

Identification of a Target Region for Specific Molecular Diagnosis of Human Papillomavirus Subtype 16

Ketly Rodrigues Barbosa dos Anjos¹, Anderson Alves da Silva Bezerra¹ and Ana Lisa do Vale Gomes²

1. Academic Center of Vitoria, Department of Nursing, Federal University of Pernambuco, Vitoria of Saint Antão 55608-250, Brazil

2. Academic Center of Vitoria, Department of Physical Education and Sport, Federal University of Pernambuco, Vitoria of Saint Antão 55608-250, Brazil

Abstract: Objective: To identify a target region for molecular diagnosis from the E6 region of the *Human papillomavirus* (HPV). Method: This was a quantitative data research. The sample was composed of sequences from the National Center for Biotechnology Information (NCBI), HPV subtype 16, E6 region, using the following search filters: viruses, genomic DNA/RNA. 3227 sequences were obtained, which were separated according to the country of origin. After selection, the total number of samples consisted of 22 sequences that were aligned in the Clustal/o program, available on the online platform at: <https://www.ebi.ac.uk/Tools/msa/clustalo/> and reading it through the Jalview program to identify consensus regions. For the primer design, the FastPCR program was used and the data were attached to a spreadsheet in Microsoft Excel 2010. Results: The consensus sequence among the countries was selected according to the consensus region for the primer design, called the target region, with 95 nucleotides. FastPCR analyzed the sequence and defined 10 pairs of primers, 5 forward and 5 reverse. The primers were selected according to the percentage CG/AT ratio (50%-60%) and the specific temperature of 58 °C, resulting in 3 primer pairs. The pairs were analyzed in the BLAST program, obtaining a 100% identification and query score, identifying other sequences of HPV 16, region E6. Conclusion: The use of a primer with greater specificity will lead to the specific diagnosis of the virus, making possible the early diagnosis and the most effective treatment.

Key words: *Papillomaviridae*, polymerase chain reaction, computational biology.

1. Introduction

The *Human Papillomavirus* (HPV) belongs to Papillomaviridae family. HPVs are doubled-stranded circular DNA viruses of about 8,000 base pairs, small with approximately 60 nm in diameter. They are non-enveloped viruses with icosahedral capsid and tropism for squamous epithelium [1]. It has several species that can be grouped into five genera (alpha, beta, gamma, mu, and nu), corresponding to the letters of the Greek alphabet [2]. Human pathogen species are classified into two groups, high-risk HPV types, closely related to cervical cancer, such as HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 68, and low-risk HPV types, mainly associated with genital warts (condyloma acuminata), such as HPV 2, 3, 10

and 57 [3].

HPV has three genomic regions: Early, Late and Long Control Region (LCR). Early HPV Open Reading Frames (ORF) include E1, E2, E4, E5, E6, E7 and E8 proteins. Late region encodes L1 and L2 capsid proteins, while LCR encompasses upstream regulatory regions (URR) or noncoding regions (NCR) [2, 4].

To infect a host cell, it is necessary to deliver the virus DNA. HPV infects basal cells in the squamous stratified epithelia, via epithelial wounding or microfissures. Some studies have shown the squamocolumnar junction in the susceptible way because of the heightened accessibility of epithelial reserve cells or stem cells in this region. And to enter the cell nucleus must have an interaction among viral capsid and different receptors in the host cells proteins [5, 6].

Corresponding author: Ketly Rodrigues, Nurse-UFPE/CAV, research fields: bioinformatics, virology and occupational safety.

There were identified 570,000 cancer new cases in women per year and 60,000 in men associated to HPV infection worldwide. This implies the need for the promotion of vaccination in countries, especially in developed countries with the aim of solving such public health problem [7].

Viral infections can promote warts and intraepithelial lesions, respiratory papillomatosis. In addition to cervical cancer, HPV infections can also lead to squamous cell carcinomas of the vagina, vulva, anus, penis tissues and mouth. And because of these changes, emotional distress, HPV infection presents as latent, subclinical and clinical symptoms [8].

Some tools for the diagnosis of HPV infection are: (1) physical examinations—warts and other skin lesions, colposcopy which allows searching for: acetowhite changes, punctuations, mosaicism, and abnormal vessels, (2) serological tests although the infection does not cause a high titer of specific antibody and (3) cervical cytology [9].

There are a few treatment options, although in some cases the infection heals itself. The warts and lesions can be treated with thermal, electrical and chemical ways. The Benign warts can be treated with: podofilox 0.5%, imiquimod 5%, sinecatechins 15%, cryotherapy, podophyllin resin 10%-25%, bichloroacetic acid or trichloroacetic acid 80%-90% and surgical procedures [9].

Genomic mutations were found in coding regions in HPV infection which includes single nucleotides polymorphisms, multiple-nucleotide polymorphisms, insertions, deletions and mixed mutation types which implies the need for a diagnosis with high specificity [10]. It is essential for virus screening molecular diagnosis based on genomic studies such as specific *primers* to the molecular target [11].

The aim of this research was to identify a target region from the E6 region of HPV for molecular diagnosis.

2. Methods and Materials

It was a descriptive research with a quantitative

approach. The sample was composed of sequences selected from the National Center for Biotechnology Information (NCBI) electronic database of the HPV subtype 16, E6 region, using the following search filters: viruses, genomic DNA/RNA. As an exclusion criterion, sequences that had areas other than E6 and E7, as well as sequences that were not HPV-related were removed.

A total number of 3,227 sequences for HPV-16 subtype were obtained and isolated by country of origin. In countries with more than one sequence, multiple alignment was performed using the Clustal X 2.1 program with standard parameters, and the sequence with the highest similarity among the other sequences was selected as representative of the country.

After multiple alignment and selection of representative sequences from each country, the total sample was composed of 22 sequences for HPV-16 subtype with the following NCBI accession numbers: KC291262.1; AF404700.1; KY306726.1; U14513.1; JN847730.1; KP313766.1; EF122285.1; KC662539.1; U34121.1; KF700124.1; KP677553.1; KP019241.1; KJ755040.1; JX847339.1; EU880264.1; KX462889.1; AF003015.1; KC904916.1; KM058595.1; FJ644959.1; KC736930.1; KR075503.1.

For the analysis of the selected samples, a multiple alignment was performed in the Clustal/o program, available on the online platform at: <https://www.ebi.ac.uk/Tools/msa/clustalo/> and the results were read in the program Jalview. For the primer design, the FastPCR program was used in default mode and the data were attached to a worksheet in Microsoft Excel 2010.

3. Results and Discussions

3.1 Preserved Regions of Sequences Representing Each Country

After the virus enters the cell, the viral cycle is initiated using the mechanism of the host cell, inducing cellular DNA replication to achieve its own

replication. Viral proteins encoded by the E6 and E7 regions deregulate the host cell cyclin interactions by two tumor suppressor genes: p53 and pRb which may lead to uncontrolled DNA repair, leading to accumulation of mutations and increased risk of malignant transformation [12, 13].

The E6 and E7 region is 450bp in size. As shown below (Fig. 1), it can be seen that the sequences had few mutations, and may contain conserved and varied regions. Conservation areas were also found in a survey of biological samples from women diagnosed with cervical cancer caused by HPV in Brazil [14]. A study carried out with cervical samples from patients with cervical cancer in Morocco, showed a larger area of conservation in the E7 region of the gene [15].

Among the conserved regions (Fig. 2), they make up position number 120-168 in the middle portion of the alignment, with predominance of thymine, the region of position 207-230 with predominance of adenine, and position number 321-415 (Fig. 3), with predominance of adenine.

3.2 Primer Drawing

To draw the primer, the position of the nucleotides

was used as criterion, so that the *primers* outside the conserved region were excluded. As the base sequence, a sequence of size and alignment similar to most countries was selected and covered the three conserved bands to avoid selection bases.

The base sequence with accession number KF700124.1 from HPV type 16 isolate 113 E6 protein (E6) gene was saved in a fasta file format. The sequence of interest was added to the FastPCR program on the PCR Primer Design tab with the PCR analysis parameters and standard program design options. Ten pairs of *primers* were obtained (Fig. 4), saved in a Microsoft Excel 2010 worksheet.

The primer was selected based on the criteria for good *primers*, such as size, percentage of Guanine and Cytosine (GC) content, melt temperature and primer stability. The GC percentage of the Forward and Reverse *primers* is 52%. The recommended percentage of GC is between 50% and 60%, because the more GC in the primer, can cause auto dimmer formation and secondary structure. Because it has more bonds between DNA nucleotides, compared to Adenine (A) and Thymine (T) binding, a higher percentage of GC will affect the melting temperature [16].

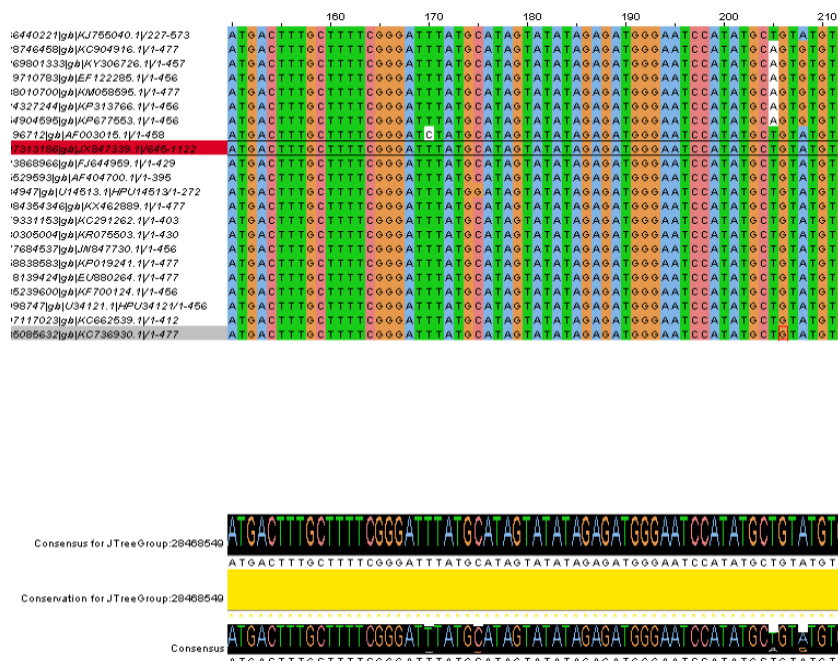


Fig. 1 Multiple alignment of the sequences of HPV subtype 16, E6 region.

Source: FastPCR.

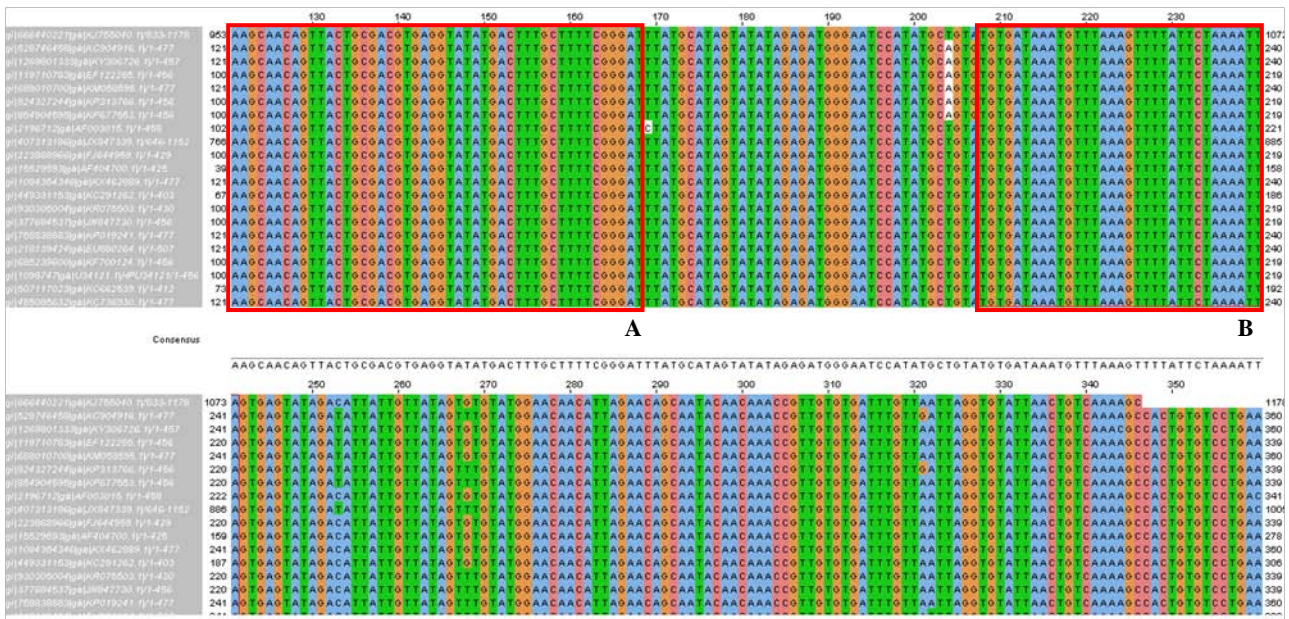


Fig. 2 Conserved regions of the multiple alignment of the HPV species. (a) conserved region spanning 48 nucleotides; (b) conserved region spanning 23 nucleotides.

Source: FastPCR.

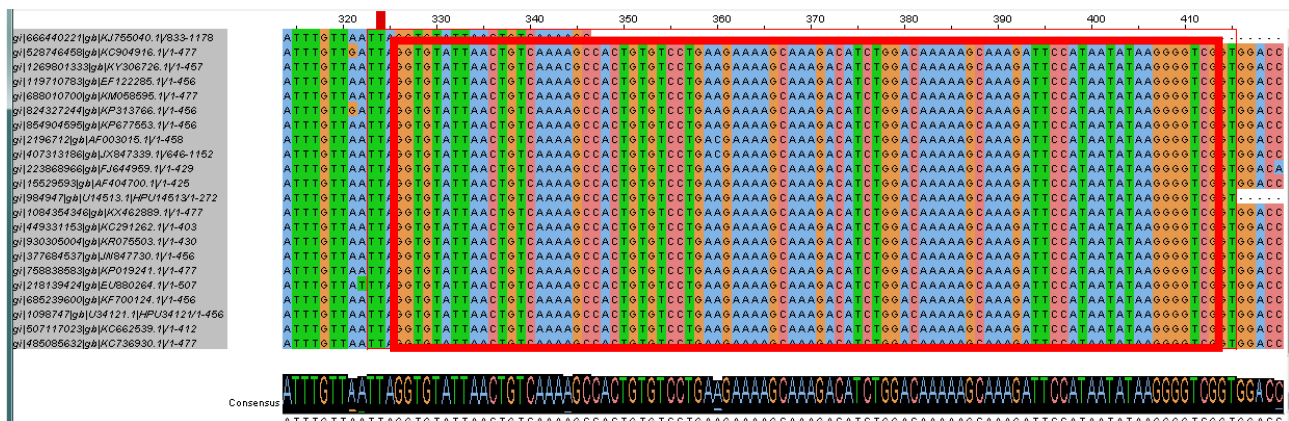


Fig. 3 Conservation range of the HPV subtype 16, E6 region, with a larger nucleotide range, corresponding to 94 nucleotides. Source: FastPCR.

The melt temperature of the *primers* was 58.1 °C. The average size of the primers was 21 bp. Long primers favor the presence of complementary bases, thus needing to be avoided [17].

3.3 Primer Testing

As authenticity check, the *primers* were analyzed in the BLAST program, in the NCBI, so that the level of identification and query score obtained 100% (Fig. 5). This result evidenced the high specificity of the primer in identifying only the virus of interest, which can also be evaluated through the *e*-value, which represents the

quality of the alignment, so that the closer to zero, the less gaps are inserted in the alignment (Fig. 6) [17].

The use of the *in silico* method allows preparation for experimentation in prediction, diagnosis and modeling, since the *in silico* studies reproduce real conditions and enable the creation of *primers* that can be used in specific situations without the need of using biological samples. One study used *in silico primers* to identify and differentiate *Candida* species in vaginal secretion and validate the *primers* used the *in silico* test and real-time PCR (qPCR), demonstrating the ability of the *primers* to differentiate between 5 different

PrimerID	Sequence(5'-3')	Tm(°C)	Primer_Quality(%)	Fragment_Size(bp)/Tm(°C)	Topt(°C)
1F4_1_40-63	cagttatgcacagctgcaaac	58,2	80		
1R13_1_105-129	atatacctcacgtcgcaagctg	57,2	88	90/69	62
1F1_1_14-35	cacaggagcagccagaaagt	58,8	80		
1R13_1_105-129	atatacctcacgtcgcaagctg	57,2	88	116/74	62
1F3_1_37-60	ccacagttatgcacagctgcaa	59,8	80		
1R13_1_105-129	atatacctcacgtcgcaagctg	57,2	88	93/70	62
1F6_1_99-120	caagcaacagttactgagcgt	58,1	87		
1R9_1_165-189	acatacagcatatggattcccatct	55,2	80	91/71	60
1F6_1_99-120	caagcaacagttactgagcgt	58,1	87		
1R4_1_411-435	tgctcttgatgatctgcaacaagac	55,8	80	337/77	62
1F6_1_99-120	caagcaacagttactgagcgt	58,1	87		
1R1_1_432-456	ttacagctgggttctctacgtgtt	57,6	80	358/77	64
1F6_1_99-120	caagcaacagttactgagcgt	58,1	87		
1R2_1_427-451	gctgggttctctacgtgttcttga	58,8	79	353/77	64
1F6_1_99-120	caagcaacagttactgagcgt	58,1	87		
1R5_1_400-424	atctgcaacaagacatactcgacc	57,2	79	326/77	63
1F6_1_99-120	caagcaacagttactgagcgt	58,1	87		
1R10_1_116-140	aaagcaagtcatactcacgtc	55,2	79	42/48	59
1F6_1_99-120	caagcaacagttactgagcgt	58,1	87		
1R3_1_417-441	tctacgtgtcttgatgatctgcaa	56,1	78	343/77	62

Fig. 4 Primers obtained after the analysis of the target region by the FastPCR program.

Download v GenBank Graphics

Human papillomavirus type 16 isolate HP_2013-128 E6 protein (E6) and E7 protein (E7) genes, complete cds
Sequence ID: [MH370229.1](#) Length: 776 Number of Matches: 1

Range 1: 453 to 477 [GenBank](#) [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
50.1 bits(25)	8e-04	25/25(100%)	0/25(0%)	Plus/Minus

Query 1 TTACAGCTGGGTTTCTCTACGTGTT 25
Sbjct 477 TTACAGCTGGGTTTCTCTACGTGTT 453

Download v GenBank Graphics

Human papillomavirus type 16 isolate HP_2013-128 E6 protein (E6) and E7 protein (E7) genes, complete cds
Sequence ID: [MH370229.1](#) Length: 776 Number of Matches: 1

Range 1: 120 to 141 [GenBank](#) [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
44.1 bits(22)	0.020	22/22(100%)	0/22(0%)	Plus/Plus

Query 1 CAAGCAACAGTTACTGCGACGT 22
Sbjct 120 CAAGCAACAGTTACTGCGACGT 141

Fig. 5 Local alignment of primer using BLAST.

Source: NCBI.

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input type="checkbox"/> Human papillomavirus type 16 isolate InDRE 163 major capsid protein (L1), gene, partial cds; E6 protein (E6), gene, complete cds; and E7 protein (E7), gene, partial cds	50.1	50.1	100%	0.001	100.00%	MH921960.1
<input type="checkbox"/> Human papillomavirus type 16 isolate InDRE 160 major capsid protein (L1), gene, partial cds; E6 protein (E6), gene, complete cds; and E7 protein (E7), gene, partial cds	50.1	50.1	100%	0.001	100.00%	MH921958.1
<input type="checkbox"/> Human papillomavirus type 16 isolate InDRE 156 major capsid protein (L1), gene, partial cds; E6 protein (E6), gene, complete cds; and E7 protein (E7), gene, partial cds	50.1	50.1	100%	0.001	100.00%	MH921957.1
<input type="checkbox"/> Human papillomavirus type 16 isolate InDRE 155 major capsid protein (L1), gene, partial cds; E6 protein (E6), gene, complete cds; and E7 protein (E7), gene, partial cds	50.1	50.1	100%	0.001	100.00%	MH921956.1
<input type="checkbox"/> Human papillomavirus type 16 isolate InDRE 39 major capsid protein (L1), gene, partial cds; E6 protein (E6), gene, complete cds; and E7 protein (E7), gene, partial cds	50.1	50.1	100%	0.001	100.00%	MH921955.1
<input type="checkbox"/> Human papillomavirus type 16 isolate InDRE 31 major capsid protein (L1), gene, partial cds; E6 protein (E6), gene, complete cds; and E7 protein (E7), gene, partial cds	50.1	50.1	100%	0.001	100.00%	MH921953.1
<input type="checkbox"/> Human papillomavirus type 16 isolate InDRE 25 major capsid protein (L1), gene, partial cds; E6 protein (E6), gene, complete cds; and E7 protein (E7), gene, partial cds	50.1	50.1	100%	0.001	100.00%	MH921952.1
<input type="checkbox"/> Human papillomavirus type 16 isolate InDRE 20 major capsid protein (L1), gene, partial cds; E6 protein (E6), gene, complete cds; and E7 protein (E7), gene, partial cds	50.1	50.1	100%	0.001	100.00%	MH921951.1
<input type="checkbox"/> Human papillomavirus type 16 isolate InDRE 9CaCu major capsid protein (L1), gene, partial cds; E6 protein (E6), gene, complete cds; and E7 protein (E7), gene, partial cds	50.1	50.1	100%	0.001	100.00%	MH921949.1
<input type="checkbox"/> Human papillomavirus type 16 isolate InDRE 7CaCu major capsid protein (L1), gene, partial cds; E6 protein (E6), gene, complete cds; and E7 protein (E7), gene, partial cds	50.1	50.1	100%	0.001	100.00%	MH921948.1
<input type="checkbox"/> Human papillomavirus type 16 isolate InDRE 7 major capsid protein (L1), gene, partial cds; E6 protein (E6), gene, complete cds; and E7 protein (E7), gene, partial cds	50.1	50.1	100%	0.001	100.00%	MH921947.1
<input type="checkbox"/> Human papillomavirus type 16 isolate InDRE 6 major capsid protein (L1), gene, partial cds; E6 protein (E6), gene, complete cds; and E7 protein (E7), gene, partial cds	50.1	50.1	100%	0.001	100.00%	MH921946.1
<input type="checkbox"/> Human papillomavirus type 16 isolate InDRE 3CaCu major capsid protein (L1), gene, partial cds; E6 protein (E6), gene, complete cds; and E7 protein (E7), gene, partial cds	50.1	50.1	100%	0.001	100.00%	MH921945.1
<input type="checkbox"/> Human papillomavirus type 16 isolate InDRE B49 major capsid protein (L1), gene, partial cds; E6 protein (E6), gene, complete cds; and E7 protein (E7), gene, partial cds	50.1	50.1	100%	0.001	100.00%	MH921944.1

Fig. 6 Local alignment of primer using BLAST with the strains of interest.

Source: NCBI.

Candida species with specificity of 90% [18].

6. Conclusions

Bioinformatics tools are essential for analyzing the high number of virus sequences and allowing design of high specificity diagnosis *primers* using online platforms and software, ensuring that the PCR has satisfactory results. Detection of HPV high-risk populations with specific consensus primers has proven to be mostly cost-effective to detect by grouping at least two or more primers sets from different regions with high sensitivity for both high and low risk HPV types [19].

References

- [1] Doorbar, J., Egawa, N., Griffin, H., Kranjec, C., and Murakami, I. 2015. "Human Papillomavirus Molecular Biology and Disease Association." *Reviews in Medical Virology* 25 (Suppl. 1): 2-23. doi:10.1002/rmv.1822.
- [2] Harden, M. E., and Munger, K. 2017. "Human Papillomavirus Molecular Biology." *Mutation Research/Reviews in Mutation Research* 772: 3-12. doi:10.1016/j.mrrev.2016.07.002.
- [3] Doorbar, J., Quint, W., Banks, L., Bravo, I. G., Stoler, M., Broker, T. R., and Stanley, M. A. 2012. "The Biology and Life-Cycle of Human Papillomaviruses." *Vaccine* 30: F55-70. doi:10.1016/j.vaccine.2012.06.083.
- [4] Nguyen, H. P., Ramírez-Fort, M. K., and Rady, P. L. 2014. "The Biology of Human Papillomaviruses." *Current Problems in Dermatology* 45 (1): 19-32. doi:10.1159/000355959.
- [5] Groves, I. J., and Coleman, N. 2015. "Pathogenesis of Human Papillomavirus-Associated Mucosal Disease." *J. Pathol.* 235 (4): 527-38. doi:10.1002/path.4496.
- [6] Aksoy, P., Gottschalk, E. Y., and Meneses, P. I. 2017. "HPV Entry into Cells." *Mutation Research/Reviews in Mutation Research* 772: 13-22. doi:10.1016/j.mrrev.2016.09.004.
- [7] De Martel, C., Plummer, M., Vignat, J., and Franceschi, S. 2017. "Worldwide Burden of Cancer Attributable to HPV by Site, Country and HPV Type." *International Journal of Cancer* 141 (4): 664-70. doi:10.1002/ijc.30716.
- [8] Juckett, G., and Hartman-Adams, H. 2010. "Human Papillomavirus: Clinical Manifestations and Prevention." *Am Fam Physician* 82 (10): 1209-13.
- [9] Hathaway, J. K. 2012. "HPV: Diagnosis, Prevention, and Treatment." *Clinical Obstetrics and Gynecology* 55 (3): 671-80. doi:10.1097/grf.0b013e31825caa36.
- [10] Kannan, A., Hertweck, K. L., Philley, J. V., Wells, R. B., and Dasgupta, S. 2017. "Genetic Mutation and Exosome Signature of Human Papilloma Virus Associated Oropharyngeal Cancer." *Scientific Reports* 7 (1). doi:10.1038/srep46102.
- [11] Venceslau, E. M., Bezerra, M. M., Lopes, A. C. M., Souza, E. V., Onofre, A. S. C., Melo, C. M., Jeraldo, V. L. S., and Onofre, F. B. M. 2014. "HPV Detection Using Primers MY09/MY11 and GP5+/GP6+ in Patients with Cytologic and/or Colposcopic Changes." *Jornal Brasileiro de Patologia e Medicina Laboratorial* 50 (4): 280-5. https://dx.doi.org/10.5935/1676-2444.20140028.
- [12] Burk, R. D., Chen, Z., and Van Doorslaer, K. 2009. "Human Papillomaviruses: Genetic Basis of Carcinogenicity." *Public Health Genomics* 12 (5-6): 281-90. doi:10.1159/000214919.
- [13] Chen, Z., Terai, M., Fu, L., Herrero, R., DeSalle, R., and Burk, R. D. 2005. "Diversifying Selection in Human Papillomavirus Type 16 Lineages Based on Complete Genome Analyses." *Journal of Virology* 79 (11). doi:10.1128/JVI.79.11.7014-7023.2005.
- [14] Oliveira, C. M., Bravo, I. G., Souza, N. C. S., Genta, M. L., Fregnani, J. H., Tacla, M., Carvalho, J. P., Longatto-Filho, A., and Levi, J. E. 2015. "High-Level of Viral Genomic Diversity in Cervical Cancers: A Brazilian Study on Human Papillomavirus Type 16." *Infect Genet Evol.* 34: 44-51. doi:10.1016/j.meegid.2015.07.002.
- [15] Assoumou, S. Z., Boumba, L. M. A., Mbiguino, A. N., Mabika, B. M., Belembaogo, E., Khattabi, A., and Ennaji, M. M. 2015. "Sequence Variations of Human Papillomavirus Type 16 E6 and E7 Genes in Cervical Cancer Isolates from Gabon." *British Microbiology Research Journal* 8 (2). doi:10.9734/BMRJ/2015/17225.
- [16] Nurjayadi, M., Efrianti, U.R., Azizah, N., and Julio, E. 2019. "Optimum Temperature of the Amplification of the fljB Gene of Salmonella Typhimurium." In *Empowering Science and Mathematics for Global Competitiveness*, edited by Rahmawati & Taylor, Taylor & Francis Group, London. ISBN 978-1-138-61666-0.
- [17] Queiroz, J. A. S., Alves, L. S., Dall'acqua, D. S. V., and Souza, L. F. B. 2017. "Design and Validation of Primers *in Silico* for Detection of Human Respiratory Syncytial Virus." *Rev FIMCA* 4 (1): 17-30.
- [18] Jackisch, E. S., Koehler, A., Toiller, B. R., Dallemole, D. R., Corbellini, V. A., and Rieger, A. 2017. "Desenvolvimento de primers para identificação e diferenciação de espécies de Candida em secreção vaginal por PCR em tempo real (qPCR)." *Revista Jovens Pesquisadores* 7 (1): 46-59. doi:10.17058/rjp.v7i1.9337.
- [19] Abhilasha, G., Mallikarjuna, R. G., Madhukar, R., Shyam,

S., and Jaya, C. 2019. "Human Papillomavirus Genome Based Detection and Typing: A Holistic Molecular

Approach." *Current Molecular Medicine* 19: 237. <https://doi.org/10.2174/1566524019666190405120441>.