

**UNIVERSIDADE FEDERAL DO AMAZONAS – UFAM**  
**INSTITUTO DE CIÊNCIAS BIOLÓGICAS – ICB**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM DIVERSIDADE BIOLÓGICA – PPG-MDB**

**CLONAGEM, SEQUENCIAMENTO E ESTUDOS  
MOLECULARES DO GENOMA DE HPV 16  
ISOLADO NA AMAZÔNIA**

Roberto Alexandre Alves Barbosa Filho

**Manaus**

**2011**

**UNIVERSIDADE FEDERAL DO AMAZONAS – UFAM**  
**INSTITUTO DE CIÊNCIAS BIOLÓGICAS – ICB**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM DIVERSIDADE**  
**BIOLÓGICA – PPG-MDB**

**CLONAGEM, SEQUENCIAMENTO E ESTUDOS MOLECULARES**  
**DO GENOMA DE HPV 16 ISOLADO NA AMAZÔNIA**

**ROBERTO ALEXANDRE ALVES BARBOSA FILHO**

Dissertação apresentada à Universidade Federal do Amazonas como requisito parcial a obtenção do grau de Mestre pelo Programa de Pós-Graduação em Diversidade Biológica, área de concentração em Caracterização da Biota Amazônica.

**Orientador:** Dr. Spartaco Astolfi Filho.

**Co-orientadora:** Dr<sup>a</sup>. Cristina Maria Borborema dos Santos.

**Manaus**

**2011**

# FICHA CATALOGRÁFICA

Barbosa Filho, Roberto Alexandre Alves

Clonagem, sequenciamento e estudos moleculares do genoma de HPV 16 isolado na Amazônia / Roberto Alexandre Alves Barbosa Filho – Manaus: UFAM, 2011.

90 p. ilustr.

Dissertação de Mestrado – Universidade Federal do Amazonas. ICB. Programa de Pós-Graduação em Diversidade Biológica / Área de concentração em Caracterização da Biota Amazônica.

1. Biologia Molecular 2. Evolução 3. Papilomavírus Humano 16

Sinopse:

O genoma do HPV 16 foi completamente sequenciado e suas regiões gênicas foram analisadas. As análises moleculares indicam que o vírus isolado na Amazônia representa uma variante ancestral das variantes africanas de HPV 16. Esse trabalho é pioneiro nos estudos genômicos do tipo de papilomavírus humano mais prevalente na região Amazônica.

## **Agradecimentos**

A Deus, por ter me dado condições de chegar até aqui e por tudo o que passei (e pelo o que ainda há de vir). Por me ensinar a discernir sabedoria de conhecimento e por me tornar sedento por ambos. Por TUDO.

Aos meus pais, pelo amor e pelo carinho. Sei que sou péssimo em expressar o que sinto por palavras ditas, no entanto, tento ao menos registrar por escrito um pouco do que vocês representam. À minha mãe, por sua total dedicação em educar seus cinco filhos e pela ajuda incondicional que tem me dado em todos os momentos da minha vida. Ao meu pai, cuja trajetória profissional norteia minha vida, pelos conselhos e pela confiança.

Aos meus irmãos, por terem me acompanhado e me apoiado durante toda a minha jornada. Podem ter certeza de que os sentimentos que temos uns pelos outros se sobrepõem a todas as diferenças que teimamos em manter entre nós.

À minha esposa, pela solitária companhia, pela compreensão incompreendida, pelas brigas, pelo carinho e, acima de tudo, por me permitir amá-la (mesmo sendo do meu jeito torto de amar).

Aos meus mestres, pelas lições, conselhos e exemplos que moldaram minha conduta em sala como professor.

Aos meus orientadores, professora Dr<sup>a</sup>. Cristina e professor Dr. Spartaco, pelo apoio e confiança em mim depositados. Pelos infinitos exemplos, não só de “Como fazer Ciência”, mas também de “Como ‘ser’ Ciência”. Obrigado por me deixar aprender. Desculpem-me se os decepcionei em algum momento, mas prometo melhorar no Doutorado.

Ao professor Dr. Edmar, por abrir meus olhos sob tantas perspectivas e pela capacidade de ensinar tanta coisa numa só conversa.

À professora Dr<sup>a</sup>. Izeni, pela atenção especial e pelos valiosos conselhos.

À professora Dr<sup>a</sup>. Isabel, pelas dicas e por me deixar ajudar em seu trabalho, onde pude amadurecer minhas idéias sobre Evolução à luz da Genética de Populações Humanas.

Aos amigos do Centro de Apoio Multidisciplinar, Jonso, Dina, Rogério, Edson, Lívio e Elza, pela companhia, pelos conselhos, pelos risos, pelas discussões, enfim, por simplesmente me dispor alguns minutos de seus dias.

Aos amigos do Laboratório de Diagnóstico Molecular, Jéssica, Mayara, André, Évelyn, Norma, Daniele, Cíntia, Júnia e Paulo pela valiosa amizade (que já não acreditava que existisse), por toda a ajuda e por me deixarem ajudar. Resumindo, não sei o que seria deste trabalho se vocês não estivessem comigo (cada um sabe a importância que teve no desenvolvimento dele).

Aos coordenadores do Programa de Pós-Graduação em Diversidade Biológica, professor Dr. Jaydione e professora Dr<sup>a</sup>. Ivone, pelos exemplos de conduta, gestão e responsabilidade e pela compreensão frente aos obstáculos que encontrei na realização deste trabalho, que fizeram do seu apoio uma ferramenta indispensável a sua conclusão.

A todos, pelos bons e maus momentos, pelos risos e prantos, pela amizade e pelo amor, mas, acima de tudo, pela companhia.

## RESUMO

O Papillomavirus Humano é responsável por lesões na mucosa oral, anal e do trato urogenital masculino e feminino, transmitidas por contato direto ou indireto com a pele infectada ou através de relações sexuais. Na mulher essas infecções podem evoluir para um câncer de colo do útero, cuja estimativa de incidência para a região Norte no ano de 2010 foi a maior do Brasil. A natureza das infecções depende do grau de integração do DNA viral com o DNA do hospedeiro associada, principalmente, aos genes das oncoproteínas E6 e E7 do HPV. A determinação dos tipos virais pode ser realizada a partir de diferenças no gene L1 do capsídeo viral e as variantes de um determinado tipo de HPV podem ser identificadas por meio do estudo da Região Não Codificadora viral. Atualmente o desenvolvimento de vacinas profiláticas contra o HPV utiliza partículas “pseudo-virais” formadas pela proteína L1 de tipos virais de alto risco, enquanto cresce o número de estudos que utilizam as oncoproteínas E6 e E7 no desenvolvimento de vacinas terapêuticas. Contudo, é necessário que o desenvolvimento de tais vacinas antivirais também considere a grande diversidade das variantes dos tipos de HPV existentes, uma vez que diferenças entre as regiões genômicas dessas variantes podem influenciar o grau de suas infecções. Este trabalho descreve o sequenciamento completo do genoma de uma variante do HPV 16, detectado no Estado do Amazonas, utilizando técnicas de Engenharia Genética, bem como a análise desse genoma por ferramentas de Bioinformática. Observou-se, pela análise de distâncias genéticas, que o genoma dessa variante apresenta grande proximidade genética dos exemplares identificados na literatura como “variantes africanas”, e as análises filogenéticas, realizadas a partir da Região Não Codificadora, reforçam essa hipótese. Além disso, também foram detectadas várias mutações ao longo do genoma obtido, resultando em alterações nas posições e na quantidade de sítios de restrição de sua sequência. As maiores diferenças entre as regiões gênicas do genoma sequenciado e as correspondentes nas variantes africanas foram observadas ao longo de E7. Espera-se, com esse trabalho, atentar os futuros projetos de pesquisa que envolvam expressão de proteínas e análises genômicas de HPV na região amazônica para as peculiaridades existentes nas variantes regionais e fornecer uma referência concisa e completa sobre o genoma do HPV 16 na região.

Palavras-chave: diversidade genética; variantes de HPV 16; Amazonas; genoma viral; regiões NCR, L1, E6 e E7.

## ABSTRACT

The Human papillomavirus is responsible for lesions in the oral mucosa, anal and urogenital tract of male and female, transmitted by direct or indirect contact with infected skin or through sexual intercourse. In women these infections can progress to cervical cancer, which is estimated incidence for the Northern region in 2010 was the largest in Brazil. The nature of the infection depends on the degree of integration of viral DNA with host DNA linked primarily to genes of oncoproteins E6 and E7 of HPV. The determination of the viral types can be held from differences in the viral capsid L1 gene and the variants of a particular type of HPV can be identified through the study of viral non-coding region. Currently the development of prophylactic vaccines against HPV particles using "pseudo-viral" formed by the L1 protein of different subtypes of high risk, while a growing number of studies that use the oncoproteins E6 and E7 in the development of therapeutic vaccines. However, it is necessary for the development of such antiviral vaccines also consider the great diversity of variants of HPV types exist, since differences between the genomic regions of these variants may influence the degree of their infections. This paper describes the complete genome sequence of a variant of HPV 16, detected in Amazonian region, using techniques of genetic engineering and the analysis of this genome by bioinformatics tools. It was observed by analysis of genetic distance that the genome of this variant has a genetic proximity of those identified in the literature as "African variants, and phylogenetic analysis, performed from the non-coding region, support this hypothesis. In addition, several mutations were detected in the genome and obtained, resulting in changes in the positions and number of restriction sites in its sequence. The major differences between the genetic regions of the genome sequenced and the corresponding variants in Africa have been observed over E7. It is expected, with that work, look for future research projects involving protein expression and genomic analysis of HPV in the Amazon region to the regional peculiarities in variants and provide a concise and complete reference on the genome of HPV 16 in the region.

Keywords: genetic diversity; HPV 16 variants; Amazonian region; viral genome; regions NCR, L1, E6 and E7.

## SUMÁRIO

<b>Introdução</b> .....	<b>9</b>
<b>Bibliografia</b> .....	<b>13</b>
<b>Objetivos</b> .....	<b>15</b>
<b>Hipóteses</b> .....	<b>16</b>
<b>Capítulo 1: Clonagem e sequenciamento do genoma de HPV 16 isolado na Amazônia</b> .....	<b>17</b>
<b>Capítulo 2: Estudos moleculares do genoma de HPV 16 isolado na Amazônia</b> .....	<b>51</b>
<b>Conclusões</b> .....	<b>75</b>
<b>Anexos</b> .....	<b>76</b>
<b>Anexo 1: Anotação do HPV 16 no GenBank</b> .....	<b>77</b>
<b>Anexo 2: Mapas de restrição</b> .....	<b>84</b>
<b>Anexo 3: Arquivo de saída do programa ModelTest 0.1.1</b> .....	<b>85</b>
<b>Anexo 4: Modelos gerados por modelagem molecular</b> .....	<b>90</b>



## INTRODUÇÃO

Os papilomavírus humanos (HPV) pertencem à família *Papillomaviridae*. São caracterizados por apresentarem genoma com dupla fita de DNA circular com aproximadamente 8.000 pares de bases e um capsídeo icosaédrico. Estão associados a lesões epiteliais e câncer cervical, podendo ser agrupados em tipos de alto risco e baixo risco. Os tipos de alto risco são comumente encontrados em câncer cervical, enquanto os considerados de baixo risco são principalmente encontrados em verrugas genitais e lesões não-malignas. Entre os vírus de alto risco encontram-se os HPV 6, 11, 44, 16, 18, 31, 33, 34, 35, 39, 45, 51, 52, 56, 58, 66, 68 e 70, e entre os de baixo risco encontram-se os tipos 6, 11, 40, 42, 43, 44, 54, 61, 72 e 81 (BURD, 2003; de VILLIERS *et al.*, 2004; RIVOIRE *et al.*, 2001).

Os HPV de alto risco são apontados como agentes de transformação maligna das células sendo, portanto, indutores de câncer, especialmente no colo uterino. Sabe-se que os tipos de HPV de alto risco estão associados a quase todos os casos de câncer no colo do útero. Desses tipos, os mais comumente encontrados neste tipo de câncer são o HPV 16 e o HPV 18. Outros fatores que contribuem para a etiologia deste tumor são: tabagismo, hipovitaminoses, multiplicidade de parceiros sexuais, iniciação sexual precoce e uso de contraceptivos orais. Segundo dados do INCA para 2011, são estimados 18 casos em cada 100.000 no Brasil; 23 casos em cada 100.000 na região Norte e 31 casos em cada 100.000 no Amazonas. (INCA, 2010).

Em 2002, Castro detectou a presença de HPV 16 em 14 amostras em um estudo no qual utilizou dois grupos de estudo: um com mulheres que procuraram os serviços de citologia da Fundação Centro de Controle de Oncologia do Amazonas (FCECON); outro com pacientes que faziam exame colpocitológico de rotina na Faculdade de Ciências Farmacêuticas da Universidade Federal do Amazonas (FCF/UFAM). No trabalho, o diagnóstico de HPV foi realizado por meio da reação em cadeia da polimerase (PCR), com iniciadores MY09 e MY11, específicos para o diagnóstico molecular de HPV. Além do HPV 16, também foi registrada a presença de HPV dos tipos 13 (relacionado à Hiperplasia Epitelial Focal, FEH, ou Doença de Heck), 33, 58, 66 e 68b (até a época descrito como “New” HPV, mas classificado posteriormente por Calleja-Macias, em 2007, como um subtipo de HPV68) (ARCHARD, 1965).

O genoma do HPV tipo 13, detectado em amostras de pacientes que apresentavam lesões na cavidade bucal e que se assemelhavam a FEH, foi isolado e sequenciado por Santos (2006) (acesso ao GenBank, DQ344807 GI: 85827580).

Em 2007, Santos detectou a presença de HPV 16 em quatro amostras, isoladas a partir de mucosa oral de pacientes atendidos na FCECON, que apresentavam carcinoma oral.

Ferreira (2007), em um estudo com amostras de mucosa cervical que apresentavam lesões pré-malignas e malignas provenientes da FCECON, detectou pela técnica de NESTED-PCR a incidência de 81,9% de HPV 16. No mesmo trabalho também foram identificados HPV 18 (4,1%), HPV 31 (3,4%), HPV 33 (4,9%), além dos tipos HPV 35, 45, 53, 56, 58 e 66 (5,7%).

O genoma do papilomavírus humano constitui-se geralmente de três regiões: uma longa região de controle (LCR ou região não codificadora, NCR), que possui função regulatória atenuante sobre os promotores dos genes E6 e E7; uma região denominada E (*Early*), devido à expressão de seus genes (E1, E2, E4, E5, E6 e E7) ocorrer durante as etapas iniciais do ciclo celular do hospedeiro; e uma região L (*Late*), com genes que se expressam durante as últimas etapas do ciclo celular do hospedeiro (L1 e L2). A região E ocupa cerca de 50% do genoma viral a partir da extremidade 5', a região L perfaz cerca de 40% e os 10% restantes correspondem à Longa Região de Controle (com aproximadamente 810 pares de bases). A identificação de novos tipos virais é feita a partir da análise da sequência gênica da proteína L1, altamente conservada entre os tipos de papilomavírus. Se duas sequências quaisquer apresentarem um grau de homologia inferior a 90% (diferenças acima de 10%) entre suas sequências de L1, poderão ser identificadas como pertencentes a dois tipos distintos de papilomavírus (BURD, 2003; de VILLIERS *et al.*, 2004; ZHENG & BAKER, 2006).

Além de sua importância na caracterização de novos subtipos virais, a proteína L1 tem sido o foco das pesquisas atuais – juntamente com a proteína L2 (outra proteína estrutural) – envolvidas no combate às infecções por HPV, através do desenvolvimento de vacinas imuno-profiláticas que induzam a produção de anticorpos capazes de neutralizar antígenos virais, utilizando-se, para isto, subunidades recombinantes dessas proteínas (STELLER, 2002).

Existem dois modos de ação imunológica das vacinas desenvolvidas para o combate ao HPV: as vacinas profiláticas, que induzem a produção de anticorpos capazes de neutralizar o antígeno viral antes de ele entrar na célula hospedeira e; as vacinas terapêuticas, que tentam induzir componentes do sistema imune a reconhecer e atacar células infectadas com HPV. As vacinas profiláticas são obtidas a partir da neutralização induzida de anticorpos antivirais antes da infecção viral, enquanto as vacinas terapêuticas são obtidas a partir da observação do destino das proteínas intracelulares virais (LOWY e SCHILLER, 2006; STELLER, 2002).

Em geral, existem quatro grandes categorias de estratégias para as vacinas terapêuticas desenvolvidas atualmente para o combate ao HPV: peptídeo-baseadas, proteína-baseadas, ácido-nucleico-baseadas e célula-baseadas. As vacinas peptídicas baseiam-se na indução de resposta imune dos linfócitos T citotóxicos (CD8+ ou CTLs), geralmente atribuída à expressão do gene alelo HLA-A\*0201 (na maioria das vezes expresso em indivíduos caucasianos), utilizando-se resíduos derivados das proteínas E6 e E7 de HPV. Vacinas protéicas, por sua vez, requerem o englobamento pelas APCs da proteína (neste caso, desnaturada) e apresentação desses resíduos peptídicos às moléculas do MHC. As vacinas de células utilizam células dendríticas para induzir resposta imune em conjunto com proteínas e ácidos nucleicos virais. E, por fim, as vacinas ácido-nucleico-baseadas, que utilizam técnicas de transferência de genes para introduzir o DNA de E6 e E7 em células-alvo, utilizando na maioria das vezes o Vaccinia vírus como vetor (LOWY e SCHILLER, 2006; STELLER, 2002).

As vacinas protéicas possuem a vantagem de incluírem todos os epítomos provavelmente imunogênicos das proteínas para todo o haplótipo do MHC, além de superarem uma limitação das vacinas peptídicas por não dependerem do conhecimento prévio do haplótipo HLA do indivíduo para a escolha do peptídeo apropriado. Os estudos com as vacinas de células dendríticas mostraram eficiência significativamente maior destas com relação às vacinas de peptídeos. A desvantagem no uso de oncogênes para o desenvolvimento é exatamente seu potencial de formação tumoral, problema que, teoricamente seria resolvido através de mutagênese induzida, mas que, na prática, comprometeriam a imunogenicidade. Além disso, os vetores de expressão viral têm causado significativo quadro de morbidez, principalmente em pacientes imunofragilizados (LOWY e SCHILLER, 2006; STELLER, 2002).

Apesar dos esforços contínuos na elaboração de vacinas terapêuticas, entretanto, os resultados têm-se demonstrado insuficientes do ponto de vista clínico, provavelmente devido a essas vacinas não induzirem a respostas imunes adequadas (LOWY e SCHILLER, 2006; STELLER, 2002).

Por outro lado, vacinas profiláticas vêm sendo desenvolvidas contra vários patógenos virais humanos. Tais vacinas têm de ser testadas visando à proteção por vários anos sem reincidências para que se torne possível sua implementação. Em face dessa relação entre o Papilomavírus humano e o câncer cervical surge a necessidade de elaborar vacinas profiláticas que interfiram nas ações, malignas ou benignas, atribuídas à infecção viral (como no caso das infecções por HPV). Porém, os esforços para desenvolver vacinas contra agentes infecciosos como o HPV, cujas doenças resultam de infecções locais, não têm tido sucesso, em contrapartida, as vacinas profiláticas aprovadas atualmente têm como alvos agentes infecciosos causadores de doenças sistêmicas (LOWY e SCHILLER, 2006; STELLER, 2002).

Recentemente, duas companhias farmacêuticas têm contribuído bastante para achar vacinas adequadas na luta contra o câncer cervical causado por HPV. Merck e GlaxoSmithKline têm desenvolvido vacinas de partículas semelhantes a subunidades virais (VLPs), compostas da proteína estrutural do capsídeo viral L1 e contendo epítomos virais de neutralização imunodominante (LOWY e SCHILLER, 2006; STELLER, 2002).

A vacina desenvolvida pela Merck é quadrivalente e possui VLPs de HPVs 6, 11, 16 e 18, enquanto a desenvolvida pela GlaxoSmithKline é bivalente e possui VLPs de HPV16 e HPV18 (LOWY e SCHILLER, 2006).

No Brasil as vacinas contra HPV vêm sendo aplicadas a altos custos (em torno de R\$ 900,00) somente pela rede privada de saúde, sendo direcionadas a homens e mulheres na faixa de nove a vinte e seis anos de idade (COLUCCI, 2011).

Os estudos de Castro (2002), Santos (2006), Santos (2007) e Ferreira (2007) são referências para a caracterização da diversidade de tipos de HPV presentes na região de Manaus e circunvizinhanças (uma vez que a FCECON serve como referência para o tratamento de oncologias na região Norte). Todavia, nenhum dos estudos citados anteriormente caracterizou o genoma do HPV 16 detectado em tais pesquisas.

Pretendeu-se, com este trabalho, caracterizar o genoma do HPV 16 encontrado no trabalho de Ferreira (2007) e compará-lo com outros da mesma espécie da família *Papillomaviridae*, descritos no banco de dados genéticos do *National Center for Biotechnology Information* - NCBI, buscando elucidar a diversidade molecular existente entre as variantes de HPV 16. Além disso, os dados genômicos advindos deste trabalho fornecerão informações de fundamental importância para estudos moleculares ao nível protéico (como desenvolvimento de fármacos e vacinas que ressaltem as peculiaridades da diversidade genética existente entre as variantes de HPV 16) e para estudos que busquem, por meio de NCR, relacionar as variantes conhecidas com novas variantes em potencial e analisar a variabilidade genética intrínseca à sequência dessas variantes (não observada pela simples identificação típica a partir de L1).

## **BIBLIOGRAFIA**

ARCHARD, H. O.; HECK, J. W.; STANLEY, H. R. Focal epithelial hyperplasia: an unusual oral mucosal lesion found in Indian children. **Oral Surg Oral Med Oral Pathol**, 20; 201-212. 1965.

BURD, E. M. Human Papilomavírus and Cervical Cancer. **Clinical microbiology Reviews**. Jan. 2003.

CALLEJA-MACIAS, I. E.; KALANTARI, M.; ALLAN, B.; WILLIAMSON, A-L.; CHUNG, L-P.; COLLINS, R. J.; ZUNA, R.E.; DUNN, S. T.; ORTIZ-LOPEZ, R.; BARRERA-SALDAÑA, H. A.; CUBIE, H. A.; CUSCHIERI, K.; VILLA, L. L.; BERNARD, H-U.; Papilomavírus Subtypes Are Natural and Old Taxa: Phylogeny of Human Papilomavírus Types 44 and 55 and 68a and -b; **Journal of virology**, 79; 6565–6569. 2005.

CASTRO, M. M. **Deteção do papilomavírus humano – HPV em esfregações endocervicais de mulheres que procuraram os serviços de citologia de Manaus**. 2002. 67p. Dissertação (Mestrado em Genética e Evolução) – Universidade Federal de São Carlos (UFSCAR) e Universidade Federal do Amazonas (UFAM), Manaus.

COLLUCCI, C. 2011. **Anvisa aprova vacina de HPV para homens**. Disponível em <<http://www1.folha.uol.com.br/equilibrioesaude>>. Acesso em 26 de jun.

- DE VILLIERS, E-M.; FAUQUET, C.; BROKER, T.R.; BERNARD, H-U.; ZUR HAUSEN, H. Classification of papillomaviruses. **Journal of Virology**, 324; 17– 27. 2004.
- FERREIRA, J. R. D. **Detecção da infecção pelo HPV e do polimorfismo da proteína p53 em pacientes com lesões cervicais atendidas na Fundação Centro de Controle de Oncologia (FCECON) do Amazonas**. 2007. Dissertação (Mestrado em Patologia Tropical) – Universidade Federal do Amazonas, Manaus.
- INCA – Instituto Nacional do Câncer. **Estimativa 2011: Incidência de câncer no Brasil**. Rio de Janeiro: Ministério da Saúde. 2010.
- LOWY, D. R.; SCHILLER, J. T. Prophylactic human papillomavirus vaccines. **Journal of Clinical Investigation**, 5(116); 1167-1173. 2006.
- RIVOIRE, W. A.; CAPP, E.; CARLETA, H. E.; SILVA, I.S.B. Bases Moleculares da Oncogênese Cervical. **Revista Brasileira de Cancerologia**, 47(2); 179-84. 2001.
- SANTOS, C. M. B. **Sequenciamento e análise do genoma do HPV 13 isolado na Amazônia**. 2006. 172p. Tese (Doutorado em Biotecnologia) – Universidade Federal do Amazonas (UFAM) e Fundação de Medicina Tropical (FMTAM), Manaus.
- SANTOS, P. J. B. **Frequência da infecção pelo vírus do papiloma humano (HPV) em lesões de carcinoma de células escamosas da boca e sua relação com o polimorfismo do códon 72 de p53 em pacientes atendidos na Fundação Centro de Controle de Oncologia (FCECON), em Manaus, Amazonas**. 2007. 160p. Tese (Doutorado em Biotecnologia) – Universidade Federal do Amazonas (UFAM) e Fundação Centro de Controle de Oncologia (FCECON), Manaus.
- STELLER, M. A. Cervical Cancer Vaccines: Progress and Prospects. **Journal of the Society of Gynecology Investigation**, 9; 254–264. 2002
- ZHENG, Z-M.; BAKER, C. C. Papillomavirus genome structure, expression and post-transcriptional regulation. **Front. Biosci**, 11; 2286-2302. 2006.

## **OBJETIVOS**

### **1. GERAL**

- Caracterizar o genoma do HPV 16 isolado no Estado do Amazonas.

### **2. ESPECÍFICOS**

- Amplificar o genoma do HPV 16 em fragmentos utilizando a técnica de PCR *overlapping*;
- Clonar e sequenciar os fragmentos do genoma de HPV 16, obtidos pela PCR;
- Realizar a análise filogenética por meio da comparação das sequências nucleotídicas da região NCR do HPV 16 isolado com as correspondentes de outras variantes de HPV 16 e demais tipos pertencentes à família *Papillomaviridae*.
- Predizer um modelo, por métodos de bioinformática, das estruturas secundárias e terciárias das proteínas L1, E6 e E7 do HPV 16 isolado e compará-las com as estruturas preditas para o protótipo geneticamente mais próximo.

## **HIPÓTESES**

### **Hipótese 1**

A caracterização do genoma do HPV 16 isolado no Amazonas poderá ser realizada pelos métodos de engenharia genética abordados no trabalho.

Os modelos obtidos *in silico* demonstrará que as estruturas secundárias e terciárias das proteínas L1, E6 e E7 apresentarão suas características morfofuncionais conservadas com relação às respectivas proteínas das outras variantes do HPV 16 descritas na literatura.

As relações filogenéticas entre o grupo formado pela variante considerada no presente trabalho e o grupo-irmão escolhido, bem como as existentes entre o formado pelo grupo-irmão e o das variantes e o grupo-externo, poderão ser inferidas com a utilização dos dados moleculares de NCR.

### **Hipótese 2**

Os métodos de engenharia genética abordados no trabalho não poderão caracterizar o genoma do HPV 16 isolado no Amazonas.

As estruturas secundárias e terciárias das proteínas L1, E6 e E7 terão suas características morfofuncionais modificadas ou perdidas com relação às respectivas proteínas das outras variantes do HPV 16 descritas na literatura.

Os dados moleculares de NCR não serão suficientes para estabelecer relações filogenéticas entre o grupo formado pela variante considerada no presente trabalho e o grupo-irmão escolhido, bem como as existentes entre o formado pelo grupo-irmão e o das variantes e o grupo-externo. Ou seja, a reconstrução da filogenia a partir de NCR resultará em topologias politômicas.



**CAPÍTULO 1**  
**CLONAGEM E SEQUENCIAMENTO DO GENOMA DE HPV 16**  
**ISOLADO NA AMAZÔNIA**  
(A ser submetido ao *Virology*)

## **Clonagem e sequenciamento do genoma de HPV 16 isolado na Amazônia**

Roberto Alexandre Alves Barbosa Filho<sup>1</sup>; Cristina Maria Borborema dos Santos<sup>1</sup>; Spartaco Astolfi Filho<sup>1</sup>; Junia Raquel Dutra Ferreira<sup>1</sup>; André Santos Silva<sup>1</sup>; Mayara Rachel Silva Vilela<sup>1</sup>; Evelyn Farias Costa<sup>1</sup>.

1 Laboratório de Diagnóstico Molecular, Divisão de Biotecnologia, Centro de Apoio Multidisciplinar, Universidade Federal do Amazonas, Manaus, Amazonas, Brasil.

Endereço eletrônico:

RAABF: [robertoaabfilho@hotmail.com](mailto:robertoaabfilho@hotmail.com)

CMBS: [crisantos@ufam.edu.br](mailto:crisantos@ufam.edu.br)

SAF: [spartaco.biotec@gmail.com](mailto:spartaco.biotec@gmail.com)

## INTRODUÇÃO

O papilomavírus humano é um vírus epiteliotrópico que infecta superfícies mucosas e cutâneas em uma larga escala de animais. Mais de duzentos tipos de papilomavírus têm sido descritos baseando-se na sequência de bases nitrogenadas do DNA, mostrando diferenças genômicas. Oitenta e cinco (85) dos genótipos de HPV são bem caracterizados, enquanto que cento e vinte (120) isolados são parcialmente sequenciados caracterizando novos genótipos em potencial (Burd, 2003; Santos, 2006).

Em todos os papilomavírus, as três regiões do genoma (E, L e NCR) são separadas por dois sítios de poliadenilação (pA): early pA e late pA. No HPV 16 existem duas regiões promotoras principais: a P97 (posicionada *upstream* da ORF E6), responsável pela expressão dos genes da região E, e; a P670 (localizada dentro da ORF E7), responsável pela expressão dos genes da região L. Outros promotores menores já foram identificados na região E do genoma, mas suas atividades ainda não foram esclarecidas. Além disso, as regiões NCR e L1 apresentam resíduos de CpG (em posições *upstream* de p97) que podem sofrer metilação e conseqüente regulação negativa (Grassmann *et al.*, 1996; Smotkin e Wettstein, 1986).

A P97 do HPV 16 equivale à P99 de HPV 31 e P105 de HPV 18, tem alta afinidade pela RNA polimerase e é regulada por quatro regiões consensuais (sítios de ligação a E2), AACC(N6)GGT, presentes na NCR. Essas quatro regiões são reconhecidas por fatores de transcrição celulares e pela proteína viral reguladora E2, que irá reprimir a transcrição dos genes E ao inibir a formação do complexo de iniciação da transcrição, interferindo na atividade da proteína ligante a região TATA (TBP) e do fator de transcrição IID (TFIID). Para poder integrar-se ao genoma da célula hospedeira, o genoma viral precisa ser linearizado através da clivagem e conseqüente

não expressão do gene E2, o que estimulará a expressão dos genes E, resultando nas oncoproteínas E6 e E7 (Hummel *et al.*, 1992; Ozbun e Meyers, 1997; Schneider e Schwarz, 1986).

A região P670 do HPV 16 equivale à P742 do HPV 31, iniciando a transcrição dos genes L. Sua atividade, contudo, é induzida somente em células diferenciadas, através da via C das proteína-quinases. A ativação da transcrição a partir do promotor da região L pode ocorrer em resposta a sinais de diferenciação celular e amplificação do DNA. A transcrição a partir desses promotores é regulada por sequências presentes nos genes E6 e E7. Como a iniciação da transcrição dos genes L ocorre ainda dentro do gene de E7, o pré-RNA resultante da transcrição é um híbrido, cuja extremidade 5' retém parte do gene de E7, enquanto a extremidade 3' apresenta os genes de L1 e L2 (Danos *et al.*, 1982; Rosenstierne *et al.*, 2003).

A região NCR não apresenta função codificadora de proteínas, mas apresenta a origem de replicação, bem como os sítios de ligação a fatores de transcrição, que atuam na regulação da ligação da RNA polimerase II aos promotores das regiões E e L. As proteínas E1, E2 e E4 atuam na regulação da replicação e da transcrição. Os oncogenes E5, E6 e E7 modulam o processo de crescimento e morte celular. As proteínas L1 e L2 formam o capsídeo viral icosaédrico (Burd, 2003).

As proteínas E1 e E2 ligam-se ao DNA viral, por domínio de ligação a DNA presentes em suas porções C-terminais, em regiões intercaladas. Os sítios de ligação a E1 são ricos em resíduos ApT. E1 e E2 são requeridas para a replicação epissomal do genoma e a afinidade de ligação de E1 ao DNA é potencializada pela presença de E2. A integração do DNA viral ao DNA do hospedeiro é feita a partir dos genes rompidos de E1 e E2 (Mcmurray *et al.*, 2001).

A oncoproteína E6 interfere no ciclo celular ao inativar a proteína p53, que impede via apoptose, a multiplicação de um DNA danificado. O gene 53 localiza-se no cromossomo 17 e codifica a proteína p53 (53 kDa), que participa do processo de morte celular programada. A p53 ativa a expressão de proteínas como a p21, que se ligará a um complexo ciclina-CDK e interagirá com a BAK, GADD45 e APCN, culminando com a interrupção de G1, induzindo a apoptose ou permitindo que a célula repare o DNA antes do início de S (Mcmurray *et al.*, 2001).

A proteína E7 interfere no ciclo celular ao ligar-se com a pRB (proteína do retinoblastoma). O gene RB está localizado no cromossomo 13 e o produto do gene, a proteína nuclear pRB (105 KDa), forma um complexo com o fator de transcrição celular E2F1 (pRB-E2F1) durante a fase G1 do ciclo celular resultando na repressão da transcrição dos genes envolvidos na proliferação celular e, conseqüentemente, na replicação do DNA. A formação do complexo E7-pRB induz a fosforilação de pRB, que, por sua vez, perde afinidade pela ligação com o receptor de membrana E2F1 (proteína que regula positivamente o processo de transcrição) (Mcmurray *et al.*, 2001).

A proteína E5 tem um potencial oncogênico nos HPV que infectam o trato genital feminino. Ela induz a ligação do fator de crescimento epidérmico ao seu receptor (EGFR) na membrana celular, facilitando sua entrada na célula. No endossomo, E5 também liga-se à ATPase inibindo a acidificação do endossomo e prolongando a atividade de EGF. Os outros tipos de HPV associados à *Epidermodysplasia Verruciformis* não apresentam E5 ou esta não apresenta similaridade com a E5 dos HPV genitais (Mcmurray *et al.*, 2001).

Duas ORF adicionais para a região E, os genes E3 e E8, são descritas em alguns vírus, mas somente E8, presente em BPV 1 e HPV 31, está relacionada com a tradução de uma proteína efetivamente funcional, que formará um complexo pós-processamento

com a E2C, atuando na regulação negativa da replicação e transcrição. Em papilomavírus que infectam coelhos, o gene E8 tem sido caracterizado como oncogene, com características similares às de E5 de BPV 1 e de vários tipos de HPV. O papel de E3 – encontrado, até o presente, somente em BPV – no ciclo viral ainda não é bem esclarecido (Choe *et al.*, 1989; Han *et al.*, 1998; Harry e Wettstein, 1996; Hubbert *et al.*, 1988; Lambert *et al.*, 1987; Stubenrauch *et al.*, 2000; Stubenrauch *et al.*, 2001).

As proteínas L1 e L2 formam o capsídeo viral icosaédrico com 360 cópias de L1 (proteína maior do capsídeo) e 12 cópias de L2 (proteína menor do capsídeo). As proteínas do capsídeo também vêm sendo estudadas no combate às infecções por HPV, com o desenvolvimento de vacinas imuno-profiláticas que induzam a produção de anticorpos capazes de neutralizar antígenos virais, utilizando-se para isto, subunidades recombinantes dessas proteínas. Além desses estudos, porém com resultados menos satisfatórios, as proteínas oncogênicas E6 e E7 também foram testadas na fabricação de vacinas terapêuticas. Os trabalhos de Harper (2004), Lowy e Schiller (2006), Steller (2002) e Villa *et al.* (2005) fornecem revisões mais detalhadas sobre estratégias de vacinas contra HPV (de Villiers *et al.*, 2004).

Este trabalho descreve o sequenciamento do genoma completo do HPV 16 isolado no Estado do Amazonas, comparando-o com as sequências de HPV 16 depositadas no GenBank, bem como a análise completa de sua sequência nucleotídica e das possíveis alterações que tenham ocorrido em sua sequência nucleotídica que possam levar a alterações na proteína expressa.

# MÉTODOS

## 1. Amostras

O presente estudo é parte integrante do projeto “Análise genômica de vírus isolados na Amazônia” aprovado no Comitê de Ética em Pesquisas da UFAM (CAAE nº. 0074.0.115.114-07).

A amostra analisada foi isolada do trabalho de Ferreira (2007) e consiste em DNA genômico isolado de esfregaço endocervical, de uma paciente manauense com 34 anos de idade, diagnosticada citologicamente com neoplasia intraepiélial cervical II (NICII). Este material encontra-se sob tutela do Laboratório de Diagnóstico Molecular do Centro de Apoio Multidisciplinar da Universidade Federal do Amazonas – LDM/CAM/UFAM, e está armazenado a -20 °C. A amostra mostrou-se positiva para HPV 16 em trabalhos anteriores.

## 2. Reações em cadeia da polimerase (PCR)

Todas as reações em cadeia da polimerase (PCR) foram realizadas no aparelho termociclador *Veriti* da *Applied Biosystems*<sup>®</sup>.

Para certificar a integridade do DNA da amostra, a mesma foi submetida à reação de amplificação utilizando-se os iniciadores ISO05 descritos por Pontes (2003) que amplificam uma região de microssatélite (GATA)<sub>13</sub> do gene da golgina (GI:728080) localizado no loco 15q13.3 do cromossomo 15 humano. O produto da amplificação consiste num fragmento de, aproximadamente, 170 pb. O sistema de reação foi realizado em volume final de 20,0 µL: 2,2 µL de água destilada e autoclavada, 0,3 µL de enzima *Taq Platinum* DNA polimerase 5 U/µL, 2,5 µL de desoxinucleotídeos trifosfatos

(dNTPs: dATP, dCTP, dGTP, dTTP) 2,5 mM, 5,0 µL do par de iniciadores ISO05 com concentração de 5 pmol/µL, 2,5 µL de cloreto de magnésio (MgCl<sub>2</sub>) 20 mM, 2,5 µL do tampão da enzima 10X e 5,0 µL do DNA molde. O termociclo da PCR com os iniciadores ISO05 ocorreu nas seguintes condições: 95 °C, por 2 minutos, para pré-aquecimento; 95 °C, por 1 minuto, para desnaturação; 55 °C, por 1 minuto, para o anelamento; 72 °C, por 1 minuto, para extensão; Repetir 39 vezes, a partir da segunda desnaturação, até a extensão; seguido de 72 °C, por 5 minutos, para extensão final e; 4 °C, até a retirada do termociclador.

Utilizando-se os iniciadores universais MY11 e MY09 realizou-se uma PCR para amplificar uma região interna de 450 pb do gene da proteína estrutural L1 do capsídeo viral e que é altamente conservada entre os diversos tipos de HPV. O sistema para PCR, utilizando os iniciadores MY11 e MY09, foi realizado em volume final de 25 µL: 13,4 µL de água destilada e autoclavada, 0,1 µL de enzima *Taq Platinum* DNA polimerase de alta fidelidade (*High Fidelity*) 5 U/µL, 0,5 µL de dNTPs 10 mM, 2,5 µL de cada um dos iniciadores citados (MY11 e MY09) com concentração de 5 pmol/µL, 1,0 µL de sulfato de magnésio (MgSO<sub>4</sub>) 50 mM, 2,5 µL do tampão da enzima 10X e 2,5 µL do DNA molde. O termociclo da PCR com os iniciadores MY11 e MY09 consistiu em: 94 °C, por 2 minutos, para pré-aquecimento; 94 °C, por 1 minuto, para desnaturação; 55 °C, por 1 minuto, para o anelamento; 68 °C, por 1 minuto, para extensão; Repetir 39 vezes, a partir da segunda desnaturação, até a extensão; seguido de 68 °C, por 5 minutos, para extensão final e; 4 °C, até a retirada do termociclador (Qu *et al.*, 1997).

As sequências nucleotídicas dos iniciadores utilizados para a amplificação dos fragmentos genômicos de HPV da amostra foram definidas a partir de sequências do HPV 16 depositadas no banco de dados do *National Center of Biotechnology*



*Information* – NCBI (<http://www.ncbi.nih.nlm.gov>). As sequências de oligonucleotídeos foram analisadas, editadas e selecionadas com o auxílio das ferramentas Perl Primer, ClustalW e Primer 3. O sistema para as reações de PCR, utilizando-se os iniciadores específicos para HPV16, foi realizado em volume final de 25 µL: 13,0 µL de água destilada e autoclavada; 0,5 µL de enzima *Taq Platinum* DNA polimerase de alta fidelidade (*High Fidelity*) 5 U/µL; 0,5 µL de dNTPs 10 mM; 2,5 µL de cada um dos iniciadores com concentração de 5 pmol/µL; 1,0 µL de sulfato de magnésio (MgSO<sub>4</sub>) 50 mM; 2,5 µL do tampão da enzima 10X e; 2,5 µL do DNA molde. Assim, o termociclo da PCR com os iniciadores específicos de HPV 16 apresentou as seguintes características: 94 °C, por 2 minutos, para pré-aquecimento; 94 °C, por 1 minuto, para desnaturação; 55 a 57 °C (dependendo dos iniciadores utilizados em cada reação) por 1 minuto para o anelamento; 68 °C, por 1 minuto, para extensão; Repetir 39 vezes, a partir da segunda desnaturação, até a extensão; seguido de 68 °C, por 5 minutos, para extensão final e; 4 °C, até a retirada do termociclador (Marshall, 2004; Rozen e Skaletsky, 2000; Thomson *et al.*, 1994).

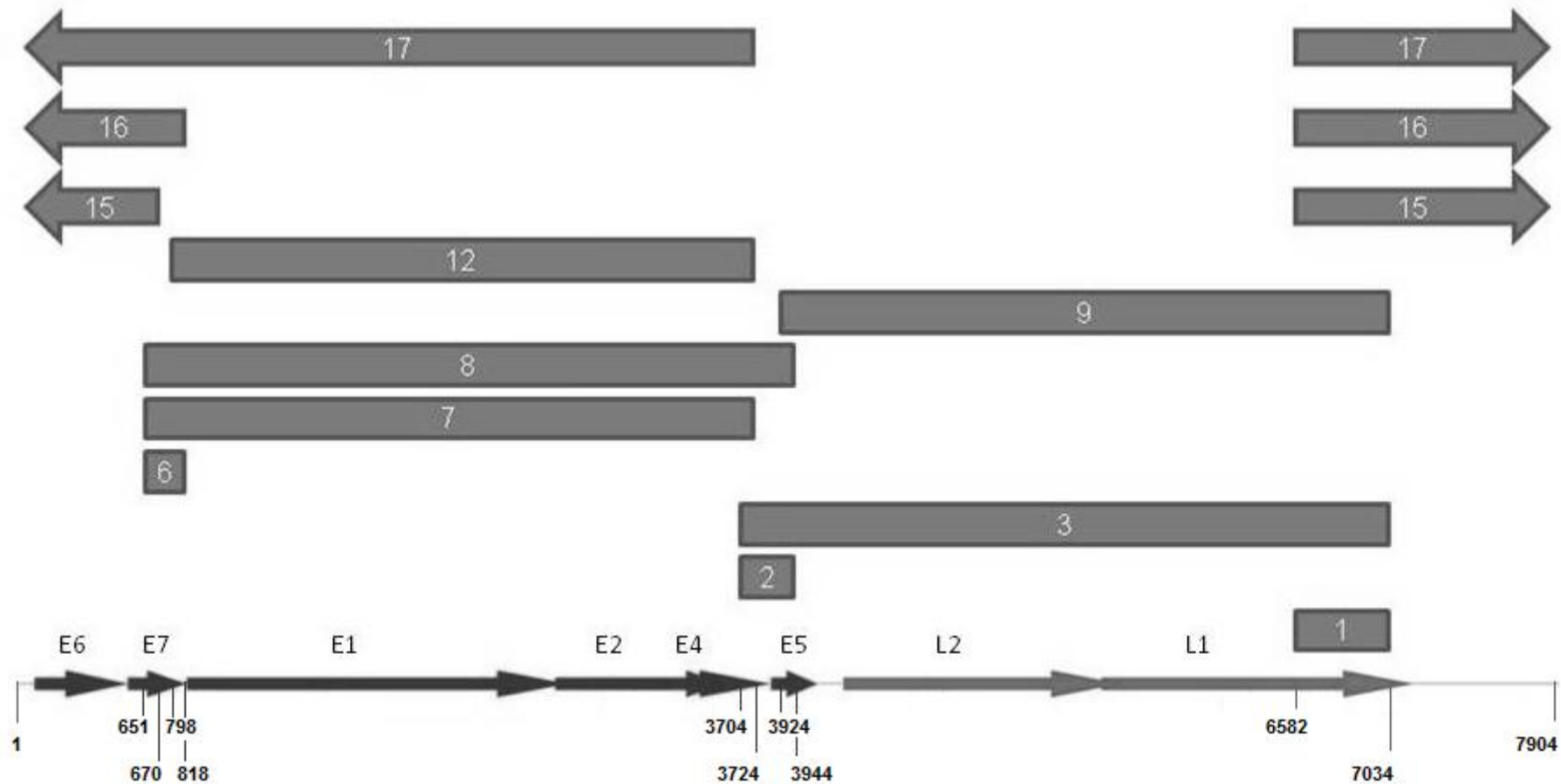
A composição de nucleotídeos e o tamanho em número de bases de cada iniciador desenvolvido no trabalho estão representados na tabela 1. As regiões em que estes iniciadores se anelam no genoma viral e o tamanho dos fragmentos amplificados por suas combinações estão representados na tabela 2 e na figura 1.

**Tabela 1 – Composição dos iniciadores utilizados no estudo.** A sequência de bases nitrogenadas está disposta no sentido 5'-3'. As letras M, Y e W dos iniciadores MY09 e MY11 representam bases degeneradas (M pode corresponder a A ou C, Y pode corresponder a C ou T e W pode corresponder a A ou T).

PRIMER	TAMANHO	SENTIDO	COMPOSIÇÃO	REFERÊNCIA
MY11	21	FORWARD	GCMCAGGGWCTATAAYAATGG	Manos <i>et al.</i> , 1989
MY09	20	REVERSE	CGTCCMARRGGAWACTGATC	Manos <i>et al.</i> , 1989
GP5+	23	FORWARD	TTTGTTACTGTGGTAGATACTAC	de Roda Husman <i>et al.</i> , 1995
GP6+	25	REVERSE	GAAAAATAAACTGTAAATCATATTC	de Roda Husman <i>et al.</i> , 1995
16F1	21	FORWARD	CTACATGGCATTGGACAGGAC	
16F2	20	FORWARD	CAGCTCAGAGGAGGAGGATG	
16F3	21	REVERSE	GCAGCGGACGTATTAATAGGC	
16F4	21	FORWARD	GGAAGACCTGTTAATGGGCAC	
16R1	21	FORWARD	GCCTATTAATACGTCCGCTGC	
16R2	21	REVERSE	GTGCCCATTAACAGGTCTTCC	
16R3	21	REVERSE	GTCCTGTCCAATGCCATGTAG	
16R4	20	REVERSE	CATCCTCCTCCTCTGAGCTG	

**Tabela 2 – Reações de PCR com os iniciadores utilizados neste estudo.** A descontinuidade observada na numeração das reações decorre de reações que foram planejadas, porém não executadas no decurso da pesquisa.

REAÇÃO	PRIMER "F"	PRIMER "R"	POSIÇÃO DE ANELAMENTO DA EXTREMIDADE 5'	POSIÇÃO DE ANELAMENTO DA EXTREMIDADE 3'	TAMANHO DO FRAGMENTO AMPLIFICADO (pb)
Reação 1	MY11	MY09	6582	7034	452
Reação 2	16F1	16F3	3704	3944	240
Reação 3	16F1	MY09	3704	7034	3330
Reação 6	16F2	16R2	651	818	167
Reação 7	16F2	16R3	651	3724	3073
Reação 8	16F2	16F3	651	3944	3293
Reação 9	16R1	MY09	3924	7034	3110
Reação 12	16F4	16R3	798	3724	2926
Reação 15	MY11	16R4	6582	670	1992
Reação 16	MY11	16R2	6582	818	2140
Reação 17	MY11	16R3	6582	3724	5046



**Figura 1 – Regiões do HPV 16 amplificadas pelos iniciadores nas reações de PCR realizadas neste estudo.** As oito setas presentes na linha de base da figura indicam as oito regiões gênicas de HPV 16 e as barras acima dessa linha representam os amplicons de cada reação. As setas que representam as reações 15, 16 e 17 mostram a continuidade dos amplicons dada a estrutura circular do genoma do HPV 16.

Os amplicons foram analisados por eletroforese em gel de agarose. Foram utilizados géis nas concentrações 0,8 % (para visualizar fragmentos maiores que 600 pb) e 2,0 % (para fragmentos de 100 a 600 pb). Os géis foram corados com brometo de etídeo 1 µg/mL, para possibilitar a visualização dos fragmentos de DNA sob luz UV.

### **3. Purificação do DNA amplificado**

O produto amplificado por PCR com iniciadores específicos foi submetido à purificação em colunas *GE Healthcare*<sup>®</sup> *Sephaglas Bandprep*<sup>TM</sup> *Purification Kit*, conforme instruções do fabricante. Neste procedimento utilizou-se uma microcentrífuga *Eppendorf-Centrifuge 5410* e as centrifugações procederam a 12.000 g. Do volume total de material purificado (30 µL), retirou-se 3 µL para análise por meio de eletroforese em gel de agarose. O restante do material purificado foi utilizado para fins de determinação da sequência nucleotídica da amostra pelo método de sequenciamento automático ou para clonagem em vetor TOPO (INVITROGEN<sup>®</sup>) de células TOP10 eletrocompetentes.

### **4. Transformação bacteriana**

Após a purificação, o DNA amplificado foi inserido em vetor de clonagem TOPO (INVITROGEN<sup>®</sup>) segundo instruções do fabricante. O sistema de ligação do DNA ao vetor de clonagem constituiu-se de: 2,0 µL de tampão 2X; 1,0 µL do vetor TOPO 50,0 ng/µL; 2,0 µL do produto purificado e; o volume necessário para completar 6,0 µL de água MilliQ livre de nucleases. A ligação ocorreu durante um período de 10 minutos a 4 °C.

Foram utilizados 2,0 µL do sistema de ligação para transformar por eletroporação 50,0 µL de células de *E. coli*, linhagem TOP10 eletrocompetentes

(Invitrogen Life Technologies). Os 52,0 µL do sistema foram aplicados em cuvetta própria para eletroporação e levados ao eletroporador programado para uma voltagem de 1.900 V. Após a eletroporação, as células foram imersas em 1,0 mL de meio LB líquido e incubadas por 1 hora a 37 °C em estufa com agitação de 170 RPM, para possibilitar a expressão do gene de resistência ao antibiótico ampicilina e recuperar as células eletroporadas que foram, a seguir, plaqueadas em meio de cultura LB/ágar com ampicilina a 200 µg/mL e incubadas por 12 horas a 37 °C. As colônias transformadas foram então submetidas a PCR, onde utilizou-se; 15,5 µL de água destilada e autoclavada; 0,5 µL de enzima *Taq Platinum* DNA polimerase de alta fidelidade (*High Fidelity*) 5 U/µL; 0,5 µL de dNTPs 10 mM; 2,5 µL de cada um dos iniciadores com concentração inicial de 5 pmol/µL; 1,0 µL de sulfato de magnésio (MgSO<sub>4</sub>) 50 mM e; 2,5 µL do tampão da enzima a 10X.

O termociclo da PCR de colônias foi realizado nas seguintes condições: 94 °C, por 2 minutos, para pré-aquecimento; 94 °C, por 1 minuto, para desnaturação; 57 °C, por 1 minuto, para o anelamento; 68 °C, por 1 minuto, para extensão; Repetir 39 vezes, a partir da desnaturação, até a extensão; 68 °C, por 5 minutos, para extensão final e; 4 °C, até a retirada do termociclador. Após a verificação por PCR da presença do inserto, os plasmídios das colônias transformadas foram extraídos utilizando-se o kit de purificação *Illustra PlasmidPrep Mini Spin Kit – GE Healthcare*, seguindo as instruções do fabricante.

## **5. Sequenciamento**

O sequenciamento consistiu, basicamente, em três etapas: 1) reação de sequenciamento e precipitação dos componentes da reação de sequenciamento; 2) leitura no sequenciador automático; 3) análise e edição inicial das sequências.

Para a reação de sequenciamento utilizou-se o seguinte sistema: 2,5 µL – 4 µL do DNA; 2,0 µL do pré-mix *DYEnamic ET Dye Terminator Cycle Sequencing Kit for MegaBACE DNA Analysis Systems* (a proporção de volume do pré-mix a ser utilizada com relação ao volume total é de 1:5); 1,0 µL de um dos iniciadores (Senso ou Antisenso) 5 pmol/µL; e o volume de água Milli-Q para completar o volume final de 10 µL. O sistema foi, então, levado ao termociclador *Veriti Applied Biosystems*<sup>®</sup>, programado para realizar o seguinte termociclo: 95°C, por 25 segundos, para pré-aquecimento; 95°C, por 15 segundos, para desnaturação; 50°C, por 20 segundos, para o anelamento; 60°C, por 1 minuto, para extensão; Repetir 29 vezes, a partir da desnaturação, até a extensão; 4°C, até a retirada do termociclador.

Após a reação de sequenciamento realizou-se a precipitação do produto da reação. Adicionou-se 1,0 µL de acetato de amônio 7,5 M (kit de sequenciamento *DYEnamic ET Dye Terminator Cycle Sequencing Kit for MegaBACE DNA Analysis Systems* da *Amersham Bioscience*) e 27,5 µL de etanol absoluto ao produto da reação de sequenciamento. Esse sistema foi homogeneizado por alguns minutos e incubado por 20 minutos à temperatura ambiente. A placa foi então envolvida em alumínio laminado a fim de evitar incidência de luz. Seguindo-se a etapa de incubação, a placa foi centrifugada a 4.000 g por quarenta minutos, a 4°C em centrífuga refrigerada *Eppendorf 5804R* e o sobrenadante foi descartado. Em seguida adicionou-se 120,0 µL de etanol 70% seguidos de agitação por alguns segundos. A placa foi novamente centrifugada a 4.000 g por 10 minutos e o sobrenadante foi novamente descartado, seguindo-se nova centrifugação com a placa invertida a 700 g por alguns segundos. Deixou-se a placa secar no fluxo laminar até que o etanol evaporasse completamente. O DNA foi, por fim, ressuspenso em 10,0 µL de *Loading buffer* (tampão de lavagem composto por formamida a 70% e EDTA 1mM). A placa foi vedada e agitada vigorosamente por dois

minutos e, a seguir, centrifugou-se a 700 g por alguns segundos. Em seguida, as amostras foram sequenciadas, utilizando-se dideoxinucleotídeos marcados e incorporados com indicador de fluorescência, nos sequenciadores automáticos “MegaBACE 1000 (Amersham Biosciences)” e ABI (Applied Biosystems®). Para a injeção das amostras no aparelho utilizou-se 3KV por 80 segundos; a corrida processou-se, a 6KV, por 300 minutos, sob uma temperatura de 44°C.

## 6. Análises

Ao final de todo seqüenciamento a edição inicial das seqüências foi realizada em três etapas: a) análise da qualidade das leituras do sequenciador com base nos gráficos dos eletroferogramas e conversão dos picos desses eletroferogramas em seqüências de bases, realizada pelas ferramentas PHRED e CAP3 (disponíveis no sítio: [www.bioinformatica.ufam.edu.br](http://www.bioinformatica.ufam.edu.br)); b) alinhamento múltiplo de seqüências e c) montagem do genoma a partir dos fragmentos sequenciados (“READS de seqüenciamento”), ambos realizados pela ferramenta ClustalW, disponível no programa BioEdit, que permite editar as seqüências de forma a disponibilizá-las para um alinhamento linear exato, possibilitando a visualização de eventuais diferenças entre elas (potenciais mutações). A associação das seqüências obtidas a seqüências de HPV 16 já conhecidas foi realizada por comparações com outras seqüências de HPV depositadas no banco de dados do NCBI (*National Center for Biotechnology Information*) utilizando-se a ferramenta BLASTn, para comparação entre seqüências nucleotídicas, e tBLASTx, para comparação entre seqüências protéicas inferidas a partir das cadeias nucleotídicas (ambas disponíveis no sítio: [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). A distância genética entre o HPV 16 isolado e as demais variantes de HPV 16 conhecidas, além da probabilidade de substituição de nucleotídeos e da freqüência de mutações

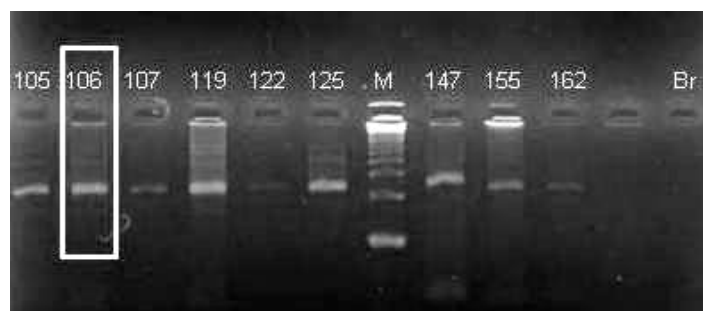
ocorridas entre o HPV seqüenciado e as variantes geneticamente mais próximas, pode ser estimada através do modelo de distância de Tamura e Nei (1993), aplicado à região NCR, disponível no programa MEGA 4.0 (Altschul *et al.*, 1997; Ewing *et al.*, 1998; Hall, 2001; Huang e Madan, 1993; Tamura e Nei, 1993; Tamura *et al.*, 2007).

A construção do mapa de restrição do genoma seqüenciado foi realizada com o auxílio da ferramenta NebCutter, disponível no sítio tools.neb.com (Vincze *et al.*, 2003).

## RESULTADOS E DISCUSSÃO

### 1. Extração de DNA genômico e controle de qualidade do DNA por PCR

O sucesso da extração de DNA pelo método de fenol-clorofórmio pode ser observado no perfil dos resultados de PCR com os iniciadores ISO05 (figura 2), no qual se observa os fragmentos correspondentes a aproximadamente 200 pb, pertencentes ao gene da golgina humana já descritos anteriormente.

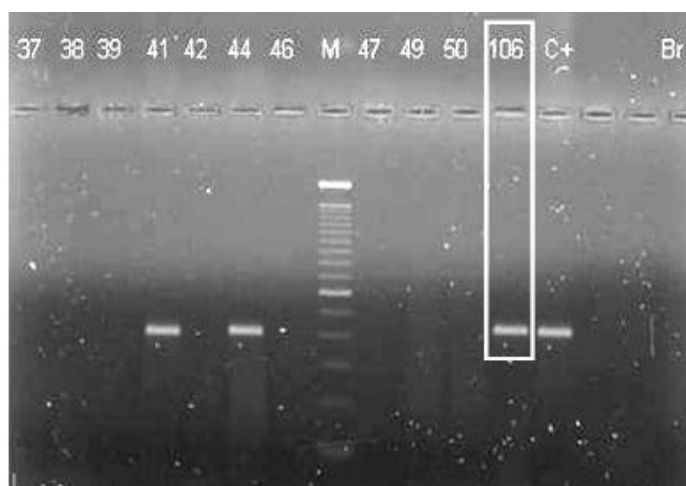


**Figura 2 – Verificação da integridade do material genético – perfil eletroforético em gel de agarose 2,0% dos fragmentos amplificados utilizando-se iniciadores ISO05.** A amostra em destaque corresponde à utilizada no estudo. As siglas Br e C- indicam controles negativos utilizados durante as PCR. A escada alélica utilizada como marcador de peso molecular neste perfil eletroforético foi o 500 pb Ladder Invitrogen®. Como mostra a figura, todos os fragmentos apresentam tamanho de aproximadamente 200 pb.



## 2. PCR com iniciadores para diagnóstico molecular de HPV

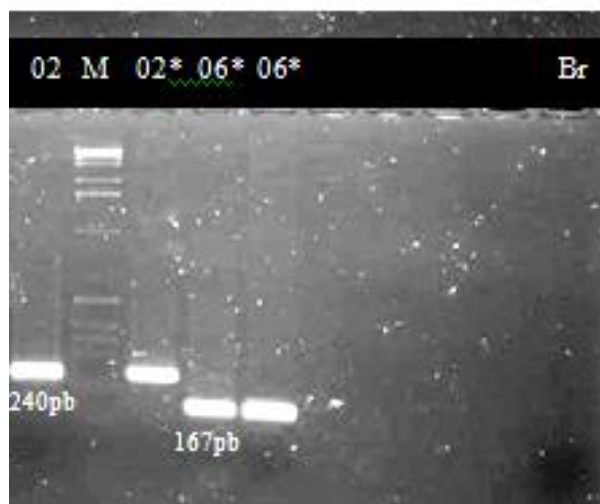
Após a verificação da qualidade do material genético por PCR com iniciadores para DNA genômico humano, verificou-se a presença de DNA viral na amostra estudada por PCR com iniciadores utilizados na detecção de HPV (MY09/11). A figura 3 mostra os resultados positivos correspondentes à presença de DNA viral com as reações de MY09/11 (reação 1 descrita na tabela 2), onde pode-se observar fragmentos equivalentes a 450 pb, correspondentes a uma região interna do gene L1 de HPV.



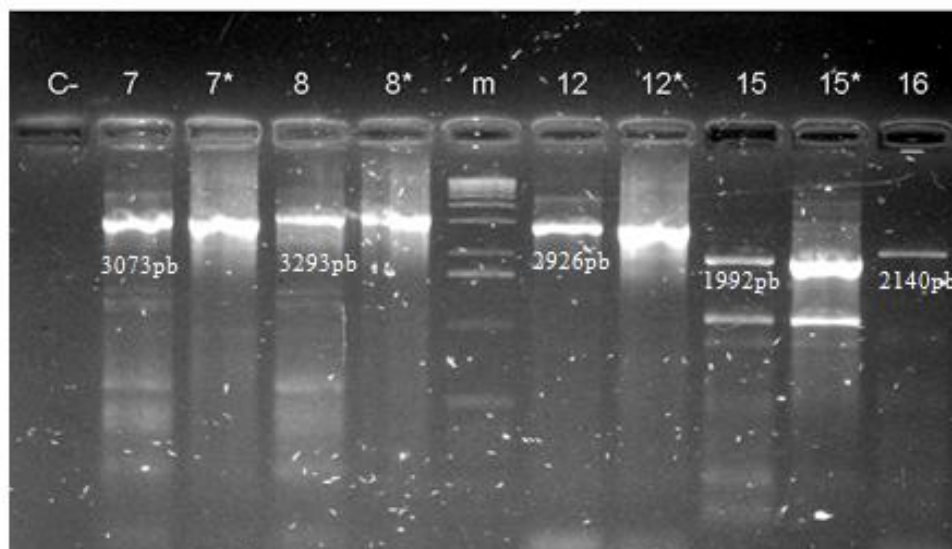
**Figura 3 – Detecção de HPV utilizando os iniciadores MY09/11 por – perfil eletroforético em gel de agarose 2,0%. O marcador de peso molecular neste perfil eletroforético foi o 500 pb Ladder Invitrogen®. Como mostra a figura, todos os fragmentos apresentam tamanho de aproximadamente 450 pb.**

## 3. PCR com iniciadores específicos e purificação dos amplicons

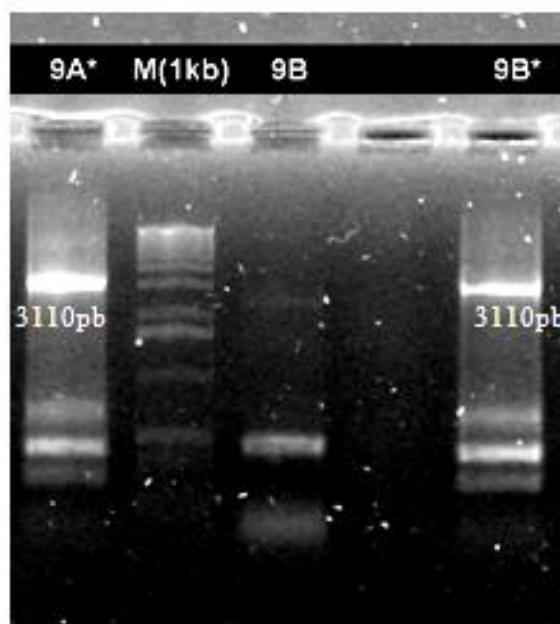
As PCR com iniciadores especificamente desenvolvidos para HPV 16 resultaram em ampliações de fragmentos de diversos comprimentos, que puderam ser observados por eletroforese em gel de agarose. As figuras 4-7 mostram os resultados de PCR contendo os fragmentos correspondentes aos produtos desejados.



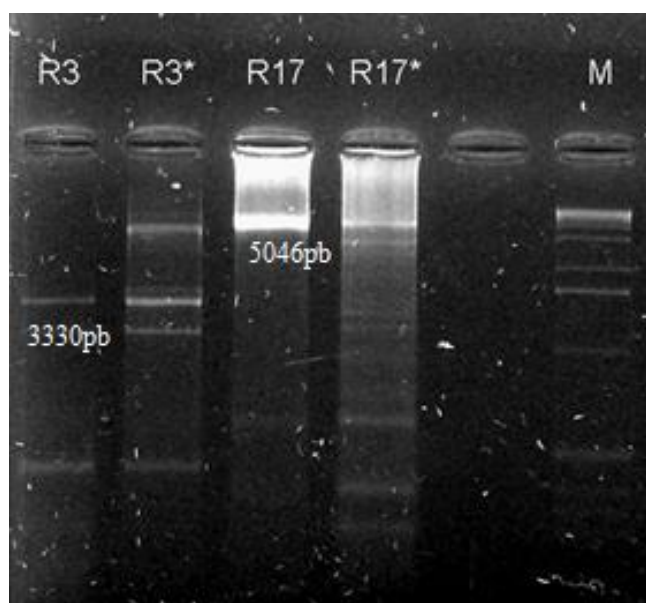
**Figura 4 – Perfil eletroforético em gel de agarose 2,0% dos produtos de PCR com iniciadores 16F1/F3 (2 e 2+) e 16F2/R2 (6 e 6+).** O marcador de peso molecular neste perfil eletroforético foi o 500 pb Ladder Invitrogen®.A figura mostra os amplicons das reações 2 e 6 com os seus respectivos tamanhos em pb.



**Figura 5 – Perfil eletroforético em gel de agarose 0,8% dos produtos de PCR com iniciadores 16F2/R3 (7 e 7\*), 16F2/F3(8 e 8\*), 16F4/R3 (12 e 12\*), MY11/16R4 (15 e 15\*) e MY11/16R2 (16 e 16\*).** O marcador de peso molecular neste perfil eletroforético foi o 1 Kb Ladder Invitrogen®. A figura mostra os amplicons das reações 7, 8, 12, 15 e 16 com os seus respectivos tamanhos em pb, além de outros fragmentos inespecíficos, que foram eliminados pela purificação descrita no texto.



**Figura 6 – Perfil eletroforético em gel de agarose 0,8% dos produtos de PCR com iniciadores 16R1/MY09 (9A, 9A\*, 9B e 9B\*).** O marcador de peso molecular neste perfil eletroforético foi o 1 Kb Ladder Invitrogen®. A figura mostra os amplicons da reação 9 com os seus respectivos tamanhos em pb.

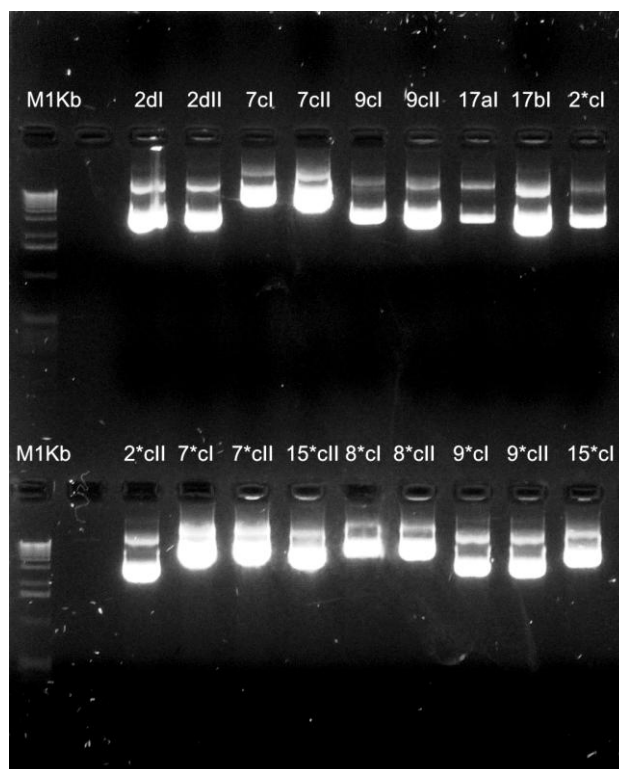


**Figura 7 – Perfil eletroforético em gel de agarose 0,8% de PCR com iniciadores 16F1/MY09 (R3 e R3\*) e MY11/16R3 (R17 e R17\*).** O marcador de peso molecular neste perfil eletroforético foi o 1 Kb Ladder Invitrogen®. A figura mostra os amplicons das reações 3 e 17 com os seus respectivos tamanhos em pb.

O tamanho do fragmento amplificado pela reação 2 é de aproximadamente 240pb, enquanto o tamanho do fragmento amplificado pela reação 6 é de aproximadamente 167pb. Os tamanhos dos fragmentos amplificados pelas reações 7, 8, 12, 15 e 16 são, respectivamente, de 3073, 3293, 2926, 1992 e 2140 pb. Observam-se ainda os fragmentos produzidos por anelamento inespecífico dos iniciadores durante as reações 7, 8 e 15. Tais fragmentos foram eliminados pelo método de purificação com *GE Healthcare® Sephaglas Bandprep™ Purification Kit*. O tamanho do fragmento amplificado pela reação 9 é de 3110 pb e observam-se ainda os fragmentos produzidos por anelamento inespecífico dos iniciadores durante a reação. Os tamanhos dos fragmentos amplificados pelas reações 3 e 17 correspondem, respectivamente a 3330 pb e 5046 pb.

#### **4. Clonagem, sequenciamento e análises**

A figura 8 mostra o perfil eletroforético dos plasmídeos recombinantes já extraídos. Após o sequenciamento dos produtos de PCR e das extrações plasmidiais, as sequências obtidas foram analisadas pelas ferramentas PHRED e CAP3 para a correção de possíveis erros de leitura do sequenciamento. Ao final das análises iniciais de qualidade das leituras do sequenciamento, as sequências obtidas (READS) foram comparadas com outras sequências existentes no GenBank através das ferramentas BLASTn e MEGA 4.0. Ao todo foram obtidos 122 READS, que puderam ser alinhados par a par entre si utilizando-se a ferramenta ClustalW. Resultando na sequência genômica completa apresentada no anexo 1.



**Figura 8 – Perfil eletroforético em gel de agarose 0,8% das extrações plasmidiais. O marcador de peso molecular neste perfil eletroforético foi o 1 Kb Ladder Invitrogen®. A figura mostra o padrão característico de deslocamento dos plasmídeos contendo os produtos de PCR.**

Como se pode observar na tabela 4, o HPV 16 sequenciado (HM057182) mantém uma distância genética maior das outras variantes de HPV 16 que as variantes africanas tipo 1 (AF472508.1 e AF536180.1) apesar de ainda manter um índice de 0,0024 de similaridade com estas seqüências. Essas variantes foram utilizadas na análise comparativa de cada região gênica do HPV sequenciado, através do BLASTn e do tBLAST (tabela 3), onde foi possível observar que o tamanho das seqüências gênicas alinhadas teve forte influência nos valores de E-value (índice estatístico que indica as possibilidades de se obter um determinado alinhamento numa única amostragem). O alinhamento do gene de E2, por exemplo, (com 1098 pb e cuja média da diferença das

variantes africanas é de 7 nucleotídeos) apresentou E-value igual a 0, enquanto o alinhamento dos genes E4 (com 288 pb e cuja média da diferença das variantes africanas é de menos de 1 nucleotídeo), E5 (com 252 pb e que difere das variantes africanas em 1 nucleotídeo) e E7 (com 297 pb e cuja média da diferença das variantes africanas é de cerca de 7 nucleotídeos) apresentaram E-values diferentes de 0.

**Tabela 3 – Comparação entre as regiões gênicas do HPV sequenciado e as variantes africanas geneticamente mais próximas.** Os valores de similaridade e e-value foram obtidos por meio de BLAST local, descrito no texto.

ORF	VARIANTE	BLASTN		TBLASTX	
		SIMILARIDADE (%)	E-VALUE	SIMILARIDADE (%)	E-VALUE
E6	AF536180.1	99	0	99	3,00E-108
	AF472508.1	99	0	99	3,00E-108
E7	AF536180.1	97	3,00E-146	96	2,00E-67
	AF472508.1	97	4,00E-145	96	2,00E-67
E1	AF536180.1	99	0	99	0
	AF472508.1	99	0	99	0
E2	AF536180.1	99	0	98	0
	AF472508.1	99	0	99	0
E4	AF536180.1	100	4,00E-151	100	2,00E-65
	AF472508.1	99	2,00E-149	98	1,00E-64
E5	AF536180.1	99	5,00E-130	98	5,00E-57
	AF472508.1	98	3,00E-127	96	1,00E-54
L1	AF536180.1	99	0	99	0
	AF472508.1	99	0	100	0
L2	AF536180.1	99	0	100	0
	AF472508.1	99	0	100	0

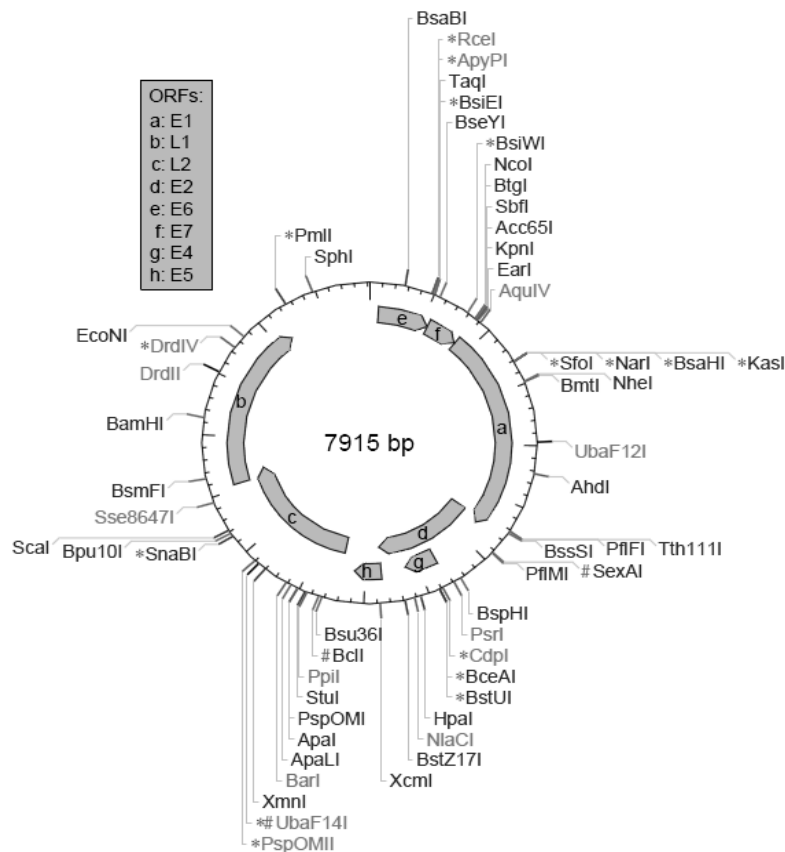
**Tabela 4 – Distância genética entre o HPV 16 isolado (HM057182) e as variantes de HPV 16 conhecidas obtida pelo programa MEGA 4.0 a partir de NCR.** Os números de acesso ao GenBank correspondem as seguintes variantes de HPV 16: isolado alemão (AF536179.1), escocês (gi:9627100), isolado do leste asiático (AF534061.1), isolado africano tipo 1 (AF472508.1), africano 1 (AF536180.1), isolado africano tipo 2 (AF472509.1) e isolado asiático-americano (AF402678.1).

	<b>AF536179.1</b>	<b>GI9627100</b>	<b>AF534061.1</b>	<b>AF472508.1</b>	<b>AF536180.1</b>	<b>HM057182</b>	<b>AF472509.1</b>	<b>AF402678.1</b>
<b>AF536179.1</b>								
<b>GI9627100</b>	0.0110							
<b>AF534061.1</b>	0.0198	0.0160						
<b>AF472508.1</b>	0.0185	0.0147	0.0236					
<b>AF536180.1</b>	0.0185	0.0147	0.0236	0.0000				
<b>HM057182</b>	0.0210	0.0172	0.0261	0.0024	0.0024			
<b>AF472509.1</b>	0.0248	0.0210	0.0299	0.0160	0.0160	0.0185		
<b>AF402678.1</b>	0.0236	0.0197	0.0287	0.0172	0.0172	0.0197	0.0184	

As análises de sequência e de sítios de restrição (figura 9) indicam que a variante HM057182 detém um sítio de restrição para a enzima BspHI (posição 3234/3238, interna ao gene E2), ausente nas duas variantes africanas (Anexo 2), além de 2.608 (32,95%) resíduos de adenina, 1.365 (17,25%) resíduos de citosina, 1.515 (19,14%) resíduos de guanina e 2.427 (30,66%) resíduos de timina (tabela 5).

**Tabela 5 – Composição de nucleotídeos do genoma do HPV sequenciado.**

<b>Nucleotídeo</b>	<b>Quantidade</b>	<b>Percentual</b>
<b>A</b>	2608	32,95
<b>C</b>	1365	17,25
<b>G</b>	1515	19,14
<b>T</b>	2427	30,66



**Figura 9 – Mapa de restrição de HPV 16 HM067182.** As siglas apresentadas na porção interna da figura que representa o genoma viral correspondem às regiões gênicas virais identificadas.

A tabela 6 mostra que as transições foram as mutações mais frequentes ocorridas dentro de regiões gênicas e que os genes E1, E2, E7 e L1 sofreram a maior parte dessas transições (o que pode ser observado pelos gráficos 2a e 2b). Nenhuma das inserções ou deleções observadas ocorreram em regiões intragênicas (o que também pode ser constatado pelos gráficos 2a e 2b) – E5/L2 representa a região entre os genes E5 e L2 e NCR tem função regulatória –, o que contribuiu para a manutenção da integridade dos produtos da expressão gênica a partir das ORF estudadas. Todas as indels observadas representaram mutações silenciosas (gráficos 3a e 3b).



**Tabela 6 – Tipos de mutações gênicas ocorridas a partir de cada protótipo.** As mutações de cada protótipo são apresentadas de acordo com as regiões genômicas em que ocorrem.

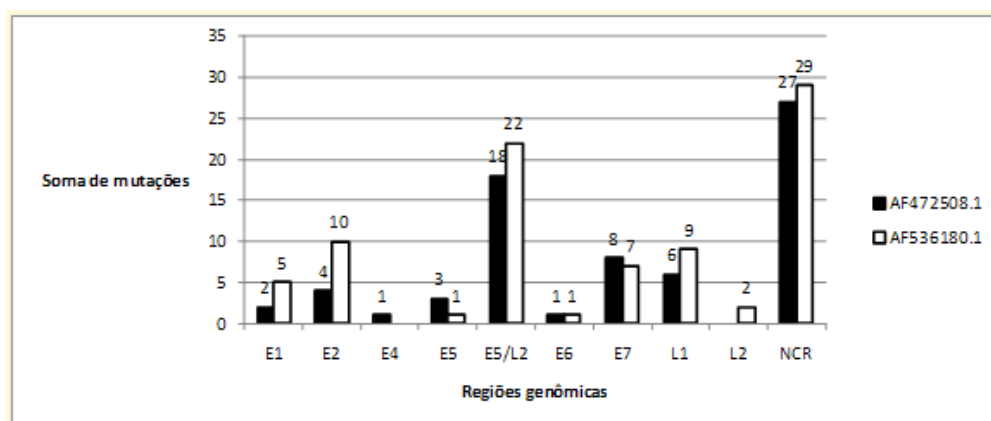
Substituição	AF472508.1									Total	AF536180.1									Total	Total geral						
	E1	E2	E4	E5	E5/L2	E6	E7	L1	NCR		E1	E2	E5	E5/L2	E6	E7	L1	L2	NCR								
<b>Deleção</b>	<b>16</b>									<b>17</b>	<b>16</b>									<b>17</b>	<b>34</b>						
A/O	5									6	5									6	12						
C/O	2									2	2									2	4						
G/O	1									1	1									1	2						
T/O	8									8	8									8	16						
<b>Inserção</b>	<b>24</b>									<b>24</b>	<b>2</b>									<b>24</b>	<b>26</b>	<b>50</b>					
O/A	4									4	1									4	5	9					
O/C	1									1										1	1	2					
O/G	7									7										7	7	14					
O/T	12									12	1									12	13	25					
<b>Transição</b>	<b>2</b>	<b>4</b>	<b>1</b>	<b>3</b>		<b>1</b>	<b>4</b>	<b>6</b>	<b>1</b>	<b>22</b>	<b>5</b>	<b>9</b>	<b>1</b>	<b>2</b>	<b>1</b>	<b>3</b>	<b>9</b>	<b>2</b>	<b>3</b>	<b>35</b>	<b>57</b>						
A/G	1			1		1	1	1	1	6	1	1			1	1	1		2	7	13						
C/T		2						3		5		4		2			3	1	1	11	16						
G/A	1			1			1	1		4	3	3					3			9	13						
T/C		2	1	1			2	1		7	1	1	1			2	2	1		8	15						
<b>Transversão</b>	<b>2</b>									<b>7</b>	<b>1</b>									<b>2</b>	<b>4</b>	<b>1</b>	<b>8</b>	<b>15</b>			
A/C										2										2				2	4		
A/T										1										1	1				2	3	
C/A											1														1	1	
C/G										1										1					1	2	
G/T										1	1											1				1	3
T/A										1													1		1	2	
<b>Total geral</b>	<b>2</b>	<b>4</b>	<b>1</b>	<b>3</b>	<b>18</b>	<b>1</b>	<b>8</b>	<b>6</b>	<b>27</b>	<b>70</b>	<b>5</b>	<b>10</b>	<b>1</b>	<b>22</b>	<b>1</b>	<b>7</b>	<b>9</b>	<b>2</b>	<b>29</b>	<b>86</b>	<b>156</b>						

O gráfico 1 mostra que a maior parte das mutações ocorreram em NCR e E5/L2, além de indicar E2 (para o protótipo AF536180.1) e E7 (para o protótipo AF472508.1) como as regiões gênicas que mais apresentaram mutações entre as sequências.

A tabela 7 e o gráfico 4 mostram que a maioria das mutações representaram mutações silenciosas em regiões não codificadoras (NCR e E5/L2) e que, ao longo de todo o genoma, ocorreram mais mutações não sinônimas que sinônimas (o que pode ser constatado pelos gráficos 3a e 3b). Essa razão entre mutações não sinônimas e sinônimas ( $\omega = d_N/d_S$ ) ocorridas nas regiões gênicas corrobora com os estudos de Chen

*et al.* (2005), onde observou-se valores de  $\omega \geq 1,0$  para as sequências de E2 e E5, indicando que essas regiões encontravam-se sobre forte pressão de seleção no sentido da mudança de aminoácidos. Além disso, os dados mostram que, além de E2 e E5, as regiões E7 e L2 também podem estar sofrendo pressão seletiva, com  $\omega_{E7} = 14$  e  $\omega_{L2} = 1$ .

**Gráfico 1 – Quantidade absoluta de mutações por protótipo utilizado.** As mutações gênicas são apresentadas de acordo com a região genômica em que ocorrem em cada protótipo.

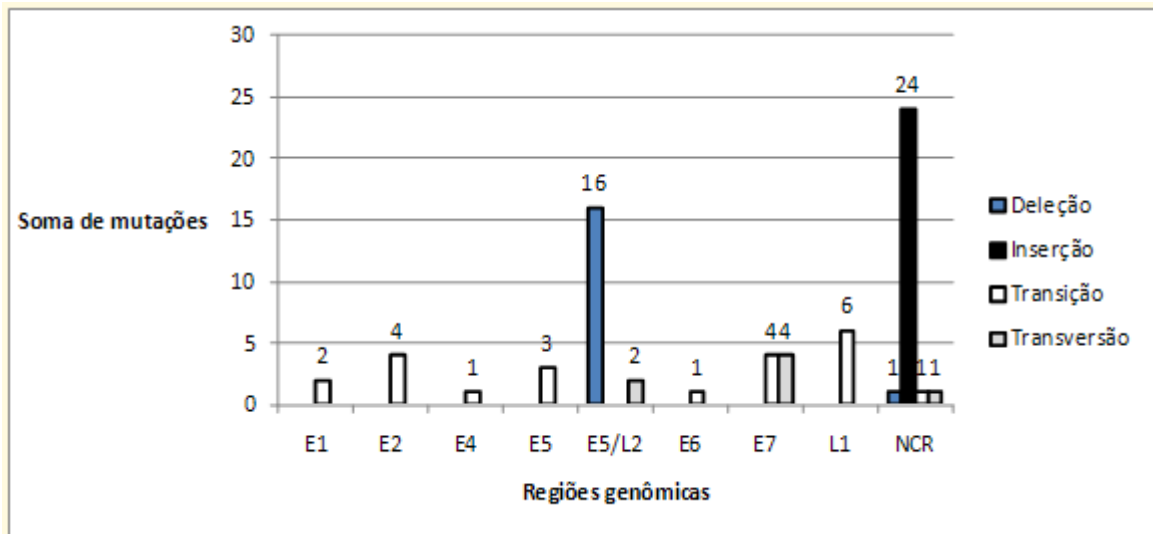


**Tabela 7 – Tipos de mutações ocorridas a partir de cada protótipo.** As classificações bioquímicas apresentadas na tabela correspondem aos aminoácidos mutantes. O símbolo “-” indicado na tabela das mutações silenciosas indica as mutações que ocorreram em regiões intergênicas (NCR ou E5/L2).

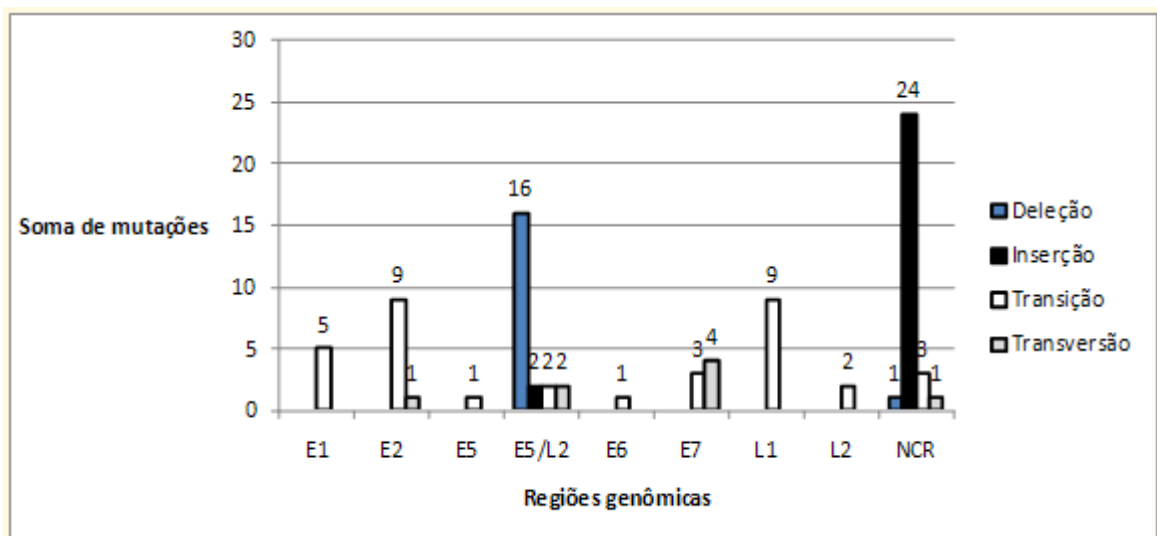
Tipos de mutação e classe bioquímica do aminoácido mutante	AF472508.1	AF536180.1	Total geral
<b>Não sinônima</b>	<b>13</b>	<b>14</b>	<b>27</b>
Apolar alifático	5	5	10
Aromático	2	3	5
Negativamente carregado		1	1
Polar não carregado	6	5	11
<b>Silenciosa</b>	<b>56</b>	<b>68</b>	<b>124</b>
-	26	27	53
Apolar alifático	5	8	13
Aromático		2	2
Negativamente carregado	2	2	4
Polar não carregado	2	2	4
Positivamente carregado	2	3	5
<b>Sinônima</b>	<b>1</b>	<b>4</b>	<b>5</b>
Apolar alifático		1	1
Negativamente carregado		1	1
Positivamente carregado	1	2	3
<b>Total geral</b>	<b>70</b>	<b>86</b>	<b>156</b>

Pode-se constatar pelos gráficos 2a e 2b que ocorreram mais transições que transversões nas regiões gênicas. Além disso, constatou-se que a região entre E5 e L2 concentrou a maioria das deleções e NCR deteve a maioria das inserções observadas (o que pode ser observado pela tabela 6).

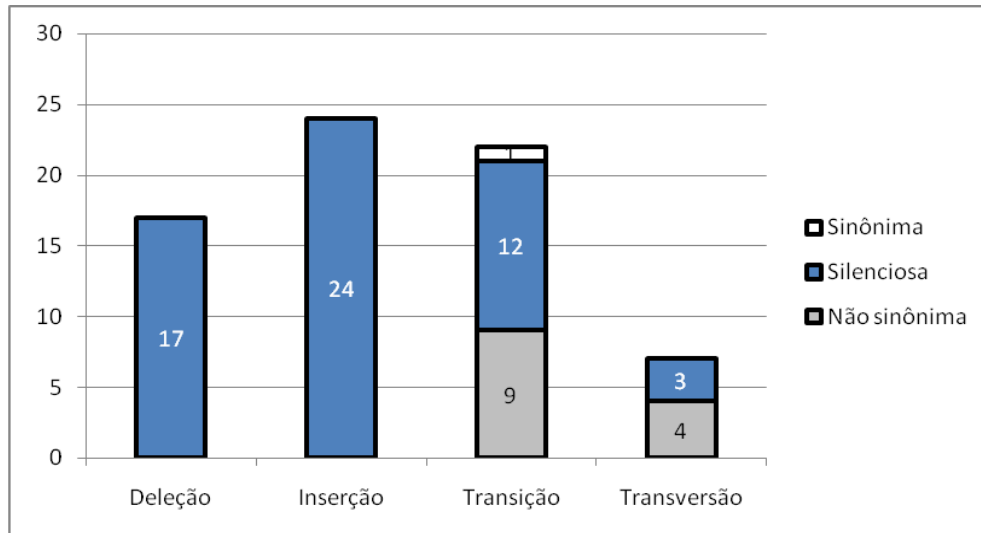
**Gráfico 2a – Quantidade absoluta de mutações por região genômica (Protótipo AF472508.1).** As mutações gênicas são apresentadas de acordo com a região genômica em que ocorrem no protótipo AF472508.1.



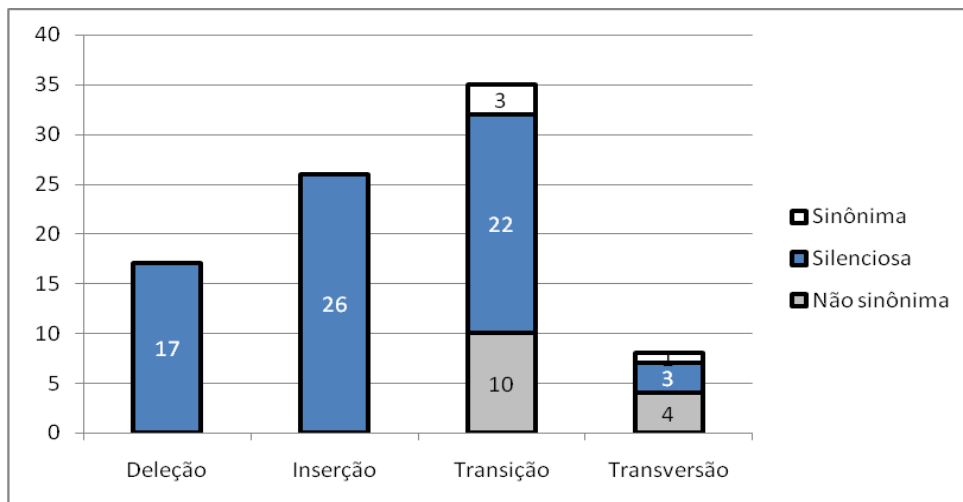
**Gráfico 2b - Quantidade absoluta de mutações por região genômica (Protótipo AF536180.1).** As mutações gênicas são apresentadas de acordo com a região genômica em que ocorrem no protótipo AF536180.1.



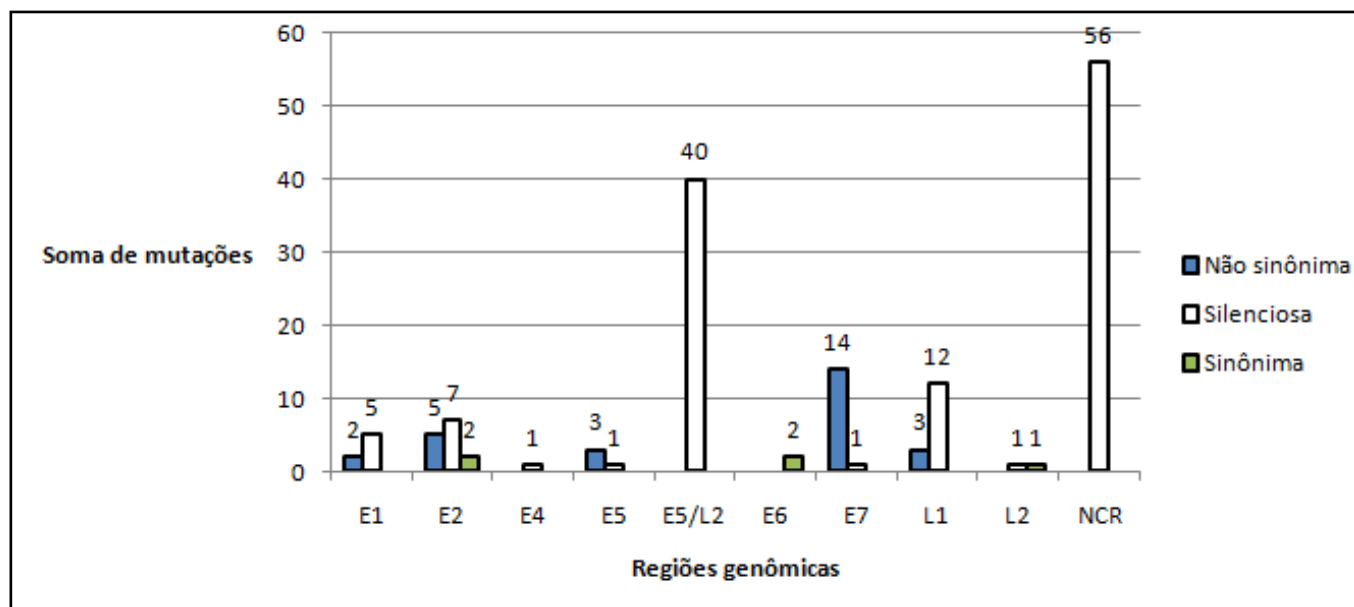
**Gráfico 3a – Quantidade absoluta de mutações por tipo de substituição (Protótipo AF472508.1).** As mutações gênicas ao nível do DNA estão sendo comparadas com as mutações ao nível proteico para o protótipo AF472508.1.



**Gráfico 3b - Quantidade absoluta de mutações por tipo de substituição (Protótipo AF536180.1).** As mutações gênicas ao nível do DNA estão sendo comparadas com as mutações ao nível proteico para o protótipo AF536180.1.



**Gráfico 4 – Relação entre as mutações no âmbito protéico e as regiões genômicas.** As mutações gênicas são apresentadas de acordo com a região genômica em que estas ocorrem em ambos os protótipos estudados.



## 5. Anotação e número de acesso ao GenBank

Após completamente anotado, o genoma do HPV 16 isolado no trabalho foi depositado no banco de dados do NCBI, sob o número de acesso HM057182.

A anotação do genoma do HPV 16 e os mapas de restrição das variantes africanas mais próximas do HPV 16 sequenciado estão presentes nos anexos 1 e 2.

## BIBLIOGRAFIA

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*. 25(17), 3389-3402.

Burd, E. M., 2003. Human Papillomavirus and Cervical Cancer. *Clinical microbiology Reviews*. 16, 1-17.

Chen, Z., Terai, M., Fu, L., Herrero, R., DeSalle, R., Burk, R. D., 2005. Diversifying Selection in Human Papillomavirus Type 16 Lineages Based on Complete Genome Analyses. *Journal of Virology*. 79, 7014–7023.

Choe, J., Vaillancourt, P., Stenlund, A., Botchan, M., 1989. Bovine papillomavirus type 1 encodes two forms of a transcriptional repressor: structural and functional analysis of new viral cDNAs. *Journal of Virology*. 63, 1743–1755.

Danos, O., Katinka, M., Yaniv, M., 1982. Human papillomavirus 1a complete DNA sequence: a novel type of genome organization among papovaviridae. *EMBO Journal*. 1, 231–236.

de Roda Husman, A.-M., Walboomers, J. M. M., Van Den Brule, A. J. C., Meijer, C. J. L. M., Snijders, P. J. F., 1995. The use of general primers GP5 and GP6 elongated at their 3' ends with adjacent highly conserved sequences improves human papillomavirus detection by PCR. *Journal of Genetic Virology*. 76, 1057–1062.

de Villiers, E.-M., Fauquet, C., Broker, T.R., Bernard, H.-U., zur Hausen, H., 2004. Classification of papillomaviruses. *Journal of Virology*. 324, 17– 27.

Ewing, B., Hillier, L., Wendl, M. C., Green, P., 1998. Base-calling of automated sequenced traces using PHRED. I - Accuracy Assessment. *Genome Research*. 8, 175-185.

Ferreira, J. R. D., 2007. Detecção da infecção pelo HPV e do polimorfismo da proteína p53 em pacientes com lesões cervicais atendidas na Fundação Centro de Controle de Oncologia (FCECON) do Amazonas. Dissertação (Mestrado em Patologia Tropical) – Universidade Federal do Amazonas, Manaus.

Grassmann, K., Rapp, B., Maschek, H., Petry, K.U., Iftner, T., 1996. Identification of a differentiation-inducible promoter in the E7 open reading frame of human papillomavirus type 16 (HPV-16) in raft cultures of a new cell line containing high copy numbers of episomal HPV-16 DNA. *Journal of Virology*. 70, 2339–2349.

Hall, T., 2001. BioEdit version 5.0.6. North Carolina State University, Department of Microbiology

Han, R., Cladel, N. M., Reed, C. A., Christensen, N. D., 1998. Characterization of transformation function of cottontail rabbit papillomavirus E5 and E8 genes. *Virology*. 251, 253–263.

Harper, D. M., 2004. Efficacy of a bivalent L1 virus-like particle vaccine in prevention of infection with human papillomavirus types 16 and 18 in young women: a randomized controlled trial. *Lancet*. 364, 1757-1765.

Harry, J. B., Wettstein, F. O., 1996. Transforming properties of the cottontail rabbit papillomavirus oncoproteins Le6 and SE6 and of the E8 protein. *Journal of Virology*. 70, 3355–3362.

Huang, X., Madan, A., 1999. CAP3: A DNA sequence assembly program. *Genome Research*. 9, 868-877.

Hubbert, N. L., Schiller, J. T., Lowy, D. R., Androphy, E. J., 1988. Bovine papilloma virus-transformed cells contain multiple E2 proteins. *Proceedings of National Academy of Science*. 85, 5864-5868.

Hummel, M., Hudson, J. B., Laimins, L. A., 1992. Differentiation-induced and constitutive transcription of human papillomavirus type 31b in cell lines containing viral episomes. *Journal of Virology*. 66, 6070–6080.

Lambert, P. F., Spalholz, B. A., Howley, P. M., 1987. A transcriptional repressor encoded by BPV-1 shares a common carboxy-terminal domain with the E2 transactivator. *Cell*. 50, 69–78.

Lowy, D. R., Schiller, J. T., 2006. Prophylactic human papillomavirus vaccines. *Journal of Clinical Investigation*. 5(116), 1167-1173.

Manos, M. M., Ting, Y., Wright, D. K., Lewis, A. J., Broker, T. R., Wolinsky, S. M., 1989. The use of polymerase chain reaction amplification for the detection of genital human papillomaviruses. *Cancer Cell* 7, 209–214.

Marshall, O. J., 2004. PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics*. 20(15), 2471-2472.

Mcmurray, H. R., Nguyen, D., Westbrook, T. F., Mcance, D. J., 2001. Biology of human Papillomavirus. *Viruses and Cancer*, 82, 15-33.

Ozbun, M. A., Meyers, C., 1997. Characterization of late gene transcripts expressed during vegetative replication of human papillomavirus type 31b. *Journal of Virology*. 71, 5161–5172.

Pontes, I. M., 2003. Desenvolvimento de novos marcadores microssatélites para análise genética em humanos. Dissertação (Mestrado em Genética e Evolução) - Universidade Federal de São Carlos/Universidade Federal do Amazonas, Manaus.

Qu, W., Jiang, G., Cruz, Y., Chang, C. J., Ho, G. Y. F., Klein, R. S., Burk, R. D., 1997. PCR Detection of Human Papillomavirus: Comparison between MY09/MY11 and GP51/GP61 Primer Systems. *Journal of clinical microbiology*. 35(6), 1304-1310.

Rosenstierne, M. W., Vinther, J., Hansen, C. N., Prydsøe, M., Norrild, B., 2003. Identification and characterization of a cluster of transcription start sites located in the E6 ORF of human papillomavirus type 16. *Journal of Genital Virology*. 84, 2909–2920.



Rozen, S., Skaletsky, H. J., 2000. Primer3 on the WWW for general users and for biologist programmers, in: Krawetz, S., Misener, S. (Eds.), *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. NJ: Humana Press, Totowa, 365-386.

Santos, C. M. B., 2006. Sequenciamento e análise do genoma do HPV 13 isolado na Amazônia. Tese (Doutorado em Biotecnologia) – Universidade Federal do Amazonas (UFAM) e Fundação de Medicina Tropical (FMTAM), Manaus.

Smotkin, D., Wettstein, F. O., 1986. Transcription of human papillomavirus type 16 early genes in a cervical cancer and a cancer-derived cell line and identification of the E7 protein. *Proceedings of National Academy of Science*. 83(13), 4680–4684.

Schneider-Gadicke, A., Schwarz, E., 1986. Different human cervical carcinoma cell lines show similar transcription patterns of human papillomavirus type 18 early genes. *EMBO Journal*. 5, 2285–2292.

Steller, M. A., 2002. Cervical Cancer Vaccines: Progress and Prospects *Journal of the Society of Gynecology Investigation*. 9, 254–264.

Stubenrauch, F., Hummel, M., Iftner, T., Laimins, L. A., 2000. The E8E2C protein, a negative regulator of viral transcription and replication, is required for extrachromosomal maintenance of human papillomavirus type 31 in keratinocytes. *Journal of Virology*. 74, 1178–1186.

Stubenrauch, F., Zobel, T., Iftner, T., 2001. The E8 domain confers a novel long-distance transcriptional repression activity on the E8E2C protein of high-risk human papillomavirus type 31. *Journal of Virology*. 75, 4139–4149.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*. 24, 1596-1599.

Thomson, J. D., Higgins, D. G., Gibson, T. J., 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids*. 22(22), 4673-4680.

Villa, L. L., Costa, R. L. R., Petta, C. A., Andrade, R. P., Ault, K. A., Giuliano, A. R., Wheeler, C. M., Koutsky, L. A., Malm, C., Lehtinen, M., Skjeldestad, F. E., Olsson, S. E., 2005. Prophylactic quadrivalent human papillomavirus (types 6, 11, 16 and 18) L1 virus-like particle vaccine in young women: a randomized double-blind placebo-controlled multicentre phase II efficacy trial. *Lancet Oncology*. 6, 271-278.

Vincze, T., Posfai, J. Roberts, R. J., 2003. NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Research*. 31, 3688–3691.

**CAPÍTULO 2**  
**ESTUDOS MOLECULARES DO GENOMA DE HPV 16 ISOLADO**  
**NA AMAZÔNIA**

## **Estudos moleculares do genoma de HPV 16 isolado na Amazônia**

Roberto Alexandre Alves Barbosa Filho<sup>1</sup>; Cristina Maria Borborema dos Santos<sup>1</sup>; Spartaco Astolfi Filho<sup>1</sup>; Izeni Pires Farias

1 Laboratório de Diagnóstico Molecular, Divisão de Biotecnologia, Centro de Apoio Multidisciplinar, Universidade Federal do Amazonas, Manaus, Amazonas, Brasil.

2 Laboratório de Evolução e Genética Animal, Departamento de Biologia, Universidade Federal do Amazonas, Manaus, AM, Brasil

Endereço eletrônico:

RAABF: [robertoaabfilho@hotmail.com](mailto:robertoaabfilho@hotmail.com)

CMBS: [crisantos@ufam.edu.br](mailto:crisantos@ufam.edu.br)

SAF: [spartaco.biotec@gmail.com](mailto:spartaco.biotec@gmail.com)

## INTRODUÇÃO

Os papilomavírus eram anteriormente associados aos polyomavírus em uma família, a *Papovaviridae*, por causa da semelhança na estrutura de seu capsídeo icosaédrico e em sua cadeia dupla-fita circular de DNA. porém, os diferentes tamanhos, diferentes organizações e ausência de similaridade entre as suas sequências de nucleotídeos e aminoácidos os separaram em duas famílias distintas, a *Papillomaviridae* e a *Polyomaviridae* (de Villiers *et al.*, 2004).

A classificação taxonômica dos papilomavírus, atualmente aceita internacionalmente pelo Comitê Internacional de Taxonomia de Vírus (ICTV), foi revista por de Villiers *et al.* (2004), e nela estão definidos os critérios para a identificação de tipos, subtipos e variantes. Tal classificação baseia-se na variabilidade genética de L1, o mais conservado gene de HPV. Um gênero dentro da família *Papillomaviridae* compartilha menos de 60% de identidade com a sequência de L1 de um gênero irmão. A família *Papillomaviridae* divide-se em dezesseis gêneros: *Alpha-papilomavirus*, *Beta-papilomavirus*, *Gamma-papilomavirus*, *Delta-papilomavirus*, *Epsilon-papilomavirus*, *Zeta-papilomavirus*, *Iota-papilomavirus*, *Pi-papilomavirus*, *Eta-papilomavirus*, *Theta-papilomavirus*, *Kappa-papilomavirus*, *Lambda-papilomavirus*, *Mu-papilomavirus*, *Nu-papilomavirus*, *Xi-papilomavirus* e *Omikron-papilomavirus*.

As espécies dentro de um gênero partilham entre 71% e 89% de identidade entre as sequências de nucleotídeos de seus genes L1, o que as agrupa em quarenta e nove taxa dentro dos gêneros. Novos tipos virais são caracterizados a partir de diferenças acima de 10% com relação a sequências já conhecidas de tipos da mesma espécie. Novos subtipos apresentam diferenças entre 2 e 10%. Novas variantes, por fim, são

caracterizadas a partir de 2% de diferença a partir de um genótipo de referência. O HPV 16 pertence, juntamente com os tipos 31, 33, 35, 52, 58 e 67, à espécie 9 do gênero *Alpha-papillomavirus*. Isso significa que a sequência de seu gene L1 apresenta entre 71 e 89% de semelhança com os tipos 31, 33, 35, 52, 58 e 67, mas possui mais de 10% de diferença dos mesmos (de Villiers *et al.*, 2004).

Baseando-se em comparações entre as sequências de E6, L1, L2 e NCR as variantes de HPV 16 são reunidas em seis principais grupos: Europeu (E), Asiático-americano (AA), Africano 1 (Af1), Africano 2 (Af2), Asiático (As) e Norte-americano (NA). Estas variantes mostraram diferentes distribuições geográficas e diferentes potenciais oncogênicos. Além disso, as diferenças entre as variantes podem, também, afetar a adsorção viral, as respostas imunológicas, a patogenicidade, a degradação de p53, a atividade de imortalização e a regulação da transcrição (Ho *et al.*, 1991; Ong *et al.*, 1993).

O papilomavírus tem sido detectado em uma larga quantidade de animais vertebrados bem como em humanos e são específicos para seus respectivos hospedeiros. Diferentes partes do genoma do grupo papilomavírus evoluem a taxas diferentes. Sequências de DNA que codificam para domínios conservados de proteínas possibilitam a comparação de tipos diferentes de papilomavírus e a análise de processos de “especiação”. Por outro lado, segmentos genômicos envolvidos na regulação da transcrição, cujas sequências são altamente variáveis entre tipos diferentes de vírus, apresentam uma diversidade entre isolados independentes do mesmo tipo de papilomavírus que possibilita a recapitulação da evolução molecular dentro de um determinado tipo de papilomavírus (Gottschling *et al.*, 2007).

Estudos envolvendo a Longa Região de Controle (LCR, ou NCR – Região Não-Codificadora) de variantes do HPV 16 mostraram que as taxas de mutação desta região

são capazes de reconstruir a história evolutiva deste tipo viral corroborando com a evolução do homem moderno. Além disso, a variabilidade genômica observada nesta região do genoma em diferentes isolados virais de HPV mostrou-se uma importante ferramenta de retrospectiva da epidemiologia molecular em testes de métodos sensíveis de diagnóstico, no entendimento de formas de infecção em indivíduos e populações e no estudo da evolução viral (Ho *et al.*, 1991; Ong *et al.*, 1993; Pande *et al.*, 2008).

Apesar da ORF L1 ser amplamente utilizada em pesquisas de diagnóstico molecular de HPV por PCR – como por exemplo os estudos de Castro (2002), Santos (2006), Santos (2007) e Ferreira (2007) na região de Manaus e circunvizinhanças –, a região de LCR apresenta-se muito mais adequada a estudos de variabilidade entre as variantes de um determinado tipo viral. A maior prova disso é que nenhum dos estudos citados anteriormente pode caracterizar as variantes de HPV 16 detectadas devido à baixa variabilidade genética existente entre as sequências obtidas durante tais estudos.

Este trabalho descreve uma análise filogenética do HPV 16 a partir da região NCR, além do estudo comparativo das estruturas secundárias e terciárias das proteínas L1, E6 e E7 de protótipos do GenBank e do genoma sequenciado.

## **MÉTODOS**

### **1. Alinhamento múltiplo de sequências**

Para a realização deste estudo foram utilizadas as sequências de NCR do HPV 16 HM057182 isolado no Estado do Amazonas além da sequência de outros tipos de HPV depositados no GenBank, descritos a seguir. O alinhamento múltiplo de sequências foi realizado pelo programa CLUSTALX e os “gaps” do alinhamento não

foram considerados como deleções, mas sim, como dados perdidos (Thomson *et al.*, 1997).

## **2. Reconstrução filogenética**

Após o alinhamento, um arquivo com extensão “txt” foi gerado e editado para tornar possível a análise desses dados posteriormente pelos programas jModelTest 0.1, PHYLIP 3.68 e/ou TOPALi 2.5. A filogenia foi inferida utilizando-se o modelo evolutivo escolhido para o alinhamento pelo programa jModelTest 0.1, que aplica métodos estatísticos de verossimilhança para selecionar o modelo evolutivo que mais se adequa à amostra. O modelo evolutivo, por sua vez foi aplicado ao alinhamento através dos programas PHYLIP 3.68 ou TOPALi 2.5 (dependendo do modelo evolutivo) para estimar as relações filogenéticas a partir dos dados moleculares (Felsenstein, 2008; Milne *et al.*, 2009; Posada *et al.*, 1998).

Os tipos de HPV 35 (M74117), HPV 58 (D90400), HPV 33 (M12732), 73 (X94165), HPV 34 (X74476) e HPV 67 (D21208) integraram o grupo-irmão (espécie ou grupo monofilético supra-específico mais próximo do provável grupo monofilético em estudo) do grupo formado pelas variantes estudadas, uma vez que também pertencem à espécie 9 de Papilomavírus (espécie à qual pertence o HPV 16, segundo estudos moleculares com a região L1). O protótipo de RhPV 1 (M60184) foi utilizado como grupo-externo para enraizar a árvore gerada (de Villiers *et al.*, 2004).

As variantes do HPV 16 [HPV 16 escocês (NC 001526.2 ou U89348), HPV 16 isolado africano 1 (AF536180.1), HPV 16 isolado alemão (AF536179.1), HPV 16 isolado do leste asiático (AF534061.1), HPV 16 isolado asiático-americano (AF402678.1), HPV 16 isolado africano tipo 2 (AF472509.1) e HPV 16 isolado africano tipo 1 (AF472508.1)] além de outros protótipos deste tipo viral depositados no



GenBank (EU918764.1, K02718.1, AY686584.1, FW344859.1, FJ006723.1, AF125673.1, FJ610151.1, FJ610148.1, FJ610147.1, AY686583.1, EU118173.1, AY686581.1, FJ610149.1, AY686580.1, FJ610146.1, FJ610152.1, AY686579.1, AY686582.1 e FJ610150.1 foram analisadas juntamente com a variante amazônica (HM057182).

A topologia da árvore filogenética obtida no estudo foi testada pelo método de *Bootstrap*, cuja base estatística consiste numa reamostragem com reposição pseudoaleatória dos dados. Em cada reamostragem, o número total de dados amostrados mantém-se constante e uma árvore é construída baseando-se em cada rearranjo. Ao final de todas as réplicas, a árvore final, um consenso de todas as árvores geradas, pôde ser estimada (Felsenstein, 1985).

### **3. Análise e modelagem molecular**

As estruturas primárias dos produtos de L1, E6 e E7 foram analisadas utilizando-se o servidor I-TASSER do sítio zhanglab.ccmb.med.umich.edu, onde foi possível inferir sobre a estrutura e a função das proteínas com base no paradigma seqüência-estrutura-função. Partindo das seqüências de aminoácidos, modelos atômicos tridimensionais (3D) foram gerados a partir de alinhamentos múltiplos de seqüências e simulações interativas de associação estrutural. A função das proteínas foi, então, inferida pela comparação estrutural com modelos 3D de outras proteínas conhecidas. O *Output* continha previsões de estrutura secundária e terciária da seqüência completa e anotações funcionais sobre prováveis sítios de ligação, números de *Enzyme Commission* e os termos do *Gene Ontology* (Roy *et al.*, 2010; Zhang, 2009; Zhang, 2008).

Uma estimativa da precisão das previsões foi fornecida com base no “*Confidence Score*” (*C-score*) de cada modelo. O *C-score* foi calculado com base na

significância dos alinhamentos com sequências moldes e nos parâmetros de convergência das simulações de montagem estrutural. O *C-score* encontrou-se tipicamente no intervalo entre -5 e 2, onde um *C-score* de maior valor significa um modelo com uma confiança elevada e vice-versa (Roy *et al.*, 2010; Zhang, 2009; Zhang, 2008).

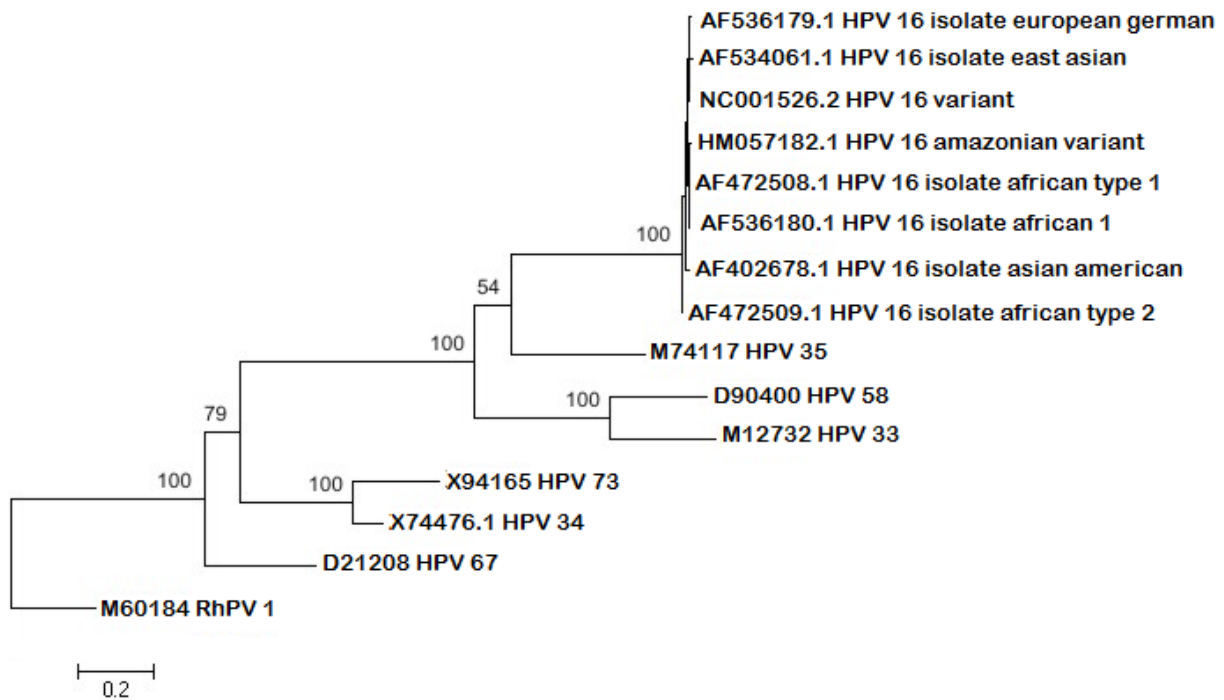
A qualidade dos alinhamentos estruturais dos modelos com outras estruturas depositadas no Banco de Dados de Proteínas (PDB) foi avaliada com base nos valores de *TM-score*, geralmente próximos a 0,5, onde valores maiores que 0,5 demonstravam alto grau de similaridade das estruturas secundárias preditas com as estruturas comparadas (Roy *et al.*, 2010; Zhang, 2009; Zhang, 2008).

A qualidade das predições de regiões com maior probabilidade de representarem sítios de ligação a substratos ou de dimerização foi avaliada com base no valor de *BS-score*, onde valores aproximados a 0,5 indicavam alto grau de significância da predição (Roy *et al.*, 2010; Zhang, 2009; Zhang, 2008).

## **RESULTADOS E DISCUSSÃO**

### **1. Reconstrução filogenética**

Após o alinhamento múltiplo das sequências através da ferramenta CLUSTALX2, o modelo evolutivo a ser adotado na análise filogenética foi o Modelo de Três Parâmetros de Kimura (1981) (anexo 3) com distribuição variável de sítios e taxas de substituição de nucleotídeos equivalentes a: [AC] = 1.2831, [AG] = 2.9964, [AT] = 1.0000, [CG] = 1.2831, [CT] = 2.9964, [GT] = 1.0000. A figura 1 apresenta as relações filogenéticas estimadas segundo o modelo adotado para NCR.

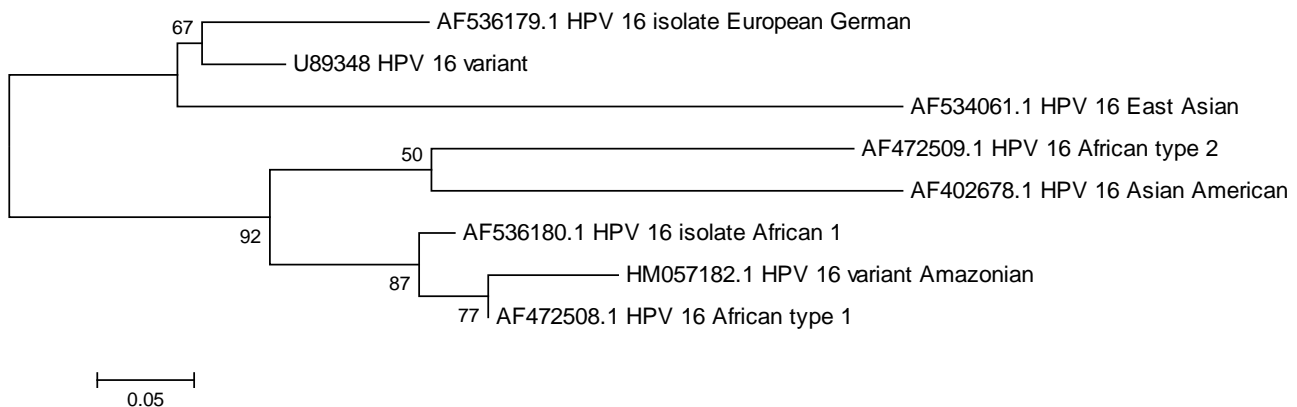


**Figura 1 – Relações filogenéticas entre diferentes variantes de HPV 16 e outros protótipos da espécie 9, segundo a análise de NCR.** Os valores apresentados sobre os ramos correspondem aos resultados em porcentagem do bootstrap com 1000 reamostragens aleatórias. Valores menores que 50% foram omitidos do resultado.

De acordo com a análise da filogenia a partir de NCR apresentada na figura 1, tem-se: o protótipo de HPV 35 formando um grupo monofilético e estatisticamente bem estruturado (bootstrap igual a 100) com o grupo formado pelas variantes de HPV 16; o grupo monofilético formado pelos HPV 58 e 33 constitui o grupo irmão do grupo supracitado; os demais tipos de HPV apresentados na figura 1 (HPV 73, 34 e 67), constituem um grupo parafilético e; a árvore foi enraizada pela análise de NCR de RhPV1, proveniente de macaco Rhesus (*Macaca mulatta*).

Como pode-se observar na figura 2, uma análise mais detalhada da filogenia de NCR das variantes de HPV 16 revelou que: o grupo formado pelas variantes do leste

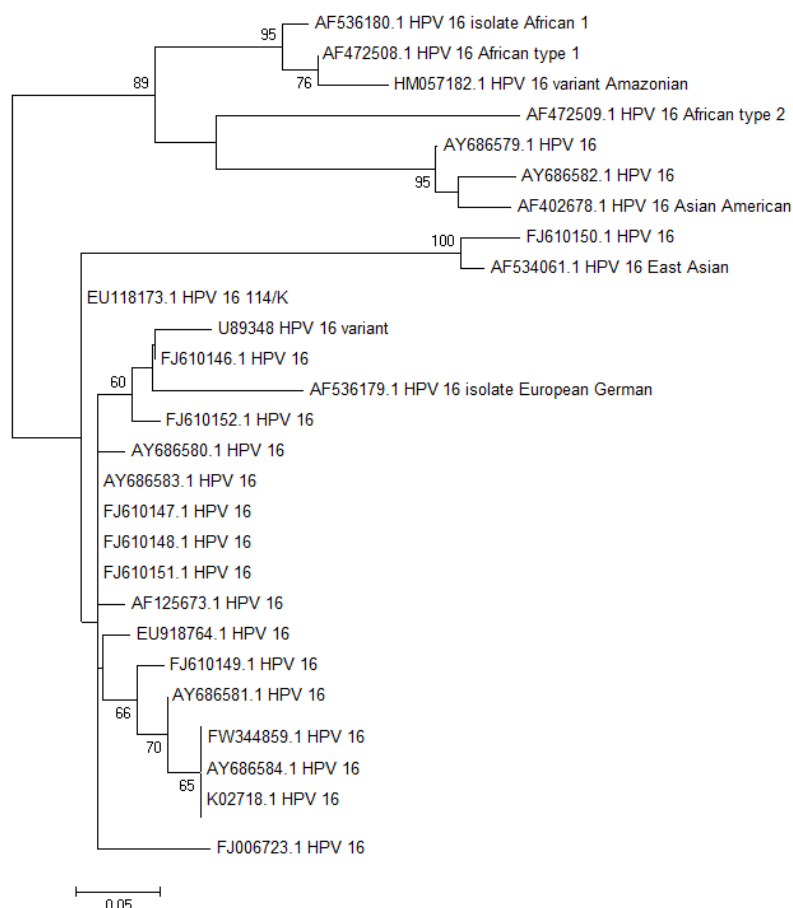
asiático (AF534061.1), escocesa (U89348) e alemã (AF536179.1) formaram um grupo monofilético e irmão do clado formado pelas variantes africana tipo 2 (AF472509.1), africanas tipo 1 (AF536180.1 e AF472508.1), amazônica (HM057182) e asiático-americana (AF402678.1); a variante amazônica formou um grupo monofilético com as variantes africanas tipo 1, mantendo um nó com relevante suporte estatístico (bootstrap igual a 87). O comprimento dos ramos sugere que a divergência entre os protótipos europeus (alemão e escocês) e o leste-asiático ocorrera antes da separação do clado formado pelos protótipos africano 2 e asiático-americano se separar do clado formado pelos africanos 1 e amazônico. Além disso, nota-se que a variante africana 1 AF536180.1 divergiu do clado formado pela africana 1 AF472508.1 e pela amazônica antes que a africana 2 divergisse da asiático-americana, mas a variante amazônica surgiu somente após a separação da asiático-americana a partir da africana 2.



**Figura 2 – Relações filogenéticas entre as principais variantes de HPV 16 segundo a análise de NCR.** Os valores apresentados sobre os ramos correspondem aos resultados em porcentagem do bootstrap com 1000 reamostragens aleatórias. Valores menores que 50% foram omitidos do resultado.

A análise filogenética de NCR dos principais protótipos de HPV 16 depositados no GenBank (figura 3) revelou que: os protótipos americanos AY686582.1 e

AY686579.1 formaram um grupo monofilético com o asiático-americano (sendo, portanto identificados como tal variante) e, por sua vez, com o africano tipo 2; o protótipo asiático (FJ610150.1) formou um grupo monofilético com o leste-asiático (sendo, por isso, caracterizado como uma segunda variante leste-asiática devido o rigor estatístico com que os dois protótipos foram agrupados); o clado leste-asiático formou um grupo politômico com o protótipo americano EU118173.1 e com o clado já politômico formado por dez protótipos asiáticos (EU918764.1, FJ006723.1, AF125673.1, FJ610146.1, FJ610147.1, FJ610148.1, FJ610149.1, FJ610151.1, FW344859.1 e FJ610152.1), quatro americanos (AY686580.1, AY686581.1, AY686583.1 e AY686584.1) e três europeus (U89348, K02718.1 e AF536179.1).

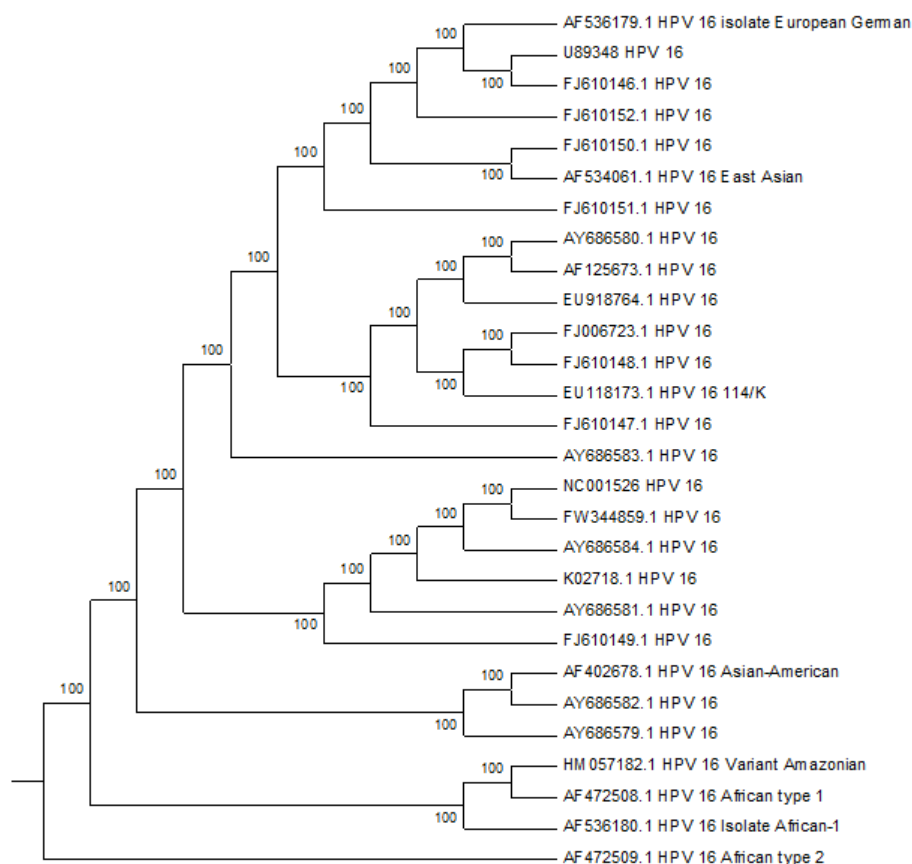


**Figura 3 – Relações filogenéticas entre diferentes protótipos de HPV 16 segundo a análise de NCR.**

Os valores apresentados sobre os ramos correspondem aos resultados em porcentagem do bootstrap com 1000 reamostragens aleatórias. Valores menores que 50% foram omitidos do resultado.

A politomia e as diferenças entre os comprimentos dos ramos observadas na árvore provavelmente decorrem da variação das diferenças genéticas existentes entre os protótipos de um mesmo continente ou de continentes diferentes, o que pode enviesar o resultado da análise. Com isso, um clado politômico pode sugerir que as mutações ocorridas entre os taxa não foram determinantes para separá-los em duas OTU's (unidades taxonômicas operacionais) distintas.

Após estimar as relações filogenéticas para NCR usando o modelo escolhido pela verossimilhança, as topologias previstas para outros modelos foram reunidas numa árvore consenso por meio de uma filogenia média gerada pelo ModelTest0.1. A figura 4 apresenta as relações filogenéticas estimadas pela análise consenso.



**Figura 4 – Relações filogenéticas entre diferentes protótipos de HPV 16 segundo a análise de NCR de acordo com a média dos modelos filogenéticos.** Os valores apresentados sobre os ramos correspondem aos resultados do bootstrap com 1000 reamostragens aleatórias.

De acordo com a análise consenso: o protótipo africano 2 formou uma OTU externa aos demais protótipos analisados; o clado formado pelos protótipos africanos tipo 1 e amazônico, manteve-se monofilético e com alto rigor estatístico (bootstrap igual a 100); o protótipo asiático-americano formou um clado monofilético com dois outros protótipos americanos (AY686582.1 e AY686579.1), o que reforça a identificação dessas sequências como variantes leste-asiáticas; os protótipos asiáticos apresentaram-se polifiléticos, bem como os outros exemplares americanos (além dos já citados); a topologia obtida sugere que a maior parte dos protótipos adicionados após a primeira análise (cujo resultado é representado pelas figuras 1 e 2) formou um clado monofilético com variantes da Ásia e da Europa e; como a topologia resulta da média estimada a partir de vários modelos evolutivos, o tamanho dos ramos não indica divergência entre as sequências.

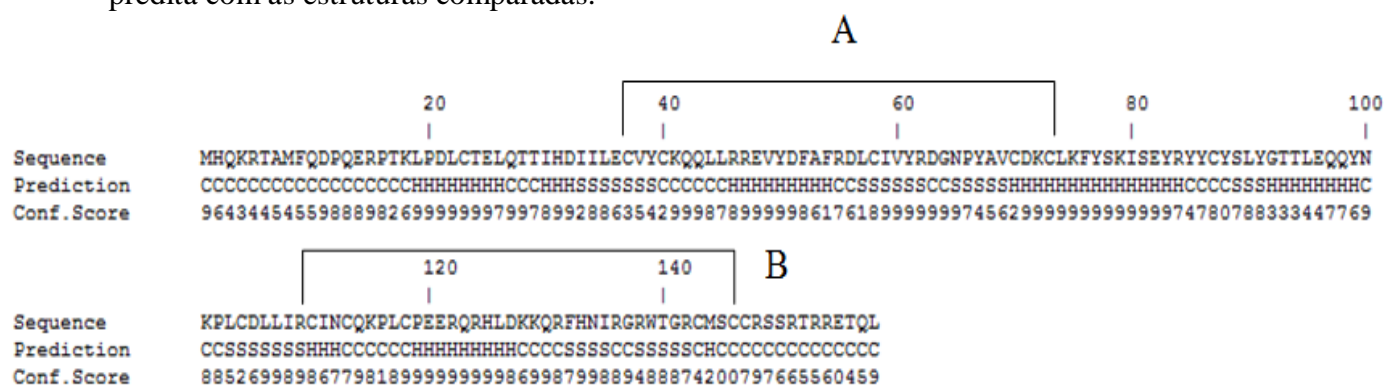
## **2. Predição e análise de estruturas**

Os modelos de estrutura terciária, gerados por homologia estrutural através do servidor I-TASSER, de cada região analisada estão presentes em anexo. As diferenças entre os modelos e as proteínas nativas poderiam ser elucidadas por estudos mais específicos de dinâmica molecular a partir de dados cristalográficos, mas como o objetivo deste trabalho foi realizar um estudo descritivo da molécula proteica a partir de sua estrutura primária a discussão será realizada sob tal perspectiva.

Os modelos teóricos obtidos neste trabalho foram comparados com outros modelos teóricos e modelos experimentais de proteínas de HPV 16 descritos na literatura. Os modelos resultantes das predições para as proteínas do protótipo HM057182 estão presentes no anexo 4.

## 2.1. E6

O *C-score* obtido com a estrutura terciária de E6 (figura 6) foi de -2,79, o que representa um baixo grau de proximidade do modelo teórico com a proteína nativa. Os resultados da análise da proteína E6 realizada pelo I-TASSER mostraram que as estruturas primárias dos domínios N e C-terminal de ligação ao zinco, descritos por Ullman *et al.* (1996), mantiveram-se conservadas, mas suas estruturas secundárias (figura 5) foram parcialmente recuperadas no modelo tridimensional obtido (figura 8). Os domínios de ligação ao zinco são compostos por duas regiões Cys-X-X-Cys-X29-Cys-X-X-Cys próximas a cada extremidade da proteína. Os alinhamentos estruturais do modelo com outras estruturas depositadas no PDB mostraram valores de *TM-score* aproximados a 0,5, demonstrando alto grau de similaridade da estrutura secundária predita com as estruturas comparadas.



**Figura 5 – Predição de estrutura secundária de E6.** Observa-se em A e B os domínios N e C-terminal, respectivamente.

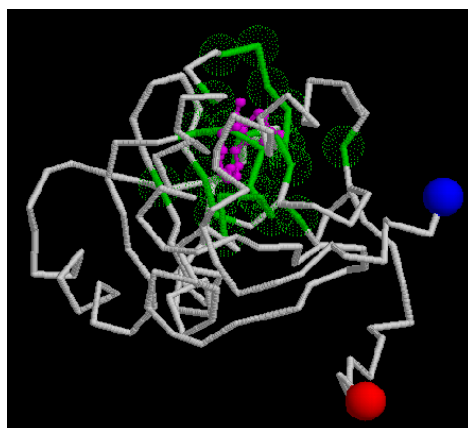


**Figura 6 – Estrutura molecular de E6.** As estruturas em  $\alpha$  hélice folha  $\beta$  estão destacadas em vermelho e amarelo, respectivamente.



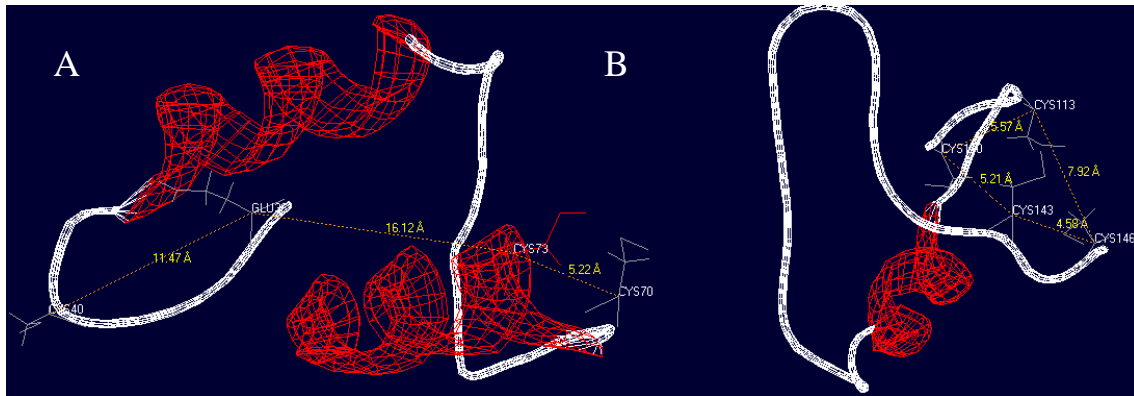
Os indicadores de possíveis ontologias mostraram que a estrutura obtida provavelmente apresentaria função ligada ao crescimento celular (GO:0008152) e que estabeleceria interações específicas com outras moléculas a partir de seu sítio de ligação (GO:0005488).

As regiões com maior probabilidade de representarem sítios de ligação à P53 obtiveram *BS-score* maior que 0,5, indicando alto grau de significância da predição. Tais regiões estão ilustradas na figura 7.



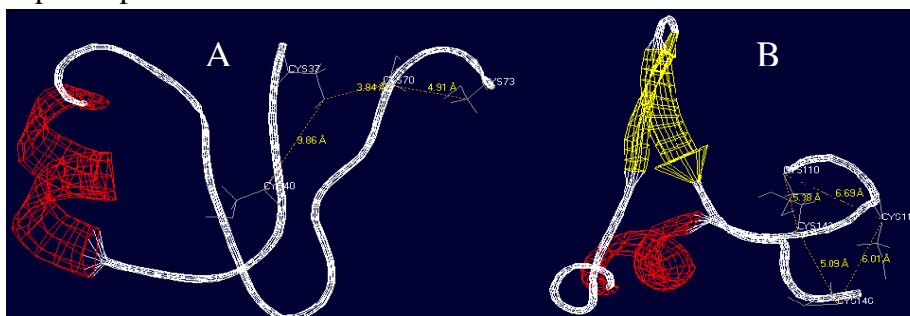
**Figura 7 – Estrutura molecular do provável sítio de ligação de E6.** A região representativa do sítio de ligação a P53 encontra-se destacada em verde. A esfera azul indica a extremidade C-terminal e a vermelha, a N-terminal.

Observa-se nas figuras 8 e 9 a formação de estruturas em  $\alpha$ -hélices e  $\beta$ -folha. Além disso, pode-se observar que os resíduos de cisteína do domínio N-terminal do modelo apresentam-se bastante afastados uns dos outros, o que reduziria a probabilidade destes resíduos coordenarem um átomo de zinco. Já o domínio C-terminal apresentou os resíduos de cisteína bem mais próximos uns dos outros (com menos de 10 Å de distância), o que contribuiria para a interação destes com um átomo de zinco.



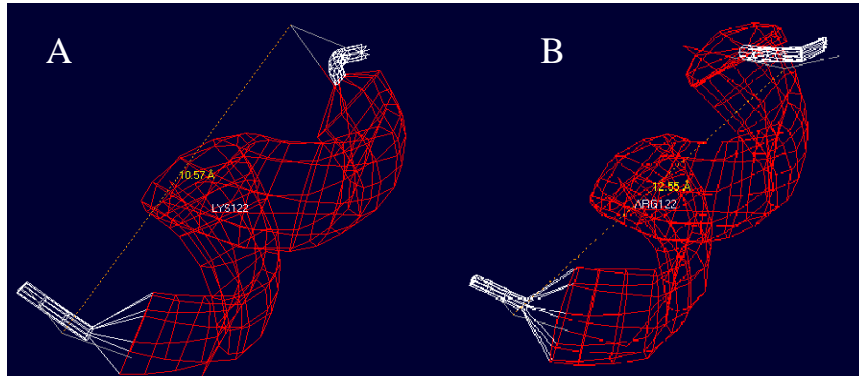
**Figura 8 – Resíduos de Cisteína dos domínios de E6.** Em A observa-se a distribuição espacial dos resíduos de Cys no domínio N-terminal. Em B observa-se distribuição espacial dos resíduos de Cys no domínio C-terminal.

A figura 9 permite estabelecer uma comparação entre o modelo de E6 obtido a partir do protótipo HM057182 e o modelo obtido a partir das variantes africanas. Comparando-se as figuras 8 e 9, tem-se que: não ocorreram desvios significativos entre a predição dos domínios de E6 das variantes africanas e do HPV HM057182, quanto a disposição espacial dos resíduos de cisteína de ambos os domínios; o domínio N-terminal do modelo africano não apresenta uma  $\alpha$ -hélice presente em HM057182 e; o domínio C-terminal de HM057182 não apresenta uma  $\beta$ -folha presente no modelo africano. Tais desvios podem decorrer de possíveis falhas ocorridas durante a predição do modelo pelo servidor ou de diferenças existentes entre as estruturas primárias de ambos os protótipos.



**Figura 9 – Resíduos de Cisteína dos domínios de E6 das variantes africanas.** Em A observa-se a distribuição espacial dos resíduos de Cys no domínio N-terminal. Em B observa-se a distribuição espacial dos resíduos de Cys no domínio C-terminal.

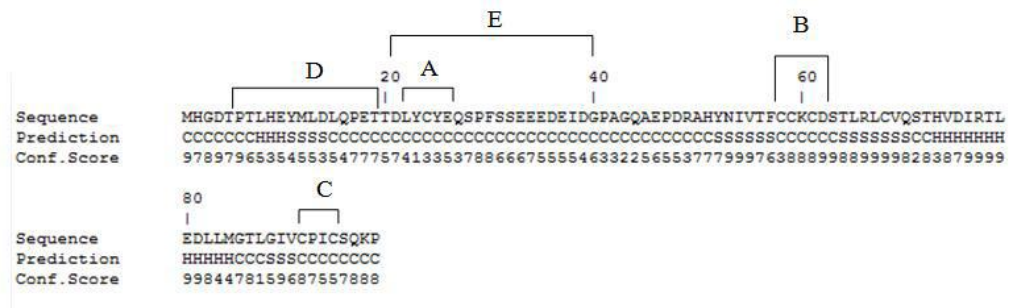
A figura 10 representa a região de E6 que apresentou mutação sinônima no sítio LYS122ARG, onde nota-se que não houve mudança significativa na estrutura terciária da molécula.



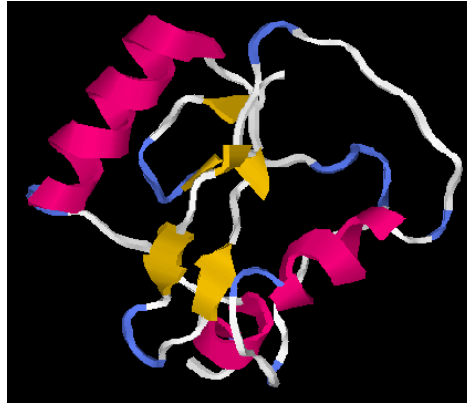
**Figura 10 – Comparação entre estruturas de E6 que apresentaram mutação.** Em A observa-se um fragmento da estrutura terciária de E6 das variantes africanas 1 (AF472508.1 e AF536180.1) e em B o fragmento correspondente na estrutura do protótipo HM057182.

## 2.2. E7

O *C-score* obtido com a estrutura terciária de E7 foi de -2,15. Contudo, tal modelo apresenta os domínios de E7 total ou parcialmente recuperados (Figuras 14 e 15). Os alinhamentos estruturais do modelo com outras estruturas depositadas no PDB mostraram valores de *TM-score* aproximados a 0,5, demonstrando alto grau de similaridade da estrutura secundária predita com as estruturas comparadas.



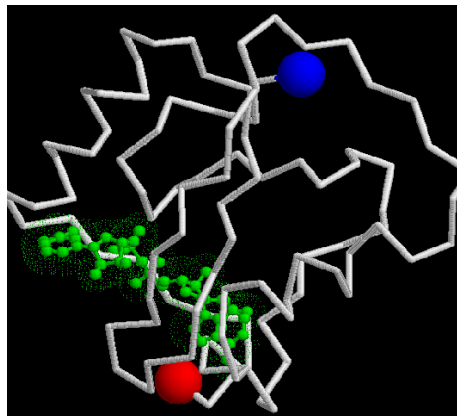
**Figura 11 – Predição de estrutura secundária de E7.** Observam-se os domínios homólogos a E1A (D e E), o domínio de ligação à pRB (A) e o domínio que forma o dedo de zinco da extremidade C-terminal (B e C).



**Figura 12 – Estrutura molecular de E7.** As estruturas em  $\alpha$  hélice folha  $\beta$  estão destacadas em vermelho e amarelo, respectivamente.

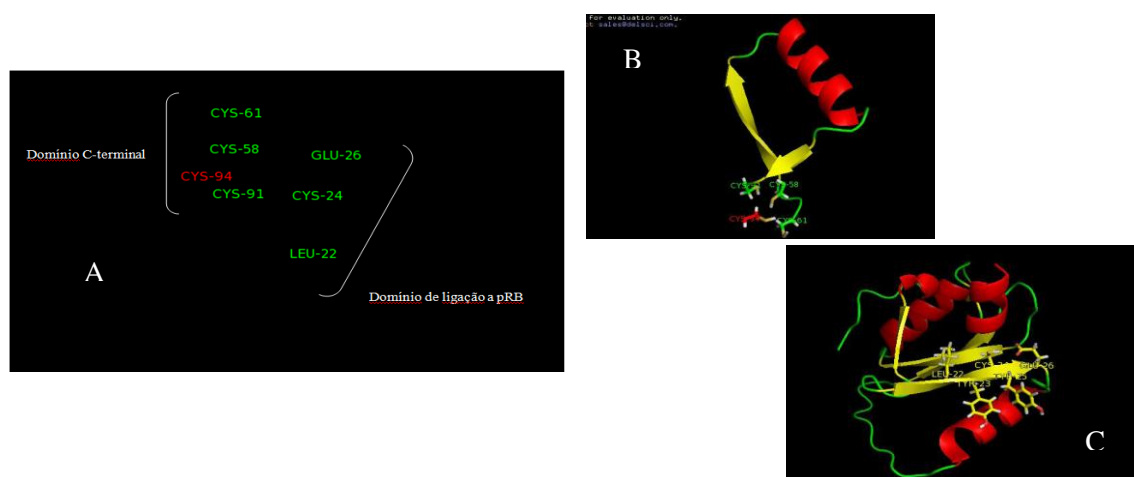
Os indicadores de possíveis ontologias mostraram que a estrutura obtida provavelmente apresentaria função ligada ao metabolismo celular de macromoléculas (GO:0044260) e que estabeleceria interações específicas com outras moléculas a partir de seu sítio de ligação (GO:0005488).

As regiões com maior probabilidade de representarem sítios de ligação obtiveram *BS-score* aproximado a 0,5. Tais regiões estão ilustradas na figura a 13.

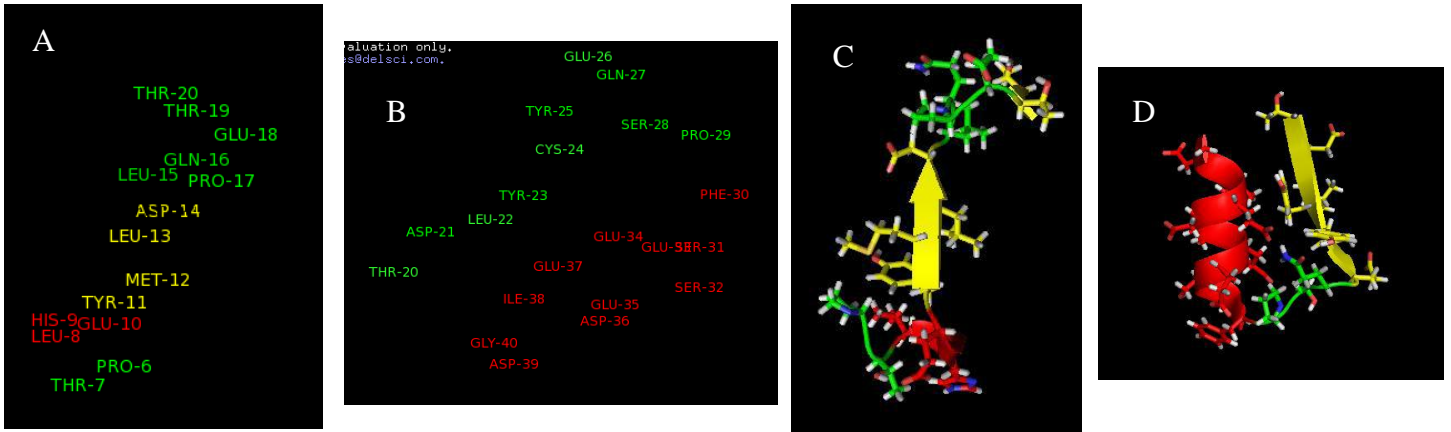


**Figura 13 – Estrutura molecular do provável sítio de ligação de E7.** A região representativa do sítio de ligação a pRB encontra-se destacada em verde. A esfera azul indica a extremidade C-terminal e a vermelha, a N-terminal.

Os resultados da análise da proteína E7 realizada pelo I-TASSER mostraram que: o domínio Cys-X-X-Cys-X29-Cys-X-X-Cys, que forma um dedo de zinco próximo ao C-terminal, o domínio de ligação à pRB (L-X-C-X-E próximo ao aminoácido 20 da cadeia primária) e os domínios homólogos à oncoproteína E1A de adenovírus mantiveram-se conservados na estrutura secundária (figura 11); as estruturas secundária e terciária apresentam a formação de estruturas em  $\alpha$ -hélices e  $\beta$ -folhas (figuras 11 e 12); o modelo de estrutura terciária obtido recuperou parcialmente a estrutura terciária do domínio C-terminal de ligação ao zinco, agrupando os resíduos de cisteína que coordenam o metal, porém, não formando a estrutura hélice-volta-hélice, característica de tais domínios (figura 14a e b); os resíduos do domínio de ligação a pRB dispuseram-se numa estrutura em  $\beta$ -folha (figura 14a e c) e; os resíduos da porção N-terminal do domínio homólogo a E1A apresentaram uma estrutura em  $\beta$ -folha, enquanto que os seus resíduos da porção C-terminal apresentaram a forma folha-alça-hélice (figura 15) (Watanabe *et al.*, 1990).

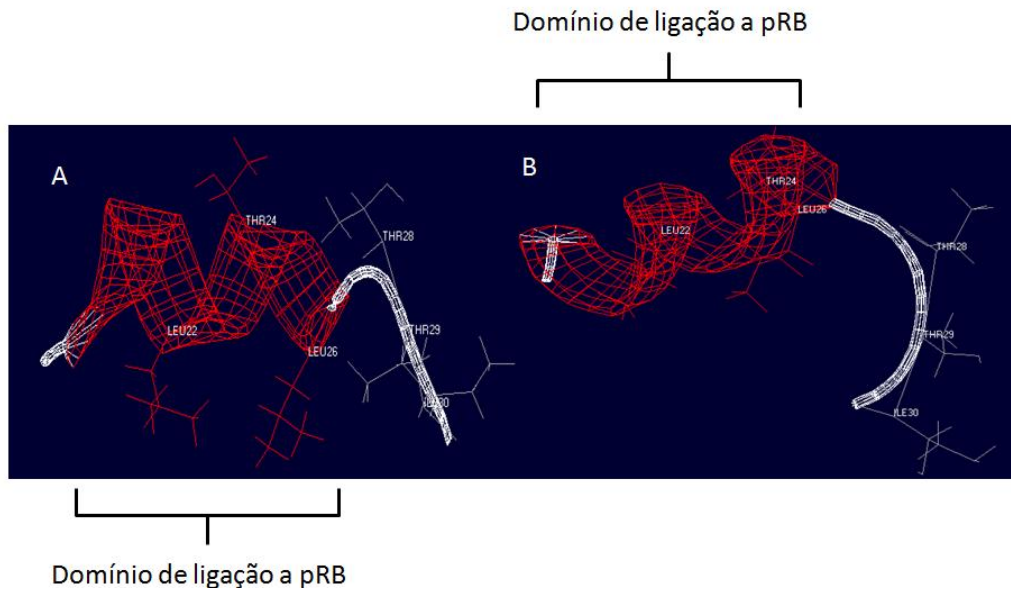


**Figura 14 – Estrutura molecular de prováveis sítios de ligação de E7.** Em A observa-se a distribuição espacial dos resíduos de Cys do domínio C-terminal e dos resíduos que formam o domínio de ligação à pRB. Em B observa-se a estrutura completa do domínio C-terminal, parcialmente recuperada pelo modelo, destacando-se os resíduos de Cys que coordenariam um átomo de zinco na estrutura. Em C observa-se a estrutura formada pelos resíduos do domínio de ligação a pRB.



**Figura 15 – Estrutura molecular do provável sítio de E7 homólogo a E1A.** Em A e B observa-se a distribuição espacial dos resíduos de aminoácidos que compõem, respectivamente, as porções N-terminal e C-terminal do domínio. Em C e D observa-se a estrutura tridimensional das duas porções do domínio.

A figura 16 apresenta a região que sofreu mutação não-sinônima a partir do protótipo africano. Observa-se que tais mutações não repercutiram em alterações significativas na estrutura terciária da proteína nesta região.



**Figura 16 – Comparação entre estruturas de E7 que apresentaram mutação.** Em A observa-se um fragmento da estrutura terciária de E6 das variantes africanas 1 (AF472508.1 e AF536180.1) próximo ao domínio de ligação a pRB e em B o fragmento correspondente na estrutura do protótipo HM057182.



O *C-score* obtido com a estrutura terciária de L1 foi de 0,38. Os alinhamentos estruturais do modelo com outras estruturas depositadas no PDB mostraram valores de *TM-score* aproximados a 0,9, demonstrando alto grau de similaridade da estrutura secundária predita com as estruturas comparadas.

Os indicadores de possíveis ontologias mostraram que a estrutura obtida provavelmente apresentaria função ligada ao metabolismo celular ligado ao crescimento celular (GO:0008152) e que apresentaria atividade enzimática (GO:0003824).

As análises de possíveis sítios de ligação da molécula indicam que não há sítios de ligação previstos para L1. Contudo, a composição de aminoácidos da porção N-terminal desta proteína (descrita na literatura como associada à sua dimerização e à entrada do vírus na célula hospedeira) é idêntica à composição de aminoácidos de L1 das outras variantes africanas já citadas anteriormente, o que nos permite supor que os sítios de dimerização e de associação a receptores de membrana não tenham sofrido alterações.

## **BIBLIOGRAFIA**

Cann, R. L., Stoneking, M., Wilson, A. C., 1987. Mitochondrial DNA and human populations. *Nature*. 325, 31-36.

de Villiers, E-M., Fauquet, C., Broker, T.R., Bernard, H-U., zur Hausen, H., 2004. Classification of papilomavíruses. *Journal of Virology*. 324, 17– 27.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. 39, 789-791.



Felsenstein, J., 2008. PHYLIP 3.68. Department of Genome Sciences and Department of Biology, University of Washington. Disponível em <<http://evolution.gs.washington.edu/phylip.html>>. Acesso em 01 de set.

Gottschling, M., Köhler, A., Stockxeth, E., Nindl, I., 2007. Phylogenetic analysis of beta-papillomaviruses as inferred from nucleotide and amino acid sequence data. *Molecular Phylogenetics and Evolution*. 42, 213–222.

Ho, L., Chan, S-Y., Chow, V., Chong, T., Tay, S-K., Villa, V. V., Bernard, H-U., 1991. Sequence Variants of Human Papillomavirus Type 16 in Clinical Samples Permit Verification and Extension of Epidemiological Studies and Construction of a Phylogenetic Tree. *Journal of Clinical Microbiology*. 29(9), 1765-1772.

Milne, I., Lindner, D., Bayer, M., Husmeier, D., McGuire, G., Marshall, D. F., Wright, F., 2009. TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. *Bioinformatics*. 25 (1), 126-127.

Ong, C-K., Chan, S-Y., Campo, M. S., Fujinaga, K., Mavromara-Nazos, P., Labropoulou, V., Pfister, H., Tay, S-K., Meulen, J. T., Villa, L. L., Bernard, H-U., 1993. Evolution of Human Papillomavirus Type 18: an Ancient Phylogenetic Root in Africa and Intratype Diversity Reflect Coevolution with Human Ethnic Groups. *Journal of Virology*. 67(11), 6424-6431.

Pande, S., Jain, N., Prusty, B. K., Bhambhani, S., Gupta, S., Sharma, S., Batra, S., Das, B. C., 2008. Human papillomavirus type 16 variant analysis of E6, E7 and L1 genes and Long Control Region in biopsy samples from cervical cancer patients in North India. *Journal of Clinical Microbiology*. 3(46), 1060-1066.

Posada, D., Crandall, K. A., 1998. Modeltest: Testing the model of DNA substitution. *Bioinformatics*. 14(9), 817-818.

- Roy, A., Kucukural, A., Zhang, Y., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*. 5, 725-738.
- Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*. 24, 1596-1599.
- Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*. 10, 512–526.
- Thomson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., Higgins, D. G., 1997. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*. 25, 4876–4882.
- Ullman, C. G., Haris, P. I., Galloway, D. A., Emery, V. C., Perkins, S. J., 1996. Predicted alpha-helix/beta-sheet secondary structures for the zinc-binding motifs of human Papillomavirus E6 and E7 proteins by consensus prediction averaging and spectroscopic studies of E7. *Biochem J*. 319, 229-239.
- Watanabe, S., Kanda, T., Sato, H., Furuno, A., Yoshiike, K., 1990. Mutational Analysis of Human Papillomavirus Type 16 E7 Functions. *Journal of Virology*. 64, 207-214.
- Zhang, Y., 2009. I-TASSER: Fully automated protein structure prediction in CASP8. *Proteins*. S9, 100-113.
- Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*. 9, 40.

## CONCLUSÕES

1. A análise genômica do HPV 16 HM057182 revelou que as regiões gênicas de E2, E5, E7 e L2 vêm sofrendo forte pressão seletiva a favor de mudanças em suas estruturas e que NCR representa uma região de grande variabilidade entre as variantes de HPV 16 estudadas.
2. O estudo das relações filogenéticas do HPV 16 HM057182 com 8 outros tipos de HPV e outros 27 protótipos de HPV 16, utilizando os dados moleculares da Região Não-Codificadora (NCR), mostrou que o HPV 16 HM057182 surgiu a partir de variantes africanas tipo 1 – após o surgimento das duas variantes a partir de uma ancestral – mantendo-se evolutivamente mais próxima de AF472508.1.
3. Os modelos teóricos de estrutura terciária das proteínas E6 e E7 obtidos neste trabalho recuperaram parcialmente os domínios de ligação destas moléculas a seus substratos, devido a possíveis falhas nos cálculos das predições causadas pela falta de estruturas resolvidas experimentalmente do Banco de Dados de Proteínas (PDB) que permitissem modelagens mais aproximadas de dados experimentais.
4. O servidor utilizado para a predição do modelo teórico de estrutura terciária da proteína L1, apesar de atribuir-lhe grande proximidade estatística a estruturas resolvidas do PDB não evidenciou domínios de dimerização ou de ligação desta proteína a receptores celulares. Contudo, devido a estrutura peptídica primária ter se conservado entre as variantes analisadas, conclui-se que os domínios citados não tenham sofrido relevantes mudanças conformacionais.

# **ANEXOS**

## ANEXO 1: Anotação do HPV 16 no GenBank

LOCUS Complete\_genome 7915 bp DNA circular 11-MAR-2010  
DEFINITION Amazonian Human Papilomavirus 16 (complete genome).  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE Human Papilomavirus 16 amazonian isolate  
ORGANISM Human Papilomavirus 16 amazonian isolate  
Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;  
Alphapapillomavirus.  
REFERENCE 1 (bases 1 to 7915)  
AUTHORS Filho,R.A.A.Barbosa., Santos,C.M.Borborema. and  
Filho,S.Astolfi.  
TITLE Amazonian Human papilomavirus 16 complete genome  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 7915)  
AUTHORS Filho,R.A.A.Barbosa., Santos,C.M.Borborema. and  
Filho,S.Astolfi.  
TITLE Direct Submission  
JOURNAL Submitted (11-MAR-2010) Biotechnology, Molecular Diagnostic's  
lab.,  
Amazonas' Federal University, Av. Gen. Rodrigo Otavio Jordao Ramos,  
Manaus, Amazonas, Brazil  
FEATURES Location/Qualifiers  
source 1..7915  
/organism="Human Papilomavirus 16 amazonian isolate"  
/proviral  
/mol\_type="genomic DNA"  
misc\_feature join(<1..81,7142..7915)  
/note="Region: Long Control Region"  
gene 82..558  
/gene="E6"  
CDS 82..558  
/gene="E6"  
/note="transforming protein"  
/codon\_start=1  
/product="E6"  
/translation="MHQKRTAMFQDPQERPTKLPDLCTELQTTIHDIILECVYCKQQL  
LRREVYDFAFRDLCIVYRDGNPYAVCDKCLKFYISKISEYRYCYSLYGTTLTLEQQYNKP

LCDLLIRCINCQKPLCPPEERQRHLDDKKQRFHNIIRGRWTGRCMSCCRSSRTRRETQL"  
 gene 561..857  
 /gene="E7"  
 CDS 561..857  
 /gene="E7"  
 /note="transforming protein"  
 /codon\_start=1  
 /product="E7"  
 /translation="MHGDTPTLHEYMLDLQPETTDLYCYEQSPFSSEEEDEIDGPAGQ  
 AEPDRAHYNIIVTFCKCDSTLRLCVQSTHVDIRLTLEDLLMGTGLGIVCPICSQKP"  
 gene 864..2813  
 /gene="E1"  
 CDS 864..2813  
 /gene="E1"  
 /note="replication protein"  
 /codon\_start=1  
 /product="E1"  
 /translation="MADPAGTNGEETGCGNGWFYVEAVVEKKTGDAISDDENENDSDT  
 GEDLVDFIVNDNDYLTQAETETAHALFTAQEAKEHRDAVQVLKRKYLGSPLSDISGCV  
 DNNISPRLKAICIEKQSRRAKRRLFESEDSGYGNTEVETQQMLQVEGRHETETPCSQY  
 SGGSGGGSSQYSSGGSGGEGVSEERHAICQTPLTNIILNVLKTNAKAAMLAKFKELYGVS  
 FSELVVRPFKSNKSTCCDWCIAAFGLTPSIADSIKTLLQQYCLYLHIQSLACSWGMVVL  
 LLVRYKCGKNRETIKLLSKLLCVSPMCMIEPPKLRSTAAALYWKYTGMSNISEVYG  
 DTPEWIQRQTVLQHSFNDCTFELSQMWAYDNDIVDDSEIAYKYAQLADTNSNASAF  
 LKSN SQAKIVKDCATMCRHYKRAEKKQMSMSQWIKYRCRVDGDDGDKQIVMFLRYQG  
 VDFMSFLTALKRFLQGIPKKNCILLYGAANTGKSLFGMSLMKFLQGSVICFVNSKSHF  
 WLQPLADAKIGMLDDATVPCWNYIDDNLRNALDGNLVSMDVKHRPLVQLKCPPLLITS  
 NINAGTDSRWPYLHNRLLVVF'TFPNEFPFDENGNPVYELNDKNWKSFFSRTWSRSLHE  
 DEDKENDGDSLPTFKCVSGQNTNTL"  
 gene 2755..3852  
 /gene="E2"  
 CDS 2755..3852  
 /gene="E2"  
 /note="regulatory protein"  
 /codon\_start=1  
 /product="E2"  
 /translation="METLCQRLNVCQDKILTHYENDSTDLRDHIDYWKQMRLECAIYY  
 KAREMGFKHINHQVVP'PLAVSKNKALQAIELQLTLETIYNSQYSNEKWTLQDVSLEVY  
 LTAPTGCICKKHGYTVEVQFDGICNTMHYTNWKYIYICEETS'VTVVEGQVDYYGLYYV  
 HEGIQT'YFVQFKDDAEKYSKNKVWEVHAGGQVILCPTSVFSSDEVSSAEIIRQHLANH

SAATHPKAVALGKETQTTIQRPSEPDTGNPCHTNKLLHRDSVDSAPILTAVNSSHK  
GRINCNSNTTPIVHLKGDANTLKCLRYRFKKHCKLYTAVSSTWHWTGHNVKHKSAIVT  
LTYDSECQREQFLSQVKIPKTITVSTGFMSI"

gene <3332..3619

/gene="E4"

CDS <3332..3619

/gene="E4"

/note="transforming protein"

/codon\_start=1

/product="E4"

/translation="YYVLHLCLAATKYPLKLLGSTWPTTPRPPIPKPSPWAPKKHRR  
LSSDQDQSQTPETPATPISCCTETQWTVLQSSLQLTAHTKDGLTVIVTLHP"

gene 3849..4100

/gene="E5"

CDS 3849..4100

/gene="E5"

/note="transforming protein"

/codon\_start=1

/product="E5"

/translation="MTNLDTTSTLLACFLLCFVLLCVCLLIRPLLSVSTYTSLLI  
LVLLWITAASAFAFRCFIVYILFVYIPLSLIHTHARFLIT"

gene 4222..5643

/gene="L2"

CDS 4222..5643

/gene="L2"

/note="minor capsid protein"

/codon\_start=1

/product="L2"

/translation="MRHKRSKRTRKASATQLYKTCKQAGTCPPDIIPKVEGKTIADQ  
ILQYGSMSGVFFGGLGIGTGSCTGGRTGYIPLGTRPPTATDTLAPVRPPLTVDPVGP  
PSIVSLVEETS FIDVGAPTPVPSIPPDVSGFSITTTSTDTTPAILDINNTVTTVTTHNN  
PTFTDPSVLQPPTPAETGGHFTLSSSTISTHNYEEIPMDTFIVSTNPNTVTSSTPIPG  
SRPVARLGLYSRTTQQVKVVDPAFVTTPTKLITYDNPAYEGIDVDNTLYFPSNDNSIN  
IAPDPDFLDIVALHRPALTSRRTGIRYSRIGNKQTLRTRSGKSIGAKVHYYYDLSTIN  
PAEEIELQTITPSTYTTASHAASPTSINNGLYDIYADDFITDTSTTPVPSIPSTSLSG  
YIPANTTIPFGGAYNIPLVSGPDIPINTDQTPSLIPIVPGSPQYTIADGGDFYLHP  
SYYMLRKRRLPYFFSDVSLAA"

gene 5546..7141

/gene="L1"

CDS 5546..7141

```

/gene="L1"
/note="major capsid protein"
/codon_start=1
/product="L1"
/translation="MEVTFIYILVITCYENDVNVYHIFQMSLWLPSEATVYLPPVPV
SKVVTDEYVARTNIYYHAGTSRLLAVGHPYFPIKKPNNNKILVPKVSGLQYRVFRIY
LPDPNKFGFPDTSFYNPDTQRLVWACVGEVGRGQPLGVGISGHPLLKLDDEENASA
YAANAGVDNRECISMDYKQTQLCLIGCKPPIGEHWGKGSPCNNVAVNPGDCPPLELIN
TVIQDGMVDTGFGAMDFTTLQANKSEVPLDICTSICKYDPYIKMVSEPYGDSLFFYL
RREQMFVRHLFNRAVAVGENVPDDLYIKGSGSTANLASSNYFPTPSGSMVTSDAQIFN
KPYWLQRAQGHNNIGCWGNQLFVTVDTRSTNMSLCAAISTSETTYKNTNFKEYLRH
GEEYDLQFIFQLCKITLTADVMTYIHSMNSTILEDWNFGLQPPPGGTLEDTYRFVTSQ
AIACQKHTPPAPKEDPLKKYTFWEVNLKEKFSADLDQFPLGRKFLQAGFKAKPKFTL
GKQKATPTTSSTSTTAKRKRKRL"

```

BASE COUNT 2608 a 1365 c 1515 g 2427 t

ORIGIN

```

   1 actacaataa ttcattgtata aactaagggt gtaaccgaaa tcggttgaac cggaaccggt
  61 tagtataaaa gcagacattt tctgcaccaa aagagaactg caatgtttca ggaccacag
121 gagcgaccca caaagttacc agatttatgc acagagctgc aaacaactat acatgatata
181 atattagaat gtgtgtactg caagcaacag ttactgcgac gtgaggtata tgactttgct
241 tttcgggatt tatgcatagt atatagagat gggaatccat atgcagtgtg tgataaatgt
301 ttaaagtttt attctaaaat tagtgagtat agatattatt gttatagttt gtatggaaca
361 acattagaac agcaatacaa caaacggtg tgtgatttgt taattagggtg tattaactgt
421 caaaagccac tgtgtcctga agaaaggcaa agacatctgg acaaaaagca aagattccat
481 aatataaggg gtcggtggac cggtcgatgt atgtcttgtt gcagatcatc aagaacacgt
541 agagaaaacc agctgtaatc atgcatggag atacacctac attgcatgaa tatatgttag
601 atttgcaacc agagacaact gatctctact gttatgagca atcgccctc agctcagagg
661 aggaggatga aatagatggt ccagctggac aagcagaacc ggacagagcc cattacaata
721 ttgtaacctt ttgttgcaag tgtgactcta cgcttcgggt gtgcgtaaca agcacacag
781 tagacatccg tacgttgtaa gacctgttaa tgggcacact aggaattgtg tgccccatct
841 gttctcagaa accataatct accatggctg atcctgcagg taccaatggg gaagagggta
901 cgggatgtaa tggatggttt tatgtagagg ctgtagtga aaaaaaaca ggggatgcta
961 tatcagatga cgagaacgaa aatgacagtg atacaggtga agatttggtg gattttatag
1021 taaatgataa tgattattta acacaggcag aaacagagac agcacaatgc ttgtttactg
1081 cacaggaagc aaaagaacat agagatgcag tacaggttct aaaacgaaag tatttggtgta
1141 gtccacttag tgatattagt ggatgtgtag acaataatat tagtcctaga ttaaagcta
1201 tatgtataga aaaacaaagt agagctgcaa aaaggagatt atttgaaagc gaagacagcg
1261 ggtatggcaa tactgaagtg gaaactcagc agatgttaca ggtagaaggc cgccatgaga
1321 ctgaaacacc atgtagtcag tatagtggtg gaagtggggg tggtagcagt cagtatagta
1381 gtggaagtgg gggagagggt gttagtgaaa gacacgctat atgccaaca ccacttaaa

```



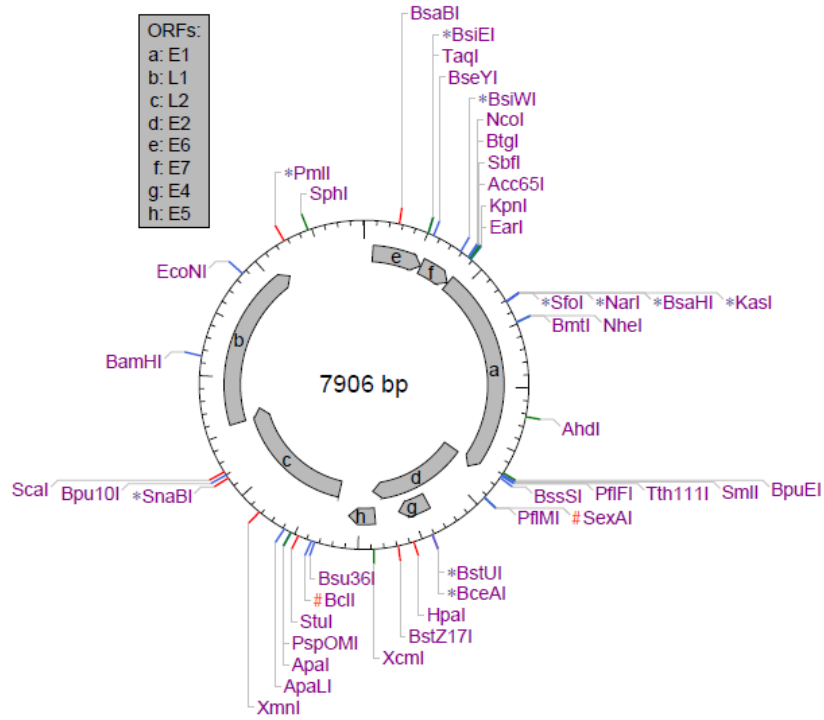
1441 atatttttaa tgtactaaaa actagtaatg caaaggcagc aatgctagca aaattttaaag  
1501 agttatacgg ggtgagtttt tcagaattag taagaccatt taaaagtaat aatcaacgt  
1561 gttgcgattg gtgtattgct gcatttggac ttacacctag tatagctgac agtataaaaa  
1621 cattattaca acaatattgt ttatatattac acattcaaag tttagcatgt tcatggggaa  
1681 tggttgtggt actattagta agatataaat gtggaaaaaa tagagaaaca attgaaaaat  
1741 tgctgtctaa actattatgt gtgtctcaa tgtgtatgat gatagagcct ccaaaattgc  
1801 gtagtacagc agcagcatta tattggtata aaacaggtat gtcaaatatt agtgaagtgt  
1861 atggagacac gccagaatgg atacaaagac aaacagtatt acaacatagt tttaatgatt  
1921 gtacatttga attatcacag atggtacaat gggcctacga taatgacata gtagacgata  
1981 gtgaaattgc atataaatat gcacaattgg cagacactaa tagtaatgca agtgcctttt  
2041 taaaaagtaa ttcacaggca aaaattgtaa aggattgtgc aacaatgtgt agacattata  
2101 aacgagcaga aaaaaaaca atgagtatga gtcaatggat aaaatataga tgtgataggg  
2161 tagatgatgg aggtgattgg aagcaaattg ttatgttttt aaggtatcaa ggtgtagact  
2221 ttatgtcatt tttactgca ttaaaaagat ttttgcaagg catacctaaa aaaaattgca  
2281 tattactata tgggtgcagct aacacaggta aatcattatt tggatgagt ttgatgaaat  
2341 ttttgcaagg gtctgtaata tgttttgtaa attctaaaag ccatttttgg ttacaacat  
2401 tagcagatgc caaaataggt atgttagatg atgctacagt gccctgttgg aactacatag  
2461 atgacaattt aagaaatgca ttggatggaa attagtttc tatggatgta aagcatagac  
2521 cattggtaca actaaaatgc cctccattat taattacatc taacattaat gctggtacag  
2581 attccagggtg gccttatttta cataatagat tgggtggtgt tacattttcca aatgagtttc  
2641 catttgacga aaacggaaat ccagtgtatg agcttaatga taagaactgg aaatcctttt  
2701 tctcaaggac gtggtccaga ttaagtttgc acgaggacga ggacaaggaa aacgatggag  
2761 actctttgcc aacgttttaa tgtgtgtcag gacaaaatac taacacatta tgaaaatgat  
2821 agtacagacc tacgtgacca tatagactat tggaaacaaa tgcgcctaga atgtgctatt  
2881 tattacaagg ccagagaaat gggatttaa catattaacc accagggtgg gccaacactg  
2941 gctgtatcaa agaataaagc attacaagca attgaactgc aactaacgtt agaaacaata  
3001 tataactcac aatataagtaa tgaaaagtgg acattacaag atgttagcct tgaagtgtat  
3061 ttaactgcac caacaggatg tataaaaaaa catggatata cagtggaagt gcagtttgat  
3121 ggagacatat gcaatacaat gcattataca aactggaaat atatatatat ttgtgaagaa  
3181 acatcagtaa ctgtggtaga ggtcaagtt gactattatg gtttatatta tgttcatgaa  
3241 ggaatacaaa catattttgt gcagtttaa gatgatgcag aaaaatatag taaaaataaa  
3301 gtatgggaag ttcattgcggg tggtcaggta atattatgct ctacatctgt gtttagcagc  
3361 gacgaagtat cctctgctga aattattagg cagcacttgg ccaaccactc cgccgagacc  
3421 catcccaaag ccgtcgcctt gggcaccaaa gaaacacaga cgactatcca gcgaccaaga  
3481 tcagagccag acaccggaaa cccctgccac accaataagt tgttgcacag agactcagtg  
3541 gacagtgctc caatcctcac tgcagttaac agctcacaca aaggacggat taactgtaat  
3601 agtaacacta cacccatagt acatttaaaa ggtgatgcta atactttaa atgtttaaga  
3661 tatagattta aaaagcattg taaattgtat acagcagtg cgtctacatg gcattggaca  
3721 ggacataatg taaaacataa aagtgcaatt gttacactta catatgatag tgaatgtcaa  
3781 cggaacaat ttttgtctca agttaaata ccaaaaacta ttacagtgtc tactggattt

3841 atgtctatat gacaaacctt gatactacat ccacaacatt actggcgtgc tttttgcttt  
3901 gcttttgtgt gcttttgtgt gtctgcctat taatacgtcc gctgcttttg tctgtgtcta  
3961 catacacatc attaatacta ttggtattat tattgtggat aacagcagcc tctgcgttta  
4021 ggtgttttat tgtatatatt ttatttgttt atataccatt atctttaata catacacatg  
4081 cacgcttctt aattacataa tgtatatgta cataatgtaa ttgttacata taattgttgt  
4141 atatgataac ttactgattt ttttttttat ttttatatat atagtttttt tttttatttg  
4201 tttgtttggt ttttaataaa catgcgacac aaacgttctg caaacgcac aaaacgtgca  
4261 tcggccaccc aactttataa aacatgcaa caagcaggta catgtccacc tgacattata  
4321 cctaagggtg aaggcaaac tattgctgat caaatattac aatatggaag tatgggtgta  
4381 ttttttgggt ggtaggaat tggaacaggg tcaggtacag gcgagcacac tgggtatatt  
4441 ccattaggaa caaggcctcc cacagctaca gatacacttg ctctgtgtaag accccttta  
4501 acagtagatc ctgtgggcc ttctgatcct tctatagttt ctttagtggga agaaactagt  
4561 tttattgatg ttggtgcacc aacacctgta ccttcattc cccagatgt atcaggattt  
4621 agtatcaciaa cttcaactga taccacacct gctatattag atattaataa tactgttact  
4681 actgttacta cacataataa tcccactttt actgacccat ctgtattgca gcctccaaca  
4741 cctgcagaaa ctggagggca ttttactt tcatcatcca ctattagtac acataattat  
4801 gaagaaattc ctatggatac atttattggt agcacaaacc ctaacacagt aactagtagc  
4861 acaccatac cagggtctcg cccagtggca cgccttggat tatatagtcg cacaacacia  
4921 caagttaaag ttgtagacc tgcttttgta accactcca ctaaacttat tacatatgat  
4981 aatcctgcat atgaaggat agatgtggat aatacattat attttcctag taatgataat  
5041 agtattaata tagctccaga tctgacttt ttggatatag ttgctttaca taggccagca  
5101 ttaacctcta ggcgtactgg cattagatac agtagaattg gtaataaaca aacactacgt  
5161 actcgtagtg gaaaatctat aggtgctaag gtacattatt attatgattt aagtactatt  
5221 aatcctgcag aagaaataga attacaaact ataacacctt ctacatatac taccgcttca  
5281 catgcagcct cacccacttc tattaataat ggattatag atatttatgc agatgacttt  
5341 attacagata cttctacaac cccagtacca tctataccct ctacatcctt atcaggttat  
5401 attcctgcaa atacaacaat tccttttgggt ggtgcataca atattccttt agtatcaggt  
5461 cctgatatac ccattaatac aactgaccaa actccttcat taattcctat agttccaggg  
5521 tctccacaat atacaattat tgctgatgga ggtgactttt atttacctc tagttattac  
5581 atgttacgaa aacgacgtaa acgtttacca ttttttttt cagatgtctc tttggctgcc  
5641 tagcgaggcc actgtctact tgctcctgt cccagtatct aaagttgtaa gcacggatga  
5701 atatgttgca cgcacaaaca tatattatca tgcaggaaca tccagactac ttgcagttgg  
5761 acatccctat tttcctatta aaaaaccta caataacaaa atattagttc ctaaagtatc  
5821 aggattacaa tacaggggat ttagaatata tttacctgac cccaataagt ttggttttcc  
5881 tgacacctca ttttacaatc cagatacaca gcggctgggt tgggcctgtg taggtgttga  
5941 ggtaggtcgt ggtcagccat taggtgtggg cattagtggc catcctttat taaataaatt  
6001 ggatgacaca gaaaatgcta gtgcttatgc agcaaagca ggtgtggata atagagaatg  
6061 tatatctatg gattacaaac aaacacaatt gtgtttaatt ggttgcaaac cacctatagg  
6121 ggaacactgg ggcaaaggat ccccatgtaa caatgttgca gtaaaccag gtgattgtcc  
6181 accattagag ttaataaaca cagttattca ggatgggtgat atgggtgata ccggctttgg

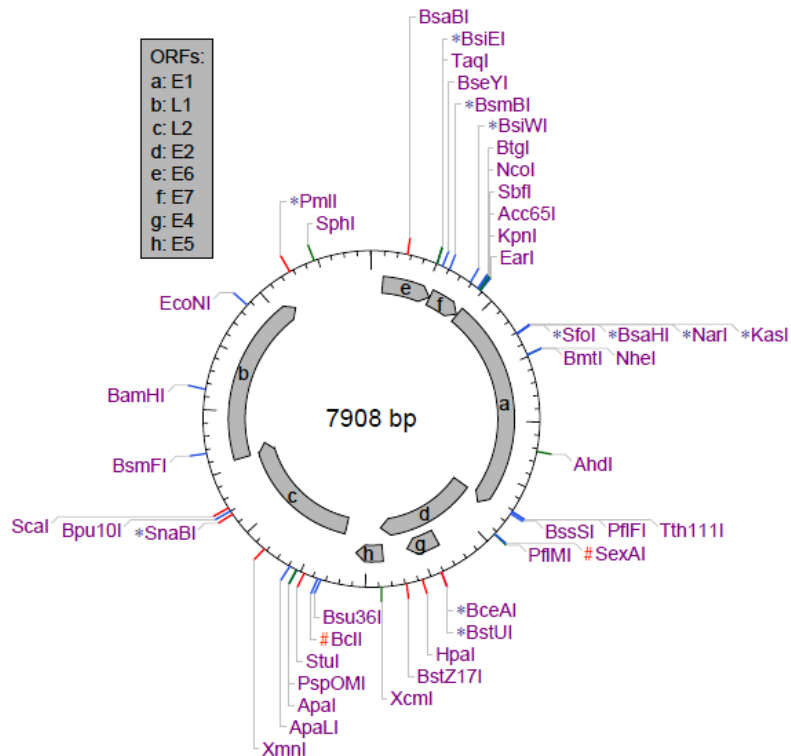
6241 tgctatggac tttactacat tacaggctaa caaaagtgaa gttccactgg atatttgtac  
6301 gtctatttgc aaatatccag attatattaa atgggtgtca gagccatatg gcgacagctt  
6361 atttttttat ttacgaaggg aacaaatggt tgtagacat ttatttaata gggctggtgc  
6421 tgttggtgaa aatgtaccag acgatttata cattaaaggc tctgggtcta ctgcaaattt  
6481 agccagttca aattattttc ctacacctag tggttctatg gttacctctg atgcccaaat  
6541 atttaataaa ccatattggt tgcaacgagc acagggtcat aacaatggca tttggtgggg  
6601 taaccaacta tttgttactg ttgttgatac tacacgcagt acaaataatgt cattatgtgc  
6661 tgccatatct acttcagaaa ctacatataa aaataactaac tttaaagagt acctacgaca  
6721 tggggaggaa tatgatttac agtttatttt tcaactgtgc aaaataacct taactgcaga  
6781 cgttatgaca tacatacatt ctatgaattc cactattttg gaggactgga attttggttt  
6841 acaacctccc ccaggaggca cactagaaga tacttatagg tttgtaacat cccaggcaat  
6901 tgcttgtcaa aaacatacac ctccagcacc taaagaagat ccccttaaaa aatatacttt  
6961 ttgggaagta aatttaaaag aaaagttttc tgacagaccta gatcagtttc ctttaggacg  
7021 caaattttta ctacaagcag gatttaaggc caaaccaaaa tttacattag gaaaacaaaa  
7081 agctacaccc accacctcat ctacctctac aactgctaaa cgcaaaaaac gtaagctgta  
7141 agtattgtat gtatggtgaa ttagtgttgt atgttgttta tatgtttgta tgtgcttgta  
7201 tgtgcttgta aatattgagt tgtatgtgtg tttgtatgta tggataata aacacgtgtg  
7261 tatgtgtttt taaatgctg tgtgtatgtg tttttaaag cttgtgtaac tattgtgtca  
7321 tgcaacataa ataaacttat tgtttcaaca cctactaatt gtgttggtgt tattcattgt  
7381 atataaacta tatttgctac atcctgtttt tgttttatat atactatatt ttgtagcgcc  
7441 agcggccatt ttgtagcttc aaccgaattc gggtgcatgc tttttggcac aaaatgtatt  
7501 tttttaaata gttctatgtc agcaactata gtttaaactt gtacgtttcc tgcttgccat  
7561 gcgtgcaaaa tccctgtttt cctgacctgc actgcttgcc aaccattcca ttgttttttg  
7621 cactgcacta tgtgcaacta ctgaatcact atgtacattg tgtcatataa aataaatcac  
7681 tatgcgcaa cgccttaa ataccgctgta ggcacatatt tttggcttgt ttttaactaac  
7741 ctaattgcat atttggcata aggtttaa ac ttttaaggcc aactaaatgt caccttagtt  
7801 catacatgaa ctgtgtaaag gttagtcata cattgttcat ttttaaaact gcacatgggt  
7861 gtgtgcaaac cgttttgggt tacaaattta caagcaactt atataataat actaa

## ANEXO 2: MAPAS DE RESTRIÇÃO

### AF536180.1



### AF472508.1



# ANEXO 3: ARQUIVO DE SAIDA DO PROGRAMA MODELTEST

## 0.1.1

### NCR

```
----- jModeltest 0.1.1 -----
(c) 2008 David Posada, Department of Biochemistry, Genetics and Immunology
University of Vigo, 36310 Vigo, Spain. e-mail: dposada@uvigo.es
-----

Mon May 24 23:27:40 GMT-04:00 2010 (Windows Vista 6.0, arch: x86)

***** NOTICE *****
This program may contain errors. Please inspect the results carefully.
*****

Reading data file "NCR.phy"... OK.

  number of sequences: 16
  number of sites: 1047

-----
*                                     *
*      COMPUTATION OF LIKELIHOOD SCORES WITH PHYML      *
*                                     *
-----

Settings:

Phyml version = 2.4.4
Candidate models = 88
  number of substitution schemes = 11
  including models with equal/unequal base frequencies (+F)
  including models with/without a proportion of invariable sites (+I)
  including models with/without rate variation among sites (+G) (nCat = 4)
Optimized free parameters (K) = substitution parameters + 29 branch lengths + topology
Base tree for likelihood calculations = ML tree

Maximum likelihood estimation for the TPM3uf+G model.

  ML optimized tree topology

  Model = TPM3uf+G
```

```

partition = 012012
-lnL = 8727.0857
K = 36
freqA = 0.2685
freqC = 0.1758
freqG = 0.1825
freqT = 0.3732
R(a) [AC] = 1.4853
R(b) [AG] = 3.0813
R(c) [AT] = 1.0000
R(d) [CG] = 1.4853
R(e) [CT] = 3.0813
R(f) [GT] = 1.0000
gamma shape = 4.1200

```

Computation time = 00h:01:03:04 (00h:28:49:03)

Computation of likelihood scores completed. It took 00h:58:10:00.

```

-----
*
*          BAYESIAN INFORMATION CRITERION (BIC)          *
*
-----

```

Settings:

sample size = 1047

Model selected:

```

Model = TPM3uf+G
partition = 012012
-lnL = 8727.0857
K = 36
freqA = 0.2685
freqC = 0.1758
freqG = 0.1825
freqT = 0.3732
R(a) [AC] = 1.4853
R(b) [AG] = 3.0813
R(c) [AT] = 1.0000
R(d) [CG] = 1.4853
R(e) [CT] = 3.0813
R(f) [GT] = 1.0000
gamma shape = 4.1200

```

\* BIC MODEL SELECTION : Selection uncertainty

Model	-lnL	K	BIC	delta	weight	cumWeight
TPM3uf+G	8727.0857	36	17704.5039	0.0000	0.8325	0.8325
TIM3+G	8726.0378	37	17709.3619	4.8580	0.0734	0.9059
HKY+G	8733.4382	35	17710.2553	5.7514	0.0469	0.9528
TPM3uf+I+G	8727.0834	37	17711.4531	6.9492	0.0258	0.9786
TrN+G	8732.0068	36	17714.3461	9.8422	0.0061	0.9847
TPM2uf+G	8732.2573	36	17714.8472	10.3433	0.0047	0.9894
TVM+G	8725.6942	38	17715.6284	11.1245	0.0032	0.9926
TIM3+I+G	8726.0388	38	17716.3176	11.8136	0.0023	0.9948
TPM1uf+G	8733.1036	36	17716.5399	12.0360	0.0020	0.9969
HKY+I+G	8733.4393	36	17717.2113	12.7073	0.0014	0.9983
TIM2+G	8730.8447	37	17718.9757	14.4718	0.0006	0.9989
GTR+G	8724.6664	39	17720.5265	16.0226	0.0003	0.9992
TIM1+G	8731.6763	37	17720.6388	16.1349	0.0003	0.9995
TrN+I+G	8732.0094	37	17721.3051	16.8012	0.0002	0.9996

TPM2uf+I+G	8732.2584	37	17721.8031	17.2991	0.0001	0.9998
TVM+I+G	8725.6912	39	17722.5761	18.0722	9.91e-005	0.9999
TPM1uf+I+G	8733.1047	37	17723.4957	18.9918	6.26e-005	1.0000
TIM2+I+G	8730.8473	38	17725.9346	21.4306	1.85e-005	1.0000
TPM3uf+I	8738.2637	36	17726.8601	22.3561	1.16e-005	1.0000
GTR+I+G	8724.6673	40	17727.4819	22.9780	8.53e-006	1.0000
TIM1+I+G	8731.6788	38	17727.5977	23.0937	8.05e-006	1.0000
HKY+I	8744.6342	35	17732.6473	28.1434	6.44e-007	1.0000
TIM3+I	8737.6986	37	17732.6836	28.1796	6.33e-007	1.0000
TPM3uf	8745.2713	35	17733.9215	29.4176	3.41e-007	1.0000
TPM2uf+I	8743.2272	36	17736.7869	32.2830	8.13e-008	1.0000
TVM+I	8736.6899	38	17737.6199	33.1159	5.36e-008	1.0000
TrN+I	8743.9749	36	17738.2824	33.7784	3.85e-008	1.0000
TPM1uf+I	8744.3918	36	17739.1162	34.6123	2.54e-008	1.0000
TIM3	8744.6467	36	17739.6261	35.1221	1.97e-008	1.0000
HKY	8752.2511	34	17740.9275	36.4235	1.03e-008	1.0000
TIM2+I	8742.5272	37	17742.3407	37.8368	5.06e-009	1.0000
GTR+I	8736.0908	39	17743.3752	38.8713	3.02e-009	1.0000
TIM1+I	8743.7435	37	17744.7733	40.2694	1.50e-009	1.0000
TVM	8743.8130	37	17744.9123	40.4084	1.40e-009	1.0000
TPM2uf	8750.9639	35	17745.3068	40.8029	1.15e-009	1.0000
TrN	8751.4429	35	17746.2647	41.7608	7.11e-010	1.0000
TPM1uf	8751.9851	35	17747.3491	42.8452	4.14e-010	1.0000
GTR	8743.1604	38	17750.5609	46.0570	8.30e-011	1.0000
TIM2	8750.1213	36	17750.5753	46.0713	8.24e-011	1.0000
TIM1	8751.1919	36	17752.7164	48.2125	2.83e-011	1.0000
F81+G	8851.0233	34	17938.4720	233.9680	1.30e-051	1.0000
F81+I+G	8851.0262	35	17945.4314	240.9274	4.01e-053	1.0000
F81+I	8859.5545	34	17955.5343	251.0304	2.57e-055	1.0000
F81	8865.8258	33	17961.1232	256.6193	1.57e-056	1.0000
TPM2+G	8880.7938	33	17991.0592	286.5553	4.96e-063	1.0000
TIM2ef+G	8879.7413	34	17995.9078	291.4039	4.39e-064	1.0000
TPM2+I+G	8880.7949	34	17998.0151	293.5112	1.53e-064	1.0000
TIM2ef+I+G	8879.7431	35	18002.8652	298.3613	1.36e-065	1.0000
TVMef+G	8880.1444	35	18003.6676	299.1637	9.07e-066	1.0000
SYM+G	8879.0276	36	18008.3878	303.8839	8.57e-067	1.0000
K80+G	8894.0076	32	18010.5331	306.0291	2.93e-067	1.0000
TVMef+I+G	8880.1454	36	18010.6235	306.1196	2.80e-067	1.0000
TPM1+G	8891.4084	33	18012.2884	307.7845	1.22e-067	1.0000
TPM2+I	8891.9497	33	18013.3710	308.8671	7.09e-068	1.0000
TrNef+G	8892.9137	33	18015.2990	310.7951	2.70e-068	1.0000
SYM+I+G	8879.0296	37	18015.3454	310.8415	2.64e-068	1.0000
TPM3+G	8893.7115	33	18016.8946	312.3907	1.22e-068	1.0000
TIM1ef+G	8890.2797	34	18016.9848	312.4808	1.16e-068	1.0000
K80+I+G	8894.0084	33	18017.4884	312.9845	9.05e-069	1.0000
TIM2ef+I	8891.2076	34	18018.8405	314.3366	4.60e-069	1.0000
TPM1+I+G	8891.4095	34	18019.2442	314.7403	3.76e-069	1.0000
TIM3ef+G	8892.5647	34	18021.5546	317.0507	1.18e-069	1.0000
TrNef+I+G	8892.9156	34	18022.2565	317.7526	8.34e-070	1.0000
TPM3+I+G	8893.7123	34	18023.8498	319.3459	3.76e-070	1.0000
TIM1ef+I+G	8890.2819	35	18023.9427	319.4388	3.59e-070	1.0000
TVMef+I	8891.5486	35	18026.4762	321.9723	1.01e-070	1.0000
TPM2	8902.0356	32	18026.5892	322.0853	9.56e-071	1.0000
TIM3ef+I+G	8892.5666	35	18028.5121	324.0082	3.66e-071	1.0000
K80+I	8904.2566	32	18031.0312	326.5272	1.04e-071	1.0000
TIM2ef	8900.8515	33	18031.1746	326.6706	9.66e-072	1.0000
SYM+I	8890.7689	36	18031.8705	327.3665	6.82e-072	1.0000
TPM1+I	8901.7362	33	18032.9440	328.4400	3.99e-072	1.0000
TrNef+I	8903.4726	33	18036.4169	331.9129	7.02e-073	1.0000
TPM3+I	8904.1829	33	18037.8373	333.3334	3.45e-073	1.0000
TIM1ef+I	8900.9155	34	18038.2562	333.7523	2.80e-073	1.0000
TVMef	8901.6953	34	18039.8159	335.3120	1.28e-073	1.0000
TIM3ef+I	8903.3736	34	18043.1724	338.6685	2.40e-074	1.0000
K80	8913.9621	31	18043.4885	338.9846	2.05e-074	1.0000
SYM	8900.4862	35	18044.3513	339.8473	1.33e-074	1.0000
TPM1	8911.3410	32	18045.1999	340.6960	8.69e-075	1.0000
TrNef	8912.6592	32	18047.8363	343.3324	2.33e-075	1.0000
TIM1ef	8909.9967	33	18049.4649	344.9610	1.03e-075	1.0000
TPM3	8913.9594	32	18050.4367	345.9327	6.34e-076	1.0000
TIM3ef	8912.6580	33	18054.7875	350.2835	7.20e-077	1.0000
JC+G	8978.4672	31	18172.4986	467.9946	1.98e-102	1.0000
JC+I+G	8978.4673	32	18179.4525	474.9485	6.12e-104	1.0000
JC+I	8985.7934	31	18187.1511	482.6471	1.30e-105	1.0000
JC	8993.6029	30	18195.8163	491.3124	1.71e-107	1.0000

```

-----
-lnL:      negative log likelihood
K:         number of estimated parameters
BIC:       Bayesian Information Criterion
delta:     BIC difference
weight:    BIC weight
cumWeight: cumulative BIC weight

```

Model selection results also available at the "Model > Show model table" menu

\* BIC MODEL SELECTION : Confidence interval

There are 88 models in the 100% confidence interval: [ TPM3uf+G TIM3+G HKY+G TPM3uf+I+G TrN+G TPM2uf+G TVM+G TIM3+I+G TPMLuf+G HKY+I+G TIM2+G GTR+G TIM1+G TrN+I+G TPM2uf+I+G TVM+I+G TPMLuf+I+G TIM2+I+G TPM3uf+I GTR+I+G TIM1+I+G HKY+I TIM3+I TPM3uf TPM2uf+I TVM+I TrN+I TPMLuf+I TIM3 HKY TIM2+I GTR+I TIM1+I TVM TPM2uf TrN TPM1uf GTR TIM2 TIM1 F81+G F81+I+G F81+I F81 TPM2+G TIM2ef+G TPM2+I+G TIM2ef+I+G TVMef+G SYM+G K80+G TVMef+I+G TPML+G TPM2+I TrNef+G SYM+I+G TPM3+G TIM1ef+G K80+I+G TIM2ef+I TPML+I+G TIM3ef+G TrNef+I+G TPM3+I+G TIM1ef+I+G TVMef+I TPM2 TIM3ef+I+G K80+I TIM2ef SYM+I TPM1+I TrNef+I TPM3+I TIM1ef+I TVMef TIM3ef+I K80 SYM TPM1 TrNef TIM1ef TPM3 TIM3ef JC+G JC+I+G JC+I JC ]

\* BIC MODEL SELECTION : Parameter importance

Parameter    Importance

```

-----
fA           1.0000
fC           1.0000
fG           1.0000
fT           1.0000
kappa       0.0484
titv        0.0484
rAC         0.9430
rAG         0.9516
rAT         0.0114
rCG         0.9399
rCT         0.9516
rGT         0.9516
pinv(I)     0.0000
alpha(G)    0.9700
pinv(IG)    0.0300
alpha(IG)   0.0300
-----

```

Values have been rounded.

- (I): considers only +I models.
- (G): considers only +G models.
- (IG): considers only +I+G models.

\* BIC MODEL SELECTION : Model averaged estimates

```

Model-averaged
Parameter      estimates
-----

```



fA	0,2682
fC	0,1764
fG	0,1823
fT	0,3731
kappa	2,6595
titv	1,2422
rAC	1,4823
rAG	3,0980
rAT	1,0946
rCG	1,4810
rCT	3,0560
rGT	1,0000
pinv(I)	0,0211
alpha(G)	4,1112
pinv(IG)	0,0009
alpha(IG)	4,1528
-----	

Numbers have been rounded.

(I): considers only +I models.

(G): considers only +G models.

(IG): considers only +I+G models.

## ANEXO 4: MODELOS GERADOS POR MODELAGEM

### MOLECULAR

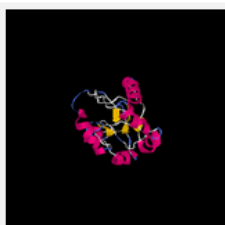
#### E6

Top 5 Models predicted by I-TASSER



[Download Model 1](#)

C-score=-2.79



[Download Model 2](#)

C-score=-2.97



[Download Model 3](#)

C-score=-2.90



[Download Model 4](#)

C-score=-3.05



[Download Model 5](#)

C-score=-3.41

Estimated accuracy of Modell:  $0.39 \pm 0.13$  (TM-score)  $11.2 \pm 4.6\text{\AA}$  (RMSD) ([Read more about C-score of generated models](#))

#### E7

Top 5 Models predicted by I-TASSER



[Download Model 1](#)

C-score=-2.15



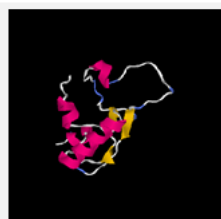
[Download Model 2](#)

C-score=-2.63



[Download Model 3](#)

C-score=-2.69



[Download Model 4](#)

C-score=-3.27



[Download Model 5](#)

C-score=-3.36

Estimated accuracy of Modell:  $0.46 \pm 0.15$  (TM-score)  $8.5 \pm 4.5\text{\AA}$  (RMSD) ([Read more about C-score of generated models](#))

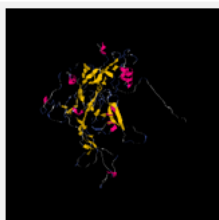
#### L1

Top 5 Models predicted by I-TASSER



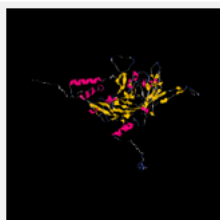
[Download Model 1](#)

C-score= 0.38



[Download Model 2](#)

C-score=-1.05



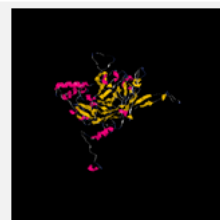
[Download Model 3](#)

C-score=-0.42



[Download Model 4](#)

C-score=-3.31



[Download Model 5](#)

C-score=-3.31

Estimated accuracy of Modell:  $0.76 \pm 0.10$  (TM-score)  $6.5 \pm 3.9\text{\AA}$  (RMSD) ([Read more about C-score of generated models](#))