# PADRÕES DE USO DE CÓDONS DO VÍRUS OROPOUCHE E SUA COMPATIBILIDADE COM HOSPEDEIROS VERTEBRADOS E INVERTEBRADOS

**Autores: Mateus Carmo de Oliveira, Agata Rossi, Edson Delatorre**

Laboratório de Genômica e Ecologia Viral, Universidade Federal do Espírito Santo – Campus de Maruípe, Av. Marechal Campos, 1468 – Prédio Básico – Maruípe, CEP 29.040-090 - Vitória - ES, Brasil,
mateus.c.oliveira@edu.ufes.br, edson.delatorre@ufes.br.

**Resumo**

O vírus Oropouche (OROV) é um arbovírus emergente cujo ciclo envolve diferentes hospedeiros vertebrados e artrópodes. Neste estudo, investigamos padrões de uso de códons do OROV e sua relação com diferentes hospedeiros, considerando os segmentos genômicos L, M e S e suas principais linhagens. Foram analisadas sequências genômicas submetidas a curadoria e classificadas filogeneticamente, seguidas de análises de composição nucleotídica, uso relativo de códons sinônimos (RSCU), número efetivo de códons (ENC), análise de correspondência e índice de adaptação de códons (CAI). Após a curadoria, foram obtidas 189 sequências do segmento L, 815 do M e 871 do S, distribuídas nas linhagens L1/L2, M1/M2 e S1/S2/S3. Os segmentos L e M apresentaram maior conteúdo de A/U e valores de ENC próximos de 47, enquanto o segmento S apresentou maior conteúdo relativo de G/C e ENC médio de 60,30. A análise de correspondência separou claramente as linhagens segundo o segmento genômico. Os valores de CAI variaram entre os hospedeiros avaliados, indicando diferentes níveis de compatibilidade entre o uso de códons viral e os repertórios de códons dos hospedeiros. Os resultados demonstram que o uso de códons constitui uma característica diferencial entre os segmentos do OROV e fornece sinais compatíveis com diferentes contextos de hospedeiro, embora a interpretação de adaptação translacional exija cautela, especialmente diante do reduzido número de sequências provenientes de invertebrados.

**Área do Conhecimento:** Microbiologia.

## Introdução

O vírus Oropouche (OROV), espécie *Orthobunyavirus oropoucheense*, é um arbovírus de RNA de fita simples e sentido negativo, com genoma trissegmentado constituído pelos segmentos L, M e S. O vírus apresenta histórico de circulação na América do Sul e Central e, nos últimos anos, adquiriu maior importância no Brasil devido à expansão para regiões anteriormente consideradas não endêmicas (Da Rosa et al., 2017; Gutierrez et al., 2020; Bai et al., 2025). Em 2024, a detecção de intensa circulação no Espírito Santo reforçou a importância de compreender não apenas a distribuição epidemiológica do vírus, mas também os processos moleculares associados à sua evolução e manutenção em diferentes contextos ecológicos (Delatorre et al., 2025; Gräf et al., 2025).

O genoma segmentado do OROV é organizado em três segmentos com funções distintas. O segmento L codifica a RNA-polimerase dependente de RNA; o M codifica a poliproteína que origina proteínas estruturais e não estruturais; e o S codifica a nucleoproteína e proteínas não estruturais. Essa organização, associada à circulação do vírus em hospedeiros vertebrados e artrópodes, fornece um sistema particularmente relevante para investigar como diferentes pressões evolutivas podem atuar sobre o genoma viral (Da Rosa et al., 2017; Naveca et al., 2024).

O uso de códons não é uniforme entre os genomas. A frequência de códons sinônimos pode ser influenciada por composição nucleotídica, pressões mutacionais, seleção natural e características do processo de tradução. Em vírus, diferenças no repertório de códons podem produzir assinaturas relacionadas ao hospedeiro, uma vez que a disponibilidade relativa de tRNAs e o padrão de uso de códons da célula hospedeira podem influenciar a eficiência translacional. Entretanto, índices como RSCU, ENC e CAI são medidas complementares e não constituem, isoladamente, prova de adaptação biológica a determinado hospedeiro (Quax et al., 2015; Subramanian et al., 2022).

Neste contexto, o presente estudo teve como objetivo caracterizar o viés de uso de códons de diferentes linhagens do OROV e avaliar se esses padrões apresentam diferenças entre segmentos genômicos e níveis distintos de compatibilidade com hospedeiros vertebrados e invertebrados.

## Metodologia

Foram utilizadas sequências genômicas de OROV provenientes de bases públicas, priorizando-se genomas com cobertura mínima de 95%. As sequências foram submetidas à curadoria para identificação das regiões codificantes dos segmentos L, M e S, utilizando sequências de referência do NCBI RefSeq. Regiões não codificantes foram removidas e sequências com deleções superiores a 30 nucleotídeos (10 aminoácidos) nas extremidades foram excluídas. Após a curadoria, foram retidas 189 sequências do segmento L, 815 do segmento M e 871 do segmento S.

As sequências foram alinhadas por MAFFT (Katoh et al., 2002) e classificadas segundo linhagens previamente estabelecidas para OROV (Naveca et al., 2024). Foram consideradas as linhagens L1 e L2 no segmento L, M1 e M2 no segmento M e S1, S2 e S3 no segmento S. As análises filogenéticas foram conduzidas por máxima verossimilhança no IQ-TREE, com seleção do modelo evolutivo pelo ModelFinder (Nguyen et al., 2015; Kalyaanamoorthy et al., 2017), e as árvores foram utilizadas para organização das sequências nas linhagens analisadas.

O viés de uso de códons foi caracterizado separadamente para cada linhagem e segmento. Foram calculados o uso relativo de códons sinônimos (RSCU) e o número efetivo de códons (ENC). A composição de nucleotídeos foi avaliada considerando A, C, G, U, GC e as posições de terceira base (GC3). A análise de correspondência foi empregada para representar a variação multivariada do uso de códons entre as linhagens.

O índice de adaptação de códons (CAI) foi calculado utilizando conjuntos de referência de uso de códons para os hospedeiros avaliados. Quando disponíveis, foram utilizados dados do Codon Statistics Database; para outros organismos foram utilizados conjuntos de referência disponíveis em bases específicas, conforme a disponibilidade de genes e estatísticas de uso de códons no projeto. O Codon Statistics Database fornece frequências de códons, RSCU, ENC, CAI e frequência de códons ótimos para espécies com genomas de referência ou representativos em RefSeq (Subramanian et al., 2022). Os hospedeiros foram agrupados em vertebrados e invertebrados para a comparação exploratória dos padrões de CAI. Devido ao reduzido número de sequências virais associadas a alguns hospedeiros invertebrados, essas comparações foram interpretadas de forma descritiva.

## Resultados

Após a curadoria, as sequências foram distribuídas em sete linhagens: L1 (n=47) e L2 (n=142), M1 (n=781) e M2 (n=33), e S1 (n=45), S2 (n=815) e S3 (n=10). A composição nucleotídica (Figura 1) revelou uma diferença consistente entre os segmentos L/M e S. As linhagens L1/L2 e M1/M2 apresentaram aproximadamente 36–38% de A, 16% de C, 28–29% de U e 18–19% de G, resultando em conteúdo AU de aproximadamente 65%. Em contraste, as linhagens S apresentaram menor proporção de A e maior contribuição relativa de G e T/U, com conteúdo AU próximo de 52% e conteúdo GC próximo de 48%.

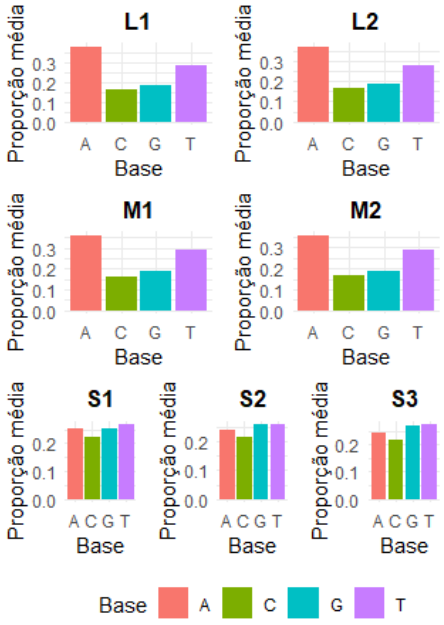


**Figura 1 – Composição nucleotídica média das linhagens do OROV nos segmentos L, M e S. As linhagens dos segmentos L e M apresentam maior contribuição de A/U, enquanto o segmento S apresenta composição mais equilibrada entre A/U e G/C.**

A análise de RSCU (Figura 2) demonstrou um viés de uso de códons mais evidente nos segmentos L e M. Nesses segmentos, vários códons terminados em adenina ou uracila apresentaram RSCU elevado, incluindo GCA, AGA, ACA, CCA, AUA e UUA, enquanto códons terminados em C ou G, como GCG, ACG, CCC, CGC e AUC, apresentaram valores geralmente inferiores a 1. No segmento S, a distribuição dos valores de RSCU foi mais homogênea e houve maior contribuição relativa de códons terminados em C/G, compatível com sua composição nucleotídica distinta.

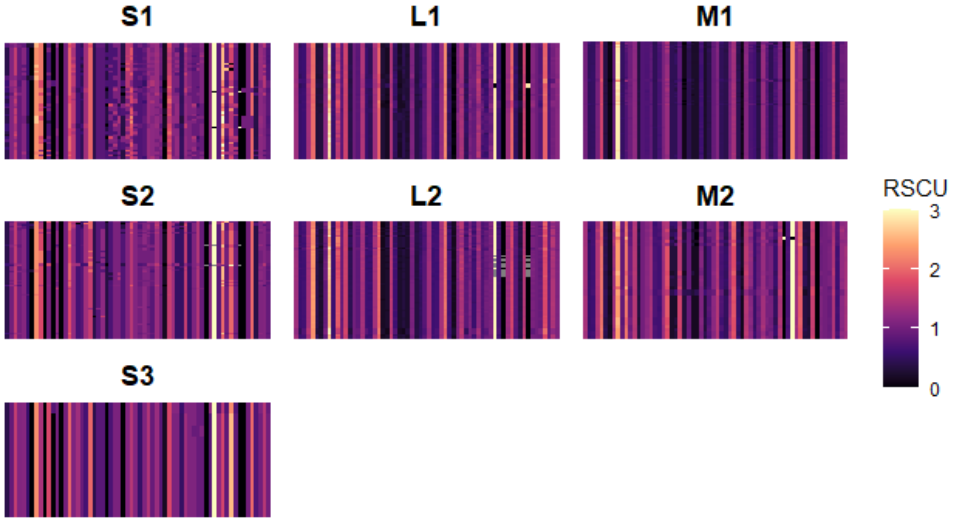



**Figura 2 – Heatmap do uso relativo de códons sinônimos (RSCU) nas linhagens dos segmentos L, M e S do OROV. Valores mais elevados indicam maior utilização relativa do códon dentro do conjunto de códons sinônimos para o mesmo aminoácido.**

O ENC reforçou a diferença entre os segmentos. Os valores médios foram de 47,43 ± 0,70 para L, 47,47 ± 0,42 para M e 60,30 ± 1,21 para S. Entre as linhagens individuais, o ENC variou de 46,32 ± 0,29 em L1 a 60,45 ± 1,00 em S2. Assim, os segmentos L e M apresentaram menor número efetivo de códons e, portanto, maior restrição relativa no repertório de códons, enquanto o segmento S apresentou valores mais elevados, compatíveis com uso mais diversificado de códons, como mostra a Tabela 1.

**Tabela 1 – Número efetivo de códons (ENC) por linhagem do OROV.**

| Segmento/linhagem | n | ENC (média ± DP) |
|---|---|---|
| L1 | 47 | 46,32 ± 0,29 |
| L2 | 142 | 47,80 ± 0,28 |
| M1 | 781 | 47,40 ± 0,21 |
| M2 | 33 | 49,19 ± 0,61 |
| S1 | 45 | 57,42 ± 1,66 |
| S2 | 815 | 60,45 ± 1,00 |
| S3 | 10 | 57,99 ± 0,57 |

Quando agrupadas por hospedeiro, as sequências provenientes de vertebrados apresentaram ENC médio de 47,46 ± 0,68 no segmento L (n=184), 47,47 ± 0,42 no M (n=809) e 60,30 ± 1,21 no S (n=862). Para invertebrados, os valores foram 46,36 ± 0,31 no L (n=3), 48,33 ± 0,55 no M (n=3) e 55,90 ± 1,80 no S (n=6). O padrão geral de maior ENC no segmento S foi mantido nos dois grupos, mas o pequeno número de sequências de invertebrados limita comparações estatísticas entre hospedeiros.

A análise de correspondência (Figura 3) evidenciou uma separação clara entre as linhagens segundo o segmento genômico. L1 e L2 formaram um agrupamento distinto na região superior direita do gráfico, M1 e M2 ocuparam uma região intermediária inferior, enquanto S1, S2 e S3 se concentraram na região esquerda. Esse padrão indica que a principal estrutura de variação do uso de códons está associada à identidade do segmento, mais do que a diferenças sutis entre as linhagens pertencentes ao mesmo segmento.

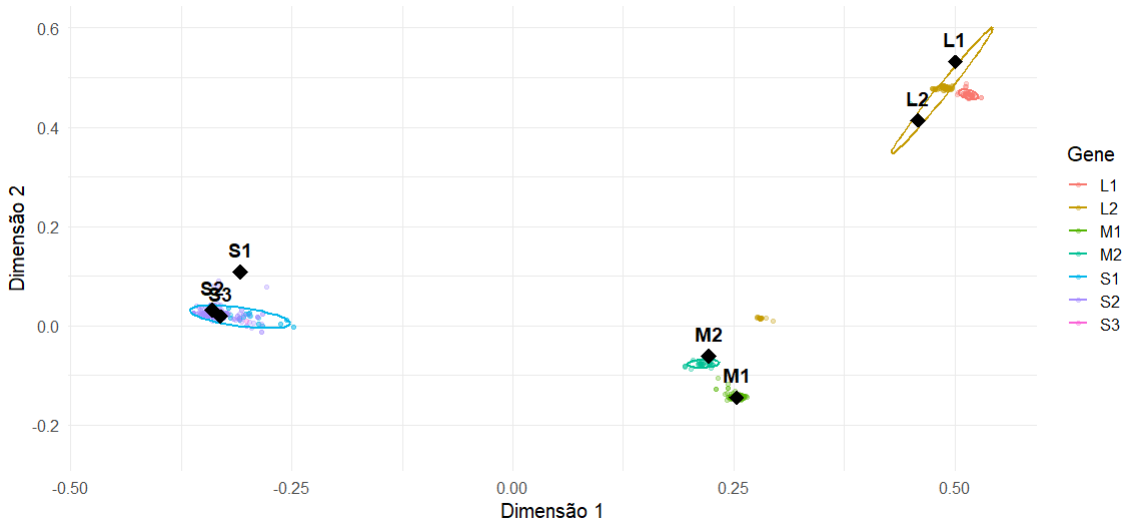



**Figura 3 – Análise de correspondência baseada no uso relativo de códons sinônimos. As linhagens são representadas pelos respectivos centroides e pela distribuição das sequências individuais. A separação entre os segmentos L, M e S indica forte contribuição da organização segmentar para a variação do uso de códons.**

A Figura 4 mostra graficamente o CAI calculado para hospedeiros vertebrados e vetores invertebrados:

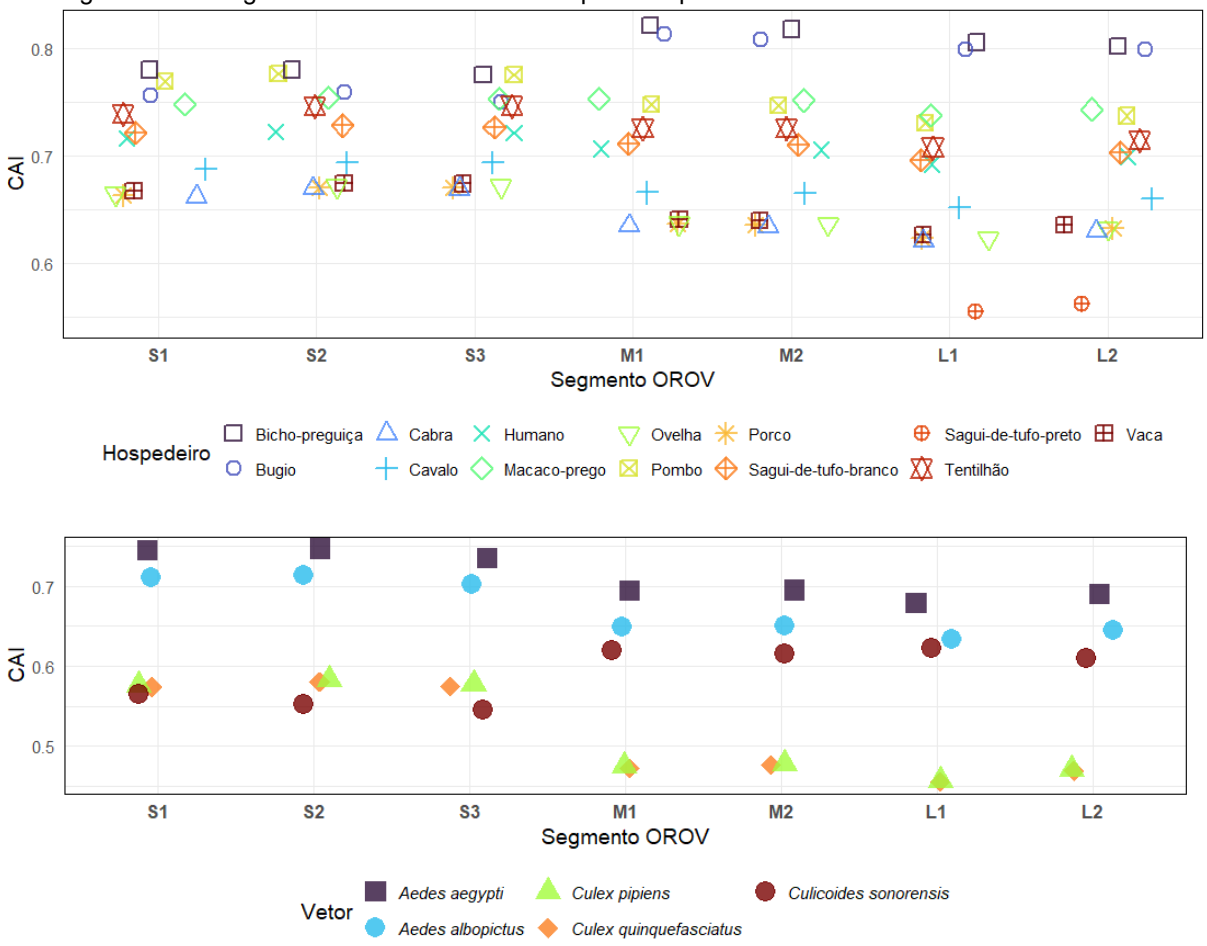



**Figura 4 – Índice de adaptação de códons (CAI) em hospedeiros vertebrados (A) e vetores invertebrados (B), segundo as linhagens dos segmentos L, M e S. Os valores diferem entre espécies hospedeiras e entre segmentos, indicando diferentes níveis de compatibilidade entre o repertório de códons viral e os conjuntos de referência utilizados.**

Os valores de CAI apresentaram variação entre as espécies hospedeiras. Entre os vertebrados, os valores permaneceram elevados para diversas espécies, principalmente Bicho-preguiça e Bugio, embora a magnitude variasse conforme o segmento e a linhagem. Entre os invertebrados também foi observado um gradiente entre as espécies avaliadas, com diferenças entre *Aedes*, *Culex* e *Culicoides*. Em geral, o segmento S apresentou valores de CAI superiores aos observados para L e M em diversos hospedeiros, embora esse padrão não tenha sido uniforme, especialmente entre os *Culicoides*.

## Discussão

Os resultados demonstram que o uso de códons do OROV é estruturado principalmente pela organização segmentar do genoma, e não pela linhagem circulante. O perfil dos segmentos L/M, rico em A/U, com preferência por códons terminados nessas bases e forte supressão de códons C/G, é semelhante ao descrito para as sequências codificadoras da glicoproteína e da RdRp por Yang et al. (2025), enquanto o perfil mais equilibrado do segmento S coincide com o observado para a nucleoproteína. A concordância entre os estudos, apesar das diferenças nas amostras e abordagens, reforça a possibilidade de que a função das proteínas codificadas em cada segmento contribua para definir seu repertório de códons, em concordância com Yang et al. (2025), que identificaram padrões relacionados à função proteica mais do que ao genótipo. Além disso, taxas evolutivas distintas já foram estimadas para os segmentos do OROV (S: $1{,}75 \times 10^{-3}$; M: $3{,}65 \times 10^{-4}$; L: $1{,}64 \times 10^{-3}$ substituições/sítio/ano), compatíveis com diferentes regimes de pressão seletiva.

A magnitude dos valores de RSCU, com códons favorecidos acima de 3 e baixa utilização de alguns códons sinônimos, como ACG, GCG e AGC, indica uma restrição importante no repertório de códons. Parte desse padrão pode estar relacionada à supressão de códons contendo CpG, característica descrita em vírus de vertebrados e associada à evasão de sensores imunes como a ZAP (Yang et al., 2025). Entretanto, a análise ENC-GC3, não realizada neste estudo, seria necessária para quantificar a contribuição relativa de mutação e seleção. Padrões distintos dentro da própria família reforçam essa interpretação: no Huaiyangshan virus, um bunyavírus transmitido por carrapatos, os códons preferidos terminam em G/C e o viés mutacional parece exercer papel predominante (Luo et al., 2015). Assim, o padrão A/U observado nos segmentos L/M do OROV pode estar relacionado às particularidades de seu ciclo envolvendo ceratopogonídeos e hospedeiros vertebrados, e não representar uma característica geral dos bunyavírus.

Os valores de ENC do OROV (~47 em L/M e ~60 em S; média ponderada de 53,4) situam o vírus próximo à média descrita para arbovírus (~52), caracterizados por uso relativamente diversificado de códons e predominância de seleção purificadora (Sexton et al., 2019). Esse contexto ajuda a interpretar os valores de CAI, uma vez que, em arbovírus, o uso de códons raramente reproduz o padrão de um hospedeiro específico. Experimentos de recodificação sinônima em DENV, ZIKV e CHIKV também demonstraram atenuação viral, indicando que escolhas sinônimas estão sujeitas à seleção, embora não sejam necessariamente otimizadas para um único hospedeiro (Sexton et al., 2019). Dessa forma, os valores de CAI segmento- e hospedeiro-específicos obtidos neste estudo devem ser interpretados como medidas de compatibilidade relativa, e não como evidência de adaptação ótima. Essa interpretação é coerente com os índices de desotimização translacional do OROV em humanos descritos por Yang et al. (2025).

Essa ressalva é particularmente importante para os invertebrados, representados por apenas três sequências dos segmentos L e M e seis do segmento S. A disponibilidade de sequências em bases públicas, concentrada principalmente em surtos humanos, também pode subrepresentar a diversidade viral em artrópodes. Além disso, a detecção do vírus em um hospedeiro não implica necessariamente papel equivalente na manutenção do ciclo viral. Apesar dessas limitações, a convergência entre composição nucleotídica, RSCU, ENC e análise de correspondência fornece uma caracterização consistente da arquitetura do uso de códons do OROV. Os resultados de CAI levantam a possibilidade de diferentes níveis de compatibilidade translacional entre segmentos e hospedeiros, hipótese que poderá ser testada com maior número de sequências de artrópodes, dados experimentais de competência vetorial e análises ENC-GC3 que controlem filogenia e composição nucleotídica. Do ponto de vista aplicado, essas assinaturas também podem contribuir para estudos futuros de recodificação viral voltados ao desenvolvimento de vírus atenuados, estratégia já explorada em outros arbovírus (Sexton et al., 2019).

## Conclusão

O OROV apresenta padrões distintos de uso de códons entre seus segmentos genômicos. Os segmentos L e M foram caracterizados por maior conteúdo de A/U, maior restrição relativa no repertório de códons e preferência por códons terminados nessas bases, enquanto o segmento S apresentou maior conteúdo relativo de G/C, valores de ENC mais elevados e distribuição mais diversificada de códons. A análise de correspondência confirmou a forte separação dos padrões segundo a organização segmentar. Os valores de CAI variaram entre diferentes hospedeiros vertebrados e invertebrados, sugerindo diferentes níveis de

compatibilidade entre o uso de códons viral e os repertórios de referência dos hospedeiros. Esses resultados reforçam o potencial das análises de uso de códons para investigar a evolução molecular do OROV, mas não permitem, isoladamente, afirmar adaptação viral a espécies específicas. Estudos com maior representação de hospedeiros invertebrados e abordagens integrativas serão necessários para testar essa hipótese.

## Referências

BAI F, et al. A comprehensive review of the neglected and emerging Oropouche virus. Viruses. 2025;17(3):439. DOI: 10.3390/v17030439.

DA ROSA JFT, et al. Oropouche Virus: clinical, epidemiological, and molecular aspects of a neglected orthobunyavirus. Am J Trop Med Hyg. 2017;96(5):1019-1030. DOI: 10.4269/ajtmh.16-0672.

DELATORRE E, et al. Emergence of Oropouche Virus in Espírito Santo State, Brazil, 2024. Emerg Infect Dis. 2025;31(6):1178-1188. DOI: 10.3201/eid3106.241946.

GRÄF T, et al. Expansion of Oropouche virus in non-endemic Brazilian regions: analysis of genomic characterisation and ecological drivers. Lancet Infect Dis. 2025;25(4):379-389. DOI: 10.1016/S1473-3099(24)00687-X.

GUTIERREZ B, et al. Evolutionary dynamics of Oropouche virus in South America. J Virol. 2020;94(5):e01127-19. DOI: 10.1128/JVI.01127-19.

KATOH K, et al. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30(14):3059-3066. DOI: 10.1093/nar/gkf436.

KALYAANAMOORTHY S, et al. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587-589. DOI: 10.1038/nmeth.4285.

LUO X, LIU Q, XIONG Y, YE C, JIN D, XU J. Genome-wide analysis of synonymous codon usage in Huaiyangshan virus and other bunyaviruses. J Basic Microbiol. 2015;55(12):1374-1383. DOI: 10.1002/jobm.201500233.

NAVECA FG, et al. Human outbreaks of a novel reassortant Oropouche virus in the Brazilian Amazon region. Nat Med. 2024;30(12):3509-3521. DOI: 10.1038/s41591-024-03300-3.

NGUYEN LT, et al. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268-274. DOI: 10.1093/molbev/msu300.

PECHMANN S, FRYDMAN J. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. Nat Struct Mol Biol. 2012;20(2):237-243. DOI: 10.1038/nsmb.2466.

QUAX TEF, et al. Codon bias as a means to fine-tune gene expression. Mol Cell. 2015;59(2):149-161. DOI: 10.1016/j.molcel.2015.05.035.

SEXTON NR, et al. Effects of arbovirus multi-host life cycles on dinucleotide and codon usage patterns. Viruses. 2019;11(7):641. DOI: 10.3390/v11070641.

SUBRAMANIAN K, PAYNE B, FEYERTAG F, ALVAREZ-PONCE D. The Codon Statistics Database: a database of codon usage bias. Mol Biol Evol. 2022;39(8):msac157. DOI: 10.1093/molbev/msac157.

YANG X, et al. Mapping evolutionary paradigm of Oropouche virus driven by dinucleotide bias and context-dependent codon bias. Virulence. 2025. DOI: 10.1080/21505594.2025.2600128.

## Agradecimentos

O presente trabalho foi realizado com apoio da Fundação de Amparo à Pesquisa e Inovação do Espírito Santo (FAPES), por meio do Edital FAPES nº 13/2025 – UNIVERSAL, processo FAPES nº 1472/2025 (SIAFEM nº 2025-G876J), e do Edital FAPES/CNPq/Decit-SECTICS-MS/SESA nº 17/2024 – PPSUS, processo FAPES nº 1342/2025 (SIAFEM nº 2025-3PBXS).