

Desenvolvimento de uma ferramenta de bioinformática para o monitoramento da adaptabilidade urbana dos *Alphavirus*

Palavras-Chave: Alphavirus, Padrões genômicos, vírus Chikungunya

Autores:

Davi Prado Menabó, Laboratório de Bioinformática e Virologia, Departamento de Genética, Evolução, Microbiologia e Imunologia, Instituto de Biologia, Universidade de Campinas

Prof. Dr. Marielton dos Passos Cunha (orientador), Laboratório de Bioinformática e Virologia, Departamento de Genética, Evolução, Microbiologia e Imunologia, Instituto de Biologia, Universidade de Campinas

INTRODUÇÃO:

A família *Togaviridae* é composta por pequenos vírus envelopados com tamanhos de genomas entre 10 kilobases e 12 kilobases, organizados em uma molécula de RNA em com polaridade positiva (*i.e.*, com orientação 5'-3'). Dentro desse grupo, encontra-se um gênero importante quanto à sua epidemiologia, com destaque aos países localizados na América Latina, os *Alphavirus*, que compreendem partículas virais com aproximadamente 70 nanômetros de diâmetro constituído de capsídeo e envelope, cuja função é proteger o genoma que está organizado em duas regiões abertas de leitura (do inglês, *open reading frame* - ORF).

A primeira ORF trata-se de uma poliproteína que será clivada em proteínas não estruturais denominadas nsP1, nsP2, nsP3 e nsP4 com base na sua posição em relação à extremidade 5' proximal da molécula, em geral a função comumente atribuída a esse conjunto proteico é a formação do complexo de replicação viral, com destaque para nsP4 que codifica a RNA polimerase RNA dependente, enzima vital para a replicação propriamente dita do RNA viral. A segunda ORF é responsável pela tradução de uma poliproteína posteriormente clivada nas proteínas estruturais C, E3, E2, 6K, E1 e TF, a primeira delas é a proteína de capsídeo C responsável pela clivagem das outras e posterior formação do capsídeo, já o complexo E (E1, E2 e E3) se responsabiliza pela formação do envelope viral e conseqüentemente pelos mecanismos de reconhecimento celular entre vírus e hospedeiro, enquanto 6K e TF atuam no brotamento do *vírion* após a replicação viral.

Os *Alphavirus* possuem uma característica biológica peculiar, que é a presença de um *opal stop codon* (UGA) entre nsP3 e nsP4, que por razões ainda não completamente elucidadas, não cessa a produção da nsP4 para a poliproteína codificada pela ORF1. Essa característica única é responsável por erros na anotação e depósito de sequências disponibilizado em banco de dados públicos, pois ao realizar o sequenciamento, o *opal stop codon* pode ser interpretado como um final prematuro para a tradução da ORF1, ou então, adiciona-se uma terceira ORF entre as duas originais, o que não corresponde ao padrão desse grupo que é sempre composto por duas ORFs.

Neste sentido, a criação de uma ferramenta única para análise de genomas virais dedicadas aos *Alphavirus* pode solucionar problemas de anotação e organização da informação genômica, sendo possível manipulá-las para obtenção de informações estatísticas quanto aos padrões genômicos desse gênero. Frente aos vírus que desempenham um importante papel epidemiológico entre os *Alphavirus*, uma importante informação a ser coletada são os padrões genômicos relevantes para competência vetorial, pois dessa forma é possível integrar dados de diferentes espécies virais e prever a probabilidade de um arbovírus adaptar-se ou não a um determinado vetor e/ou hospedeiro vertebrado.

Atualmente, o principal vetor de relevância urbana no Brasil é o *Aedes aegypti*, que domina quase completamente o nicho. Portanto a adaptabilidade eficiente a esse vetor, somada a seu

potencial de replicação em um hospedeiro urbano competente, resulta em eventos biológicos, onde um vírus exclusivamente silvestre passa a desenvolver ciclos urbanos, a exemplo do vírus Chikungunya (CHIKV), consolidado globalmente em ciclos urbanos e o mais amplamente documentado *Alphavirus* nesse ambiente.

METODOLOGIA:

Para o desenvolvimento da ferramenta, foram utilizadas sequências depositadas no Centro Nacional de Informações Biotecnológicas (NCBI) como arquivos no formato Genbank. Ao todo, foram recolhidas 1528 sequências completas para 14 espécies de *Alphavirus*, sendo elas: *Alphavirus mayaro*, *Alphavirus chikungunya*, *Alphavirus barmah*, *Alphavirus getah*, *Alphavirus rossriver*, *Alphavirus semliki*, *Alphavirus middelburg*, *Alphavirus mucambo*, *Alphavirus tonate*, *Alphavirus ndumu*, *Alphavirus una*, *Alphavirus onyong*, *Alphavirus bebaru* e *Alphavirus everglades*. O processamento das sequências e dos dados associados ao depósito das sequências foi realizado utilizando a linguagem de programação *Python3*, onde as sequências foram submetidas a filtros que descartaram genomas incompletos ou que possuíam nenhuma informação sobre as características da coleta (como local, data, cepa viral ou hospedeiro) restando ao final do processamento, 1265 sequências. Os dados foram utilizados para desenvolvimento de uma ferramenta que analisa a funcionalidade do *stop codon* indicado pelo arquivo com base em sua posição, caso se trate de uma posição incomum essa indicação é descartada e substituída pela ORF1 verdadeira caso essa seja identificada pelo programa, e no caso de falta de dados no arquivo sobre o real CDS1 o software procura na sequência o *stop codon* mais próximo baseado no padrão da espécie e escreve no arquivo o real *stop codon* referente a ORF1.

Para testar a funcionalidade da ferramenta, árvore filogenéticas oriundas desses dados foram reconstruídas utilizando o programa *Iq-tree* na versão 3.0.1, e a edição para visualização da árvore foi realizada utilizando o programa *FigTree* na versão 1.4, posteriormente tendo seus resultados comparados ao de árvores semelhantes produzidas sem o uso do software para averiguação do funcionamento do programa.

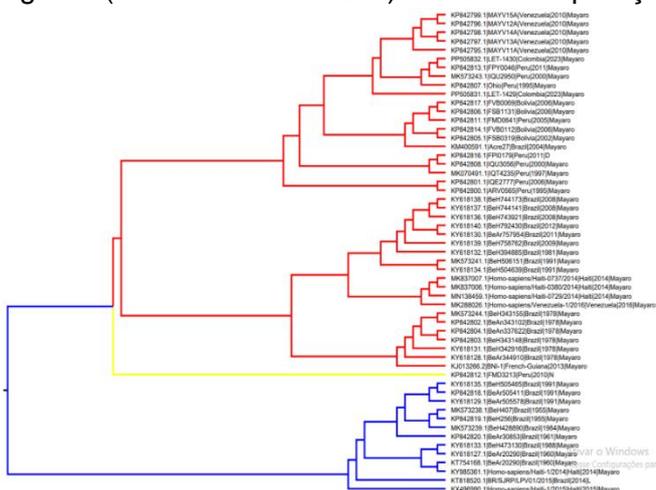
Os dados coletados pela ferramenta são escritos no formato FASTA contendo a sequência codificante encabeçada por suas informações de coleta, e esses *datasets* passaram pelo cálculo dos valores de uso relativo de códons sinônimos (RSCU), tendo seus resultados utilizados para comparação entre os vírus com intuito de observar um gradiente de semelhança genética entre os *Alphavirus* em diferentes níveis de urbanização, e para tal, foram realizado a análise de componentes principais (PCA) utilizando *Python3*, e a reconstrução de árvores filogenéticas.

RESULTADOS E DISCUSSÃO:

A árvore filogenética reconstruída com auxílio da ferramenta aqui proposta (Figura 1) obteve resultados semelhantes à referência bibliográfica reconstruída sem a utilização da ferramenta, mas com a adição de um número maior de sequências, na árvore gerada com a ferramenta, em razão das datas de obtenção das sequências e do espectro mais amplo abordado em comparação ao estudo original (Forato *et al.* 2024). Essa comparação direta de resultados aponta a eficiência na

automatização do processo para obtenção dos arquivos, pois os dados extraídos com a ferramenta formaram resultados semelhantes quanto à categorização genotípica das sequências de MAYV disponibilizadas sem a necessidade da correção manual dos erros de sequenciamento.

Figura 1. Árvore filogenética reconstruída com uso da ferramenta aqui proposta, onde os genótipos D, N e L são representados pelas cores vermelho amarelo e azul respectivamente.



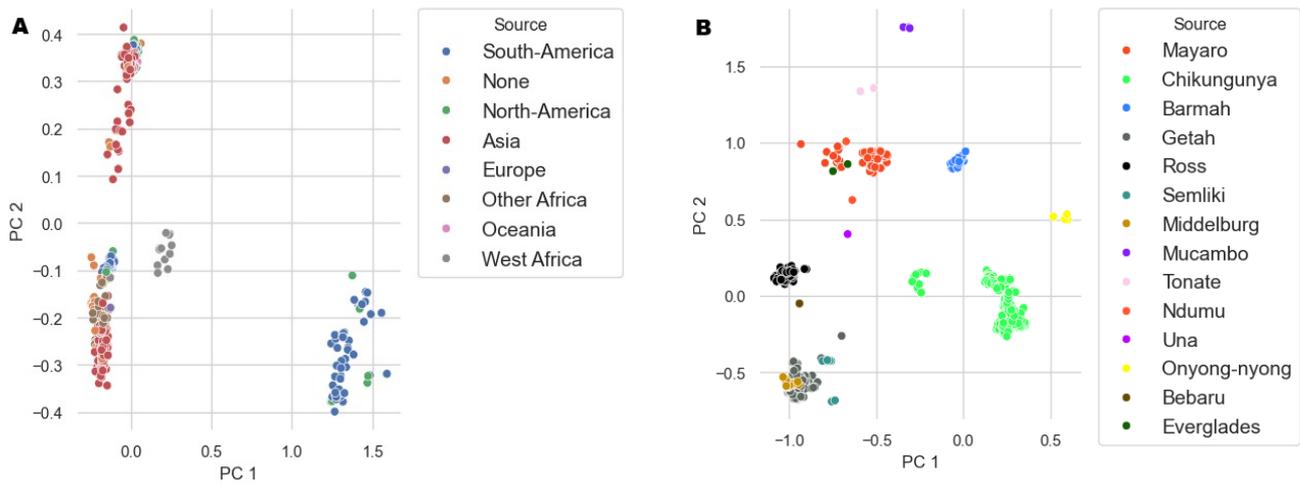


Figura 2. (A) PCA realizado entre MAYV e CHIKV com legenda baseada no local de coleta das seqüências. (B) PCA baseada em RSCU realizada entre os 14 espécimes de Alphavirus analisados. Em A, as cores representam as espécies virais, e em B, as cores representam os locais de isolamento das seqüências.

Os dados obtidos na análise de componentes principais (Figura 2B) indicam uma maior proximidade das espécies virais de um dos clusters do vírus Chikungunya, localizado entre -0.5 e 0.0 em PC1, subgrupo esse que possui ampla dispersão na África Ocidental (Figura 2A).

O CHIKV possui três genótipos distintos, dos quais dois deles - principalmente ECSA, mas também em menor escala o genótipo asiático - possuem clara adaptação ao meio urbano em razão de seu principal vetor, o mosquito *Aedes aegypti*, porém um de seus genótipos contempla até o seguinte momento somente o ciclo silvestre, sem dados que comprovem sua adaptação ao meio urbano (o genótipo exclusivo da África Ocidental que está mais próximo à maioria das outras espécies de *Alphavirus* em comparação com os outros genótipos de CHIKV).

Os dados observados no estudo corroboram o cenário atual, no qual alguns dos genótipos de CHIKV até o momento foram os únicos *Alphavirus* com ciclos urbanos amplamente documentados, o que não significa a estabilidade no panorama geral epidemiológico desse grupo, pois é possível observar certa proximidade preocupante do vírus silvestre *Onyong-nyong* com genótipos tipicamente urbanos de CHIKV.

Ao reconstruir uma árvore filogenética (Figura 3) é possível constatar novamente a similaridade de ONNV com CHIKV, embora se encontrem relativamente distantes dos demais grupos.

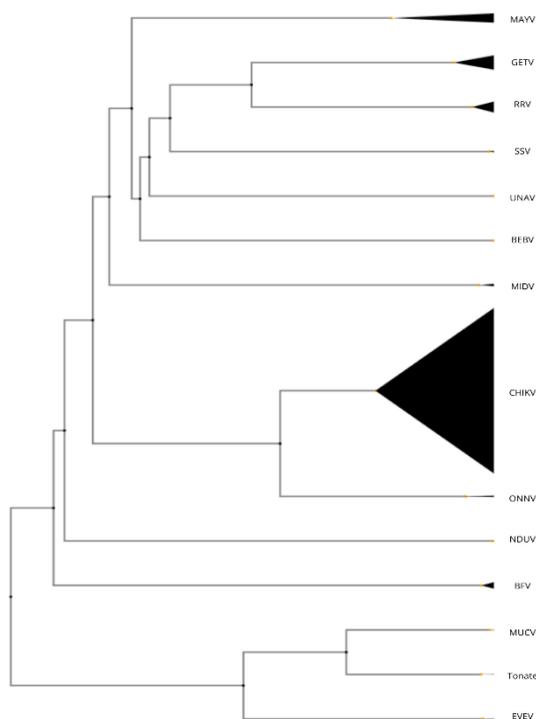


Figura 3. Árvore filogenética dos 14 espécimes de Alphavirus analisados.

CONCLUSÕES:

Os resultados obtidos até o momento apontam que a ferramenta já se encontra funcional e com ela é possível desenvolver estudos com maior facilidade sobre os padrões genômicos presentes nos

Alphavirus, dessa forma informações relevantes como por exemplo o monitoramento da adaptabilidade vetorial desse grupo podem ser extraídas.

Os padrões comparados entre os vírus foram características genômicas gerais como uso relativo de códons, o que é de extrema importância para que uma vez dentro do vetor e/ou do hospedeiro o vírus possa se replicar com eficiência, mas para o desenvolvimento de um padrão relacionado a urbanização viral características genômicas específicas também devem ser estudadas, como por exemplo a similaridade entre as ORF2s responsáveis pelas proteínas estruturais do vírion que serão reconhecidas pela célula hospedeira.

Somado a isso os valores do índice de adaptação de códons (do inglês, *Codon adaptation index* - CAI) devem ser extraídos para mapeamento do uso de códons por parte dos vetores (com foco em *Aedes aegypti*) e hospedeiros em relação ao uso de códons dos *Alphavirus*. Com essas informações construídas espera-se um aprofundamento sobre os padrões genômicos associados a adaptabilidade urbana por meio do vetor *Aedes aegypti* para *Alphavirus*, para que dessa forma o monitoramento de possíveis surtos epidemiológicos possa ser feito para essa categoria.

BIBLIOGRAFIA

01. Cunha M dos P, Ortiz-Baez AS, Freire CC de M, Zanotto PMDA. Codon adaptation biases among sylvatic and urban genotypes of Dengue virus type 2. *Infect Genet Evol* 2018; 64: 207–211.
02. Bahir I, Fromer M, Prat Y, Linial M. Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Mol Syst Biol* 2009; 5: 311.
03. Diagne CT, Bengue M, Choumet V, Hamel R, Pompon J, Missé D. Mayaro virus pathogenesis and transmission mechanisms. *Pathogens*. 2020; 9. doi:10.3390/pathogens9090738.
04. Esposito DLA, Fonseca BAL da. Will Mayaro virus be responsible for the next outbreak of an arthropod-borne virus in Brazil? *Brazilian J. Infect. Dis.* 2017; 21. doi:10.1016/j.bjid.2017.06.002.
05. Pereira TN, Carvalho FD, De Mendonça SF, Rocha MN, Moreira LA. Vector competence of *Aedes aegypti*, *Aedes albopictus*, and *Culex quinquefasciatus* mosquitoes for Mayaro virus. *PLoS Negl Trop Dis* 2020; 14. doi:10.1371/journal.pntd.0007518.
06. Li, Raymond et al. A structural and functional analysis of opal stop codon translational readthrough during Chikungunya virus replication. **Journal of General Virology**, v. 104, n. 10, p. 001909, 2023.
07. Tamanash Bhattacharya et al. A conserved opal termination codon optimizes a temperature-dependent trade-off between protein production and processing in alphaviruses. *Sci. Adv.* 11, eads7933(2025).
08. ZHOULI, Julien BZ et al. Urbanization is a main driver for the larval ecology of *Aedes* mosquitoes in arbovirus-endemic settings in south-eastern Côte d'Ivoire. **PLoS neglected tropical diseases**, v. 11, n. 7, p. e0005751, 2017.
09. GOULD, Ernest et al. Emerging arboviruses: Why today?. **One health**, v. 4, p. 1-13, 2017.
10. WEAVER, Scott C. Urbanization and geographic expansion of zoonotic arboviral diseases: mechanisms and potential strategies for prevention. **Trends in microbiology**, v. 21, n. 8, p. 360-363, 2013.

11. Forato, Julia et al. "Molecular Epidemiology of Mayaro Virus among Febrile Patients, Roraima State, Brazil, 2018-2021." *Emerging infectious diseases* vol. 30,5 (2024): 1013-1016. doi:10.3201/eid3005.231406