed.2025.03.001

used for genomic DNA extraction. The L1 region was amplified using GP5+ and GP6+ primers to detect HPV DNA. Real-time PCR was conducted to genotype high-risk HPV (HR-HPV). Whole genome sequencing was used for phylogenetic analysis of HPV-16 and to detect mutations/single nucleotide polymorphisms (SNPs).

Results: Among the 85 samples, 7.1 % were positive for HPV, where 4.7 % were positive for HPV-16 and 2.4 % were positive for HPV-18. A significant association was found between HR-HPV positivity and histological grade ($p < 0.05$). The HPV-16 genome sequence obtained in this study was closely related to those from Thailand, the United States, India, China, and Europe, and 11 mutations/SNPs were detected in the sequenced genome, where four were novel.

Conclusion: The findings obtained in the present study demonstrate the low prevalence of HR-HPV associated HNSCC in Pakistan. Phylogenetic analysis showed that HPV-16 genome isolated and sequenced in this study had a distinct genetic structure and it also shared similarities with genomes reported from Thailand, the United States, India, China, and Europe.

Keywords: FFPE; HNSCC; HPV-16; OPSCC; OSCC; Phylogenetic analysis; SNPs

Introduction

Head and neck squamous cell carcinoma (HNSCC) is a highly malignant tumor that arises in the squamous epithelial linings of the oral cavity, oropharynx, hypopharynx, and larynx, and each of these tumors is linked to various risk factors and prognoses. The latest GLOBOCAN estimates (2022) show that malignancies in the head and neck region, including the lip, oral cavity, salivary glands, larynx, nasopharynx, oropharynx, and hypopharynx, constitute approximately 4.7 % of all cancer cases worldwide. These cancers collectively account for about 4.9 % of global cancer-related deaths.[1] Tobacco and alcohol use are closely linked to the development of HNSCC, and a specific subset of HNSCCs is associated with high-risk human papillomavirus (HR-HPV) infection.[2] Evidence suggests that HR-HPV plays a role in the pathogenesis of a subset of HNSCCs, particularly in young patients. The oropharynx, tonsils, and base of the tongue are the most common sites for the development of HPV-associated HNSCC.[3]

It is well established that around 18 high risk HPV genotypes have oncogenic potential and are linked to the development of cancers.[4] In 2007, the International Agency for Research on Cancer recognized HPV-16 as the sole carcinogenic type implicated in non-cervical sites, including the anus, penis, vagina, vulva, oral cavity, and oropharynx.[5] This HPV genotype has been reported as a main causative factor for a subset of HNSCC called oropharyngeal squamous cell carcinoma (OPSCC). However, the prevalence of HPV associated HNSCC is unclear in Asian populations.[6]

HPVs are small, epitheliotropic, double-stranded DNA viruses with a genome comprising approximately 8000 base pairs (bp). The genome is characterized by two late genes (L1 and L2) that encode virus capsid proteins, and seven early genes (E1 to E7) encoding proteins primarily involved in host genome replication and transcription regulation.[7] The proteins encoded by E6 and E7 are the main triggering factors for HPV-16 associated carcinogenesis. These two viral proteins disrupt host cell cycle regulation pathways by degrading tumor suppressor proteins p53 and retinoblastoma.[8]

The detection rate of HR-HPV in HNSCC varies considerably due to differences in the frequency of anatomical site of involvement, geographic locations, age of incidence, and detection methods.[9] HPV infection is an increasingly common risk factor, especially for the OPSCC subset. HPV-related HNSCC exhibits unique immune and genetic characteristics, including sensitivity to chemoradiotherapy and immune checkpoint blockade treatments compared with HPV-negative tumors. Key features include TRAF3 mutations, PD-L1/PD-1 immune resistance markers, tumor microenvironment differences, epigenetic alterations, and genome instability linked to HPV integration. APOBEC gene expression and driver mutations further highlight HPV's role in oncogenesis.[10] However, very few studies have investigated the prevalence of HPV associated HNSCC in Pakistan. Furthermore, no data are available regarding HR-HPV based on whole genome sequencing (WGS) and analysis of mutations/single nucleotide polymorphisms (SNPs) in HNSCC patients from Pakistan. Therefore, the present study aimed to investigate the prevalence of HPV associated HNSCC and the genetic features of HPV-16 in the Southern Punjab region of Pakistan.

Materials and Methods

Study design and participants

The present study was conducted in the Immunology Department and Oral Pathology Department, University of Health Sciences Lahore, Pakistan. Tissue sections were obtained from 85 HNSCC patients after obtaining written informed consent, as well as information about their age, sex, tumor location, histological type, tumor grade, etc. Patients who lacked clinical information or whose paraffin blocks lacked sufficient tumor tissue were excluded from the study. The important baseline demographic, clinical, and histological characteristics are presented in Table 1.

Hematoxylin and Eosin (H&E) staining

To confirm the initial diagnosis, H&E staining was performed for 4 μm sections after cutting from all tumors. Anatomo-pathological variables were obtained after histological slides were reviewed by three histopathologists who specialized in oral histopathology. The study was blinded, so each examiner/histopathologist was unaware of the evolution of the cases analyzed. Tumor grade was reclassified following Anneroth's system of histological grading for

squamous cell carcinoma.[11] The largest subgroup (n = 44, 51.8 %) consisted of moderately differentiated squamous cell carcinoma, followed by well-differentiated squamous cell carcinoma (WDSCC) (n = 23, 27.1 %) and poorly differentiated squamous cell carcinoma (n = 18, 21.2 %).

Genomic DNA extraction from formalin-fixed paraffin-embedded (FFPE) tissue sections

DNA was extracted from FFPE tissue sections using a Gene JET FFPE DNA Purification Kit (catalog number K-0881, Thermo Fisher Scientific, Wilmington, DE, USA), following the manufacturer's instructions. The concentration and quality of the extracted DNA were then measured using a NanoDrop ND-2000c spectrophotometer (Thermo Fisher Scientific, USA).

HPV genotyping

HR-HPV genotyping was conducted sequentially for the precise detection and typing of HR-HPV infections. To screen for the occurrence of HPV, conventional PCR amplification of the universal L1 region (150 bp) was conducted using GP5+ (5′ TTT GTT ACT GTG GTA GAT ACT AC 3′) and GP6+ (5′ GAA AAA TAA ACT GTA AAT CAT ATT C-3′) primers.[12] The 25 μL reaction volume for conventional PCR contained 5 μL of extracted DNA, 1 μL each of the primers (GP5+/GP6), 0.5 μL Taq polymerase (Thermo Fisher Scientific, USA), 1 μL dNTP (25 mM; Thermo Fisher Scientific, USA), 1 μL $MgCl_2$ (25 mM), and 2.5 μL Taq buffer, and the remaining volume was adjusted with injection water. The tubes were closed and processed in a Bio-Rad thermal cycler with the following parameters: 95 °C $^{3:00}$ [94 °C $^{0:15}$; 42 °C $^{0:30}$; 72 °C $^{0:45}$]$_{35}$; 72 °C $^{10:00}$. The amplified product was separated by 2 % agarose gel electrophoresis, where 6 × loading dye (3 μL) was mixed with 5 μL of DNA sample and loaded into the second well and a 100 bp/1 KB DNA ladder (Thermo Fisher Scientific, USA) was added to the first well. This step was conducted to identify a wide range of HPV DNA in numerous samples as an initial evaluation before genotyping. In vitro real-time amplification for the quantitative or qualitative detection and genotyping of HR-HPV types was performed using an HPV Genotypes 14 Real-TM Quant kit (Sacace Biotechnologies, Como, Italy). This real-time PCR (RT-PCR)-based technique enabled the identification of 14 HR-HPV genotypes in genomic DNA samples, including types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68. Each 25 μL RT-PCR reaction mixture contained 15 μL PCR mix (PCR-mix + buffer-FRT) and 10 μL of DNA sample, where controls (negative control DNA, K2 standard) and calibrators (K1 and K2) were included. The temperature profile was generated with a Sacace thermo-cycler (SaCycler-96 Real-Time PCR System) according to the manufacturer's instructions. RT-PCR cycling was conducted at 95 °C for 15 min (initial denaturation), followed by 40 cycles at 95 °C for 5 s, 60 °C for 20 s (fluorescence detection), and 72 °C for 15 s. The fluorophores for detecting HPV genotypes were FAM (16, 39, 33, 58), HEX/JOE (31, 45, 35, 52), ROX (18, 59, 68, 66), and Cy5 (56, 51, internal control). Qualitative analysis identified positives based on fluorescence curves crossing the threshold (Ct values), whereas negative controls produced no signals and positive controls/calibrators generated the expected signals with calibration correlation >0.98.

WGS using next-generation sequencing (NGS)

DNA from all positive HPV-16 samples was quantified by using a Qubit Fluorometric Quantification System (Invitrogen, California, USA), where the DNA quantity and quality were measured using a Multiskan Sky Microplate (Figures S1−S3). One sample of HPV-16 (HNC49, Accession No. OQ 911727) with good purity (260/280 ratio ≈ 1.8) was used for library construction in a DNA-sequencing Microplate. The latest Thermo Fisher Scientific Technology "Ion Torrent" technique was used for the WGS of HPV-16 from genomic DNA extracted from HNSCC patients. A custom panel with ID: IAD186133-197 was utilized for amplifying the whole genome of HPV-16 (Accession no: K02718.1) containing 7906 bp. The process covered 99 % of the HPV-16 genome and generated 47 overlapping amplicons. Using Ion AmpliSeq Designer (https://www.ampliseq.com/, Life Technologies), custom AmpliSeq libraries were designed, which yielded 175−375 amplicons dispersed in two primer pools. Following the manufacturer's instructions, IT AmpliSeq 2.0 kit (Life Technologies, Carlsbad, Canada) libraries were amplified using long reads of DNA per primer pool. Barcoded libraries were quantified using an Ion Library Quantitation Kit (Life Technologies, Carlsbad, Canada), diluted to 100 pM, and pooled together, where this process verified that the sample fulfilled the required sequence coverage depth (30 ×). Next, 25 μL of the pool was used for processing with an Ion Chef Template Kit (Life Technologies). The sample was sequenced on E530 chips with an Ion S5 XL Sequencing Kit (Life Technologies) (Table S1).

Bioinformatics analysis of HPV-16

The quality of raw reads generated by Ion Torrent methods was evaluated using FastQC (Version 0.12.0), before filtering and quality trimming using the Trimmomatic tool (Version v0.39) to obtain clean reads with consistent lengths (Supplemental file S1).[13,14] Clean reads were mapped to the HPV-16 reference genome (AF125673) using the Burrows−Wheeler aligner (BWA Version 0.6.2). Mapped reads were extracted using samtools (Version 1.20) and fed into Velvet assembler (Version 1.2.10) for *de novo* assembly by using the default parameters.[15] SPAdes (Version 4.0.0) assembler was also used with the default parameters for *de novo* assembly.[16] Both assemblies were employed to produce scaffolds in Geneious Prime (version 2024.0), before aligning with the reference genome using MAFFT consensus sequence (Version 7) to construct a complete genome.[17,18] Incorrect bases were corrected manually. The newly constructed genome was annotated in Geneious Prime (version 2024.0). The phylogenetic tree was created in IQ-Tree (Version 1.6.12) using the maximum likelihood method. The phylogenetic tree in Newick format was further improved in TreeDyn (Version 198.3).[19]

Statistical analysis

The Statistical Package for Social Sciences (SPPS Inc., Chicago, USA) 25.0 was utilized to generate and analyze the data. The Chi-square test, Chi-square test for linear trends, and Fisher's exact test were employed to compare percentages. A *p*-value ≤ 0.05 was considered to indicate a statistically significant difference.

Results

Table 1 presents the detailed demographic, clinical, histological, and PCR-related characteristics of HNSCC patients. The overall prevalence of HR-HPV in HNSCC was determined as 7.1 % (n = 6), where 4.7 % (n = 4) of cases were positive for HPV-16 and 2.4 % (n = 2) tested positive for HPV-18. No other HR genotypes were detected in the collected samples. Furthermore, we examined the distributions of the HR-HPV status and genotypes among different clinical variants of HNSCC. Among the six positive HR-HPV cases, five were observed in OPSCC and larynx cancer, but only one case of oral squamous cell carcinoma (OSCC) was HR-HPV positive. A strong statistical association ($p = 0.008$) was found between HR-HPV positivity and histological grading, where 5/6 HR-HPV positive HNSCC cases were at the well differentiated grade and only one was at the moderately differentiated grade. However, no significant statistical associations were found between HPV positivity status and the other baseline parameters.

Table 1: Baseline characteristics of patients with HNSCC and comparison with their HR-HPV positivity status.

Variables	n (%)	*p*-value
HR-HPV positivity status		
Positive cases	6 (7.1 %)	
Negative cases	85 (92.9 %)	
Age group		
≤40 years	16 (18.8 %)	0.887
>40 years	69 (81.2 %)	
Sex		
Males	31 (36.5 %)	0.664
Females	54 (63.5 %)	
Tumor size		
T1	65 (76.5 %)	0.340
T2	16 (18.8 %)	
T3	3 (3.5 %)	
T4	1 (1.2 %)	
Biopsy type		
Incision	59 (69.4 %)	0.881
Excision	25 (30.6 %)	
Site involved		
Oral SCC	35 (41.17 %)	0.139
Oro-pharyngeal SCC	32 (37.64 %)	
Nasopharyngeal SCC	3 (3.5 %)	
Facial skin/neck SCC	13 (15.3 %)	
HNSCC grade		
WDSCC	18 (21.2 %)	**0.008**
MDSCC	44 (51.8 %)	
PDSCC	23 (27 %)	
HNSCC histological subtypes		
Conventional	79 (92.9 %)	0.822
Verrucous	3 (3.5 %)	
Baso-squamous	2 (2.4 %)	
Basaloid	1 (1.2 %)	

HR-HPV, high-risk human papillomavirus; HNSCC, head and neck squamous cell carcinoma; SCC, squamous cell carcinoma; WDSCC, well-differentiated SCC; MDSCC, moderately differentiated SCC; PDSCC, poorly differentiated SCC.

In this study, we successfully assembled one complete genome of HPV-16 (designated as HNC49), which was subsequently submitted to the NCBI genome submission portal GenBank and assigned accession number OQ911727. HNC49 was derived from a 48-year-old male diagnosed with OPSCC in the oropharynx region. The tumor was classified as T1 based on its size and extent. The sample was obtained via excisional biopsy and histopathologically identified as WDSCC of the conventional subtype. PCR analysis confirmed the presence of HPV type 16 and the status of HPV was reported as positive. Table S3 provides an overview of the size of the assembled HPV-16 genome and highlights specific genes.

Figure 1 presents the annotations for the WGS HPV-16 sample (HNC49). All genes within the genome are depicted unidirectionally. We successfully assembled a nearly full-length genome spanning 7171 bp, encompassing all coding regions. Non-coding regions are absent at both ends of the circular genome of HPV-16. In the annotations in Figure 1, coding regions are highlighted by yellow lines and gene regions are represented by green lines. It should be noted that the E4 gene is situated within the E2 gene.

Figure 2 shows the phylogenetic tree obtained in this study. The Pakistani genome (HNC49) assembled in this study clustered together with genomes from Thailand, Germany, and several from the United States. Interestingly, clustering did not have a country-specific pattern and samples from the same countries were distributed across multiple branches of the tree.

We identified 11 mutations/SNPs within the HPV-16 sample sequenced from HNSCC cases, where seven were previously reported mutations/SNPs distributed across the genome and the remaining four mutations/SNPs were novel (https://www.ncbi.nlm.nih.gov/clinvar/?term=Human20Papillomavirus16). Most of the variant regions in the HPV-16 genome were located

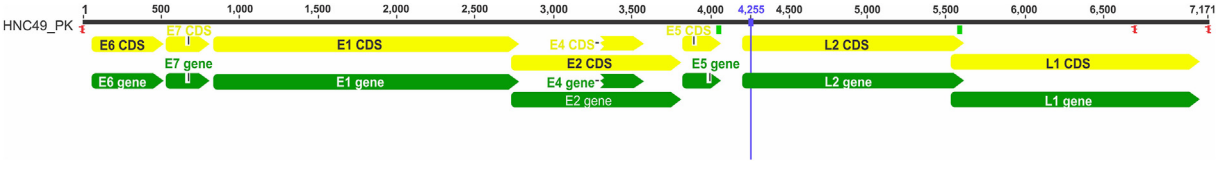

Figure 1: Assembled genome represented by idiograms depicting all genes in a unidirectional manner. The E4 gene is fully contained within the E2 gene.

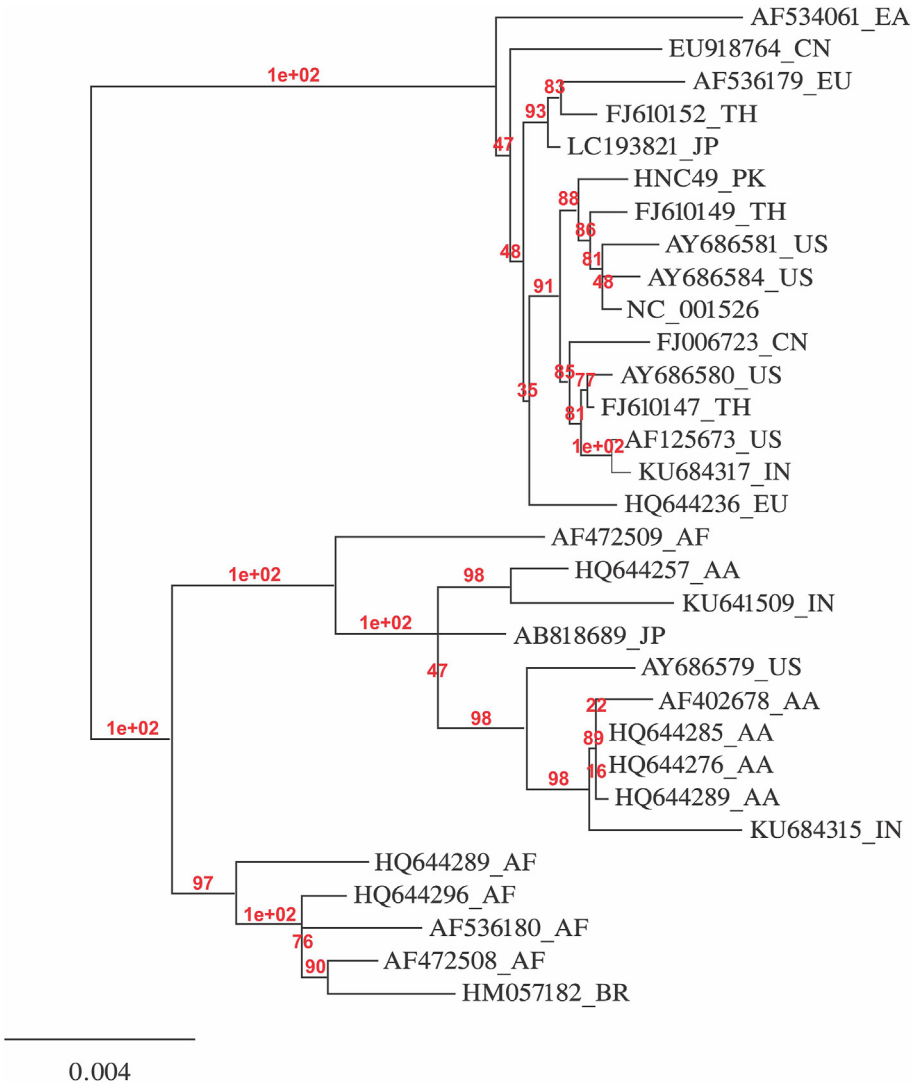

Figure 2: Phylogenetic tree created for the whole genome of HPV-16 based on next generation sequencing (NGS) data. The tree was generated using an HPV-16 reference genome obtained from the National Center for Biotechnology Information (NCBI) database (AF125673). Both the reference sequence and previously reported HPV-16 whole genome sequences are identified by their country of origin and GenBank accession numbers. Details of the abbreviations of countries presented in the figure are given in Table S2.

Table 2: Analysis of HPV-16 genome mutations/SNPs with early and late region coordinate positions, reference alleles, variant alleles, attributes, reference and variant amino acids, and allele sources (novel/reported) in patients with HNSCC.

Coordinate Position	Reference Allele	Variant Allele	Attribute	Reference Amino Acid	Variant Amino Acid	Allele Source
1389 (E1)	T	C	Non-synonymous	Isoleucine	Threonine	**Novel**
1483 (E1)	G	A	Synonymous	Valine	Valine	Previously reported
2191 (E1)	T	C	Synonymous	Phenylalanine	Phenylalanine	**Novel**
3378 (E2)	C	T	Non-synonymous	Proline	Serine	Previously reported
3815 (E2)	T	C	Synonymous	Serine	Serine	**Novel**
4010 (E5)	A	G	Non-synonymous	Isoleucine	Valine	Previously reported
4195	T	C	Non-coding			Previously reported
4905 (L2)	G	A	Synonymous	Glutamine	Glutamine	Previously reported
5193 (L2)	A	C	Non-synonymous	Leucine	Phenylalanine	Previously reported
6401 (L1)	A	G	Non-synonymous	Threonine	Alanine	Previously reported
6958 (L1)	A	G	Synonymous	Leucine	Leucine	**Novel**

HNSCC, head and neck squamous cell carcinoma; SNP, single nucleotide polymorphism; E, early; L, late; A, adenine; T, thymine; C, cytosine; G, guanine.

within the E1 gene, followed by the E2, L2, and L1 genes. Equal numbers of synonymous and non-synonymous mutations/SNPs (n = 5 each) were found in the E1, E2, E5, L2, and L1 regions. Among the novel mutations/SNPs, most were identified in the E1 region (2/4), followed by the E2 region (1/4) and L1 region (1/4). The most prevalent mutations/SNPs were T-C (4/11; 36.3 %), followed by A-G (3/11; 27.2 %) and G-A (2/11; 18.1 %) (Table 2).

Discussion

This study examined various aspects of HNSCC, including the presence and characteristics of HR-HPV genotypes, and molecular and mutational analysis of HPV-16. Among 85 cases of HNSCC, only six (7.1 %) were positive for HR-HPV by RT-PCR. Many studies conducted throughout the world have confirmed that the prevalence of HPV infections in human-related cancers is linked to particular geographic regions. According to a study conducted by Junor et al. at the Edinburgh Cancer Center, the percentage of HNSCC cases that tested positive for HPV was 41 % from 1999 to 2001, but 63 % from 2003 to 2005.[20] The Surveillance, Epidemiology, and End Results (SEER) program reported that from 1988 to 2004, the incidence of cancers not caused by HPV decreased by 50 %, whereas that of OPSCC caused by HPV increased by 225 %.[21]

A study of HNSCC FFPE tissue sections in the Central African Republic found that only one OPSCC biopsy sample was positive for HPV-16. The overall HPV prevalence in the HNSCC biopsy series was 0.74 %.[22] Another study conducted using HNSCC FFPE tissue samples in Bangladesh detected HPV DNA in 36/174 samples (~21 %), where HPV-16 was identified as the most prevalent HR-HPV genotype, accounting for 33 % of cases.[23]

A limited number of local studies have determined the prevalence of HPV in HNSCC. A study conducted by Naqvi et al. using 58 OSCC FFPE tissue samples found that no samples were positive for HPV DNA.[24] However, another study by Awan et al. using 47 OSCC FFPE tissue samples determined that 32 (68.1 %) were positive for HPV DNA.[25] This difference in the prevalence of HPV infection between previous studies and the present study may be attributable to differences among regions because the previous studies used samples from Karachi whereas the samples tested in the present study were collected from the Southern region of Punjab province. Moreover, the previous studies used conventional PCR, whereas we applied the gold standard real-time PCR technique.

It is important to note that the detection of HPV depends on the type of sample, quality of genomic DNA, and identification technique utilized. According to previous studies, the detection rate may be high in blood and saliva because good quality DNA can be extracted from both samples but they are not truly representative media in the case of HNSCC. A tissue biopsy containing cancerous tissue is a true representative of malignancy but the quality of the DNA extracted from FFPE tissue is not good due to the formation of dimers and degradation of the DNA double helix during tissue processing steps. Therefore, the rate of HPV detection may be lower in FFPE tissue samples compared with saliva and blood.[26]

Interestingly, phylogenetic analysis of the HPV-16 genome sequenced in this study showed that it clustered closely with HPV-16 genomes from Thailand and the United States, as well as sharing similarities with genomes from India, China, and Europe (Figure 2). The closeness of the Indian and Pakistani genomes was due to their common ancestral history. A study conducted in China also found that HPV-16 samples predominantly clustered with the Asian lineage, which is consistent with the findings obtained in the present study.[27] The similarities between the Pakistani HPV-16 genome and those from other Asian regions could be explained by the migration of individuals.

No previous locally based studies have conducted phylogenetic analyses of HPV-16 from HNSCC patients. In addition, very few studies have reported phylogenetic analyses of HPV-16 from cervical samples. Aziz et al. conducted a study in Pakistan based on cervical samples and found that 4.74 % of the patients had HPV, where HPV-16 accounted for 4.16 % of the cases. Similar HPV genotypes and close clusters were observed in China, India, Mexico, Iran, Slovenia, and Germany.[28] Based on these observations, it was concluded that the prevalence of HR-HPV types is low and that the L1 gene sequences are highly similar. However, another study by Abdullah et al. based on 100 cervical samples found that HPV-16 samples (n = 2) clustered closely with those from Japan and Costa Rica.[29]

Many studies have classified HPV-16 sublineages by sequencing the E6, E7, and long control region (LCR) regions, and most of them included sufficient information on the relevant SNPs required to accurately assign them to appropriate sublineages.[30] Hence, sequencing the entire viral genome is the most appropriate approach for identifying the sublineage classification that is most closely related.

In the present study, four novel mutations/SNPs were detected and they were located in the E1, E2, and L1 regions of the HPV-16 genome. These regions are known to be the targets of neutralizing antibodies and they play roles in viral replication. Moreover, some of these mutations/SNPs may not result in changes in the amino acid sequences of viral proteins, whereas others may lead to a single amino acid changing. For example, the conversion of a T allele into a C allele at position 1389 in the E1 region may result in the substitution of one amino acid with another in the E1 protein. These mutations/SNPs may indicate the development of abnormal viral clones, which could have implications for the pathogenesis and/or transmission of the virus. Further studies are needed to determine the functional significance of the mutations/SNPs identified in the present study. Pathogenic variations in the HPV-16 genome may lead to changes in amino acid sequences, which could affect the function of each viral subunit. For instance, changes in L1 could impact the viral efficiency of infection or subsequent antigenicity. However, the presence of several variations in multiple targets is necessary to alter the overall pathophysiology of the disease, and it cannot be concluded that changes in the biological functions or mode of action of HPV may occur due to these specific mutations/SNPs.[31] Studies have investigated the impacts of different polymorphic sites on cancer development and the results have been conflicting.

Increased diversity was also indicated by their geographic and sequence dissimilarity.

Data obtained from research and clinical studies highlight the importance of continuing to investigate the underlying mechanisms that lead to the development of HPV associated cancer. In order to improve early detection, risk stratification, and treatment outcomes for patients with HPV associated HNSCC, it is essential to obtain a better understanding of the mechanisms involved in HPV associated carcinogenesis, such as viral oncogenes, epigenetic modifications, and immune system evasion. Given the inconsistent data available regarding the prevalence of HPV-related HNSCC in Pakistan, we recommend conducting further studies with larger sample sizes using molecular methods for DNA extraction and HPV detection in HNSCC patients, instead of relying solely on p16 immunohistochemistry, which is only a cell cycle marker. Directly detecting HPV through RT-PCR is more sensitive and specific. In Pakistan, the results obtained in the present study could help to raise public health awareness, and facilitate the development of HPV screening and future vaccination programs targeting specific HR-HPV genotypes to reduce the burden of HPV related HNSCC.

The present study had several limitations. First, the small sample size (85 tissue samples) may limit the generalizability of our findings, and the cross-sectional design provides only a snapshot of HPV prevalence without long-term insights. In addition, the present study was limited to Southern Punjab, which may not reflect the prevalence of HPV in other regions. Using FFPE tissue samples could have affected the detection of HPV due to the possible degradation of DNA, and focusing primarily on HPV-16 and HPV-18 excluded other HR-HPV types. Moreover, the selection of the HPV sample for NGS/phylogenetic analyses based only on the DNA quality was a weakness of this study because the genetic constitutions (sequence variants) of other sample could have been dissimilar. Furthermore, the absence of clinical outcome data limited our understanding of the prognostic significance of HPV in HNSCC patients. Finally, the novel mutations/SNPs in HPV-16 identified in the present study require further investigation to assess their functional relevance.

Conclusion

Very few cases of HR-HPV associated HNSCC were identified in this study and HPV-16 was the most common genotype. Phylogenetic analysis showed that the HPV-16 genome isolated and sequenced in this study had a distinct genetic structure as well as sharing similarities with genomes reported from Thailand, the United States, India, China, and Europe. These findings enhance our knowledge of the epidemiological and biological impacts of genetic variations in HPV, which could be valuable for preventing and treating HR-HPV-associated HNSCC.

Source of funding

University of Health Sciences Lahore provided necessary funds and logistics to conduct this research (No. UHS/Edu/126-16/1226).

Conflict of interest

The authors have no conflict of interest to declare.

Ethical approval

Approval for this research was granted by the Ethical Review Committee and Advanced Studies & Research Board of the University of Health Sciences Lahore (Approval No. UHS/Edu/126-16/1226). The research was conducted in compliance with the principles outlined in the Declaration of Helsinki.

Authors contributions

Conceptualization, M.K., S.M., S.J., and N.A.; methodology, M.K., S.M., F.S., R.T., A.A., M.I., and N.A; software, M.K., A.A., S.M., and F.S.; validation, M.K., A.A., R.T., S.J., M.I., and N.A.; formal analysis, M.K, F.S., A.A., and S.M.; investigation, M.K., S.M., A.A., and S.J.; resources, N.A., S.J., M.I., R.T., and M.K.; data curation, M.K. and F.S.; writing—original draft preparation, M.K.; writing—review and editing, M.K., S.M., S.J., R.T., F.S., M.I., A.A., and N.A. All authors have critically reviewed and approved the final draft and are responsible for the content and similarity index of the manuscript.

Acknowledgment

We would like to express our gratitude to the laboratory technicians and assistants in the Immunology Department, Morbid Anatomy & Histopathology Department, and Oral Pathology Department at UHS Lahore, as well as the GENOME laboratory, for their technical support. In addition, we would like to thank the oral and maxillofacial surgeons at government hospitals for their help recruiting participants for our study.

Data availability statement

The whole genome sequence of the HPV-16 isolate is available via the NCBI genome submission portal GenBank under accession number OQ911727). No other data are associated with this article.

NCBI GenBank: Human papillomavirus type 16 isolate HNC49, complete genome [HPV16]. Accession number OQ911727.

Extended data are available in the supplementary file.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jtumed.2025.03.001.

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022:

GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **CA Cancer J Clin 2024 May-Jun**; 74(3): 229–263. https://doi.org/10.3322/caac.21834. Epub 2024 Apr 4. PMID: 38572751.
2. Leemans CR, Braakhuis BJ, Brakenhoff RH. The molecular biology of head and neck cancer. **Nat Rev Cancer 2011 Jan**; 11(1): 9–22. https://doi.org/10.1038/nrc2982. Epub 2010 Dec 16. PMID: 21160525.
3. Deschler DG, Richmon JD, Khariwala SS, Ferris RL, Wang MB. The "new" head and neck cancer patient-young, nonsmoker, nondrinker, and HPV positive: evaluation. **Otolaryngol Head Neck Surg 2014 Sep**; 151(3): 375–380. https://doi.org/10.1177/0194599814538605. Epub 2014 Jun 12. PMID: 24925311; PMCID: PMC4894657.
4. Muñoz N, Bosch FX, de Sanjosé S, Herrero R, Castellsagué X, Shah KV, et al. International agency for research on cancer multicenter cervical cancer study group. Epidemiologic classification of human papillomavirus types associated with cervical cancer. **N Engl J Med 2003 Feb 6**; 348(6): 518–527. https://doi.org/10.1056/NEJMoa021641. PMID: 12571259.
5. IARC. *IARC monographs on the evaluation of carcinogenic risks to humans*; 2007. V 90. França. [cited on 2023 April 1]. Available from: https://monographs.iarc.who.int/.
6. Chaudhary AK, Singh M, Sundaram S, Mehrotra R. Role of human papillomavirus and its detection in potentially malignant and malignant head and neck lesions: updated review. **Head Neck Oncol 2009 Jun 25**; 1: 22. https://doi.org/10.1186/1758-3284-1-22. PMID: 19555477; PMCID: PMC2706235.
7. Coca-Pelaz A, Rodrigo JP, Suárez C, Nixon IJ, Mäkitie A, Sanabria A, et al. The risk of second primary tumors in head and neck cancer: a systematic review. **Head Neck 2020 Mar**; 42(3): 456–466. https://doi.org/10.1002/hed.26016. Epub 2019 Nov 21. PMID: 31750595.
8. Westra WH, Taube JM, Poeta ML, Begum S, Sidransky D, Koch WM. Inverse relationship between human papillomavirus-16 infection and disruptive p53 gene mutations in squamous cell carcinoma of the head and neck. **Clin Cancer Res 2008 Jan 15**; 14(2): 366–369. https://doi.org/10.1158/1078-0432.CCR-07-1402. PMID: 18223210.
9. Kreimer AR, Clifford GM, Boyle P, Franceschi S. Human papillomavirus types in head and neck squamous cell carcinomas worldwide: a systematic review. **Cancer Epidemiol Biomarkers Prev 2005 Feb**; 14(2): 467–475. https://doi.org/10.1158/1055-9965.EPI-04-0551. PMID: 15734974.
10. Dong H, Shu X, Xu Q, Zhu C, Kaufmann AM, Zheng ZM, et al. Current status of human papillomavirus-related head and neck cancer: from viral genome to patient care. **Virol Sin 2021 Dec**; 36(6): 1284–1302. https://doi.org/10.1007/s12250-021-00413-8. Epub 2021 Jun 21. PMID: 34152564; PMCID: PMC8692589.
11. Akhter M, Hossain S, Rahman QB, Molla MR. A study on histological grading of oral squamous cell carcinoma and its co-relationship with regional metastasis. **J Oral Maxillofac Pathol 2011 May**; 15(2): 168–176. https://doi.org/10.4103/0973-029X.84485. PMID: 22529575; PMCID: PMC3329698.
12. de Roda Husman AM, Walboomers JM, van den Brule AJ, Meijer CJ, Snijders PJ. The use of general primers GP5 and GP6 elongated at their 3' ends with adjacent highly conserved sequences improves human papillomavirus detection by PCR. **J Gen Virol 1995 Apr**; 76(Pt 4): 1057–1062. https://doi.org/10.1099/0022-1317-76-4-1057. PMID: 9049358.
13. Andrews S, Babraham Bioinformatics. FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ **2010**. [Accessed 12 February 2025].
14. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics 2014 Aug 1**; 30(15): 2114–2120. https://doi.org/10.1093/bioinformatics/btu170. Epub 2014 Apr 1. PMID: 24695404; PMCID: PMC4103590.
15. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. **Genome Res 2008 May**; 18(5): 821–829. https://doi.org/10.1101/gr.074492.107. Epub 2008 Mar 18. PMID: 18349386; PMCID: PMC2336801.
16. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. **J Comput Biol 2012 May**; 19(5): 455–477. https://doi.org/10.1089/cmb.2012.0021. Epub 2012 Apr 16. PMID: 22506599; PMCID: PMC3342519.
17. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. **Bioinformatics 2012 Jun 15**; 28(12): 1647–1649. https://doi.org/10.1093/bioinformatics/bts199. Epub 2012 Apr 27. PMID: 22543367; PMCID: PMC3371832.
18. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. **Mol Biol Evol 2013 Apr**; 30(4): 772–780. https://doi.org/10.1093/molbev/mst010. Epub 2013 Jan 16. PMID: 23329690; PMCID: PMC3603318.
19. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. **Mol Biol Evol 2015 Jan**; 32(1): 268–274. https://doi.org/10.1093/molbev/msu300. Epub 2014 Nov 3. PMID: 25371430; PMCID: PMC4271533.
20. Junor E, Kerr G, Oniscu A, Campbell S, Kouzeli I, Gourley C, et al. Benefit of chemotherapy as part of treatment for HPV DNA-positive but p16-negative squamous cell carcinoma of the oropharynx. **Br J Cancer 2012 Jan 17**; 106(2): 358–365. https://doi.org/10.1038/bjc.2011.542. Epub 2011 Dec 6. PMID: 22146523; PMCID: PMC3261669.
21. Pytynia KB, Dahlstrom KR, Sturgis EM. Epidemiology of HPV-associated oropharyngeal cancer. **Oral Oncol 2014 May**; 50(5): 380–386. https://doi.org/10.1016/j.oraloncology.2013.12.019. Epub 2014 Jan 22. PMID: 24461628; PMCID: PMC4444216.
22. Kofi B, Mossoro-Kpinde CD, Mboumba Bouassa RS, Péré H, Robin L, Gresenguet G, et al. Infrequent detection of human papillomavirus infection in head and neck cancers in the Central African Republic: a retrospective study. **Infect Agents Cancer 2019 Apr 8**; 14: 9. https://doi.org/10.1186/s13027-019-0225-x. PMID: 31007710; PMCID: PMC6454762.
23. Shaikh MH, Khan AI, Sadat A, Chowdhury AH, Jinnah SA, Gopalan V, et al. Prevalence and types of high-risk human papillomaviruses in head and neck cancers from Bangladesh. **BMC Cancer 2017 Nov 25**; 17(1): 792. https://doi.org/10.1186/s12885-017-3789-0. PMID: 29178862; PMCID: PMC5702125.
24. Naqvi SU, Khan S, Ahmed A, Lail A, Gul S, Ahmed S. Prevalence of EBV, CMV, and HPV in oral squamous cell carcinoma patients in the Pakistani population. **J Med Virol 2020 Dec**; 92(12): 3880–3883. https://doi.org/10.1002/jmv.25791. Epub 2020 Apr 10. PMID: 32219873.
25. Awan MS, Irfan B, Zahid I, Mirza Y, Ali SA. Comparison of polymerase chain reaction and immunohistochemistry assays for analysing human papillomavirus infection in oral squamous cell carcinoma. **J Clin Diagn Res 2017 Jun**; 11(6): XC10–XC13. https://doi.org/10.7860/JCDR/2017/24742.10119. Epub 2017 Jun 1. PMID: 28764278; PMCID: PMC5535468.
26. Einaga N, Yoshida A, Noda H, Suemitsu M, Nakayama Y, Sakurada A, et al. Assessment of the quality of DNA from various formalin-fixed paraffin-embedded (FFPE) tissues and the use of this DNA for next-generation sequencing (NGS) with no artifactual mutation. **PLoS One 2017 May 12**; 12(5):e0176280. https://doi.org/10.1371/journal.pone.0176280. PMID: 28498833; PMCID: PMC5428915.

27. 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 Mar 21**; 7(1): 263. https://doi.org/10.1038/s41598-017-00364-9. PMID: 28325903; PMCID: PMC5428204.
28. Aziz H, Iqbal H, Mahmood H, Fatima S, Faheem M, Sattar AA, et al. Human papillomavirus infection in females with normal cervical cytology: genotyping and phylogenetic analysis among women in Punjab, Pakistan. **Int J Infect Dis 2018 Jan**; 66: 83–89. https://doi.org/10.1016/j.ijid.2017.11.009. Epub 2017 Nov 11. PMID: 29138009.
29. Abdullah A, Qasim M, Shafiq M, Ijaz M, Parveen S, Murtaza S, et al. Molecular diagnosis and phylogenetic analysis of human papillomavirus type-16 from suspected patients in Pakistan. **Infect Agents Cancer 2016 Jan 12**; 11: 1. https://doi.org/10.1186/s13027-016-0047-z. PMID: 26766963; PMCID: PMC4711104.
30. 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 Apr 29**; 108(9):djw100. https://doi.org/10.1093/jnci/djw100. PMID: 27130930; PMCID: PMC5939630.
31. Schiffman M, Herrero R, Desalle R, Hildesheim A, Wacholder S, Rodriguez AC, et al. The carcinogenicity of human papillomavirus types reflects viral evolution. **Virology 2005 Jun 20**; 337(1): 76–84. https://doi.org/10.1016/j.virol.2005.04.002. PMID: 15914222.

How to cite this article: Kashif M, Minhas S, Jahan S, Shahzad F, Tahir R, Abbas A, Idrees M, Afzal N. Exploring HPV-linked head and neck cancer in Southern Punjab, Pakistan: Insights from HPV-16 phylogenetic analysis. J Taibah Univ Med Sc 2025;20(2):242–250.