

Universidade Federal de São Carlos (UFSCar)
Centro de Ciências Biológicas e da Saúde (CCBS)
Departamento de Genética e Evolução (DGE)

JUNIOR OLÍMPIO MARTINS

**ANÁLISE FILOGEOGRÁFICA DO SEGMENTO GENÔMICO DA PROTEÍNA
NEURAMINIDASE DO H1N1 NO BRASIL**

SÃO CARLOS - SP

2021

UNIVERSIDADE FEDERAL DE SÃO CARLOS

JUNIOR OLÍMPIO MARTINS

**ANÁLISE FILOGEOGRÁFICA DO SEGMENTO GENÔMICO DA PROTEÍNA
NEURAMINIDASE DO H1N1 NO BRASIL**

Trabalho apresentado à banca examinadora da
Universidade Federal de São Carlos como requisito para
a obtenção do título de Bacharel em Biotecnologia.

Área de Concentração: Bioinformática

Orientador: Prof. Dr. Caio César de Melo Freire

SÃO CARLOS -- SP

2021

RESUMO

A gripe sazonal, causada pelos vírus Influenza, é uma das infecções virais mais prevalentes no mundo, causando 290-650 mil óbitos anuais. Tais vírus, cujo genoma é composto por segmentos de RNA fita simples negativamente orientada, possuem elevadas taxas mutacionais e a capacidade de rearranjo entre segmentos, permitindo-lhes rápido escape imune e mesmo o surgimento de linhagens pandêmicas. Após a implantação do SINAN Influenza no Brasil, em 2009, houve um grande aumento no número de genomas sequenciados e a unificação dos casos notificados, permitindo a realização de análises filodinâmicas sob as sequências dos vírus coletados no país e sua correlação com dados epidemiológicos. Durante o presente Trabalho de Conclusão de Curso, objetivou-se estudar a evolução molecular e a epidemiologia do vírus Influenza A H1N1 no Brasil, usando como modelo o segmento genômico que codifica a neuraminidase. Por meio desta abordagem, foi possível: (i) calcular a taxa mutacional do segmento (3.02×10^{-3} substituições/sítio/ano) em acordo com a literatura existente; (ii) identificar variações nos tamanhos populacionais correlatas a estações gripais com aumento no número de casos; e (iii) mesmo identificar a região Sul como semeadora de linhagens gripais em 2009. Tais resultados, apesar de não incluírem o impacto do rearranjo, fornecem importantes informações sobre a epidemiologia e parâmetros evolutivos da gripe sazonal por Influenza A H1N1 no Brasil.

Palavras-chave: Gripe, Influenza A, neuraminidase, filogenética, filogeografia.

LISTA DE ILUSTRAÇÕES

- Figura 1 - Amostras positivas de síndrome gripal para vírus respiratório dentre os casos analisados em 2019 (BRASIL, 2020). Entre a 11^a (março) e a 30^a semana epidemiológica (julho), pode-se notar não só o aumento dos casos associado à temporada de gripe (outono-inverno) como também a dominância dos vírus Influenza A e B..... 8
- Figura 2 - Esquema de nomenclatura para vírus Influenza (Adaptado de CDC, 2021). O nome é dado pelo tipo do vírus (A-D), seguido pelo local de isolamento, número da cepa viral (geralmente um identificador da amostra), o ano de isolamento e, no caso do Influenza A, o subtipo viral, conforme definido pelas proteínas hemaglutinina e neuraminidase..... 9
- Figura 3 - Segmentos genômicos do vírus Influenza A e seus produtos proteicos (VIRALZONE, 2021).....10
- Figura 4 - Frameshift +1 no segmento 3 (PA) de Influenza A (VIRALZONE, s.d.).....11
- Figura 5 - Filogenia das sequências de Influenza A(H5N1) isoladas na Europa, Oriente Médio e África (RUSSELL et al., 2012)..... 12
- Figura 6 - Representação de um evento de rearranjo (SIMON-LORIERE; HOLMES, 2011). Durante a etapa de montagem dos capsídeos virais, é possível o recrutamento de segmentos de ambas as linhagens, permitindo o surgimento de linhagens com diferentes combinações de segmentos genômicos.....13
- Figura 7 - Linhagens de gripe suína que levaram à linhagem Influenza A(H1N1)pdm2009 (GARTEN et al., 2009). Acredita-se que os segmentos que deram origem a essa linhagem adentraram diferentes populações suínas em três momentos diferentes (1918, 1979 e 1998), sendo que o último segmento, codificante para a proteína PB2, circulou na população humana entre 1968 e 1998 antes de adentrar populações suínas, em que ficou circulando anteriormente à pandemia..... 14
- Figura 8 - Diferentes tipos de mutação e suas taxas associadas (Elaborado pelo autor). As setas indicadas por Tv são taxas de transversão relativas às transversões entre purinas e pirimidinas, enquanto aquelas indicadas por Ts são relativas às taxas de transição numa mesma classe de nucleotídeos.....16
- Figura 9 - Diferentes modelos populacionais e seus impactos nas filogenias (VOLZ; KOELLE; BEDFORD, 2013)..... 17
- Figura 10 - Sequências obtidas por ano e local de coleta (Elaborado pelo autor). Pode-se notar uma grande representatividade da região Sul, com mais de 89% das sequências, seguida pela região Nordeste com aproximadamente 6% e as restantes disputando 4%..... 22
- Figura 11 - Gráfico comparando o tamanho efetivo populacional (NeG) pelo tempo (elaborado pelo autor). Em vermelho estão destacados o ano de 2009 e o período de junho de 2011 a junho de 2012, durante os quais ocorreram aumento no número de casos e correspondente aumento no tamanho efetivo populacional. 23
- Figura 12 - Número de notificações de SRAG confirmadas para Influenza A (H1N1)pdm2009 por ano (Adaptado de BRASIL, 2012, 2013, 2014, 2015, 2016, 2017). Durante o ano de 2009, devido à pandemia, foram notificados mais de cem mil casos de SRAG por influenza, de forma que o gráfico foi truncado para melhor visualização dos outros anos..... 24
- Figura 13 - Árvore filogenética baseada nas sequências nucleotídicas do segmento 6 (elaborada pelo autor). A escala temporal dos ramos é anual, sendo que os nós internos foram coloridos de acordo com a localização mais provável do ancestral dos nós descendentes. Além disso, estão exibidas as probabilidades posteriores superiores a 80%, sendo que ramos compartilhando o mesmo local de coleta e baixa probabilidade entre nós foram colapsados, com o número de nós terminais indicado.....26
- Figura 14 - Mapa do Brasil com as quatro principais migrações do vírus entre 1995 e 2016 (elaborado pelo autor). O mapa foi elaborado a partir do shapefile das fronteiras estaduais do país, disponível em <<https://nereus.webhostusp.sti.usp.br/?dados=brasil>>, em conjunto com os dados da animação do Spread3. As setas indicam a mudança da localização mais provável de um nó ancestral para a de seu descendente, sendo que as se concentraram no primeiro semestre de cada ano.....28
- Figura 15 - Porcentagens relativas das mortes (confirmadas por laboratório) por semana por região durante a pandemia. (SCHUCK-PAIM et al., 2012). Pode-se notar o domínio das regiões Sul e Sudeste no início e final da pandemia, sugerindo uma região de origem para a linhagem pandêmica no país.....29

Figura 16 - Taxas de migração do vírus HIV-1 subtipo C entre regiões após sua introdução no país na década de 1970 (DELATORRE et al., 2013). A largura das setas é proporcional à taxa de transição viral entre regiões, com grande participação de Santa Catarina (SC) e Paraná como fontes.....30

Figura 17 - Taxas de migração do vírus Influenza A H3N2 entre regiões entre 1999 e 2012 (BORN et al., 2016). A largura das setas é proporcional ao número médio estimado de transições virais entre regiões, com elevada participação das regiões Sul (BR-S) e Sudeste (BR-SE) em detrimento do Nordeste (BR-NE).....30

LISTA DE TABELAS

Tabela 1 - Resultados da inferência coalescente Bayesiana (Elaborado pelo autor). MRCA equivale ao Ancestral Comum Mais Recente (Most Recent Common Ancestor).....	23
--	----

LISTA DE ABREVIATURAS E SIGLAS

AIC	<i>Akaike Information Criterion</i>
AIC _c	<i>Corrected Akaike Information Criterion</i>
BEAST	<i>Bayesian Evolutionary Analysis Sampling Trees</i>
BEAUti	<i>Bayesian Evolutionary Analysis Utility</i>
CDC	<i>Centers for Disease Control and Prevention</i>
ESS	<i>Effective Sample Size</i>
GTR	<i>Generalized time-reversible</i>
HA, H	Hemaglutinina
H5N1	<i>Highly Pathogenic Avian Influenza</i>
MRCA	<i>Most Recent Common Ancestor</i>
HPD	<i>Highest Posterior Density</i>
MCMC	<i>Monte Carlo Markov Chains</i>
M1	Proteína de matriz 1
M2	Proteína de matriz 2
MCC	<i>Maximum Clade Credibility</i>
NA, N	Neuraminidase
NeG	Tamanho efetivo populacional
NEP	Proteína de Exportação Nuclear
NP	Nucleoproteína
NS1	Proteína não estrutural
RDP	<i>Recombination Detection Program</i>
TVM	<i>Transversion Model</i>
OMS, WHO	Organização Mundial da Saúde
SINAN	Sistema de Informação de Agravos de Notificação
Spread	<i>Spatial Phylogenetic Reconstruction of Evolutionary Dynamics</i>
RAxML	<i>Randomized Axelerated Maximum Likelihood</i>

SUMÁRIO

1. INTRODUÇÃO.....	8
1.1. Influenza A.....	8
1.2. <i>Drift</i> e <i>Shift</i> antigênico.....	11
1.3. Filodinâmica e Filogeografia.....	14
2. OBJETIVOS.....	18
2.1. Objetivo Geral.....	18
2.2. Objetivos Específicos.....	18
3. MATERIAIS E MÉTODOS.....	19
3.1. Aquisição de sequências.....	19
3.2. Pré-processamento dos dados.....	19
3.3. Análise filogenética Bayesiana.....	20
3.4. Diagnóstico e análise dos dados.....	20
4. RESULTADOS E DISCUSSÃO.....	22
6. CONCLUSÃO.....	32
7. PERSPECTIVAS FUTURAS.....	33

1. INTRODUÇÃO

1.1. Influenza A

A gripe sazonal é uma das infecções virais mais prevalentes no mundo em números absolutos, infectando aproximadamente um bilhão de pessoas anualmente, dentre as quais 290-650 mil evoluem a óbito (IULIANO *et al.*, 2018). Além disso, o repentino surgimento de novas linhagens pandêmicas evidenciou o potencial devastador desses vírus, em especial durante as pandemias de 1918, 1957 e 2009, que ceifaram 17 milhões, 1-4 milhões e 150-570 milhares de vidas, respectivamente (DAWOOD *et al.*, 2012; ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2013; SPREEUWENBERG; KRONEMAN; PAGET, 2018).

Tanto a gripe sazonal quanto a pandêmica possuem como agentes etiológicos os vírus Influenza, membros da família Orthomyxoviridae, os quais podem ser classificados nos tipos A, B, C e D (MCCAULEY *et al.*, 2011). Tais vírus, cujo genoma é dividido em segmentos de RNA de fita simples negativamente orientado, são caracterizados por elevada variabilidade genética, tanto intra quanto entre hospedeiros.

Dentre tais tipos, os mais relevantes para a saúde pública são o A e B (WONG; WEBBY, 2013), presentes na maioria dos casos de síndrome gripal causada por vírus respiratório, conforme pode ser visto na Figura 1.

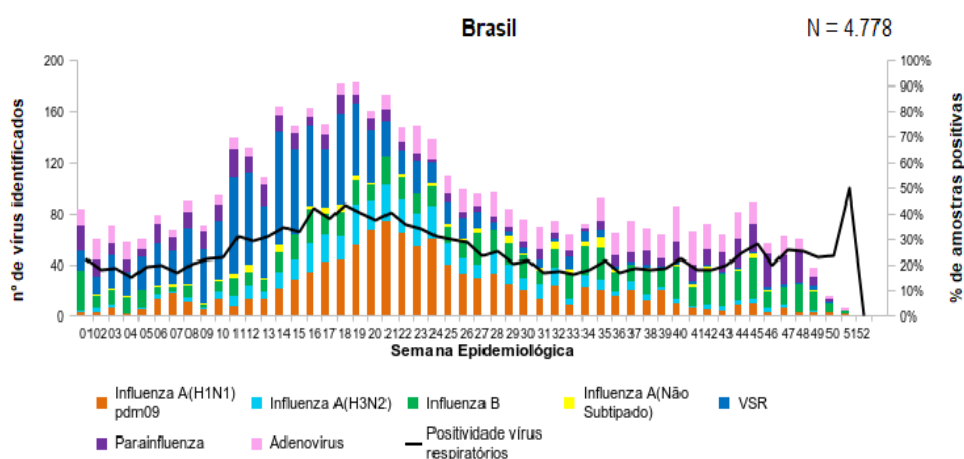


Figura 1 - Amostras positivas de síndrome gripal para vírus respiratório dentre os casos analisados em 2019 (BRASIL, 2020). Entre a 11^a (março) e a 30^a semana epidemiológica (julho), pode-se notar não só o aumento dos casos associado à temporada de gripe (outono-inverno) como também a dominância dos vírus Influenza A e B.

Além disso, há um destaque maior nos estudos do Influenza A devido ao seu potencial pandêmico, visto que diversas linhagens deste vírus presentes em espécies reservatório não só possuem a capacidade de causar doença como também de adquirir transmissão eficiente entre humanos em pouco tempo (TAUBENBERGER; MORENS, 2010).

Dada a alta variedade intraespecífica do Influenza A, o mesmo é classificado em subtipos de acordo com suas glicoproteínas de superfície, a hemaglutinina (H) e a neuraminidase (N). Apesar de já terem sido descritos 18 tipos de hemaglutinina e 11 tipos de neuraminidase, pouco mais de 130 das 198 possíveis combinações de subtipos foram encontradas na natureza (CENTERS FOR DISEASE CONTROL AND PREVENTION (CDC), 2021), sendo que os mais prevalentes no Brasil são H1N1 e H3N2 (BRASIL, 2020).

Considerando essa diversidade de subtipos dos vírus Influenza, em 1980 a Organização Mundial da Saúde definiu um sistema de nomenclatura para isolados desse vírus (CDC, 2021), conforme o modelo da Figura 2.

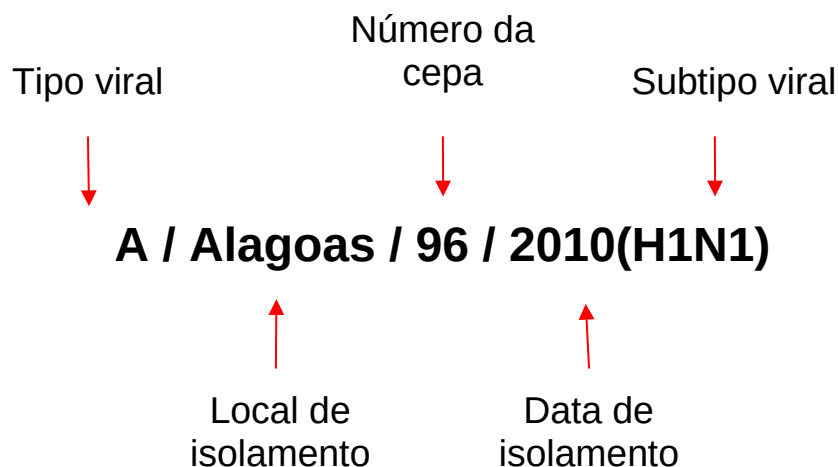


Figura 2 - Esquema de nomenclatura para vírus Influenza (Adaptado de CDC, 2021). O nome é dado pelo tipo do vírus (A-D), seguido pelo local de isolamento, número da cepa viral (geralmente um identificador da amostra), o ano de isolamento e, no caso do Influenza A, o subtipo viral, conforme definido pelas proteínas hemaglutinina e neuraminidase.

As partículas virais do Influenza A possuem uma estrutura complexa e de tamanho variável entre 80 e 120 nm. Cada partícula possui um envelope lipídico no qual estão incrustadas três tipos de proteínas: hemaglutinina (H ou HA), neuraminidase (N ou NA) e um canal iônico, que é a proteína de matriz 2 (M2). Esse

envelope recobre o nucleocapsídeo, composto pela proteína de matriz 1 (M1), pela proteína de exportação nuclear (NEP) e um conjunto de ribonucleoproteínas, que consistem nas oito moléculas de RNA fita simples de sentido negativo, nucleoproteínas (NP), além do complexo RNA polimerase (PB1, PB2 e PA) (KILBOURNE, 1987).

Cada um dos oito segmentos genômicos do vírus Influenza A possui um produto gênico principal, havendo também segmentos que codificam para duas ou mais proteínas (YAMAYOSHI *et al.*, 2016). Os segmentos, numerados de 1 a 8, têm como produto principal, respectivamente, PB2, PB1, PA, HA, NP, NA, M1 e NS1, conforme a figura abaixo (Figura 3).

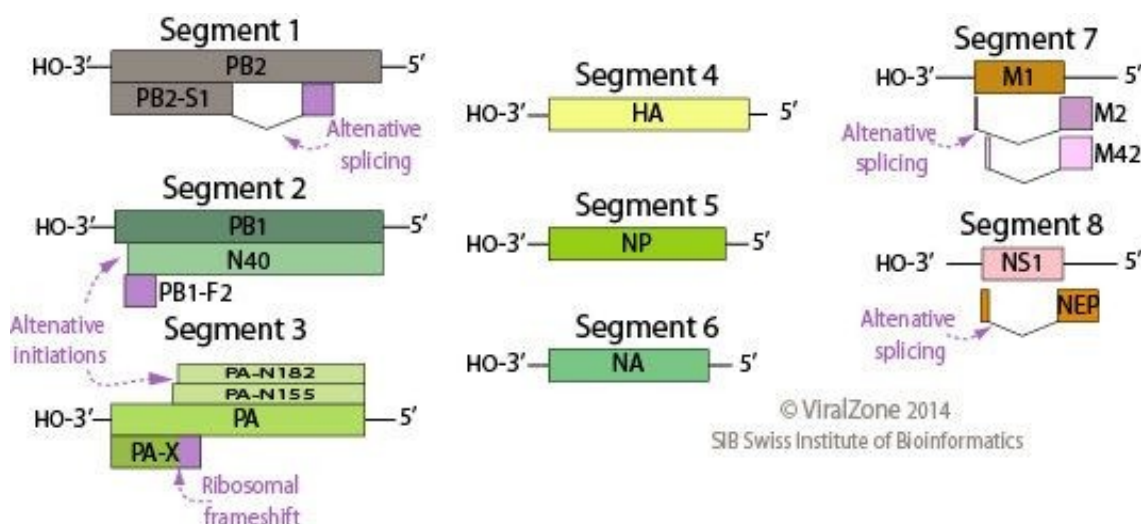


Figura 3 - Segmentos genômicos do vírus Influenza A e seus produtos proteicos (VIRALZONE, 2021).

São vários os mecanismos que podem ser utilizados para a expressão de mais de uma proteína num mesmo segmento genômico, dentre os quais utilizar os mecanismos celulares de *splicing*, como é o caso das proteínas PB2-S1, M2, M42 e NEP (NS2). Ademais, o vírus pode utilizar um códon alternativo para início da tradução, como acontece nas proteínas N40 e PB1-F2 do segmento PB1 (YAMAYOSHI *et al.*, 2016).

Há também um terceiro processo conhecido como *ribosomal frameshift*, que ocorre quando o ribossomo tem sua leitura de códons afetada, mudando seu quadro de leitura em +1 ou em -1 nucleotídeo. Este processo geralmente resulta numa

proteína radicalmente diferente, como é o caso do *frameshift* +1 da proteína PA-X, cujos códons estão exemplificados na figura abaixo (Figura 4) (YEWDELL; INCE, 2012).

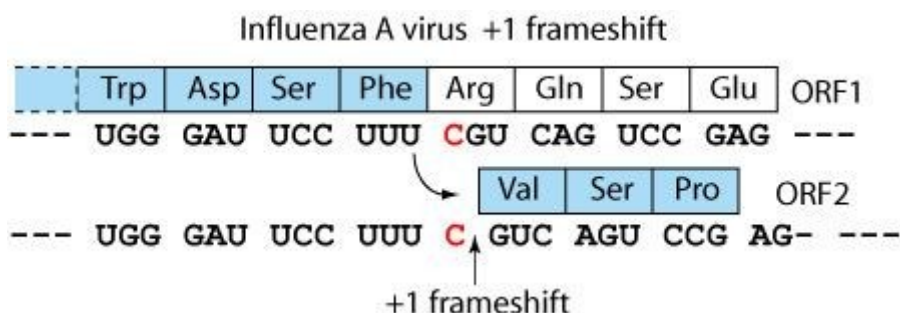


Figura 4 - Frameshift +1 no segmento 3 (PA) de Influenza A (VIRALZONE, s.d.).

O processo de infecção celular pelo vírus inicia-se com a adsorção do mesmo a receptores celulares contendo em sua região terminal um ácido siálico alfa-2,3Gal (em aves) ou α2,6Gal (em mamíferos) (MA; KAHN; RICHT, 2009). Essa adsorção, mediada pela hemaglutinina, é seguida pela fagocitose da partícula viral, que libera seu conteúdo no interior celular e permite a replicação dos segmentos genômicos (DOU *et al.*, 2018). Ao final do processo, a liberação das partículas virais nascentes é mediada pela neuraminidase, que cliva o ácido siálico, porém sem exibir um nível tão grande de especificidade quanto a hemaglutinina (MCAULEY *et al.*, 2019).

Esta proteína, com estrutura homotetramérica bastante conservada, é o principal alvo para tratamentos via inibidores análogos ao ácido N-acetilneuramínico, como o oseltamivir e zanamivir, porém é menos antigênica que a hemaglutinina, posto que poucas mutações próximas ao seu sítio ativo são capazes de eliminar as interações com anticorpos (AIR, 2012).

1.2. *Drift* e *Shift* antigênico

Existem dois mecanismos antigênicos principais que resultam em melhor evasão ao sistema imune pelo vírus Influenza, sendo eles o *drift* antigênico e o *shift* antigênico.

O *drift* antigênico é o resultado de mutações pontuais no genoma, que são positivamente selecionadas pela menor resposta imune que causam, renovando

anualmente a susceptibilidade da população e por conseguinte requerendo alteração da composição vacinal (WONG; WEBBY, 2013).

Tais mutações são especialmente relevantes quando afetam as subunidades da RNA polimerase, alterando a capacidade de replicação; ou as proteínas de superfície antigênicas, de forma que a ligação alterada aos anticorpos ou às moléculas receptoras pode causar alterações no tropismo celular, comprometimento da memória imunológica, além de mudanças na virulência e patogenicidade (LONG *et al.*, 2019).

Dada a continuidade de tais mutações, pode-se observar seu impacto na filogenia característica dos vírus Influenza, a qual possui uma aparência em “escada” com linhagens surgindo e sendo extintas continuamente, conforme a Figura 5 (GRENFELL *et al.*, 2004; PYBUS; RAMBAUT, 2009).

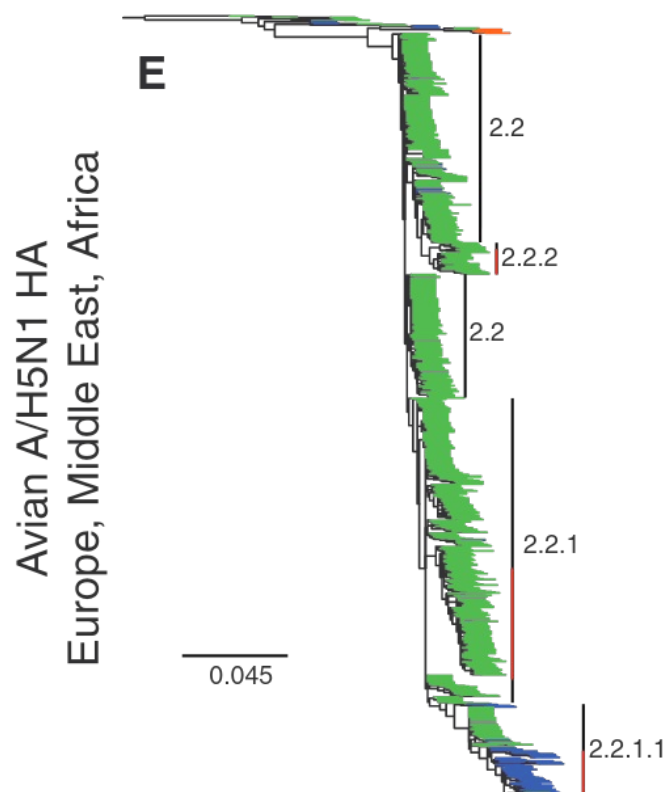


Figura 5 - Filogenia das sequências de Influenza A(H5N1) isoladas na Europa, Oriente Médio e África (RUSSELL *et al.*, 2012).

O segundo mecanismo é o *shift* antigênico, uma mudança abrupta nas características virais, que envolve o rearranjo (*reassortment*) de segmentos

genômicos durante a coinfeção por diferentes cepas, possibilitando o surgimento de linhagens com a capacidade de expressar proteínas de ambas (Figura 6) (VIJAYKRISHNA; MUKERJI; SMITH, 2015).

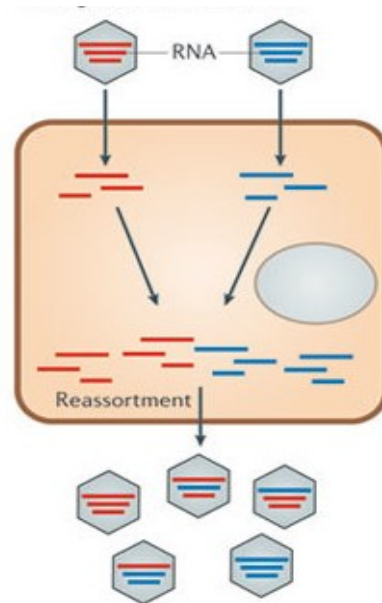


Figura 6 - Representação de um evento de rearranjo (SIMON-LORIERE; HOLMES, 2011). Durante a etapa de montagem dos capsídeos virais, é possível o recrutamento de segmentos de ambas as linhagens, permitindo o surgimento de linhagens com diferentes combinações de segmentos genômicos.

Este processo tem maior potencial disruptivo, pois pode gerar novas combinações das proteínas imunogênicas H e N em um curto intervalo de tempo, cuja combinação os humanos podem não ter memória imunológica, ou mesmo vantagens na transmissibilidade de linhagens patogênicas, como no caso do HPAIV H5N1, considerado um candidato à próxima linhagem pandêmica (RUSSELL *et al.*, 2012).

Além disso, a existência de determinados hospedeiros, como os suínos, que são suscetíveis à infecção por vírus de humanos e aves, torna o *shift* antigênico mais frequente e ainda menos previsível que o normal. Essa susceptibilidade se deve à presença de ambos os sacarídeos para linhagens aviárias e humanas em seu trato respiratório, facilitando a coinfeção (MA; KAHN; RICHT, 2009). Um dos exemplos mais relevantes disso é a linhagem que causou a pandemia de 2009, cujo

genoma é o produto de diversos rearranjos em hospedeiros suínos (Figura 7) (GARTEN *et al.*, 2009; TRIFONOV; KHIABANIAN; RABADAN, 2009).

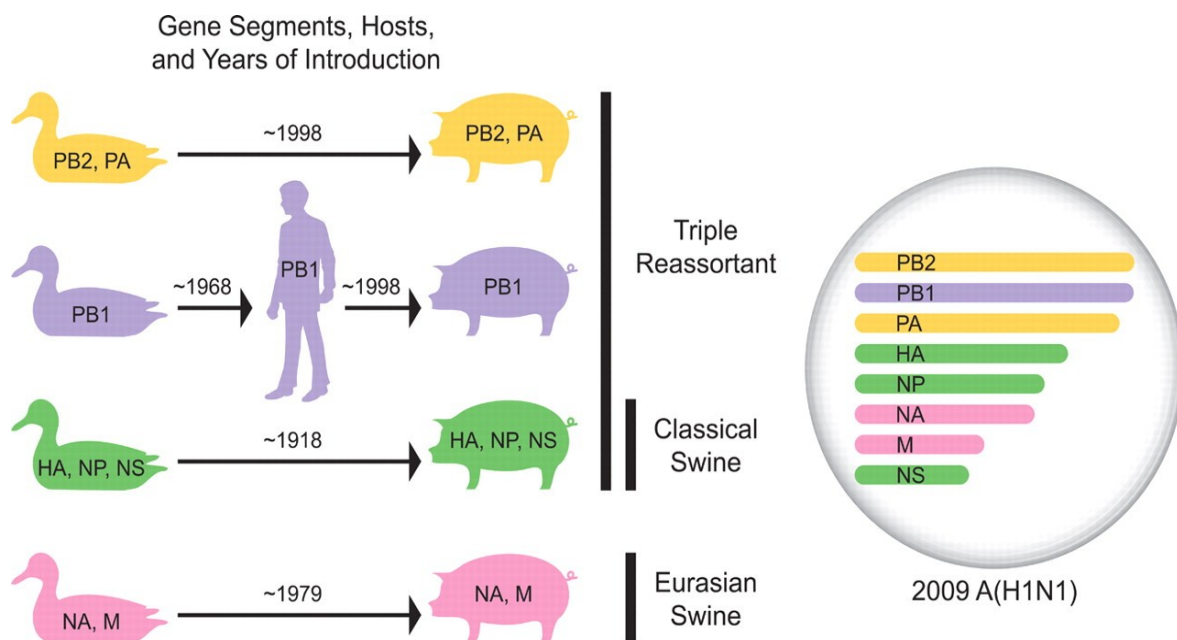


Figura 7 - Linhagens de gripe suína que levaram à linhagem Influenza A(H1N1)pdm2009 (GARTEN *et al.*, 2009). Acredita-se que os segmentos que deram origem a essa linhagem adentraram diferentes populações suínas em três momentos diferentes (1918, 1979 e 1998), sendo que o último segmento, codificante para a proteína PB2, circulou na população humana entre 1968 e 1998 antes de adentrar populações suínas, em que ficou circulando anteriormente à pandemia.

1.3. Filodinâmica e Filogeografia

Um aspecto muito importante do estudo dos vírus de RNA é a sua grande variabilidade genética. Essa variedade, causada tanto pelas altas taxas mutacionais quanto pelo curto tempo geracional, também é modulada por fatores epidemiológicos, como o número de hospedeiros, dinâmicas populacionais e mesmo a forma e taxa de transmissão, além de fatores imunológicos, como a intensidade e o tipo de resposta imune (HOLMES; GRENFELL, 2009).

Com o aumento nos esforços e tecnologias de amostragem e sequenciamento virais nos últimos anos, houve uma consequente melhora na capacidade de amostragem da variabilidade genética viral. Essa melhora, por sua vez, permitiu o desenvolvimento de novas técnicas de análise, dentre elas a filodinâmica, que é a área que se dedica a estudar como fatores epidemiológicos,

imunológicos e evolutivos interagem para definir a filogenia de um determinado grupo em estudo (PYBUS; RAMBAUT, 2009; VOLZ; KOELLE; BEDFORD, 2013).

Tal área encontra aplicações relevantes na saúde pública, pois o estudo das características únicas de cada vírus, como sua distribuição geográfica e tamanho da população suscetível, possibilita avaliar também a efetividade de vacinas e tratamentos antivirais (CHAO *et al.*, 2012). Adicionalmente, o estudo das informações contidas nas sequências permite estimar a origem de epidemias ou até mesmo de pandemias, tanto no tempo quanto no espaço (PYBUS *et al.*, 2003; PYBUS; RAMBAUT, 2009).

Além disso, ao se atribuir as coordenadas geográficas de cada táxon na árvore filogenética, é possível, através da filogeografia, estimar as probabilidades de um determinado nó da mesma estar associado a uma localização específica, permitindo estudar suas dinâmicas temporais e espaciais simultaneamente (HOLMES, 2004).

O uso de algoritmos computacionais para inferir filogenias, seja de máxima verossimilhança ou estatística Bayesiana, requer, dentre outros parâmetros, um modelo de evolução dos dados. No caso de sequências nucleotídicas, isso se traduz em modelos de substituição nucleotídica, que podem assumir variação tanto nas frequências nucleotídicas em si quanto nas taxas de variação entre as mesmas.

Dentre estes modelos se destaca o modelo *Generalized Time Reversible* (GTR) (TAVARÉ, 1986), devido à sua capacidade de tanto permitir variação nas frequências nucleotídicas como também ao assumir uma matriz simétrica com valores diferentes para cada um dos seis tipos diferentes de mutação (Figura 8). Essas propriedades são especialmente úteis para o estudo de sequências com padrões evolutivos mais complexos.

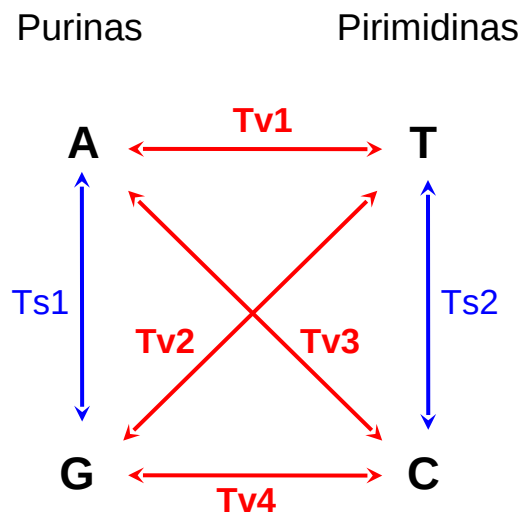


Figura 8 - Diferentes tipos de mutação e suas taxas associadas (Elaborado pelo autor). As setas indicadas por Tv são taxas de transverso relativas às transições entre purinas e pirimidinas, enquanto aquelas indicadas por Ts são relativas às taxas de transição numa mesma classe de nucleotídeos.

Além disso, há a possibilidade de se adicionar parâmetros a tais modelos, permitindo situações como uma proporção de sítios invariantes (I) e/ou a variação nas taxas de mutação entre os sítios (Γ).

O processo de escolha do modelo de substituição adequado, por sua vez, pode ser realizado pela comparação da relação do quanto um dado modelo explica os dados *versus* a quantidade de parâmetros livres que o mesmo adiciona, minimizando-se assim o *over-fitting*. Uma forma comum de estimar essa relação é através do cálculo de *Akaike Information Criterion* (Critério de Informação de Akaike (AIC)) de cada modelo, ou sua versão corrigida para pequenos tamanhos amostrais (*corrected AIC* (AIC_c)).

Outro parâmetro importante para a inferência filogenética é o modelo de relógio molecular, que define a distribuição a partir da qual as taxas de mutação dos ramos são estimadas. Dependendo da intensidade e forma de variação entre ramos, o mesmo pode ser classificado como modelo estrito, sob o qual todos os ramos compartilham da mesma taxa evolutiva, ou modelo relaxado, em que ramos podem possuir diferentes estimativas obtidas a partir de uma distribuição subjacente.

Por fim, um outro parâmetro de grande importância é o modelo demográfico, que mede as variações no tamanho populacional efetivo (N_e) ao longo do tempo, podendo ser categorizados em coalescentes, como os modelos de tamanho

populacional constante (paramétrico) e *Skyride* (não paramétrico), e não coalescentes, como o processo de nascimento e morte.

Tais modelos afetam a topologia geral da árvore (Figura 9) e sua estimativa também é útil para identificar a origem de pandemias e epidemias (VOLZ; KOELLE; BEDFORD, 2013).

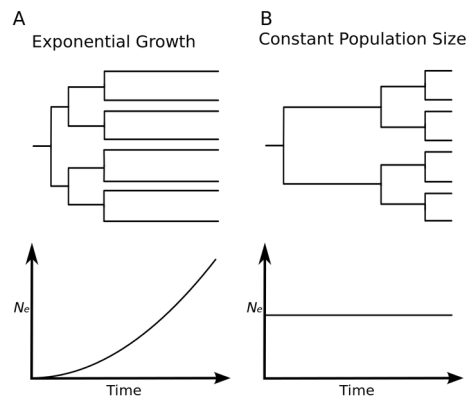


Figura 9 - Diferentes modelos populacionais e seus impactos nas filogenias (VOLZ; KOELLE; BEDFORD, 2013).

2. OBJETIVOS

2.1. Objetivo Geral

O presente Trabalho de Conclusão de Curso tem por objetivo estudar a evolução molecular e a epidemiologia do vírus Influenza A H1N1 no Brasil, utilizando como modelo o segmento genômico que codifica a neuraminidase.

2.2. Objetivos Específicos

- Comparar diferenças na taxa mutacional no tempo;
- Calcular o tamanho populacional efetivo e tempo para coalescência;
- Investigar o fluxo gênico no Brasil.

3. MATERIAIS E MÉTODOS

3.1. Aquisição de sequências

As sequências foram baixadas, junto às suas anotações, como local e data de isolamento, do banco de dados público Influenza Virus Resource (<https://www.ncbi.nlm.nih.gov/genomes/FLU/Database/nph-select.cgi?go=database>). Como critérios de inclusão, foram selecionadas apenas sequências completas de hospedeiros humanos, do segmento 6 do subtipo A H1N1 e coletadas no Brasil.

3.2. Pré-processamento dos dados

Os cabeçalhos das sequências foram padronizados de forma a incluir também local (estado) e ano de isolamento. Após isso, as sequências foram então alinhadas usando o software Clustal Ω v1.2.4 (SIEVERS *et al.*, 2011; SIEVERS; HIGGINS, 2018). Os alinhamentos foram então visualizados no software AliView v1.27 (LARSSON, 2014), permitindo a remoção de sequências de baixa qualidade, isto é, que contivessem gaps internos muito grandes ou nucleotídeos ambíguos, seguida pelo realinhamento das restantes.

As sequências restantes foram então analisadas com o software Recombination Detection Program v5.05 (RDP5) (MARTIN *et al.*, 2021) em busca de evidências de eventos de recombinação, os quais violariam os princípios necessários às simulações filodinâmicas, sendo eliminadas aquelas identificadas por pelo menos três métodos diferentes do mesmo.

Além disso, visando eliminar sequências com baixa coerência temporal (*clock-likeness*), ou seja, aquelas com pouca relação entre sua data de amostragem e diversidade relativa à raiz da árvore, foi estimada uma árvore de Máxima Verossimilhança com o software RAxML v8.2.12+dfsg-6 (STAMATAKIS, 2014), a qual foi analisada com o software TempEst v1.5.3 (RAMBAUT *et al.*, 2016).

Após essa etapa, foram removidas as regiões UTR 3' e 5' do segmento, levando-se em conta a proteína principal do mesmo, visando assim reduzir os ruídos nas estimativas do BEAST.

3.3. Análise filogenética Bayesiana

Foi utilizado o BEAST v1.10.4 (SUCHARD *et al.*, 2018) com a biblioteca BEAGLE (AYRES *et al.*, 2012) para a estimativa dos parâmetros, sendo o software BEAUTI v1.10.4 (<https://beast.community/beauti>) utilizado para elaborar os arquivos de configuração .xml do BEAST.

Em tais arquivos, foram definidas 100 milhões de iterações com 10% de burn-in devido à complexidade das sequências, com a distribuição da média do relógio molecular baseada na média estimada em corrida prévia.

Para a seleção de modelos, foram definidos após levantamento bibliográfico o relógio relaxado lognormal (DRUMMOND *et al.*, 2006) para o modelo de relógio molecular e o Coalescente (*Bayesian Skyline*) (DRUMMOND *et al.*, 2005) selecionado como modelo populacional.

Quanto à escolha do modelo de substituição nucleotídica, foi realizada uma comparação usando o software jModelTest2 (DARRIBA *et al.*, 2012), tendo como base o melhor critério de informação de Akaike corrigido para pequenas amostras (AICc). Para a estimativa dos parâmetros do modelo, foram utilizados os valores empíricos para as frequências nucleotídicas, e os parâmetros restantes (taxas de transição e transversão) foram estimados pelo BEAST.

Foram realizadas duas corridas, com duas réplicas independentes cada, uma sem e a outra contendo os locais de isolamento indicados para a elaboração da árvore filogeográfica.

3.4. Diagnóstico e análise dos dados

O diagnóstico das corridas foi feito tanto pela análise dos arquivos dos operadores de cadeia como também através do software Tracer v1.7.1 (RAMBAUT *et al.*, 2018), após o qual foi realizada a combinação dos logs e das árvores das replicatas das corridas usando o software LogCombiner v1.10.4 (<https://beast.community/logcombiner>).

Além disso, o módulo "*Bayesian Skyline Reconstruction*" do Tracer foi utilizado para a reconstrução e sumarização da dinâmica populacional do segmento. Por sua vez, a árvore de máxima credibilidade de clado (MCC) foi estimada usando

o TreeAnnotator v1.10.4 (<https://beast.community/treeannotator>) com as corridas combinadas.

Para a análise filogeográfica, foi utilizado o software SpreaD3 (BIELEJEC *et al.*, 2016) com a árvore MCC anotada e coordenadas obtidas pelo site <<https://www.geody.com/>>.

A elaboração dos gráficos desta e das outras etapas foi realizada majoritariamente pelo R (<https://www.r-project.org/>).

4. RESULTADOS E DISCUSSÃO

Ao todo, foram obtidas 211 sequências e seus metadados do banco de dados público Influenza Virus Resource. A partir dos metadados disponíveis, foi possível identificar o ano e o local de isolamento da maioria das sequências com base em sua nomenclatura, e com isso normalizar os cabeçalhos para o modelo “{ID}_{LOCAL}_{ANO}”.

Dado que diversas sequências possuíam indicado o nome da cidade de isolamento em detrimento do estado, foram realizadas buscas usando a ferramenta IBGE Cidades (<https://cidades.ibge.gov.br/>) para classificar tais locais, sendo definido como não encontrados as sequências cuja cidade de isolamento tivesse homônimos em mais de um estado, como por exemplo São Gabriel, que é encontrada em ambos os estados da Bahia e Rio Grande do Sul.

Durante a etapa de pré-processamento dos dados, foram removidas 22 sequências. Dessas, uma foi removida por mau alinhamento (KY926084_RS_2016), visto que possuía um grande gap não justificado no meio da sequência, em comparação a algumas outras sequências com gaps *in frame* (que portanto não levariam a grandes mudanças na proteína de interesse); outra sequência foi removida por evidência de recombinação (KY926065_RS_2012), conforme determinado pelo RDP; e as 20 restantes foram removidas por falha na identificação do local (estado) de isolamento, seja pela ambiguidade na cidade (10) ou pelo não fornecimento de informação a nível mais preciso que país (10).

Além disso, vale ressaltar que não foi necessária a remoção de sequências por ambiguidade de nucleotídeos ou por ausência de coerência temporal. Dessa forma, as 189 sequências restantes foram classificadas quanto ao ano e estado de isolamento conforme a Figura 10.

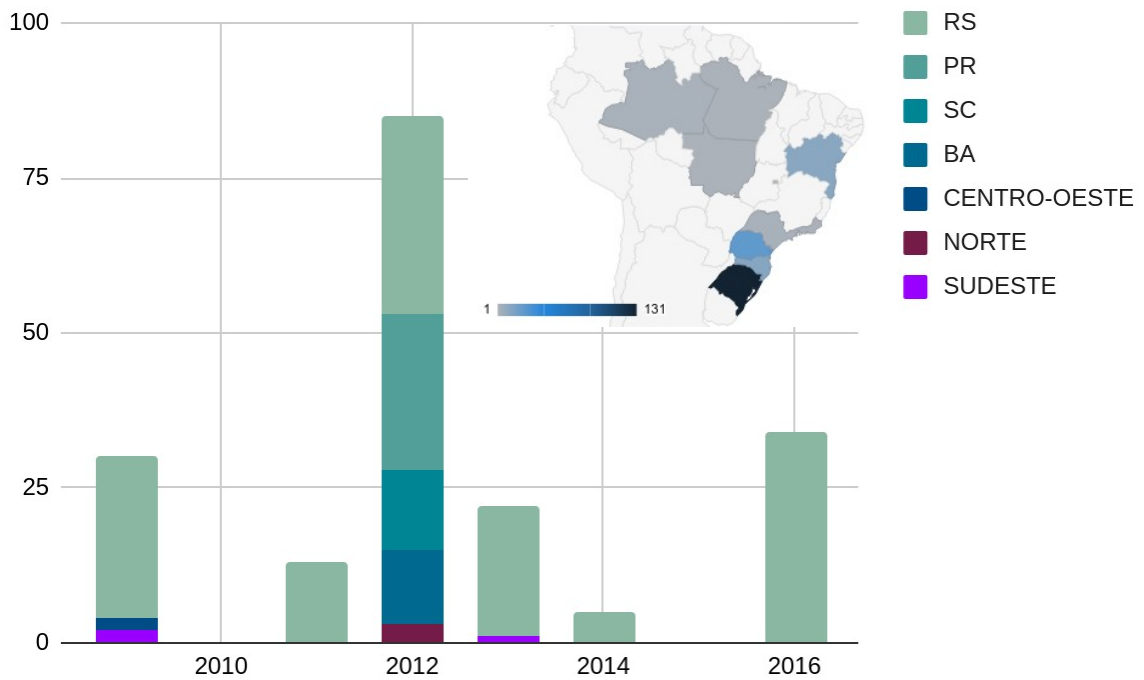


Figura 10 - Sequências obtidas por ano e local de coleta (Elaborado pelo autor). Pode-se notar uma grande representatividade da região Sul, com mais de 89% das sequências, seguida pela região Nordeste com aproximadamente 6% e as restantes disputando 4%.

Durante a etapa de seleção do modelo de substituição nucleotídica, foi identificado pelo jModelTest2 como melhor modelo o GTR+ Γ (AIC_c igual a 11648.66), seguido por GTR+ Γ +I (11652.25) e TVM+ Γ (*Transversion Model*) (11652.63).

Após as corridas, uma análise cursória dos arquivos com os operadores de cadeia mostrou que nenhum dos mesmos possuía uma probabilidade nula de aceitar a transição entre estados, demonstrando um comportamento normal das mesmas.

Na Tabela 1 estão resumidas as estimativas dos resultados da corrida contendo os parâmetros geográficos, modelo de substituição GTR+ Γ 4, relógio lognormal relaxado e modelo populacional coalescente *Bayesian Skyline*.

Parâmetro	Média	Intervalo HPD 95%	ESS
Log likelihood	-5595.39	-5623.49 a -5568.20	3786
Prior	-5051.78	-5104.55 a -4999.11	1044
Posterior	-10647.17	-10705.92 a -10587.35	1194
Taxa média de substituição (em subst/sítio/ano)	3.02×10^{-3}	2.54×10^{-3} a 3.52×10^{-3}	1230
Idade da raiz em relação ao topo da árvore (em anos)	24.86	14.62 a 38.58	149
Ano do MRCA	1991.14	1977.42 a 2001.38	149

Tabela 1 - Resultados da inferência coalescente Bayesiana (Elaborado pelo autor). MRCA equivale ao Ancestral Comum Mais Recente (Most Recent Common Ancestor)

A taxa média de substituição do segmento, 3.02×10^{-3} subst/sítio/ano, apesar de dentro dos valores esperados para um vírus de RNA (HOLMES; GRENFELL, 2009), se encontra um pouco abaixo dos valores esperados em relação à literatura, como 5.41×10^{-3} (RAMBAUT *et al.*, 2008) e 4.98×10^{-3} (GOÑI *et al.*, 2012).

Apesar do ESS dos parâmetros do modelo populacional obtidos estarem em valores razoáveis, embora possam ser considerados aceitáveis, com sete dos 10 acima de 100 e apenas quatro desses acima de 200 (um ESS superior a 200 é um indicativo de convergência da cadeia conforme o site <https://beast.community/ess_tutorial> do BEAST), foi elaborada a reconstrução da dinâmica populacional do segmento usando o Tracer, conforme a Figura 11.

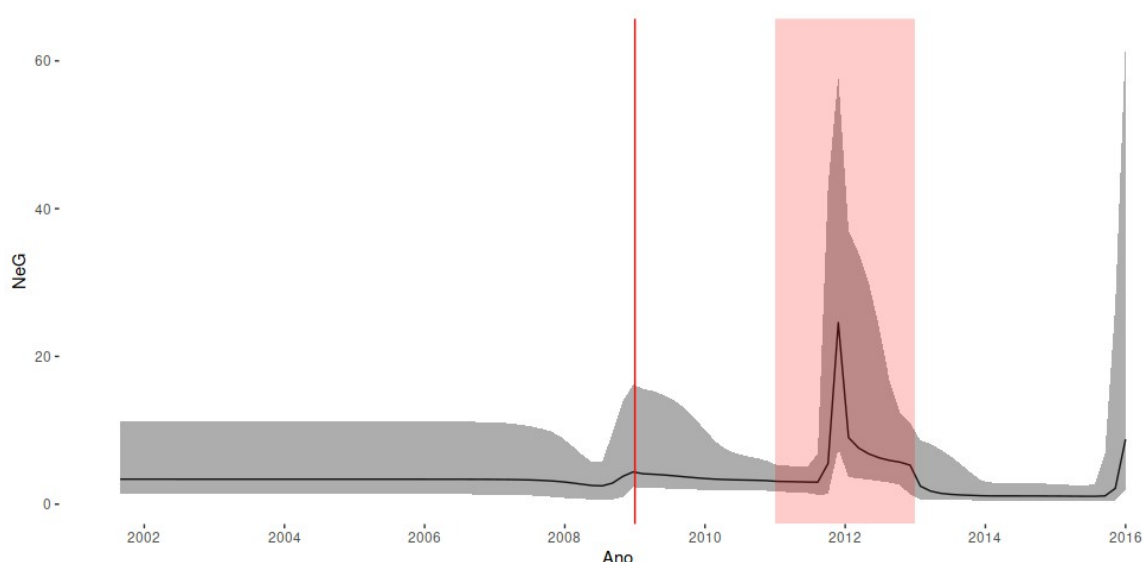


Figura 11 - Gráfico comparando o tamanho efetivo populacional (NeG) pelo tempo (elaborado pelo autor). Em vermelho estão destacados o ano de 2009 e o período de junho de 2011 a junho de 2012, durante os quais ocorreram aumento no número de casos e correspondente aumento no tamanho efetivo populacional.

Através da análise do *plot* do tamanho efetivo populacional (N_e) pelo tempo (sob o modelo coalescente *Bayesian Skyline*) (Figura 11), foi possível a identificação de aumentos na biodiversidade precedendo estações com número relativamente elevado de casos de Síndrome Respiratória Aguda Grave (SRAG) por Influenza A (Figura 12), porém sem correlação direta com tais números.

Pode-se notar três períodos indicando aumento da diversidade genética. O primeiro deles coincide com a pandemia de 2009, durante a qual a linhagem sazonal foi substituída pela A(H1N1)pdm2009.

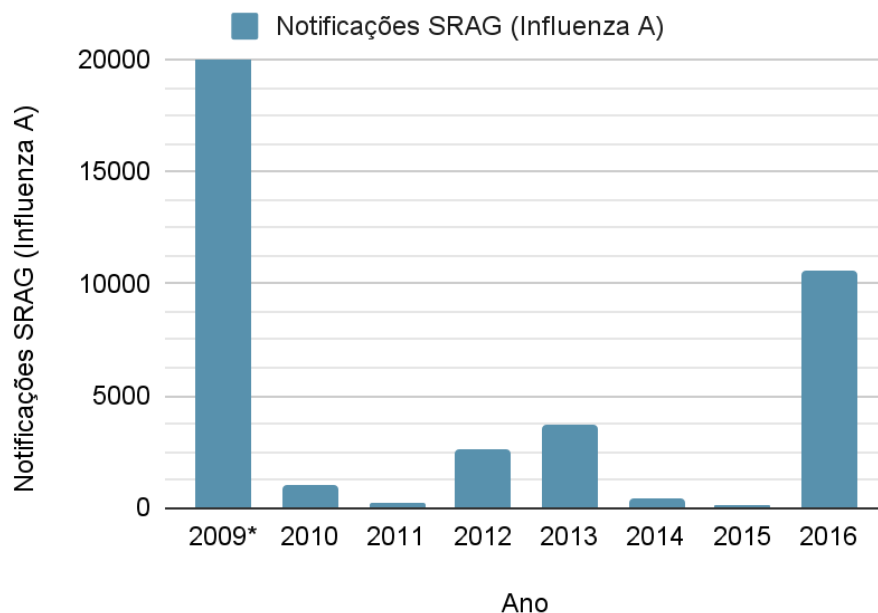


Figura 12 - Número de notificações de SRAG confirmadas para Influenza A (H1N1)pdm2009 por ano (Adaptado de BRASIL, 2012, 2013, 2014, 2015, 2016, 2017). Durante o ano de 2009, devido à pandemia, foram notificados mais de cem mil casos de SRAG por influenza, de forma que o gráfico foi truncado para melhor visualização dos outros anos.

No período de 2011-2013, houve um novo aumento no número dos casos, com o número de notificações de SRAG por Influenza A saltando de 181, em 2011 (BRASIL, 2012), para 3733 em 2013 (BRASIL, 2014) (Figura 12).

O ano de 2016, por sua vez, correspondeu a uma epidemia três meses antes de época e com o maior número de casos e óbitos desde 2009, o que se justifica pela baixa sobreposição entre a mesma e a campanha anual de vacinação. Dentre as possíveis causas para o fenômeno, são mencionados fatores como a época

gripal tardia no hemisfério norte (CDC, 2016) e/ou aumento nas viagens internacionais.

Por outro lado, pode-se notar a ausência de picos de biodiversidade em outros anos, o que pode ser devido em parte à subamostragem dos dados, visto que não foram analisadas sequências anteriores a 2008, as quais não possuíam a identificação de local de coleta a nível de estado, além de sequências dos anos 2010 e 2015, os quais não possuíam sequências durante a etapa de aquisição.

É interessante notar que os aumentos nos tamanhos efetivos populacionais precederam suas respectivas estações gripais em alguns meses, mesmo no caso da pandemia de 2009. Além disso, devido à expansão clonal da mesma, resultado da baixa quantidade de eventos de salto a partir dos suínos, a linhagem pandêmica apresentava diversidade genética menor que as linhagens sazonais (GARTEN *et al.*, 2009), o que pode ser associado a uma redução no NeG.

Além do *plot* de tamanho efetivo pelo tempo, as árvores inferidas pelo BEAST foram sumarizadas em uma árvore de máxima credibilidade de clado (MCC), a qual foi anotada com as localizações e posterior conforme a Figura 13.

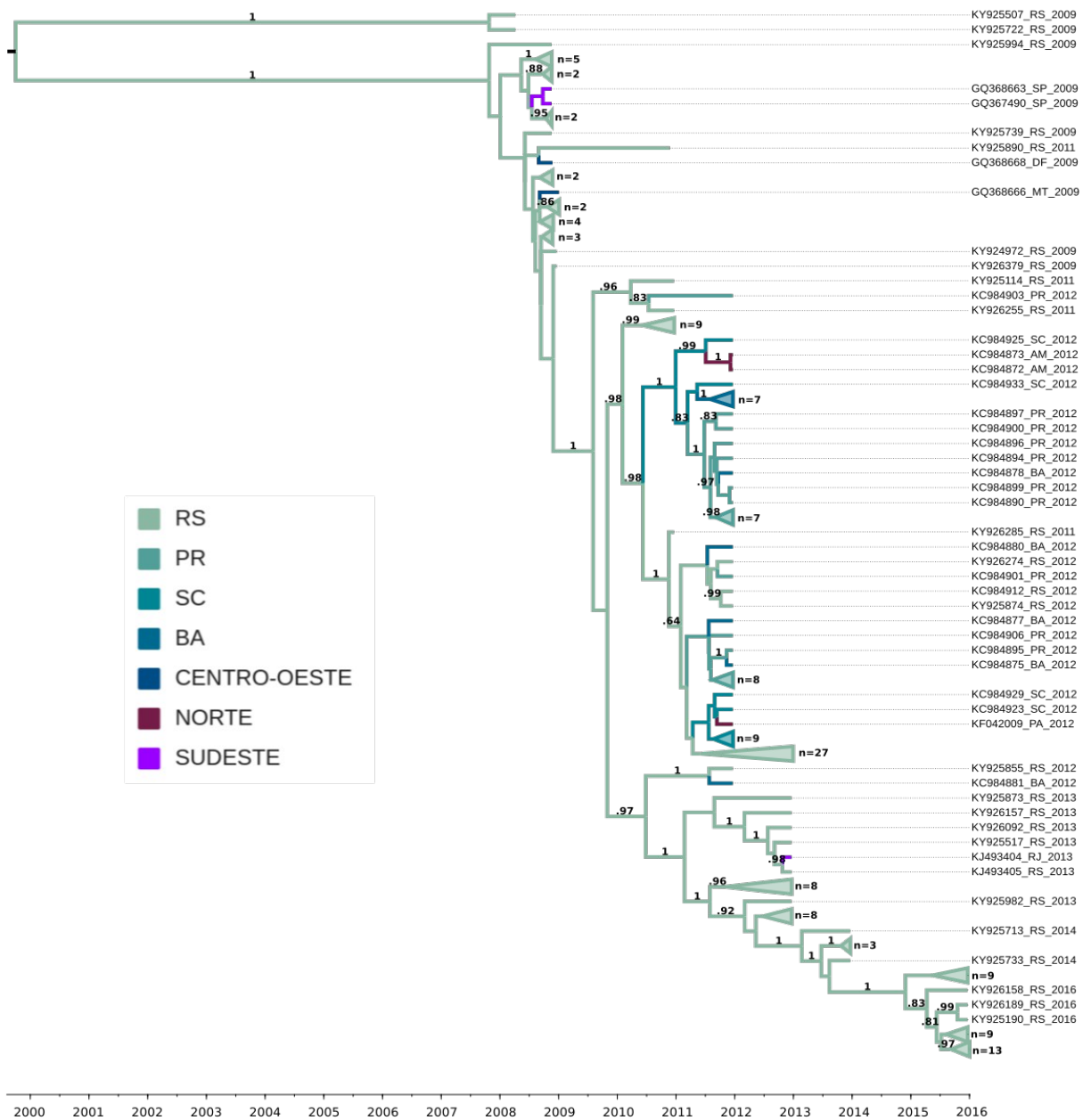


Figura 13 - Árvore filogenética baseada nas sequências nucleotídicas do segmento 6 (elaborada pelo autor). A escala temporal dos ramos é anual, sendo que os nós internos foram coloridos de acordo com a localização mais provável do ancestral dos nós descendentes. Além disso, estão exibidas as probabilidades posteriores superiores a 80%, sendo que ramos compartilhando o mesmo local de coleta e baixa probabilidade entre nós foram colapsados, com o número de nós terminais indicado.

Com base na mesma, pode-se notar a topologia característica dos vírus Influenza, isto é, um constante surgimento e extinção de linhagens, dentre as quais uma única linhagem atua como a origem das mais recentes (BUSH *et al.*, 1999). Além disso, é interessante notar que a esmagadora maioria das sequências coletadas em 2009 apresenta um ancestral comum bem próximo do início da pandemia, corroborando a hipótese de que ocorreram poucos eventos de saltos interespecíficos a partir dos suínos (GARTEN *et al.*, 2009).

Por outro lado, pode-se notar a presença de dois táxons formando um grupo externo, KY925507_RS_2009 e KY925722_RS_2009, cujo ancestral comum mais recente (MRCA) com a raiz da árvore ocorreu em 1995. Uma possível explicação para o ocorrido é tais sequências sejam de uma das linhagens sazonais, dado que a linhagem pandêmica circulava em porcos desde, pelo menos, 1992 (TRIFONOV; KHIABANIAN; RABADAN, 2009), estando presente no Brasil em 2009 (RAJÃO *et al.*, 2013), e portanto possuía baixa similaridade em relação às mesmas.

Vale ressaltar também que, conforme esperado para os vírus Influenza, o tempo até o MRCA de linhagens contemporâneas está sempre dentro do intervalo de 1-3 anos, que, além de ser a expectativa de vida média de uma linhagem qualquer (BUSH *et al.*, 1999), é superior à duração de cada época gripal. Tal fato pode ser relacionado à presença de uma reserva dessas linhagens em hospedeiros não humanos, que permitem a manutenção da biodiversidade mesmo após extinções locais ao final de uma estação (RAMBAUT *et al.*, 2008).

Com base na árvore MCC, também foram elaboradas uma animação com o software Spread3 e um mapa das principais migrações (Figura 14), permitindo a identificação de dois grandes eventos migratórios simultâneos das sequências, durante os quais diversas migrações ocorreram em um curto intervalo de tempo.

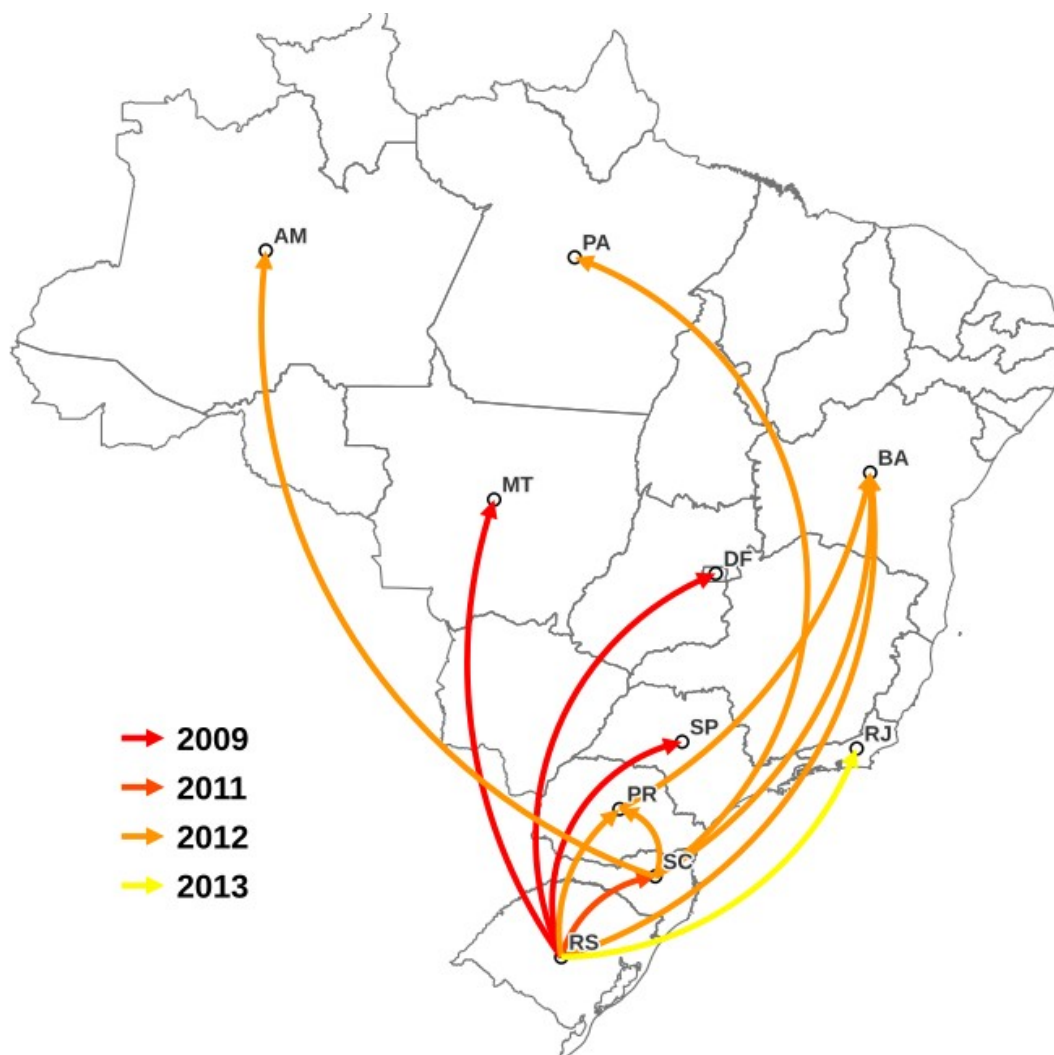


Figura 14 - Mapa do Brasil com as quatro principais migrações do vírus entre 1995 e 2016 (elaborado pelo autor). O mapa foi elaborado a partir do shapefile das fronteiras estaduais do país, disponível em <<https://nereus.webhostusp.sti.usp.br/?dados=brasil>>, em conjunto com os dados da animação do SpreaD3. As setas indicam a mudança da localização mais provável de um nó ancestral para a de seu descendente, sendo que as se concentraram no primeiro semestre de cada ano.

Dentre as migrações identificadas no período, pode-se notar como fator em comum o ancestral na região do Rio Grande do Sul. Essa região, que serviu como fonte da migração viral para Santa Catarina no primeiro semestre de 2011, também foi a origem na migração ao Rio de Janeiro no primeiro semestre de 2013.

Quanto aos eventos migratórios simultâneos, o primeiro é associado ao início de 2009, pouco antes da declaração, em junho, do início da pandemia pela Organização Mundial da Saúde, porém meses após a primeira identificação da doença (OMS, 2009). A outro evento, por sua vez, ocorreu no período entre outubro

de 2011 e maio de 2012, isto é, antes da época gripal no país (maio a outubro do mesmo ano) (SÃO PAULO, 2013; SISTEMA DE INFORMAÇÕES DE AGRAVOS DE NOTIFICAÇÃO, 2016). Esse, diferentemente dos outros eventos, partiu de ambos Rio Grande do Sul e Santa Catarina rumo ao Paraná e às regiões Norte e Nordeste.

Um outro contexto importante, relativo à atuação da região Sul como origem das linhagens durante a pandemia de 2009, é sua concentração do números de casos e óbitos, especialmente no início da mesma (Figura 15) (SCHUCK-PAIM *et al.*, 2012). Além disso, dentre os vírus com semelhante via de infecção, como o vírus sincicial respiratório ou o adenovírus, nota-se um diferente padrão na dispersão de casos e óbitos, com picos temporalmente mais achatados e melhor distribuídos geograficamente (BRASIL, 2014, 2015).

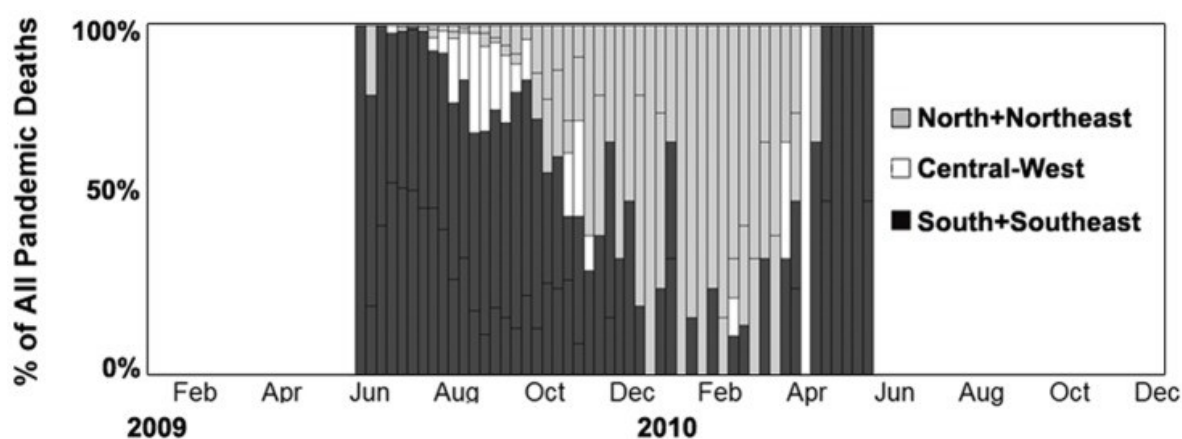


Figura 15 - Porcentagens relativas das mortes (confirmadas por laboratório) por semana por região durante a pandemia. (SCHUCK-PAIM *et al.*, 2012). Pode-se notar o domínio das regiões Sul e Sudeste no início e final da pandemia, sugerindo uma região de origem para a linhagem pandêmica no país.

Outro fator relevante é a direcionalidade do fluxo gênico, que é distinta de outros vírus prevalentes na região, como o HIV-1 subtipo C, que apresenta um fluxo bidirecional em que o RS atua como região importadora (Figura 16) (DELATORRE *et al.*, 2013); e mesmo o Influenza A H3N2, que apesar da maior participação da região Sul na rede migratória, não apresenta direção preferencial durante uma época gripal (Figura 17) (BORN *et al.*, 2016).

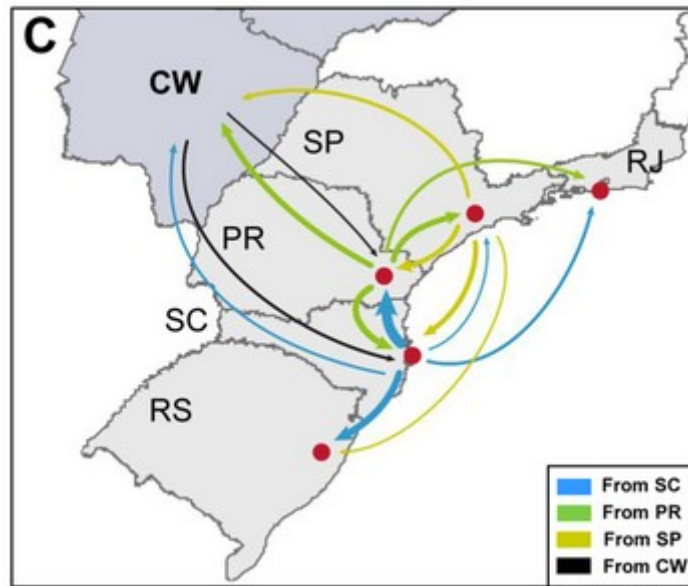


Figura 16 - Taxas de migração do vírus HIV-1 subtipo C entre regiões após sua introdução no país na década de 1970 (DELATORRE et al., 2013). A largura das setas é proporcional à taxa de transição viral entre regiões, com grande participação de Santa Catarina (SC) e Paraná como fontes.



Figura 17 - Taxas de migração do vírus Influenza A H3N2 entre regiões entre 1999 e 2012 (BORN et al., 2016). A largura das setas é proporcional ao número médio estimado de transições virais entre regiões, com elevada participação das regiões Sul (BR-S) e Sudeste (BR-SE) em detrimento do Nordeste (BR-NE).

6. CONCLUSÃO

A gripe sazonal é uma doença de manifestação anual, cuja manutenção entre estações se deve tanto ao escape imune dos vírus Influenza A e B quanto ao suprimento de reservatórios animais, principalmente suínos e aves aquáticas. No caso dos vírus Influenza A, os subtipos H1N1 e N3N2 são os mais comuns, sendo responsáveis pela maioria dos casos no Brasil e no mundo.

Apesar disso, sabe-se que ambos os subtipos apresentam padrões diferentes de dispersão, com interações complexas entre *drift* e *shift* antigênicos, além da relevância dos *selective sweeps*, que levam à extinção de linhagens locais entre temporadas (RAMBAUT *et al.*, 2008).

Tais interações e padrões conseguem se manifestar diferencialmente tanto nas árvores filogenéticas quanto nos *plots* de tamanho populacional pelo tempo, permitindo a análise filodinâmica das sequências virais.

Com base no presente trabalho, pode-se confirmar, além das elevadas taxas mutacionais, características de vírus de RNA, variações no tamanho populacional efetivo e fluxo gênico específicos aos vírus Influenza, permitindo sua correlação com épocas gripais particularmente intensas. Nesse contexto, a identificação da região Sul como semeadora de linhagens, em especial sua intensa participação durante a pandemia de 2009, é um fator marcante para a biodiversidade observada e sua correlação com a expansão clonal em reação ao salto a partir dos suínos.

Além disso, pode-se notar outros padrões característicos das filogenias de H1N1, como a baixa longevidade e manutenção da biodiversidade entre épocas gripais. Tais padrões podem ser explicados pela seleção imune negativa, extinção das linhagens locais e fornecimento a partir de espécies reservatório não humanas, como aves aquáticas e suínos.

Apesar de não incluírem o impacto do *shift* antigênico, dado que foi estudado apenas um segmento genômico em isolamento, o conjunto desses resultados fornecem importantes *insights* quanto à epidemiologia e parâmetros evolutivos da gripe sazonal por Influenza A H1N1 no Brasil.

7. PERSPECTIVAS FUTURAS

No atual contexto pandêmico, tornou-se evidente a importância dos serviços de monitoramento epidemiológico, em especial do monitoramento da evolução genética de vírus de tão rápida mutação como os vírus de RNA. Esse monitoramento é especialmente importante para os vírus Influenza, posto que temos exemplos não só de pandemias devastadoras como também de linhagens em circulação com grande potencial pandêmico, como a HPAIV H5N1, dentre outras (CDC, 2019).

Além disso, dadas as alterações nos padrões mundiais de movimentação nos últimos anos, podem ter ocorrido mudanças na epidemiologia e evolução dos vírus Influenza que são mais prontamente observadas através de análises filodinâmicas e filogeográficas, como é o caso da possível extinção da linhagem de Influenza B/Yamagata (KOUTSAKOS *et al.*, 2021) ou mesmo a epidemia fora de época de H3N2 (GOMES, 2021).

Outro fator importante para os padrões de movimentação da Influenza A é justamente a migração de aves que atuam como reservatório da mesma, o que pode facilitar o surgimento de pandemias ou acelerar a dispersão de epidemias (XU *et al.*, 2016). Dessa forma, uma análise que correlacione tais padrões migratórios e a filogeografia viral na região, tanto de linhagens de aves quanto humanas, seria muito interessante, visto que a mesma poderia auxiliar na identificação de eventos de emergência viral.

No Brasil, apesar da criação do Sistema de Informação de Agravos de Notificação (SINAN) em 1993, foi apenas em 2006 que se tornou compulsória a notificação de casos de Influenza (BRASIL, 2006). Dessa forma, espera-se que o acúmulo de dados genéticos e epidemiológicos permita um maior subsídio à tomada de decisões no que tange o combate e a prevenção de epidemias de gripe (WEBSTER; GOVORKOVA, 2014).

REFERÊNCIAS

AIR, G. M. Influenza neuraminidase. **Influenza and other Respiratory Viruses**, [s. l.], v. 6, n. 4, p. 245–256, 2012. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/22085243>.

AYRES, D. L. *et al.* BEAGLE: An application programming interface and high-performance computing library for statistical phylogenetics. **Systematic Biology**, [s. l.], v. 61, n. 1, p. 170–173, 2012. Disponível em: <https://doi.org/10.1093/sysbio/syr100>.

BIELEJEC, F. *et al.* Spread3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes. **Molecular Biology and Evolution**, [s. l.], v. 33, n. 8, p. 2167–2169, 2016. Disponível em: <https://doi.org/10.1093/molbev/msw082>.

BORN, P. S. *et al.* Phylodynamics of influenza A(H3N2) in South America, 1999–2012. **Infection, Genetics and Evolution**, [s. l.], v. 43, p. 312–320, 2016. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1567134816302362>.

BRASIL. MINISTÉRIO DA SAÚDE. SECRETARIA DE VIGILÂNCIA EM SAÚDE. **Informe epidemiológico sobre influenza (2009 a 2011)**. [S. l.], 2012. Disponível em: https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/g/gripe-influenza-1/arquivos/informe-tecnico-de-influenza_2009-2010-2011_janeiro-2012.pdf/view. Acesso em: 21 mar. 2022.

BRASIL. MINISTÉRIO DA SAÚDE. SECRETARIA DE VIGILÂNCIA EM SAÚDE. **Informe epidemiológico sobre influenza (2012)**. [S. l.], 2013. Disponível em: https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/g/gripe-influenza-1/arquivos/boletim-informativo_influenza_se52-de-2012.pdf/view. Acesso em: 21 mar. 2022.

BRASIL. MINISTÉRIO DA SAÚDE. SECRETARIA DE VIGILÂNCIA EM SAÚDE. **Informe epidemiológico sobre influenza (2013)**. [S. l.], 2014. Disponível em: https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/g/gripe-influenza-1/arquivos/boletim-epidemiologico_influenza_se52_2013.pdf/view. Acesso em: 21 mar. 2022.

BRASIL. MINISTÉRIO DA SAÚDE. SECRETARIA DE VIGILÂNCIA EM SAÚDE. **Informe epidemiológico sobre influenza (2014)**. [S. l.], 2015. Disponível em: https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/g/gripe-influenza-1/arquivos/boletim-epidemiologico_influenza_se53_2014.pdf/view. Acesso em: 21 mar. 2022.

BRASIL. MINISTÉRIO DA SAÚDE. SECRETARIA DE VIGILÂNCIA EM SAÚDE. **Informe epidemiológico sobre influenza (2015)**. [S. l.], 2016. Disponível em: https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/g/gripe-influenza-1/arquivos/boletim-epidemiologico_influenza-se52-2015.pdf/view. Acesso em: 21 mar. 2022.

BRASIL. MINISTÉRIO DA SAÚDE. SECRETARIA DE VIGILÂNCIA EM SAÚDE. **Informe epidemiológico sobre influenza (2016)**. [S. l.], 2017. Disponível em: https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/g/gripe-influenza-1/arquivos/informe-epidemiologico_influenza-2016-se-52.pdf/view. Acesso em: 21 mar. 2022.

BRASIL. MINISTÉRIO DA SAÚDE. SECRETARIA DE VIGILÂNCIA EM SAÚDE. **Informe epidemiológico sobre influenza (2019)**. [S. l.], 2020. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/g/gripe-influenza-1/arquivos/informe-epidemiologico-sobre-influenza-2019.pdf>. Acesso em: 21 mar. 2022.

BRASIL. MINISTÉRIO DA SAÚDE. SECRETARIA DE VIGILÂNCIA EM SAÚDE. **Portaria nº 5, de 21 de fevereiro de 2006**. 2006. Disponível em: https://bvsms.saude.gov.br/bvs/saudelegis/svs/2006/prt0005_21_02_2006_comp.html.

BUSH, R. M. *et al.* Predicting the evolution of human influenza A. **Science**, [s. l.], v. 286, n. 5446, p. 1921–1925, 1999. Disponível em: <https://www.science.org/doi/10.1126/science.286.5446.1921>.

CENTERS FOR DISEASE CONTROL AND PREVENTION. **Summary of the 2015-2016 Influenza Season**. [S. l.], 2016. Disponível em: <https://www.cdc.gov/flu/about/season/flu-season-2015-2016.htm>.

CENTERS FOR DISEASE CONTROL AND PREVENTION. **Types of Influenza Viruses**. [S. l.], 2021. Disponível em: <http://www.cdc.gov/flu/about/viruses/types.htm>.

CENTERS FOR DISEASE CONTROL AND PREVENTION. **Viruses of Special Concern**. [S. l.], 2019. Disponível em: <https://www.cdc.gov/flu/pandemic-resources/monitoring/viruses-concern.html>.

CHAO, D. L. *et al.* The global spread of drug-resistant influenza. **Journal of the Royal Society Interface**, [s. l.], v. 9, n. 69, p. 648–656, 2012. Disponível em: <https://doi.org/10.1098/rsif.2011.0427>.

DARRIBA, D. *et al.* JModelTest 2: More models, new heuristics and parallel computing. **Nature Methods**, [s. l.], v. 9, n. 8, p. 772, 2012. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/22847109>.

DAWOOD, F. S. *et al.* Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study. **The Lancet Infectious Diseases**, [s. l.], v. 12, n. 9, p. 687–695, 2012. Disponível em: [https://doi.org/10.1016/S1473-3099\(12\)70121-4](https://doi.org/10.1016/S1473-3099(12)70121-4).

DELATORRE, E. *et al.* Tracing the Origin and Northward Dissemination Dynamics of HIV-1 Subtype C in Brazil. **PLoS ONE**, [s. l.], v. 8, n. 9, p. e74072, 2013.

DOU, D. *et al.* Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement. **Frontiers in immunology**, [s. l.], v. 9, n. JUL, p. 1581, 2018. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/30079062>.

DRUMMOND, A. J. *et al.* Bayesian coalescent inference of past population dynamics from molecular sequences. **Molecular Biology and Evolution**, [s. l.], v. 22, n. 5, p. 1185–1192, 2005. Disponível em: <https://doi.org/10.1093/molbev/msi103>.

DRUMMOND, A. J. *et al.* Relaxed phylogenetics and dating with confidence. **PLoS Biology**, [s. l.], v. 4, n. 5, p. 699–710, 2006. Disponível em: <https://doi.org/10.1371/journal.pbio.0040088>.

GARTEN, R. J. *et al.* Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. **Science**, [s. l.], v. 325, n. 5937, p. 197–201, 2009. Disponível em: <http://science.sciencemag.org/content/325/5937/197.abstract>.

GOMES, M. H3N2 Darwin: saiba mais sobre o tipo do vírus influenza em circulação no país. **Instituto Oswaldo Cruz**, [s. l.], 2021. Disponível em: <https://portal.fiocruz.br/noticia/h3n2-darwin-saiba-mais-sobre-o-tipo-do-virus-influenza-em-circulacao-no-pais>.

GOÑI, N. *et al.* Bayesian coalescent analysis of pandemic H1N1 influenza A virus circulating in the South American region. **Virus Research**, [s. l.], v. 170, n. 1–2, p. 91–101, 2012. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0168170212003206>.

GRENFELL, B. T. *et al.* Unifying the Epidemiological and Evolutionary Dynamics of Pathogens. **Science**, [s. l.], v. 303, n. 5656, p. 327–332, 2004. Disponível em: <http://science.sciencemag.org/content/303/5656/327.abstract>.

HOLMES, E. C. The phylogeography of human viruses. **Molecular Ecology**, [s. l.], v. 13, n. 4, p. 745–756, 2004. Disponível em: <https://doi.org/10.1046/j.1365-294X.2003.02051.x>.

HOLMES, E. C.; GRENFELL, B. T. Discovering the phylodynamics of RNA viruses. **PLoS Computational Biology**, [s. l.], v. 5, n. 10, p. e1000505, 2009. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/19855824>.

IULIANO, A. D. *et al.* Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. **The Lancet**, [s. l.], v. 391, n. 10127, p. 1285–1300, 2018. Disponível em: [https://doi.org/10.1016/S0140-6736\(17\)33293-2](https://doi.org/10.1016/S0140-6736(17)33293-2).

KILBOURNE, E. D. Viral Structure and Composition. *In*: KILBOURNE, E. D. (org.). **Influenza**. Boston, MA: Springer US, 1987. p. 33–56. *E-book*. Disponível em: https://doi.org/10.1007/978-1-4684-5239-6_3.

KOUTSAKOS, M. *et al.* Influenza lineage extinction during the COVID-19 pandemic?. **Nature Reviews Microbiology**, [s. l.], v. 19, n. 12, p. 741–742, 2021. Disponível em: <https://doi.org/10.1038/s41579-021-00642-4>.

LARSSON, A. AliView: A fast and lightweight alignment viewer and editor for large datasets. **Bioinformatics**, [s. l.], v. 30, n. 22, p. 3276–3278, 2014. Disponível em: <https://doi.org/10.1093/bioinformatics/btu531>.

LONG, J. S. *et al.* Host and viral determinants of influenza A virus species specificity. **Nature Reviews Microbiology**, [s. l.], v. 17, n. 2, p. 67–81, 2019. Disponível em: <https://doi.org/10.1038/s41579-018-0115-z>.

MA, W.; KAHN, R. E.; RICHT, J. A. The pig as a mixing vessel for influenza viruses: Human and veterinary implications. **Journal of Molecular and Genetic Medicine**, [s. l.], v. 03, n. 01, p. 158–166, 2009. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/19565018>.

MARTIN, D. P. *et al.* RDP5: A computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. **Virus Evolution**, [s. l.], v. 7, n. 1, p. veaa087, 2021. Disponível em: <https://doi.org/10.1093/ve/veaa087>.

MCAULEY, J. L. *et al.* Influenza Virus Neuraminidase Structure and Functions. **Frontiers in microbiology**, [s. l.], v. 10, p. 39, 2019. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/30761095>.

MCCAULEY, J. *et al.* **ICTV 9th Report (2011): Orthomyxoviridae**. [S. l.], 2011. Disponível em: https://talk.ictvonline.org/ictv-reports/ictv_9th_report/negative-sense-rna-viruses-2011/w/negrna_viruses/209/orthomyxoviridae. Acesso em: 24 nov. 2021.

ORGANIZAÇÃO MUNDIAL DA SAÚDE. **Pandemic Influenza Risk Management: WHO Interim Guidance**. [S. l.: s. n.], 2013. Disponível em: https://www.who.int/influenza/preparedness/pandemic/influenza_risk_management/en/. Acesso em: 17 mar. 2022.

ORGANIZAÇÃO MUNDIAL DA SAÚDE. **Statements 2009**. [S. l.], 2009. Disponível em: <https://web.archive.org/web/20091015120058/http://www.who.int/mediacentre/news/statements/2009/en/>. Acesso em: 21 mar. 2021.

PYBUS, O. G. *et al.* The epidemiology and iatrogenic transmission of hepatitis C virus in egypt: A Bayesian coalescent approach. **Molecular Biology and Evolution**, United States, v. 20, n. 3, p. 381–387, 2003. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/12644558>.

PYBUS, O. G.; RAMBAUT, A. Evolutionary analysis of the dynamics of viral infectious disease. **Nature Reviews Genetics**, [s. l.], v. 10, n. 8, p. 540–550, 2009. Disponível em: <https://doi.org/10.1038/nrg2583>.

RAJÃO, D. S. *et al.* Genetic characterization of influenza virus circulating in Brazilian pigs during 2009 and 2010 reveals a high prevalence of the pandemic H1N1 subtype. **Influenza and other Respiratory Viruses**, [s. l.], v. 7, n. 5, p. 783–790, 2013. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/irv.12072>.

RAMBAUT, A. *et al.* Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). **Virus Evolution**, [s. l.], v. 2, n. 1, p. vew007, 2016. Disponível em: <https://doi.org/10.1093/ve/vew007>.

RAMBAUT, A. *et al.* Posterior summarization in Bayesian phylogenetics using Tracer 1.7. **Systematic Biology**, [s. l.], v. 67, n. 5, p. 901–904, 2018. Disponível em: <https://doi.org/10.1093/sysbio/syy032>.

RAMBAUT, A. *et al.* The genomic and epidemiological dynamics of human influenza A virus. **Nature**, [s. l.], v. 453, n. 7195, p. 615–619, 2008. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/18418375>. Acesso em: 16 maio 2021.

RUSSELL, C. A. *et al.* The potential for respiratory droplet-transmissible A/H5N1 influenza virus to evolve in a mammalian host. **Science**, [s. l.], v. 336, n. 6088, p. 1541–1547, 2012. Disponível em: <https://www.science.org/doi/10.1126/science.1222526>.

SÃO PAULO. SECRETARIA DE ESTADO DA SAÚDE. COORDENADORIA DE CONTROLE DE DOENÇAS. CENTRO DE VIGILÂNCIA EPIDEMIOLÓGICA. DIVISÃO DE DOENÇAS DE TRANSMISSÃO RESPIRATÓRIA. **Informe Técnico: Influenza ESP**. [s. l.], 2013. Disponível em: https://www.saude.sp.gov.br/resources/cve-centro-de-vigilancia-epidemiologica/areas-de-vigilancia/doencas-de-transmissao-respiratoria/influenza-ah1n1/pdf/if13_influ_11marco.pdf. Acesso em: 5 mar. 2022.

SCHUCK-PAIM, C. *et al.* Were Equatorial Regions Less Affected by the 2009 Influenza Pandemic? The Brazilian Experience. **PLOS ONE**, [s. l.], v. 7, n. 8, p. e41918, 2012. Disponível em: <https://doi.org/10.1371/journal.pone.0041918>.

SIEVERS, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. **Molecular Systems Biology**, [s. l.], v. 7, n. 1, p. 539, 2011. Disponível em: <https://doi.org/10.1038/msb.2011.75>.

SIEVERS, F.; HIGGINS, D. G. Clustal Omega for making accurate alignments of many protein sequences. **Protein Science**, [s. l.], v. 27, n. 1, p. 135–145, 2018. Disponível em: <https://doi.org/10.1002/pro.3290>.

SIMON-LORIERE, E.; HOLMES, E. C. Why do RNA viruses recombine?. **Nature Reviews Microbiology**, [s. l.], v. 9, n. 8, p. 617–626, 2011. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/21725337>.

SISTEMA DE INFORMAÇÕES DE AGRAVOS DE NOTIFICAÇÃO. **Calendário Epidemiológico 2012**. [S. l.], 2016. Disponível em: <http://portalsinan.saude.gov.br/calendario-epidemiologico-2012>. Acesso em: 21 mar. 2022.

SPREEUWENBERG, P.; KRONEMAN, M.; PAGET, J. Reassessing the global mortality burden of the 1918 influenza pandemic. **American Journal of Epidemiology**, [s. l.], v. 187, n. 12, p. 2561–2567, 2018.

STAMATAKIS, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. **Bioinformatics**, [s. l.], v. 30, n. 9, p. 1312–1313, 2014. Disponível em: <https://doi.org/10.1093/bioinformatics/btu033>.

SUCHARD, M. A. *et al.* Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. **Virus Evolution**, [s. l.], v. 4, n. 1, p. vey016, 2018. Disponível em: <https://doi.org/10.1093/ve/vey016>.

TAUBENBERGER, J. K.; MORENS, D. M. Influenza: the once and future pandemic. **Public health reports (Washington, D.C. : 1974)**, [s. l.], v. 125 Suppl, n. Suppl 3, p. 16–26, 2010. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/20568566>.

TAVARÉ, S. Some probabilistic and statistical problems in the analysis of DNA sequences. **American Mathematical Society**, Providence, R.I., v. 17, p. 57–86, 1986.

TRIFONOV, V.; KHIABANIAN, H.; RABADAN, R. Geographic Dependence, Surveillance, and Origins of the 2009 Influenza A (H1N1) Virus. **New England Journal of Medicine**, [s. l.], v. 361, n. 2, p. 115–119, 2009. Disponível em: <https://doi.org/10.1056/NEJMp0904572>.

VIJAYKRISHNA, D.; MUKERJI, R.; SMITH, G. J. D. RNA Virus Reassortment: An Evolutionary Mechanism for Host Jumps and Immune Evasion. **PLoS Pathogens**, [s. l.], v. 11, n. 7, p. e1004902, 2015. Disponível em: <https://doi.org/10.1371/journal.ppat.1004902>.

VIRALZONE. **Alphainfluenzavirus**. [S. l.], 2021. Disponível em: https://viralzone.expasy.org/6?outline=all_by_species. Acesso em: 24 nov. 2021.

VIRALZONE. **Ribosomal frameshifting**. [S. l.], [s. d.]. Disponível em: <https://viralzone.expasy.org/860>. Acesso em: 24 nov. 2021.

VOLZ, E. M.; KOELLE, K.; BEDFORD, T. Viral Phylodynamics. **PLoS Computational Biology**, [s. l.], v. 9, n. 3, p. e1002947, 2013. Disponível em: <https://doi.org/10.1371/journal.pcbi.1002947>.

WEBSTER, R. G.; GOVORKOVA, E. A. Continuing challenges in influenza. **Annals of the New York Academy of Sciences**, [s. l.], v. 1323, n. 1, p. 115–139, 2014. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/24891213>.

WONG, S. S.; WEBBY, R. J. Traditional and new influenza vaccines. **Clinical Microbiology Reviews**, [s. l.], v. 26, n. 3, p. 476–492, 2013. Disponível em: <http://cmr.asm.org/content/26/3/476.abstract>.

XU, Y. *et al.* Southward autumn migration of waterfowl facilitates cross-continental transmission of the highly pathogenic avian influenza H5N1 virus. **Scientific Reports**, [s. l.], v. 6, n. 1, p. 30262, 2016. Disponível em: <https://doi.org/10.1038/srep30262>.

YAMAYOSHI, S. *et al.* Identification of a Novel Viral Protein Expressed from the PB2 Segment of Influenza A Virus. **Journal of Virology**, [s. l.], v. 90, n. 1, p. 444–456, 2016. Disponível em: <https://jvi.asm.org/content/90/1/444>.

YEWDELL, J. W.; INCE, W. L. Frameshifting to PA-X influenza. **Science**, [s. l.], v. 337, n. 6091, p. 164–165, 2012. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/22798590>.