INFERINDO RELAÇÕES EVOLUTIVAS ENTRE DIFERENTES GENÓTIPOS DE PAPILOMAVÍRUS HUMANO A PARTIR DE MÚLTIPLOS ALINHAMENTOS DA PROTEÍNA L1 DO CAPSÍDEO MAIOR

INFERRING EVOLUTIONARY RELATIONSHIPS AMONG DIFFERENT HUMAN PAPILLOMAVIRUS GENOTYPES FROM MULTIPLE ALIGNMENTS OF THE MAJOR CAPSID PROTEIN L1

Jane Eyre GABRIEL[1], Kárita Cláudia FREITAS-LIDANI[2], Rodrigo Pessoa de FARIAS[3]

1- Professora da Universidade Federal do Vale do São Francisco UNIVASF, Petrolina, PE.
2- Pós-doutoranda da Universidade Federal do Paraná UFPR, Curitiba, PR
3- Técnico de Nível Superior do Instituto do Câncer de São Paulo Octavio Frias de Oliveira ICESP, São Paulo, SP
Autor para correspondência: jane.gabriel@univasf.edu.br

ABSTRACT:
The major capsid protein L1 constitutes the entire exterior surface of the stabilized mature human papillomavirus (HPV), mediating initial attachment to host tissues or cells, and become pliable enough to ultimately allow release of the viral genome into a new target cell. The purpose of this study was to infer evolutionary relationships among different variable-risk HPV genotypes from comparative alignments of multiple sequences of the protein L1 deposited previously in biological information database. First, sequences of the protein L1 of 20 HPV genotypes were searched and selected from a non-redundant protein sequence database UniProtKB/Swiss-Prot. Next, a phylogenetic dendogram was constructed by comparing multiple sequences of the protein L1 using molecular evolutionary genetics analyses by Mega software. The dendogram generated from comparative alignments of the L1 protein sequences of different HPV types revealed the presence of two main clusters: a first cluster containing 12 HPV types linked intimately in several sub-branches and a second cluster grouping 8 HPV types linked in another sub-branches. Evolutionary groupings generated from L1 capsid protein sequences of variable-risk HPV genotypes demonstrated weak association between pathogenicity and phylogenetic proximity in the types analyzed, accompanied by low identity among their amino acid residues. The findings described herein reveal important insights into evolutionary patterns and phylogenetic relationships among variable-risk HPV genotypes for malignant conversion of virally infected cells from multiple alignments of the major viral capsid protein L1.
Key words: Bioinformatics; Evolutionary relationships; HPV; L1 protein

RESUMO:
A proteína L1 do capsídeo maior constitui a superfície exterior do papilomavírus humano (HPV), mediando a liberação do genoma viral para uma célula-alvo hospedeira. O objetivo da presente estudo foi inferir relações evolutivas entre diferentes genótipos de HPV a partir de alinhamentos comparativos de sequências da proteína L1 do capsídeo maior depositadas previamente em bancos de dados de informação biológica. Sequências da proteína L1 de 20 genótipos de HPV de riscos variáveis de patogenicidade foram pesquisadas e selecionadas a partir de um banco de dados de sequências não redundantes de proteínas UniProtKB/Swiss-Prot. Em seguida, um dendograma filogenético foi

construído pela comparação direta das sequências da proteína L1 a partir de análises moleculares empregando o programa Mega6.0. O dendograma dos alinhamentos comparativos das sequências da proteína L1 de diferentes tipos de HPV revelou a presença de dois agrupamentos principais: um primeiro agrupamento com 12 tipos de HPV intimamente ligados em vários sub-ramos e um segundo agrupamento contendo 8 tipos de HPV ligados em outros sub-ramos. Os agrupamentos evolutivos gerados a partir de sequências proteicas de capsídeo L1 de genótipos de HPV de riscos variáveis demonstraram fraca associação entre patogenicidade e proximidade filogenética nos genótipos analisados, acompanhada de baixa identidade entre seus resíduos de aminoácidos. Os achados aqui descritos revelam importantes subsídios acerca dos padrões evolutivos e relações filogenéticas entre genótipos de HPV de riscos variáveis para conversão maligna de células por análises de múltiplos alinhamentos da proteína L1 do capsídeo viral.
Palavras-chave: Bioinformática; HPV; Proteína L1; Relações evolutivas

1. INTRODUCTION

Papillomaviruses are a diverse group of nonenveloped DNA viruses that replicate exclusively in specialized keratinocytes near the apical surface of the epithelium by expressing viral late genes, such as those encoding the major capsid protein L1. According to DOORBAR (2005), this protein constitutes the entire exterior surface of the stabilized mature virion, mediating initial attachment to host tissues or cells, and become pliable enough to ultimately allow release of the viral genome into a new target cell. Recently, DIGIUSEPPE et al*.* (2017) reported that protein L1 stabilizes the viral genome within the subviral complex during intracellular trafficking, accompanying the viral genome beyond the endosomal compartment. Over the last decade, prophylactic vaccines for human papillomavirus (HPV) in late stages of clinical testing are composed of L1 capsid protein that self-assemble into virus-like particles when expressed in recombinant systems, resulting in strong adaptive immune responses to prevent progression of HPV infection, induce regression of intraepithelial neoplasia, or eradicate residual cervical cancer (VILLA, 2006).

Dynamic functions, structural aspects, prophylactic and therapeutic implications concerning stability maintenance of the major capsid protein L1 in the papillomavirus life cycle have extensively been described in the literature, and some studies have already focused on molecular evolutionary genetic analyses of this viral capsid protein in HPVs. According to NTOVA et al*.* (2012), the genomic characterization of HPV variants provides important evidences about geographical relatedness, biological differences and pathogenicity risk. Thus, the purpose of this study was to infer evolutionary relationships among different variable-risk HPV genotypes from comparative alignments of multiple

sequences of the major capsid protein L1 deposited previously in biological information database.

2. MATERIAL AND METHODS

Under these molecular evolutionary genetic analyses, sequences of the capsid protein L1 of 20 HPV types (HPV07, 09, 10, 11, 15, 16, 17, 18, 19, 25, 30, 31, 32, 35, 45, 52, 56, 58, 61, 67) were searched and selected from a high quality annotated and non-redundant protein sequence database designated UniProtKB/Swiss-Prot. Among the variable-risk HPV genotypes selected in current study, HPV16, 18, 31, 35, 45, 52, 56 and 58 are admittedly associated with cervical cancer, laryngeal and oral papillomatosis and anogenital warts (COGLIANO et al., 2005). A phylogenetic dendogram was constructed by direct comparing multiple sequences of the protein L1 using molecular evolutionary genetics analyses by Mega (version 6.0) software (TAMURA et al., 2013). Relevant evaluative parameters concerning identity and similarity values were also generated using Clustalo program (SÖDING, 2005), as described in details by GABRIEL and LIDANI (2015).

3. RESULTS

The dendogram generated from comparative alignments of the major capsid protein L1 of different HPV genotypes revealed the presence of two main clusters: a first cluster containing 12 HPV types linked intimately in several sub-branches [15 (0.06), 17 (0.07), 09 (0.07), 25 (0.65), 31 (0.31), 61 (0.42), 35 (0.95), 11 (0.99), 07 (0.85), 32 (0.64), 18 (1.38) and 30 (1.31)], and a second cluster grouping 8 HPV types linked in another sub-branches [19 (0.88), 67 (0.86), 56 (1.46), 10 (1.17), 16 (0.51), 58 (0.38), 45 (0.61) and 52 (0.61)] (Figure 1). Examining the amino acid average amount of the L1 protein sequences in the HPV types analyzed in current study, high and low contents of leucine (7,992 residues) and histidine (1,741 residues), respectively, were found in the comparative alignments of the L1 protein of these HPV types. Multiple alignments of the amino acid residues of the L1 protein of all genotypes selected demonstrated a 23.826% identity with the presence of similar and identical positions corresponding to 132 and 142 residues, respectively, in a total amino acid amount of 519.

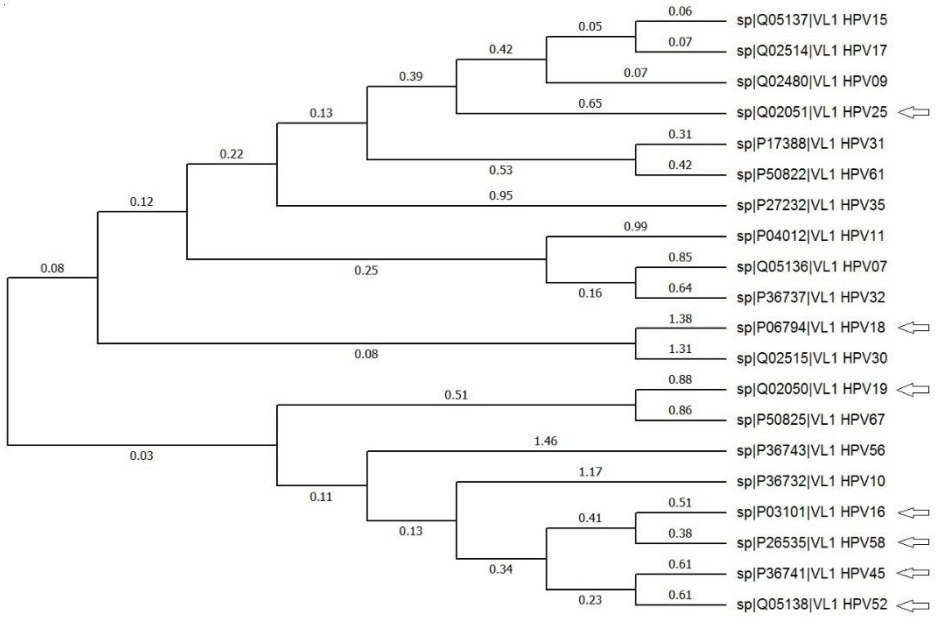

FIGURE 1. EVOLUTIONARY RELATIONSHIPS FROM COMPARATIVE ALIGNMENTS OF THE AMINO ACID RESIDUES OF THE MAJOR CAPSID PROTEIN L1 IN DIFFERENT VARIABLE-RISK HPV GENOTYPES. SEQUENCE IDENTIFIERS AND ACCESSION NUMBERS OF ALL HPV GENOTYPES ANALYZED ARE INDICATED TO THE RIGHT OF THE DENDOGRAM. ARROWS INDICATE GROUPINGS AMONG DIFFERENT HPV GENOTYPES WITH VARIABLE-RISKS.

Interestingly, HPV16 (0.51) and 18 (1.38) types, known as highest-risk strains for genital cancers and malign tumors, were grouped in different clusters from multiple alignments of the L1 protein sequences with evolutionary divergence estimated of 0.397 (arrows in Figure 1). On the other hand, some HPV genotypes characterized as highest- and high-risks for development of genital cancer were strongly clustered in sub-branches related, as shown in the groupings between HPV16 (0.51) and 58 (0.38), and between HPV45 (0.61) and 52 (0.61) (arrows in Figure 1). The greatest evolutionary divergence (0.630) was estimated between HPV types 16 and 19 classified as I class carcinogenic and unknown-risk genotypes, respectively (arrows in Figure 1), whereas the smallest evolutionary divergence (0.059) was detected between highest-risk HPV16 and unknown-risk HPV25 for genital cancer (arrows in Figure 1).

4. DISCUSSION

Evolutionary groupings generated from amino acid residues of the capsid protein L1 of variable-risk HPV genotypes demonstrated weak association between pathogenicity and phylogenetic proximity, accompanied by low identity among their amino acid residues in the types selected (Figure 1). These evidences may presumably be a consequence of mechanisms of selective pressure for papillomaviruses to accumulate mutations and prevent the binding of neutralizing antibodies raised by prior infections. YUE et al. (2013), investigating polymorphisms and intratypic variations of HPV-16 and HPV-58 L1/L2 genes originating in Southwest China, demonstrated that most of the mutations were of positive selection, suggesting that such amino acid changes were beneficial to accommodate the human papillomavirus to its environment. Although it is tempting to speculate that all HPV types classified as high-risk have similar properties, few experimental evidences have supported this argument. Especially not congruent risk classifications reported for some HPV types that are phylogenetically closely related to accepted high-risk HPV types underline the need for additional molecular data for types with low prevalence rates or unclear risk status, as suggested by HILLER et al. (2006).

Sequence comparison of different human L1 types has revealed a conserved region of the carboxyl terminus containing clustered basic amino acids that bear resemblance to proposed heparin-binding motifs in unrelated proteins (JOYCE et al., 1999). In the meantime, BUCK et al. (2013) demonstrated that the surface loops are poorly conserved, even among closely related papillomavirus types, as well as their gene sequences are considered highly polymorphic. Especially, papillomaviruses are classified according to the sequence identity in the capsid gene L1, and findings have indicated a discontinuity in the evolutionary history of the L1 genes in this virus, giving rise to differences in the phylogenetic reconstructions of the early and of the late genes (BRAVO & ALONSO, 2007). *In silico* analyses have supported that some high-risk HPV genotypes have similar functional and structural properties (GABRIEL et al., 2013), and the identification of polymorphisms in the coding region of the major capsid protein L1 within the HPV genome may discriminate the infectious potential of different variants (CHIESA et al., 2016). AHMED et al. (2013) detected that HPV6 and HPV11 L1 sequence variations were low with 17% of sequences incorporating an E431Q substitution in the α4-βJ region and 7% of sequences incorporating an A235S substitution in the α1 region, respectively. Thus, the genomic characterization of HPV variants from protein L1 is crucial for a better understanding of the biological differences

involved to evolutionary adaptation and pathogenicity risks of these viruses.

In conclusion, comparative assessments from amino acid residues of the major capsid protein L1 using Bioinformatics' tools described in current study provide new insights into phylogenetic relationships among different human papillomavirus genotypes, evidencing accentuated evolutionary divergence among highest-risk variants for malignant conversion of virally infected cells.

5. ACKNOWLEDGMENTS

This research was supported by the Bioinformatics and Computational Biology Group, designated "BIO in BYTES".

6. REFERENCES

AHMED AI, BISSETT SL, BEDDOWS S. 2013. Amino acid sequence diversity of the major human papillomavirus capsid protein: implications for current and next generation vaccines. *Infection, Genetics and Evolution*, 18, 151-159.

BRAVO IG, ALONSO A. 2007. Phylogeny and evolution of papillomaviruses based on the E1 and E2 proteins. *Virus Genes*, 34, 249-262.

BUCK CB, DAY PM, TRUS BL. 2013. The papillomavirus major capsid protein L1. *Virology*, 445, 2169–2174.

CHIESA IJ, PEREZ MS, NUÑEZ GG, *et al*. 2016. Genetic variability and phylogeny analysis of partial L1 gene of human papillomavirus variants in Buenos Aires, Argentina. *Virus Disease*, 27, 41-47.

COGLIANO V, BAAN R, STRAIF K, *et al*. 2005. Carcinogenicity of human papillomaviruses (external site). *Lancet Oncology*, 6, 204.

DIGIUSEPPE S, BIENKOWSKA-HABA M, GUION LGM, *et al*. 2017. Human papillomavirus major capsid protein L1 remains associated with the incoming viral genome throughout the entry process. *Journal of Virology*, 91, e00537-17.

DOORBAR J. 2005. The papillomavirus life cycle. *Journal of Clinical Virology*, 32, S7-S15.

GABRIEL JE, FIGUEIREDO DD, FARIAS RP. 2013. Revealing highly conserved regions in the E6 protein among distinct human papillomavirus types using comparative analysis of multiple sequence alignments. *Brazilian Journal of Biology*, 73, 449-450.

GABRIEL JE, LIDANI KC. 2015. Molecular conservation of the mammalian leptin protein. *Genetics and Molecular Research*, 14, 253-258.

HILLER T, POPPELREUTHER S, STUBENRAUCH F, *et al*. 2006. Comparative analysis of 19 genital human papillomavirus types with regard to p53 degradation, immortalization, phylogeny, and epidemiologic risk classification. *Cancer Epidemiology, Biomarkers & Prevention*, 15, 1262-1267.

JOYCE JG, TUNG JS, PRZYSIECKI CT, *et al*. 1999. The L1 major capsid protein of human papillomavirus type 11 recombinant virus-like particles interacts with heparin and cell-surface glycosaminoglycans on human keratinocytes. *Journal of Biology Chemistry*, 274, 5810-5822.

NTOVA CK, KOTTARIDI C, CHRANIOTI A, *et al*. 2012. Genetic variability and phylogeny of high risk HPV type 16, 18, 31, 33 and 45 *L1* gene in Greek women. *International Journal of Molecular Science*, 13, 1-17.

SÖDING J. 2005. Protein homology detection by HMM–HMM comparison. *Bioinformatics*, 21, 951-960.

TAMURA K, STECHER G, PETERSON D, *et al*. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30, 2725-2729.

VILLA LL. 2006. Prophylactic HPV vaccines: reducing the burden of HPV-related diseases. *Vaccine*, 24 Suppl 1, S23-S28.

YUE Y, YANG H, WU K, *et al*. 2013. Genetic variability in L1 and L2 genes of HPV-16 and HPV-58 in Southwest China. *PLoS One*, 8:e55204.