Molecular and Phylogenetic Analysis of Human Papillomavirus Using L1 Gene in Oral Squamous Cell Carcinoma Patients in Baghdad, Iraq

Maysaa Kadhim Al-Malkey1*, Zainab Abdul Jabbar Aldhaher2, Rasha Abbas Azeez2, Sumaiah Ibrahim Hussein2, Sinai Waleed Mohammed1, Munira Ch Ismeeal3 and Khawla Ibrahim Mesheal1

1Tropical-Biological Research Unit, College of Science, University of Baghdad, Baghdad-Iraq.
2Department of Basic Sciences, College of Dentistry, University of Baghdad, Baghdad-Iraq.
3Department of Microbiology, College of Science, Al-Karhk University of Science, Baghdad-Iraq.
*Coresponding author E-mail: maysakadhim@uobaghdad.edu.iq

http://dx.doi.org/10.13005/bpj/1426

(Received: 21 February 2018; accepted: 28 April 2018)

Oral squamous cell carcinoma (OSCC) is the most common malignant neoplasm of the oral mucosa. Human papillomavirus (HPV) virus cause a broad scope of diseases from benign to invasive tumors, types 16 and 18 classified as carcinogenic to humans. This study aimed to provide the first molecular characterization of HPV types in Iraq. Thirty-five unstimulated whole saliva samples were collected from histopathologically confirmed patients with oral cancer were enrolled in this study. Genomic DNA was extracted from exfoliating cells to amplify HPV-DNA using HPV-L1 gene sequence primers by polymerase chain reaction method (PCR), the viral genotyping was performed using direct sequencing method. HPV genotypes identified were deposited in GenBank. HPV DNA was detected in 20 of 35 OSCC patients representing (57%). The most frequent HPV genotypes were HPV-18 accounting for (75%) (15 out of 20 patients) followed by HPV-16 accounting for (20%) (4 out of 20), and HPV-11 accounting for (5%) (5 out of 20 patients). This study highlights the high-risk HPV genotypes in OSCC patients and their phylogenetic analysis tree and their homology to the ancestral sequence which may indicate emerging of a new biological entity of HPV-positive OSCC with a potential sexually transmission.

Keywords: OSCC; HPV; L1 gene; phylogenetic tree.

Human papillomavirus (HPV) is a circular double-stranded DNA virus, its viral genome composed of three parts: proteins necessary for viral replication and transcription are encoded at the early (E) region; the structural proteins of the viral capsid (L1 and L2) are encoded at the late (L) region; the elements that regulate the viral deoxyribonucleic acid (DNA) replication and transcription are content of the non-coding region [1].

The L1 gene, which encodes the major capsid protein is conserved, which make it particularly suitable for to be used as a family taxonomic criterion [2]. Establishing a distinct HPV “type” is selected when there is at least 10% difference between the DNA sequence of the L1 open reading frame (ORF) and any other characterized type in the cloned viral genome; meanwhile, “subtypes” are referred to the isolates of the same HPV type when the nucleotide
sequences of the L1 ORF differ by 2-10%. When the nucleotide sequences of their L1 ORF differ by less than 2%, referred as a “variants” [3, 4]. Oral cancer (OC) accounted for 300,000 cases and 45,000 deaths occurred worldwide, of which 77% were in the less developed regions [5]. In United State, approximately 7% of the population will have an oral or oropharyngeal cancer which is generally caused by HPV infection at any given time [6]. The estimated of sexually active adults is 65%–100%, which have been exposed to HPV at any anatomic site such as oral, genital, or anal, the exposure rate of oral HPV during a lifetime is unknown [7, 8]. The literature had been firmly established etiological agents of anogenital carcinomas with the high-risk human papillomavirus type 16 (HPV-16) and type 18; the morphological similarities, epitheliotropic homing of HPV and its oncogenic potential encouraged the logical assumption of the association between OC and HPV infection [9].

HPV-16 as an independent risk factor in 30%-50% of oral squamous cell carcinoma (OSCC) has been assigned according to International Agency for Research on Cancer (IARC) since 2007 [10]. The most common genotype in almost 90% of the HPV positive oropharyngeal cancers (OPC) is HPV-16 [11].

The aim of the current study is to detect the HPV genotypes using the L1 gene obtained from OSCC patients and construct a phylogenetic tree.

MATERIALS AND METHODS

Patients

The current study was approved and underwent to the terms of ethical considerations of the Iraqi Ministry of Health. Thirty-five newly diagnosed patients were histopathologically confirmed with OC by two independent pathologists (24 patients were grade I tumors, 4 patients were grade II tumors, and 7 patients were grade III tumors); these patients were attended to maxillofacial surgery clinic of Ghazi Al-Hariri for Specialized Surgery Hospital in Baghdad, Iraq which was enrolled in this study during the period from April 2014 to Jun 2015. An informed consent was taken from each patient. The inclusion criteria for this study were: a) presence of oral cavity cancer (including oral tongue, the floor of mouth, gingiva, lips, buccal mucosa); b) no previous head and neck cancer; c) no prior oncological therapy. The exclusion criteria were: a) patients with a history of autoimmune disorders or systemic diseases; b) previous tumor resection.

Saliva Samples Processing

Unstimulated whole saliva sample up to 5 mL was taken from each patient, which was collected in a 50 mL centrifuge tube and remains on ice whole time during collection. The samples were centrifuged at 2,500 rpm for 15 min at 4 °C to spin down exfoliated cells, the saliva supernatant was discarded. Cell pellets were stored at -80 °C until further processing [12].

DNA Extraction and PCR Analysis

AccuPrep® Genomic DNA extraction kit (Cat# K-3032, Bioneer, Korea) for viral DNA extraction from 200 μl of saliva samples (cell pellet) was used according to the manufacturer’s instructions. HPV-DNA was amplified by conventional PCR assay using specifically designed primer set Forward 5′-ACTGGAAAGGTGCTTGTACC-3′ and Reverse 5′- ACAGGGTTCACAGCCAACAA-3′, to obtain amplicon size (321bp) to target the conserved region according to GenBank reference for alignment of the partial sequence of HPV-L1 gene (GenBank: JX316023.1) using National Center for Biotechnology Information (NCBI) website via Basic Local Alignment Search Tool (BLAST) as previously demonstrated by Agoston et al. (2010) [13]. The amplification was performed using AccuPower® PCR PreMix Kit (Cat# K-2012, Bioneer, Korea) to prepare master mix according to manufacturer’s instructions as follows: 5 μl of template DNA, 1.5 μl of (10 pmol from both primers), and 12 μl of PCR water. The 20 μl reactions were incubated in Thermocycler (MyGene, Korea). Thermocycler condition is listed in Table 1. As a negative control, all the components of the mixture except the DNA template was used. Amplification products of viral DNA were analyzed in 2% polyacrylamide gel electrophoresis containing ethidium bromide, which was placed in TBE buffer and runs at a constant voltage (100V) for approximately 35min. A UV transilluminator was used to visualize the DNA (VISION Gel Documentation System, SciPlas, UK). A digital image was captured for the result documentation.
Sequencing and Phylogenetic Analysis
The amplification products were purified by EZ-10 Spin Column DNA Gel Extraction Kit (Cat#BS353, Biobasic, Canada) following the manufacturer’s instructions. HPV genotyping based on Sanger sequencing PCR fragments using Applied Biosystems DNA sequencing analysis was performed in Bioneer Corporation, Korea. The products were deposited in GenBank database and analyzed via BLAST search (http://blast.ncbi.nlm.nih.gov/) with Popset submission (94454367) under the following accession numbers (KT365828–KT365847). Phylogenetic analysis was performed via the neighbor-joining method using the Molecular Evolutionary Genetics Analysis (MEGA) software version 6.0 to identify the evolutionary relationships among the analyzed sequences and a phylogenetic tree was constructed [14].

Protein Structure
To study the protein structures corresponding to the sequenced genes, we used the I-TASSER server, which is an Internet-based software product that enables protein structure and function predictions. I-TASSER allows automated generation of high-quality predictions of the 3D structures and biological functions of protein molecules based on their amino acid sequence [15, 16].

RESULTS

Patients
In the current study 35 newly diagnosed patients (24 males and 11 females). The major histological subtype was squamous cell carcinoma (SCC). This subtype was seen in 26 out of 35 patients accounting for (74%). Adenocarcinoma accounted for (14%) was seen in 5 out of 35 patients and (12%) were other histological subtypes. Grade I Tumors (69%) (24 out of 35). The most common tumor location was the tongue (40%) (14 out of 35) followed by mouth floor (14%) (5 out of 35) and finally, other tumor locations such as maxilla, cheeks were (46%) (16 out of 35).

PCR and Sequencing Analysis
HPV-1 DNA was detected in 20 out of 35 collected saliva samples representing (57%). These HPV-positive samples produced DNA fragments of 321 bp in length for L1 gene, which was amplified using HPV specific primers for partial L1 (GenBank: JX316023.1) (Figure 1). Direct sequencing results from L1 amplicons were analyzed using the BLAST on the NCBI website to

<table>
<thead>
<tr>
<th>Step</th>
<th>Temperature</th>
<th>Duration</th>
<th>Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial denaturation</td>
<td>95 °C</td>
<td>5 minutes</td>
<td>1</td>
</tr>
<tr>
<td>Denaturation</td>
<td>95 °C</td>
<td>30 seconds</td>
<td>30</td>
</tr>
<tr>
<td>Annealing</td>
<td>58°C</td>
<td>30 seconds</td>
<td></td>
</tr>
<tr>
<td>Extension</td>
<td>72 °C</td>
<td>30 seconds</td>
<td></td>
</tr>
<tr>
<td>Final extension</td>
<td>72 °C</td>
<td>5 minutes</td>
<td>1</td>
</tr>
</tbody>
</table>

Fig. 1. PCR product analysis of L1 gene in HPV-positive samples. Where M: marker (2000-100bp), lane (1-10) positive L1 gene at (321bp) PCR product.
confirm HPV-18, HPV-16, and HPV-11 homology. PCR and sequencing analysis revealed that 75% (15 of 20 HPV-positive DNA) were HPV-18 followed by HPV-16 recorded 20% (4 out of 20) and finally 5% (only one out of 20 HPV-positive DNA) was HPV-11.

Phylogenetic Analysis

A phylogenetic tree was generated using retrieved genome sequences that were deposited under accession number (KT365828–KT365847) and analyzed against 7 high risk and low risk HPV genotype references sequences (HPV type 18: KF225496.1; HPV type 16: AF140365.1; HPV type 11: AF217526.1; HPV type 33: GQ479017.1; HPV type 40: HE793060.1; HPV type 58: HM6397161; HPV type 5: AM922325.1); the local HPV isolates in this study were referred as HPV-1, HPV-2, ……, and HPV-20. HPV-18 was the common genotype identified (15 out of 20 samples) representing 75%. The aligned sequences showed high homology to the nucleotide sequence of HPV type 18 (accession number KF225496.1) (Figure 2). HPV-16 was identified in 4 samples representing (20%), the aligned sequences showed high homology to the nucleotide sequence of HPV type 16 (accession number AF140365.1) (Figure 3). HPV-11 was identified in only one sample, the aligned sequences showed high homology to the nucleotide sequence of HPV type 11 (accession number AF217526.1) (Figure 4), the detailed of human papillomavirus local isolates homology to the nucleotide sequence of references from GenBank database are shown in Table 2.

![Phylogenetic tree analysis](image)

**Fig. 2.** Phylogenetic tree analysis based on the capsid protein (L1) gene partial sequence that used for local Human papillomavirus typing detection. The phylogenetic tree was constructed using Unweighted Pair Group method with Arithmetic Mean (UPGMA tree) in (MEGA 6.0 version). The local Human papillomavirus isolates (HPV1, HPV2, HPV3, HPV5, HPV6, HPV7, HPV11, HPV12, HPV13, HPV14, HPV15, HPV17, HPV18, HPV19, and HPV20) were show closed related to NCBI-Blast Human papillomavirus type 18 (KF225496.1), and whereas other NCBI-Blast Human papillomavirus were show out off tree.
The NCBI–BLAST genotypes identity with this study local isolates are revealed in Table 2. The HPV-18 identity match was found with HPV-18 L1 sequence (accession number KF225496.1) isolated from an Indonesian patient infected with cervical cancer. HPV-16 identity match was found with HPV-16 L1 sequence (accession number AF140365.1) isolated from a Chinese patient infected with Condyloma acuminatum. The HPV-11 identity match was found with HPV-11 L1 sequence (accession number AF217526.1) isolated from Chinese patients.

**Protein structure**

The protein structure of the L1 gene that isolated from Iraqi HPV-18, HPV-16 was analyzed using the I-TASSER server based on their amino acid sequences as shown in Figure 5 (a, b). Based on this analysis, a ligand-binding site was predicted on L1 that facilitates its binding to heparan sulfate proteoglycan (HSPG) as shown in Figure 5 (c, d).

**DISCUSSION**

According to Iraqi cancer registry [17], approximately 4.5% of all cancer cases are oral...
cancer and about 91.5% of OC are OSCC. In this study, SCC is the most common histological malignant type found in the mouth which agrees with results by (Museedi and Younis, 2014) [18] which reported that 91% of OC were SCC; also agrees with earlier reports [19, 20–27]. Nevertheless, disagrees with Al-Kawaz, (2010) which in his study on the prevalence of OC during 2003-2006 in Iraqi governorates in relation to gender, age, and site found that the OSCC was (41.08%) in Baghdad [28]. The findings of the current study for cancer staging have disagreed with a recent study by Le Campion et al., (2017) in a Brazilian cohort which, reported that (85.1%) were in an advanced stage [29]. Unfortunately, a high percentage of oral cancers are only discovered when they become symptomatic, at which point they may be at an advanced stage [21, 30–34].

Tongue was the primary tumor location in this study which, agrees with Al-Kawaz, (2010) [28], the highest prevalence of tumor location was observed in tongue (55.81%), while the lowest (3.87%) was found in the floor of the mouth also agree with a study by Al Jaber et al., (2016) [35]. The primary location of OC is an important prognostic factor because the affected anatomic area can determine the accessibility and extension of surgery [36].

The papillomavirus major capsid protein, L1, is a ~55 kD protein with the ability to spontaneously self-assemble into virus-like particles (VLPs) which, are potent immunogens, likely due to innate B-cell recognition of the regular icosahedrally displayed spacing of surface epitopes (Bachmann et al., 1993; von Bubnoff, 2012) [37–38]. The conserved and structural features of L1 make it a good model for viral taxonomy. To the best of our knowledge, the current study is the first in Iraq concerning detection of HPV from saliva exfoliates cells by identification of viral DNA with sets of primers based on HPV partial capsid protein (L1) gene sequence, then genotyping was performed using direct sequencing method.

The phylogenetic analysis of the current study shows that most of the isolates were HPV-18 representing (75%), a 100% identity match was

<table>
<thead>
<tr>
<th>No.</th>
<th>Local HPV isolate No.</th>
<th>Accession number (deposited to GenBank)</th>
<th>NCBI – BLAST Genotypes Identity (%)</th>
<th>Identification HPV Genotype (current study)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>HPV-1</td>
<td>KT365828.1</td>
<td>99%</td>
<td>Genotype 18</td>
</tr>
<tr>
<td>2</td>
<td>HPV-2</td>
<td>KT365829.1</td>
<td>100%</td>
<td>Genotype 18</td>
</tr>
<tr>
<td>3</td>
<td>HPV-3</td>
<td>KT365830.1</td>
<td>100%</td>
<td>Genotype 18</td>
</tr>
<tr>
<td>4</td>
<td>HPV-4</td>
<td>KT365831.1</td>
<td>100%</td>
<td>Genotype 18</td>
</tr>
<tr>
<td>5</td>
<td>HPV-5</td>
<td>KT365832.1</td>
<td>99%</td>
<td>Genotype 16</td>
</tr>
<tr>
<td>6</td>
<td>HPV-6</td>
<td>KT365833.1</td>
<td>100%</td>
<td>Genotype 18</td>
</tr>
<tr>
<td>7</td>
<td>HPV-7</td>
<td>KT365834.1</td>
<td>100%</td>
<td>Genotype 18</td>
</tr>
<tr>
<td>8</td>
<td>HPV-8</td>
<td>KT365835.1</td>
<td>99%</td>
<td>Genotype 16</td>
</tr>
<tr>
<td>9</td>
<td>HPV-9</td>
<td>KT365836.1</td>
<td>98%</td>
<td>Genotype 16</td>
</tr>
<tr>
<td>10</td>
<td>HPV-10</td>
<td>KT365837.1</td>
<td>100%</td>
<td>Genotype 11</td>
</tr>
<tr>
<td>11</td>
<td>HPV-11</td>
<td>KT365838.1</td>
<td>99%</td>
<td>Genotype 18</td>
</tr>
<tr>
<td>12</td>
<td>HPV-12</td>
<td>KT365839.1</td>
<td>100%</td>
<td>Genotype 18</td>
</tr>
<tr>
<td>13</td>
<td>HPV-13</td>
<td>KT365840.1</td>
<td>100%</td>
<td>Genotype 18</td>
</tr>
<tr>
<td>14</td>
<td>HPV-14</td>
<td>KT365841.1</td>
<td>100%</td>
<td>Genotype 18</td>
</tr>
<tr>
<td>15</td>
<td>HPV-15</td>
<td>KT365842.1</td>
<td>100%</td>
<td>Genotype 18</td>
</tr>
<tr>
<td>16</td>
<td>HPV-16</td>
<td>KT365843.1</td>
<td>98%</td>
<td>Genotype 16</td>
</tr>
<tr>
<td>17</td>
<td>HPV-17</td>
<td>KT365844.1</td>
<td>100%</td>
<td>Genotype 18</td>
</tr>
<tr>
<td>18</td>
<td>HPV-18</td>
<td>KT365845.1</td>
<td>100%</td>
<td>Genotype 18</td>
</tr>
<tr>
<td>19</td>
<td>HPV-19</td>
<td>KT365846.1</td>
<td>100%</td>
<td>Genotype 18</td>
</tr>
<tr>
<td>20</td>
<td>HPV-20</td>
<td>KT365847.1</td>
<td>100%</td>
<td>Genotype 18</td>
</tr>
</tbody>
</table>
found with an HPV-18 sequence isolated from an Indonesian patient with (accession number KF225496.1), followed by HPV-16 representing (20%) from a Chinese patients with (accession number AF140365.1). The results of current study disagree with an Indian study by Kabekkodu \textit{et al}, (2015) and with a study in Argentina by Chiesa \textit{et al}, (2016), in that HPV-16 was the most common genotype[39–40]. Only a few studies indicated that HPV-18 was the commonest type as in Iranian study by Kermani \textit{et al}, (2012) [41].

Papillomavirus capsid initial interaction with the host is largely attributable to L1 interactions with heparan sulfate (HS) carbohydrates displayed on proteoglycans, which is supported by virion binding inhibition and infectious entry by soluble heparin (a highly sulfated form of HS) or by enzymatic removal of HS with heparinase both \textit{in vivo}, utilizing a murine vaginal challenge model as well as \textit{in vitro} with cultured cell lines [42–43]. \textit{In vivo}, the heparan sulfate proteoglycan (HSPG) interaction occurs on the extracellular basement membrane, whereas \textit{in vitro} HPV interact with HSPG on the cell surface and, then to the extracellular matrix (ECM) at a lesser degree [44]. Complete dependence upon HSPG binding was found in vivo studies [43]. As a result, it appears that all papillomaviruses require initial attachment to HSPGs for successful infection in vivo [45].

**CONCLUSION**

The genomic characterization of HPV genotypes is pivotal for a deeper understanding of the biological differences of these viruses. There are certain molecular profiles of HPV-positive OSCCs that differentiate them from HPV-negative cancers, as a result, they could be considered as a biological entity separated from the rest of head and neck cancers, and of likely sexual transmission.
Unlike most published literature, the highest frequent HPV genotype was HPV-18, which may reflect an intrinsic geographical relatedness and biological differences of HPV-18/HPV-16 and contributes further to research on their infectivity, pathogenicity, and vaccine strategy.

ACKNOWLEDGMENT

We would like to express the sincere gratitude to all seniors of Maxillofacial Surgery Clinic of Ghazi Al-Hariri Hospital for their help, the thanks are extended to all patients who donate their samples in spite of the affliction.

REFERENCES


