

Reconhecimento de padrões genômicos relacionados a especificidade vírus-hospedeiro para o vírus Influenza A (H3N2)

Palavras-Chave: Vírus Influenza A (H3N2), Padrões genômicos, Especificidade vírus-hospedeiro

Autores:

João Pedro Balzano Junqueira Guimarães, 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:

O vírus Influenza é um vírus de RNA fita simples de polaridade negativa e segmentado, pertencente à família *Orthomyxoviridae*. É classificado em quatro tipos (A, B, C e D) com base em sua diversidade genética e antigênica. Os vírus do tipo A (do inglês, *Influenza A Virus* - IAV) são classificados de acordo com a atividade de suas glicoproteínas hemaglutinina (HA) e neuraminidase (NA), que estão presentes na superfície do envelope viral.

O genoma do IAV é formado por 8 segmentos de RNA viral (vRNA), que está presente nos *vírions* (partícula viral infecciosa) na forma de complexos de ribonucleoproteínas (vRNPs). Esses segmentos de vRNA contém toda a informação necessária para a replicação viral, que ocorre no núcleo da célula hospedeira. O subtipo H3N2 representa um problema para a saúde pública há mais de cinquenta anos, sendo destaque no cenário epidemiológico brasileiro (e mundial) e se encontra, atualmente, em circulação em cadeias de transmissão sustentadas envolvendo humanos.

A adaptação do vírus ao hospedeiro é fundamental para o estabelecimento de uma infecção produtiva. Entender esse processo é fundamental para a compreensão da evolução viral, permitindo a análise temporal do processo evolutivo. Dessa forma, o objetivo deste trabalho foi caracterizar os padrões genômicos associados ao comportamento biológico do H3N2, buscando associar os padrões genômicos às características relacionadas ao processo de adaptação vírus-hospedeiro, desde a especificidade, até a patogenicidade viral associada.

METODOLOGIA:

Os dados genômicos utilizados nas análises foram obtidos do Centro Nacional de Informações Biotecnológicas (<https://www.ncbi.nlm.nih.gov/>) como um *dataset* composto pelas sequências em nucleotídeos e suas respectivas informações associadas. Para o desenvolvimento deste trabalho, foram utilizadas 28357 sequências de H3N2 (para cada um dos 8 segmentos), com a aplicação de filtros específicos a fim de remover sequências consideradas de baixa qualidade.

As sequências foram organizadas em 8 *datasets* (um para cada segmento), onde cada registro possui um cabeçalho (com informações da sequência) e a sequência nucleotídica na linha de baixo, integrados em arquivos no

formato FASTA. Para a análise desses dados, foram utilizadas ferramentas distintas por meio da linguagem de programação *Python3*, com bibliotecas como *biopython*, *pandas*, *numpy* e *matplotlib*, essenciais para o cálculo, geração de tabelas e gráficos como o de Uso Relativo de Códon Sinônimos (do inglês, *Relative Synonymous Codon Usage - RSCU*) e do Índice de Adaptação de Códon (do inglês, *Codon Adaptation Index - CAI*).

RESULTADOS E DISCUSSÃO:

Para confirmar a tipagem dos vírus, inicialmente realizadas pelos autores dos depósitos das sequências, estas foram submetidas à classificação taxonômica utilizando o classificador taxonômico *kaiju 1.10.1* (uma ferramenta que utiliza um banco de dados do NCBI e atribui um *txid*, ou *taxon id*, para cada sequência analisada, associado a um subtipo específico, nesse caso, de IAV) permite visualizar a preservação principalmente do segmento 4 (Hemaglutinina - HA), sendo a grande maioria das sequências analisadas verdadeiramente como H3 do subtipo H3N2 (apenas 4 - 0,01% - tratam-se de H1 do H1N1, possivelmente identificadas como H3N2 por algum erro na anotação das sequências). Além disso, é possível visualizar uma maior preservação dos segmentos 6 e 2 (Neuraminidase - NA e Polimerase Básica 1 - PB1, respectivamente) frente aos outros segmentos, onde todos apresentam mais de 3400 sequências (mais de 12,3% para cada um desses outros cinco segmentos) pertencentes a outros subtipos de IAV, mas não ao H3N2 (aqui, serão tratadas como falsos H3N2) (**Fig. 01**).

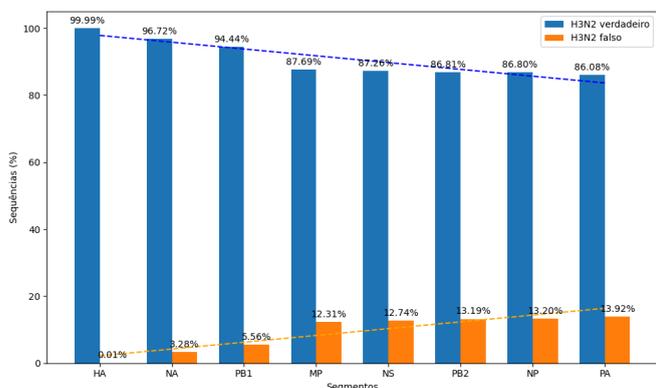


Figura 01. Gráfico de barras indicando as sequências dos segmentos que são verdadeiramente do subtipo H3N2 (em azul) e das que são falsas H3N2 (em laranja), após análise utilizando o classificador taxonômico *kaiju*. É importante destacar que esse sistema de classificação baseia-se apenas na tipagem dos segmentos HA e NA, e não na classificação de todos os segmentos.

Para entender melhor como é a organização genômica ao longo dos segmentos para os vírus H3N2, utilizamos o diagrama de sankey (**Fig. 02**) que ilustra esse comportamento observado por meio da classificação taxonômica (**Fig. 01**), onde observamos a preservação dos segmentos HA, NA e PB1 presentes no subtipo H3N2, e um perfil heterogêneo quanto aos subtipos de IAV para os outros segmentos. Como representado no diagrama, o segmento 4 (HA – H3) é altamente específico para este subtipo (H3N2), enquanto os segmentos 6 e 2 (NA - N2 e PB1), apesar de apresentarem uma maior variabilidade, com alguns dos vírus, que chamaremos de H3Nx (ou seja, o segmento NA dos isolados que são “falsos” H3N2 não são os N2 desse subtipo, mas tratam-se de outra neuraminidase, com diferentes características genéticas e antigênicas), não serem verdadeiros H3N2, são mais preservados. Analisando os outros segmentos (além de HA que é altamente específico, e o NA e o PB1 que apresentam uma certa especificidade também para suínos), estes, são específicos apenas para *Homo sapiens*, indicando que o subtipo H3N2, de uma forma geral (englobando todos os segmentos, que compõem o genoma completo do vírus), apresenta uma especificidade vírus-hospedeiro que é particular ao *Homo sapiens*.

No caso dos isolados de suínos, é bem evidente essa diferenciação, sendo, para os segmentos de HA, NA e PB1, a maioria das sequências efetivamente de H3N2, mas para os outros segmentos, as sequências são, na verdade, correspondentes ao subtipo H1N1 (*txid* 641809). Analisando a história evolutiva do H3N2, que surgiu em 1968, causando uma pandemia (inicialmente em Hong Kong), é possível encontrar na literatura informações e análises sobre a

conservação do genoma viral ao longo do tempo (provenientes de isolados virais antigos) que corroboram com os padrões identificados nas análises realizadas neste trabalho. Inicialmente, o subtipo H1N1, causador da gripe espanhola em 1918, teve origem suína e, posteriormente, um rearranjo entre um vírus derivado do H1N1 pandêmico e H2N2 derivado de aves deram origem ao H2N2 causador da pandemia de 1957 (que ocorreu inicialmente na Ásia). Anos depois, um outro rearranjo genômico entre o H2N2 pandêmico e H3Nx derivado de aves, finalmente resultou na emergência do subtipo H3N2 que causou uma pandemia em 1968, e até hoje o seu agente etiológico continua em circulação em cadeias de transmissão sustentadas envolvendo seres humanos.

Uma característica notória ao longo da trajetória de evolução molecular do vírus, é a ocorrência desses eventos de rearranjo, e para esses vírus IAV que apresentam a capacidade de infectar humanos e continuar em circulação, é a conservação de alguns dos segmentos, e a troca de outros. Os segmentos 4, 6 e 2 (HA, NA e PB1, respectivamente), foram sempre alterados nos eventos de rearranjo em relação ao H1N1 suíno “original” de 1918, ou seja, nos eventos discutidos anteriormente, esses segmentos, presentes nos vírus circulantes entre humanos, foram provenientes de IAVs aviários. Já todos os outros cinco segmentos (NP, NS, MP, PA e PB2), após os eventos de rearranjo, continuaram sendo os mesmos do H1N1 suíno de 1918. Portanto, esse padrão parecido de comportamento desses segmentos do ponto de vista evolutivo, podem explicar os padrões encontrados neste trabalho.

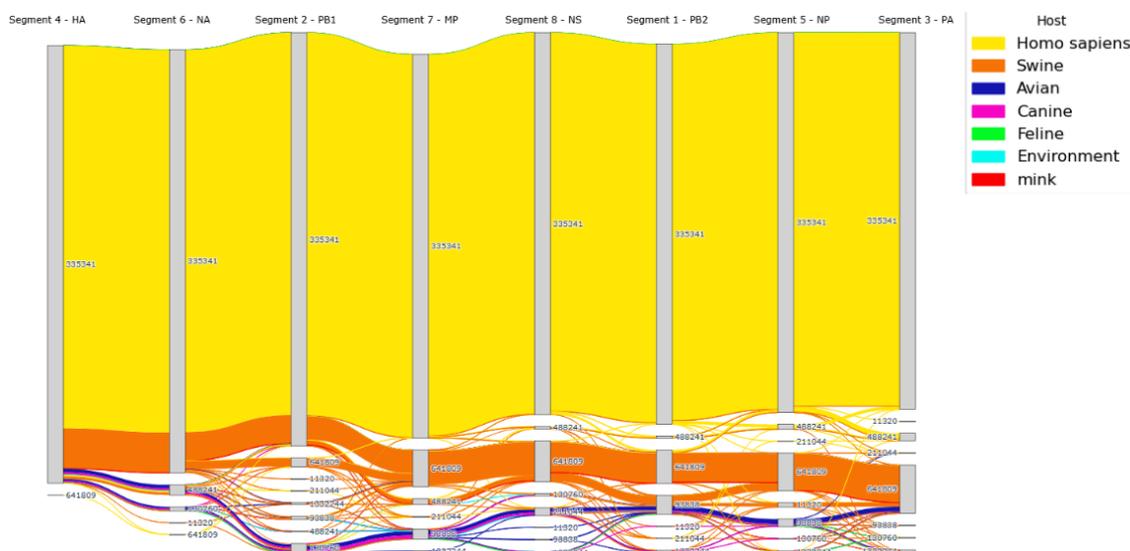


Figura 02. Diagrama de sankey mostrando o fluxo da informação, com os segmentos de H3N2 mais preservados à esquerda (HA, NA e PB1, principalmente) e os menos preservados à direita, em que ocorre uma maior “mistura” entre eles (neste caso, as colunas que são referentes aos subtipos estão identificadas com o txid, atribuídos pelo *kaiju*, sendo 335341 referente ao H3N2 e o 641809 referente ao H1N1, por exemplo).

A análise dos gráficos de PCA do RSCU de cada um dos segmentos (**Fig. 03**) ajuda a explicar essa especificidade diferencial dos segmentos, como já apresentado e discutido anteriormente. Primeiramente, é importante ressaltar que após as primeiras análises, oito sequências foram removidas para as análises consideradas apresentadas à seguir, sendo elas: 4 de HA (que não eram H3 de H3N2), além de outras 3 isoladas de “Environment” (ambiente) e 1 sequência de “mink”, totalizando em 28349 sequências para cada segmento.

No caso dos segmentos 4, 6 e 2 (HA, NA e PB1, respectivamente, referentes aos gráficos **A**, **C** e **G**), é possível visualizar uma sobreposição entre as sequências de *Homo sapiens* e suínos, sem uma diferenciação clara no uso relativo de códons sinônimos entre esses hospedeiros (considerando PC1 e PC2). Já para os segmentos 7, 5, 8, 3 e 1 (MP, NP, NS, PA e PB2, respectivamente, referentes aos gráficos **B**, **D**, **E**, **F** e **H**) há um claro agrupamento considerando os padrões de RSCU para *Homo sapiens* e os outros hospedeiros vertebrados (suínos, aves, caninos e felinos), sendo possível visualizar essa separação entre grupos nesses gráficos. Quando analisada a variância explicada para os dois

primeiros componentes principais (PC1 e PC2, de um total de 59), a variabilidade dos dados é maior (melhor explicada) para o caso dos segmentos 7, 5, 8, 3 e 1, corroborando com o que foi visto até então.

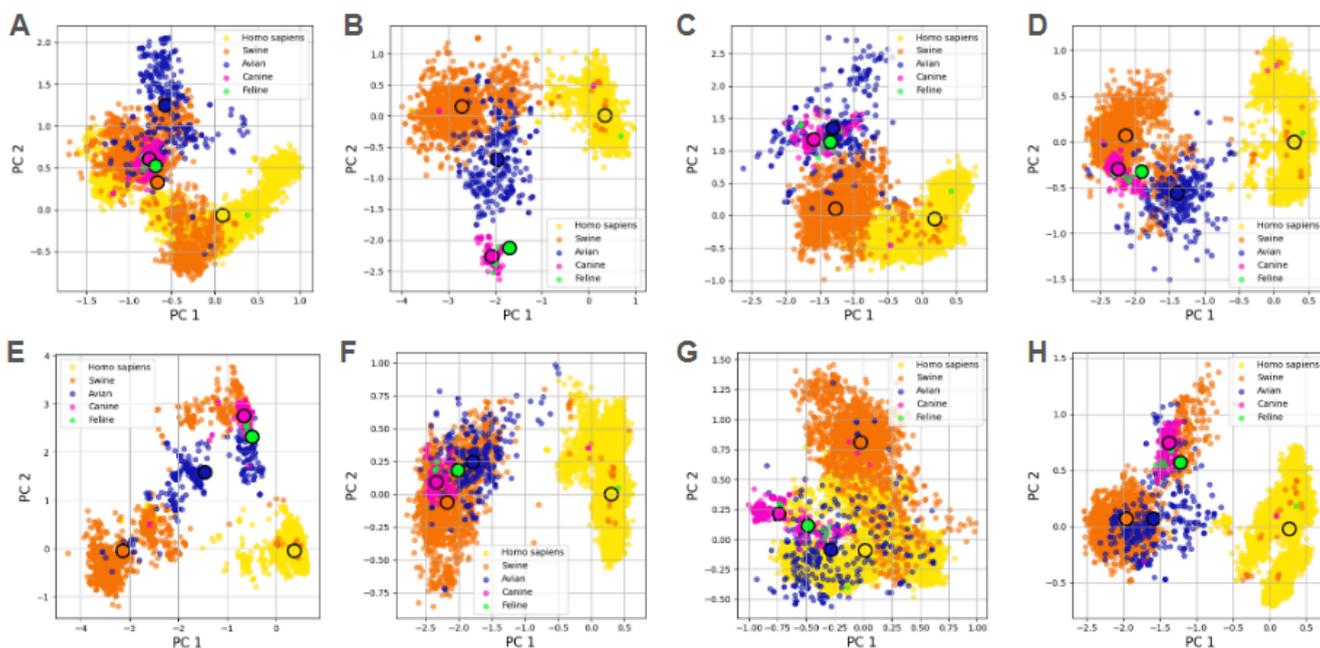


Figura 03. Gráficos de PCA do RSCU de cada um dos segmentos. **(A)** Segmento 4 - HA; **(B)** Segmento 7 - MP; **(C)** Segmento 6 - NA; **(D)** Segmento 5 - NP; **(E)** Segmento 8 - NS; **(F)** Segmento 3 - PA; **(G)** Segmento 2 - PB1; **(H)** Segmento 1 - PB2. As cores distintas representam as fontes/hospedeiros de obtenção das sequências analisadas (amarelo = *Homo sapiens*; laranja = suínos; azul = aves; rosa = caninos; verde = felinos).

Ao observar os gráficos gerados a partir dos resultados de CAI para os segmentos 4, 6 e 2 (HA, NA e PB1, respectivamente), tendo como base a tabela de uso de códons de hospedeiros *Homo sapiens* e aves (*Avian*), obtidas a partir do CoCoPUTs (do inglês, *Codon and Codon Pair Usage Tables*), é possível visualizar que em todos os casos (para esses três segmentos), a maioria das sequências estão acima da linha do eCAI (ou *expected CAI*, que é o CAI esperado para um sequência nucleotídica randômica). Isso indica uma adaptação ao uso de códons do hospedeiro, ou seja, à maquinaria celular hospedeira. Apesar de, para alguns casos, ser possível reconhecer sequências de suínos abaixo da linha do eCAI (ou seja, não serem tão bem adaptadas), para o caso de humanos (sequências de *Homo sapiens*), se mostraram praticamente todo o tempo (desde 1968, quando o H3N2 começou a circular entre humanos até os dias de hoje) acima do eCAI, ou seja, bem adaptados à tradução em células hospedeiras.

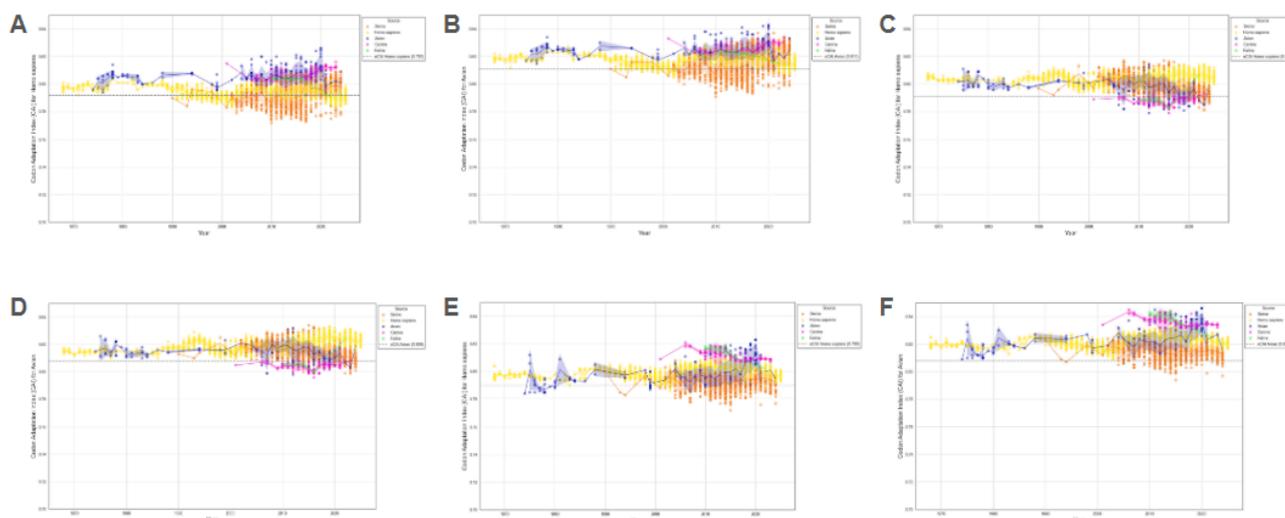


Figura 04. Gráficos de CAI x Tempo com eCAI para os segmentos 4, 6 e 2 (HA, NA e PB1 respectivamente), com a tabela de uso de códons de hospedeiros *Homo sapiens* e Aves (*Avian*). **(A)** Segmento 4 - HA - *Homo sapiens*; **(B)** Segmento 4 - HA - *Avian*; **(C)** Segmento 6 - NA - *Homo sapiens*; **(D)** Segmento 6 - NA - *Avian*; **(E)** Segmento 2 - PB1 - *Homo sapiens*; **(F)** Segmento 2 - PB1 - *Avian*. As cores distintas representam as fontes/hospedeiros de obtenção das sequências analisadas (amarelo = *Homo sapiens*; laranja = suínos; azul = aves; rosa = caninos; verde = felinos).

CONCLUSÕES:

Dessa forma, à partir do reconhecimento de diferentes padrões genômicos característicos dos segmentos virais do subtipo H3N2, os quais apresentam níveis variados de conservação e adaptação ao hospedeiro, torna-se possível compreender os processos históricos e evolutivos da relação entre o vírus Influenza A e os diferentes hospedeiros, em especial considerando o *Homo sapiens*. Assim, o avanço desse tipo de estudo, pautado em análises como RSCU, CAI e classificação taxonômica, pode elucidar as bases moleculares que sustentam o surgimento, a permanência e, eventualmente, a reemergência de variantes virais em diferentes contextos epidemiológicos, destacando a importância da investigação integrada dos padrões genômicos na vigilância e controle de vírus respiratórios como o Influenza.

BIBLIOGRAFIA:

1. Long, J. S., Mistry, B., Haslam, S. M., & Barclay, W. S. (2018). Host and viral determinants of influenza A virus species specificity. *Nature Reviews Microbiology* 17:2, 17(2), 67–81. <https://doi.org/10.1038/s41579-018-0115-z>
2. Wendel, I., Rubbenstroth, D., Doedt, J., Kochs, G., Wilhelm, J., Staeheli, P., Klenk, H.-D., & Matrosovich, M. (2015). The Avian-Origin PB1 Gene Segment Facilitated Replication and Transmissibility of the H3N2/1968 Pandemic Influenza Virus. *Journal of Virology*, 89(8), 4170–4179. <https://doi.org/10.1128/JVI.03194-14/ASSET/AECAFA8A-8A7A-492C-BD42-4F5319F5592A/ASSETS/GRAPHIC/JV9990902320009.JPEG>
3. Sun, T., Guo, Y., Zhao, L., Fan, M., Huang, N., Tian, M., Liu, Q., Huang, J., Liu, Z., Zhao, Y., Ji, Z., & Ping, J. (2022). Evolution of the PB1 gene of human influenza A (H3N2) viruses circulating between 1968 and 2019. *Transboundary and Emerging Diseases*, 69(4), 1824–1836. <https://doi.org/10.1111/TBED.14161;REQUESTEDJOURNAL:JOURNAL:18651682;WGROUPE:STRING:PUBLICATION>
4. Smith, B. L., Chen, G., Wilke, C. O., & Krug, R. M. (2018). Avian influenza virus PB1 gene in H3N2 viruses evolved in humans to reduce interferon inhibition by skewing codon usage toward interferon-altered tRNA pools. *MBio*, 9(4). <https://doi.org/10.1128/MBIO.01222-18/ASSET/A5A9D78E-C695-4998-A64C-2E1322B61E00/ASSETS/GRAPHIC/MBO0041839610006.JPEG>
5. Allen, J. D., & Ross, T. M. (2018). H3N2 influenza viruses in humans: Viral mechanisms, evolution, and evaluation. *Human Vaccines and Immunotherapeutics*, 14(8), 1840–1847. <https://doi.org/10.1080/21645515.2018.1462639;CTYPE:STRING:JOURNAL>
6. Chauhan, R. P., & Gordon, M. L. (2022). An overview of influenza A virus genes, protein functions, and replication cycle highlighting important updates. *Virus Genes* 2022 58:4, 58(4), 255–269. <https://doi.org/10.1007/S11262-022-01904-VV>
7. Resende, P. C., Junqueira, D. M., Tochetto, C., Ogrzewalska, M., Motta, F. C., Lopes, J., Appolinario, L., Macedo, L., Caetano, B., Matos, A., Silva, T., Pereira, E. C., Lima, L. F., Riediger, I., Debur, M. do C., Becker, G. N., Andrade, A., Nasr, A. M. L. F., Piler, R. A., ... Siqueira, M. M. (2024). Zoonotic transmission of novel Influenza A variant viruses detected in Brazil during 2020 to 2023. *Nature Communications* 2024 15:1, 15(1), 1–12. <https://doi.org/10.1038/s41467-024-53815-z>
8. Zhou, J., Xing, Y., Zhou, Z., & Wang, S. (2023). A comprehensive analysis of Usutu virus (USUV) genomes revealed lineage-specific codon usage patterns and host adaptations. *Frontiers in Microbiology*, 13, 967999. <https://doi.org/10.3389/FMICB.2022.967999/BIBTEX>
9. Puigbò, P., Bravo, I. G., & Garcia-Vallve, S. (2008). CAIcal: A combined set of tools to assess codon usage adaptation. *Biology Direct*, 3(1), 1–8. <https://doi.org/10.1186/1745-6150-3-38/FIGURES/2>
10. Alexaki, A., Kames, J., Holcomb, D. D., Athey, J., Santana-Quintero, L. v., Lam, P. V. N., Hamasaki-Katagiri, N., Osipova, E., Simonyan, V., Bar, H., Komar, A. A., & Kimchi-Sarfaty, C. (2019). Codon and Codon-Pair Usage Tables (CoCoPUTs): Facilitating Genetic Variation Analyses and Recombinant Gene Design. *Journal of Molecular Biology*, 431(13), 2434–2441. <https://doi.org/10.1016/J.JMB.2019.04.021>
11. Kawashima, I. Y., Lopez, M. C. N., Cunha, M. dos P., & Hashimoto, R. F. (2022). SARS-CoV-2 host prediction based on virus-host genetic features. *Scientific Reports*, 12(1), 1–9. <https://doi.org/10.1038/S41598-022-08350-6.SUBJMETA=114.1305.326.4130.596.631;KWRD=MACHINE+LEARNING.SARS-COV-2>
12. dos Passos Cunha, M., Ortiz-Baez, A. S., de Melo Freire, C. C., & de Andrade Zanotto, P. M. (2018). Codon adaptation biases among sylvatic and urban genotypes of Dengue virus type 2. *Infection, Genetics and Evolution*, 64, 207–211. <https://doi.org/10.1016/J.MEEGID.2018.05.017>
13. Linster, M., Schrauwen, E. J. A., van der Vliet, S., Burke, D. F., Lexmond, P., Bestebroer, T. M., Smith, D. J., Herfst, S., Koel, B. F., & Fouchier, R. A. M. (2019). The Molecular Basis for Antigenic Drift of Human A/H2N2 Influenza Viruses. *Journal of Virology*, 93(8). <https://doi.org/10.1128/JVI.01907-18/ASSET/DD123B9B-A305-4674-94DE-91A902CE8F80/ASSETS/GRAPHIC/JVI.01907-18-F0004.JPEG>
14. Lindstrom, S. E., Cox, N. J., & Klimov, A. (2004). Genetic analysis of human H2N2 and early H3N2 influenza viruses, 1957–1972: evidence for genetic divergence and multiple reassortment events. *Virology*, 328(1), 101–119. <https://doi.org/10.1016/J.VIROL.2004.06.009>
15. Joseph, U., Linster, M., Suzuki, Y., Krauss, S., Halpin, R. A., Vijaykrishna, D., Fabrizio, T. P., Bestebroer, T. M., Maurer-Stroh, S., Webby, R. J., Wentworth, D. E., Fouchier, R. A. M., Bahl, J., & Smith, G. J. D. (2015). Adaptation of Pandemic H2N2 Influenza A Viruses in Humans. *Journal of Virology*, 89(4), 2442–2447. https://doi.org/10.1128/JVI.02590-14/SUPPL_FILE/JV999090067SO1.PDF
16. Guo, F., Roy, A., Wang, R., Yang, J., Zhang, Z., Luo, W., Shen, X., Chen, R. A., Irwin, D. M., & Shen, Y. (2021). Host Adaptive Evolution of Avian-Origin H3N2 Canine Influenza Virus. *Frontiers in Microbiology*, 12, 655228. <https://doi.org/10.3389/FMICB.2021.655228/BIBTEX>