Sequenciamento genômico do SARS-CoV-2

Guia de implementação para máximo impacto na saúde pública

8 de janeiro de 2021

OPAS
Organização Pan-Americana da Saúde
Organização Mundial da Saúde
Escritório Regional para as Américas

Sequenciamento genômico do SARS-CoV-2

Guia de implementação para máximo impacto na saúde pública

8 de janeiro de 2021

OPAS
Organização Pan-Americana da Saúde
Organização Mundial da Saúde
ESCRITÓRIO REGIONAL PARA AS Américas

Versão oficial em português da obra original em Inglês
Genomic sequencing of SARS-CoV-2: a guide to implementation for maximum impact on public health
© World Health Organization 2021
ISBN 978-92-4-001844-0 (electronic version)

Sequenciamento genômico do SARS-CoV-2. Guia de implementação para máximo impacto na saúde pública. 8 de janeiro de 2021

ISBN: 978-92-75-72388-3 (impresso)
ISBN: 978-92-75-72389-0 (pdf)

Referência bibliográfica sugerida. Sequenciamento genômico do SARS-CoV-2. Guia de implementação para máximo impacto na saúde pública. 8 de janeiro de 2021. Brasília, D.F.: Organização Pan-Americana da Saúde; 2021. Licença: CC BY-NC-SA 3.0 IGO. https://doi.org/10.37774/9789275723890.

Dados da catalogação na fonte (CIP). Os dados da CIP estão disponíveis em http://iris.paho.org.

BRA/2021

Sumário

Prefácio

O ano de 2020 foi um marco na história e na saúde global. A pandemia da COVID-19 destacou o potencial que as doenças com tendência epidêmica mortal têm para dominar nosso mundo globalizado. Aprendemos uma dura lição sobre a vulnerabilidade intrínseca de nossas sociedades a um único patógeno.

Embora a COVID-19 tenha suscitado uma tragédia incalculável, também mostrou como a ciência pode reagir quando desafiada por uma emergência global massiva. Em suma, a pandemia abriu grandes oportunidades científicas que foram bem aproveitadas. Uma revolução tecnológica, que ocorreu na última década, proveu várias capacidades novas para uma resposta à pandemia. O desenvolvimento de vacinas na velocidade da luz é uma delas. O sequenciamento genômico é outra.

O sequenciamento permitiu que o mundo identificasse rapidamente o SARS-CoV-2; e o conhecimento da sequência genômica permitiu o rápido desenvolvimento de testes de diagnóstico e outras ferramentas para a resposta. O sequenciamento contínuo do genoma dá apoio ao monitoramento da propagação da doença e da atividade e evolução do vírus.

A pandemia da COVID-19 ainda está em andamento e novas variantes virais estão surgindo. A resposta global terá que ser continuada no futuro previsível. O progresso feito desde o início da pandemia com o uso do sequenciamento genômico pode ser consolidado e expandido para novas situações e novos usos.

À medida que mais países se movem para implementar programas de sequenciamento, haverá mais oportunidades para entender melhor o mundo dos patógenos emergentes e suas interações com seres humanos e animais em uma variedade de climas, ecossistemas, culturas, estilos de vida e biomas. Esse conhecimento moldará uma nova visão do mundo e abrirá novos paradigmas na prevenção e controle de epidemias e pandemias.

O aumento da urbanização e da mobilidade humana estão proporcionando as condições para futuras epidemias e pandemias. A integração acelerada do sequenciamento genômico nas práticas da comunidade global de saúde é uma necessidade, se quisermos estar mais bem preparados para as ameaças futuras. Esperamos que esta orientação ajude a pavimentar o caminho para essa preparação.

Sylvie Briand, Diretora

Programa Mundial de Preparação Global para Riscos Infecciosos e Emergências de Saúde da Organização Mundial da Saúde

Agradecimentos

Este guia de implementação foi desenvolvido em consulta com especialistas com experiência em vários campos de sequenciamento genômico da Aliança Laboratorial Global para Patógenos de Alta Ameaça (GLAD-HP), dos laboratórios de referência da OMS que fornecem testes de confirmação para COVID-19 e da Rede Global de Alerta e Resposta a Surtos (GOARN). Após discussões iniciais por um grupo de redação técnica liderado por um consultor temporário e membros da Equipe Laboratorial para COVID-19 da OMS, foram solicitadas contribuições de outros especialistas dentro e fora da OMS, e duas reuniões online foram realizadas para resolver questões pendentes. As sugestões de melhorias e correções que podem ser incorporadas em uma segunda edição deste guia devem ser enviadas para WHElab@who.int.

Grupo líder de redação e edição

Sarah C. Hill, Royal Veterinary College, Londres, e Universidade de Oxford, Oxford, Reino Unido
Mark Perkins, Doenças e Zoonoses Emergentes, Programa de Emergências de Saúde, OMS, Genebra, Suíça
Karin J. von Eije, Doenças e Zoonoses Emergentes, Programa de Emergências de Saúde, OMS, Genebra, Suíça

Grupo de redação

Kim Benschop, Instituto Nacional de Saúde Pública e Meio Ambiente da Holanda (RIVM), Bilthoven, Holanda
Nuno R. Faria, Imperial College, Londres, e Universidade de Oxford, Oxford, Reino Unido
Tanya Golubchik, Universidade de Oxford, Oxford, Reino Unido
Edward Holmes, Universidade de Sydney, Sydney, Austrália
Liana Kafetzopoulou, KU Leuven – Universidade de Leuven, Bélgica
Philippe Lemey, KU Leuven – Universidade de Leuven, Bélgica
Tze Minn Mak, Centro Nacional de Doenças Infecciosas, Singapura
Meng Ling Moi, Universidade de Nagasaki, Nagasaki, Japão
Bas Oude Munnink, Erasmus MC, Rotterdam, Holanda
Leo Poon, Universidade de Hong Kong, Região Administrativa Especial (RAE) de Hong Kong, China
James Shepherd, Universidade de Glasgow, Glasgow, Reino Unido
Timothy Vaughan, Eidgenössische Technische Hochschule Zurich (ETH Zurich), Zurique, Suíça
Erik Volz, Imperial College, Londres, Reino Unido

Revisores

Kristian Andersen, Scripps Research, La Jolla, CA, EUA
Julio Croda, Ministério da Saúde, Rio de Janeiro, Brasil
Simon Dellicour, Universidade Livre de Bruxelas, Bruxelas, Bélgica
Túlio de Oliveira, Universidade de KwaZulu-Natal, Durban, África do Sul
Nathan Grubaugh, Universidade de Yale, New Haven, CT, EUA

Marion Koopmans, Erasmus MC, Rotterdam, Holanda
Tommy Lam, Universidade de Hong Kong, RAE de Hong Kong, China
Marcio Roberto Nunes, Instituto Evandro Chagas, Ananindeua, Pará, Brasil
Gustavo Palacios, Agência dos Estados Unidos para o Desenvolvimento Internacional, Washington, DC, EUA
Steven Pullan, Public Health England, Londres, Reino Unido
Josh Quick, Universidade de Birmingham, Birmingham, Reino Unido
Andrew Rambaut, Universidade de Edimburgo, Edimburgo, Reino Unido
Chantal Reusken, Instituto Nacional de Saúde Pública e Meio Ambiente da Holanda (RIVM), Bilthoven, Holanda
Etienne Simon-Loriere, Institut Pasteur, Paris, França
Tanja Stadler, Eidgenössische Technische Hochschule Zurich (ETH Zurich), Suíça
Marc Suchard, Universidade da Califórnia em Los Angeles, Los Angeles, CA, EUA
Huaiyu Tian, Universidade Normal de Pequim, Pequim, China
Lia van der Hoek, Amsterdam Medical Center, Amsterdam, Holanda
Jantina de Vries, Professora Associada em Bioética, Departamento de Medicina, Universidade da Cidade do Cabo, África do Sul

Outros contribuidores

Kazunobu Kojima, Interface de Biossegurança e Segurança em Saúde, Programa de Emergências de Saúde, OMS, Genebra, Suíça
Lina Moses, Operações de Emergência, Programa de Emergências de Saúde, OMS, Genebra, Suíça
Lane Warmbrod, Equipe de Epidemiologia, Programa de Emergências de Saúde, OMS, Genebra, Suíça
Vasee Sathyamoorthy, Pesquisa em Saúde, Divisão de Ciências, OMS, Genebra, Suíça
Katherine Littler, Health Ethics & Governance, OMS, Genebra, Suíça
Mark Perkins, Doenças e Zoonoses Emergentes, Programa de Emergências de Saúde, OMS, Genebra, Suíça

Abreviaturas

ACE	enzima conversora de angiotensina
BDSKY	Pacote do modelo skyline de nascimento e morte
pb	par de bases
CDC	Centros para Controle e Prevenção de Doenças (EUA)
CoV	coronavírus
Ct	limiar do ciclo
DDBJ	Banco de dados de DNA do Japão
E	envelope
EBI	Instituto Europeu de Bioinformática
EMBL	Laboratório Europeu de Biologia Molecular
ENA	Arquivo Europeu de Nucleotídeos
HIV	vírus da imunodeficiência humana
INSDC	Colaboração internacional de dados de sequência de nucleotídeos
	membrana
MERS	Síndrome respiratória do Oriente Médio
MRCA	ancestral comum mais recente nucleocapsídeo
NAAT	teste de amplificação do ácido nucleico
NCBI	Centro Nacional de Informações sobre Biotecnologia (EUA)
NGS	sequenciamento de nova geração
nt	nucleotídeo
ORF	fase de leitura aberta
PCR	reação em cadeia da polimerase
R_0	número de reprodução
RACE	amplificação rápida de extremidades de cDNA
RBD	domínio de ligação ao receptor
RNA	ácido ribonucleico
S	espícula
SARS	síndrome respiratória aguda grave
SARS-CoV-2	coronavírus 2 da síndrome respiratória aguda grave
SRA	Arquivo de leitura de sequência
TMRCA	tempo desde o ancestral comum mais recente
OMS	Organização Mundial da Saúde

Sumário executivo

Avanços recentes permitiram que os genomas do coronavírus 2 da síndrome respiratória aguda grave (SARS-CoV-2) – o agente causador da COVID-19 – fossem sequenciados horas ou dias após a identificação de um caso. Como resultado, pela primeira vez, o sequenciamento genômico em tempo real foi capaz de orientar a resposta da saúde pública a uma pandemia. O sequenciamento metagenômico foi fundamental para a detecção e caracterização do novo patógeno. O compartilhamento precoce das sequências do genoma do SARS-CoV-2 permitiu que ensaios de diagnóstico molecular fossem desenvolvidos rapidamente, o que melhorou a preparação global e contribuiu para o projeto de contramedidas. O sequenciamento rápido e em grande escala do genoma do vírus está contribuindo para a compreensão da dinâmica das epidemias virais e para a avaliação da eficácia das medidas de controle.

O crescente reconhecimento de que o sequenciamento genômico viral pode contribuir para a melhoria da saúde pública está levando mais laboratórios a investir nessa área. No entanto, o custo e o trabalho envolvidos no sequenciamento de genes são substanciais e os laboratórios precisam ter uma ideia clara dos retornos esperados para a saúde pública em relação a esse investimento. Este documento fornece orientação para os laboratórios sobre como maximizar o impacto das atividades de sequenciamento do SARS-CoV-2 agora e no futuro.

Objetivos pretendidos do sequenciamento

Antes de iniciar um programa de sequenciamento, é importante ter uma compreensão clara dos objetivos do sequenciamento, uma estratégia para análise e um plano de como os resultados serão usados para orientar as respostas de saúde pública. Cada fase da pandemia da COVID-19 levantará diferentes questões que são fundamentais para a saúde pública, algumas das quais exigem estratégias distintas de amostragem de genoma. O sequenciamento do gene do SARS-CoV-2 pode ser usado em muitas áreas diferentes, incluindo diagnósticos aprimorados, desenvolvimento de contramedidas e investigação da epidemiologia da doença. Apesar do óbvio poder do sequenciamento, é importante que aqueles que definem os objetivos, conduzem análises genômicas e usam os dados resultantes estejam cientes das limitações e de possíveis fontes de viés.

Considerações ao implementar um programa de sequenciamento

As decisões sobre as metas de sequenciamento devem ser tomadas em um esquema multidisciplinar que inclua representantes seniores de todas as partes interessadas. As fontes de financiamento devem ser identificadas para garantir apoio sustentável, incluindo o custo de pessoal especializado, dispositivos de sequenciamento e consumíveis, e a arquitetura computacional necessária para processar e armazenar dados. Os aspectos éticos do projeto devem ser avaliados cuidadosamente. Os laboratórios devem realizar avaliações de risco de bioproteção e biossegurança para cada etapa do protocolo escolhido.

Os objetivos do sequenciamento devem orientar as considerações técnicas sobre os métodos a serem usados para o sequenciamento e a seleção das amostras. Vários dispositivos estão

disponíveis para sequenciar genomas do SARS-CoV-2 e cada qual pode ser mais ou menos apropriado em circunstâncias particulares, como resultado de diferenças na precisão por leitura, no volume de dados gerados e no tempo de resposta. Para a maioria dos objetivos, são necessários dados de sequência viral e metadados de amostra. A aquisição e tradução desses dados no formato correto para análise podem exigir muitos recursos, mas ajudarão a maximizar o impacto em potencial do sequenciamento. Muitas análises dependem da habilidade para comparar as sequências de vírus adquiridas localmente com a diversidade genômica global do vírus. Portanto, é crucial que as sequências genômicas virais sejam compartilhadas de maneira apropriada. Esse compartilhamento está ocorrendo a uma velocidade impressionante por meio de repositórios como o GISAID e o GenBank.

A decisão de quais amostras devem ser sequenciadas dependerá da pergunta a ser respondida e do contexto. Deve-se levar em conta também a logística da amostra, por exemplo, qual é a melhor maneira de transporte e qual é o melhor modo de realizar a extração e o sequenciamento do RNA sem arriscar sua integridade. Quando várias organizações realizam sequenciamentos e análises, deve ser desenvolvido um sistema prático e compartilhado de identificação de amostras.

Assim que uma amostra for sequenciada e os metadados apropriados forem coletados, é necessária uma análise bioinformática. O pipeline de bioinformática dependerá das etapas laboratoriais de pré-sequenciamento, da plataforma de sequenciamento e dos reagentes usados. O alinhamento da sequência e a análise filogenética exigirão poder computacional de alto desempenho, que pode ser dispendioso. A análise e a interpretação dos dados exigirão uma equipe altamente treinada. Os resultados e as conclusões devem ser compartilhados com as partes interessadas relevantes de maneira clara e consistente para evitar interpretações errôneas.

Como maximizar o impacto na saúde pública

Não importa quantas sequências do genoma do SARS-CoV-2 sejam geradas, elas somente terão um impacto positivo na saúde pública se forem definidas estratégias subsequentes para produção e comunicação de resultados úteis e oportunos. Os programas devem sempre ponderar como os resultados da análise da sequência do SARS-CoV-2 podem ampliar, complementar ou substituir outras abordagens existentes e decidir se o sequenciamento é o método mais apropriado ou efetivo em termos de recursos para atingir os objetivos desejados. Os resultados devem ser comunicados de maneira oportuna e clara às partes interessadas que podem usar as informações diretamente para o benefício da saúde pública. Isso pode ser alcançado de forma mais eficiente se os laboratórios de análise e sequenciamento genômico estiverem intimamente integrados aos programas de diagnóstico e epidemiológicos de saúde pública existentes.

O desenvolvimento de uma rede de sequenciamento global forte e resiliente pode maximizar o impacto do sequenciamento na saúde pública, não apenas para o SARS-CoV-2, mas também para futuros patógenos emergentes. Várias redes de laboratórios específicos para patógenos têm investido na capacidade de sequenciamento como parte de suas atividades de vigilância. Como os custos de sequenciamento ainda são substanciais e muitas partes do fluxo de trabalho de sequenciamento podem ser usadas para vários patógenos ou objetivos de sequenciamento, incentiva-se a colaboração nacional para garantir o uso ideal da capacidade existente. É

necessário investimento de longo prazo para fortalecer a capacidade de análise bioinformática e filogenética, já que agora ela está muito aquém da capacidade laboratorial molecular em muitos lugares. Os programas de capacitação devem se concentrar em uma abordagem gradual para o desenvolvimento de competências. O ponto focal da capacitação dependerá do contexto: alguns países podem precisar desenvolver sua capacidade de laboratório úmido, ao passo que outros podem decidir terceirizar o sequenciamento real e se concentrar na bioinformática e no gerenciamento e na interpretação de dados. A colaboração entre grupos de sequenciamento será facilitada por protocolos de sequenciamento compartilhados, pela padronização da estrutura de banco de dados e formatos de metadados, por reuniões e treinamento conjuntos e pelo acesso a auditorias e testes de proficiência usando padrões de referência.

1 Introdução

As sequências genômicas do coronavírus 2 da síndrome respiratória aguda grave (SARS-CoV-2) – o vírus que causa a COVID-19 – estão sendo geradas e compartilhadas em uma velocidade sem precedentes. Avanços tecnológicos recentes permitiram que os genomas do SARS-CoV-2 fossem sequenciados horas ou dias após a identificação de um caso. O uso desses genomas para orientar a política de saúde pública durante um surto em andamento significa uma revolução nas investigações genômicas virais. Pela primeira vez, o sequenciamento genômico pode ajudar a orientar a resposta da saúde pública a uma pandemia em tempo quase real.

O sequenciamento genômico viral já se mostrou fundamental na identificação do SARS-CoV-2 como o agente causador da COVID-19 e na investigação de sua disseminação global. Além disso, as sequências do genoma do vírus podem ser usadas para investigar a dinâmica do surto, incluindo alterações na dimensão da epidemia ao longo do tempo, na disseminação espaço-temporal e nas rotas de transmissão. Além disso, as sequências genômicas podem ajudar no desenho de ensaios diagnósticos, nos projetos de medicamentos e vacinas, e no monitoramento, caso alterações hipotéticas em sua eficácia ao longo do tempo sejam atribuídas a alterações no genoma do vírus. A análise dos genomas do vírus SARS-CoV-2 pode, portanto, complementar, aumentar e apoiar estratégias para redução da carga da COVID-19.

O aumento da compreensão do potencial do sequenciamento genômico para melhorar a saúde pública está levando mais laboratórios a investir nesse processo. No entanto, o custo potencialmente alto e o trabalho envolvido exigem clareza sobre os retornos esperados desse investimento, sobre como os dados da sequência genômica podem ser melhor utilizados e sobre os meios pelos quais um impacto benéfico na saúde pública e na política pode ser alcançado.

Este guia tem como objetivo ajudar os técnicos e laboratórios de saúde pública responsáveis por programas de sequenciamento genômico para o SARS-CoV-2 ou que estejam cogitando o estabelecimento desses programas. Ele fornece informações sobre as considerações a serem feitas ao se planejar ou conduzir um programa de sequenciamento do SARS-CoV-2, de modo a garantir o melhor uso dos resultados na melhoria da saúde pública. Além disso, levanta questões práticas, detalha as possíveis aplicações e limitações das análises genômicas e fornece uma breve orientação sobre estratégias técnicas para sequenciamento e análise.

2 Retrospectiva

2.1 Progresso do sequenciamento genômico viral

As primeiras duas décadas do século XXI trouxeram uma mudança transformacional no uso da genômica viral em surtos de doenças, substituindo os longos protocolos e análises retrospectivas do passado por uma nova capacidade de investigação da epidemiologia genômica em tempo quase real. A aplicação generalizada do sequenciamento foi facilitada por rápidas reduções no custo por base e no tempo de resposta desde a amostra até o resultado, por aumentos no volume de dados gerados e na capacidade computacional necessária para processá-los, e pelo desenvolvimento de equipamentos de sequenciamento de bancada facilmente implementáveis e custo-efetivos *(1)*. Consequentemente, o sequenciamento tornou-se uma ferramenta crítica em microbiologia clínica para detectar e caracterizar patógenos virais em amostras clínicas *(2),* para apoiar o controle de infecção, para orientar investigações epidemiológicas e para caracterizar respostas virais evolutivas a vacinas e tratamentos *(3, 4)*.

O aumento da importância do sequenciamento genômico viral nas investigações clínicas e epidemiológicas é exemplificado pelas diferenças na velocidade e escala entre as respostas genômicas durante a epidemia de 2002-2003 da síndrome respiratória aguda grave (SARS) e as da pandemia da COVID-19 atual. Durante a epidemia de SARS, apenas três genomas virais foram compartilhados publicamente no primeiro mês após a identificação de um coronavírus como o patógeno causador, e apenas 31 estavam disponíveis em três meses. A genômica foi usada para desenhar ensaios moleculares que pudessem estabelecer uma associação entre a doença e o novo coronavírus em questão (5 - 7), mas não estava suficientemente desenvolvida para permitir que a epidemiologia viral fosse estudada em tempo real em grande escala. Em contraste, durante a pandemia da COVID-19, o sequenciamento metagenômico foi usado para identificar o patógeno causador da pneumonia inexplicada dentro de uma semana após a doença ser relatada (8, 9). O patógeno foi anunciado como sendo um novo coronavírus (SARS-CoV-2, anteriormente conhecido como 2019-nCoV) no início de janeiro de 2020 *(9)*. Seis genomas foram compartilhados publicamente antes de meados de janeiro, permitindo o rápido desenvolvimento de testes de diagnóstico e estratégias para um extenso sequenciamento genômico viral. Os esforços de sequenciamento continuaram à medida que o vírus se espalhou pelo mundo, resultando em um conjunto de dados em constante crescimento com mais de 60.000 genomas virais quase completos no prazo de seis meses após a identificação do SARS-CoV-2. Frequentemente, os genomas são gerados poucos dias após a identificação do caso e usados para entender a disseminação do vírus durante a pandemia.

2.2 Progresso das aplicações genômicas virais

Nos últimos anos, as emergências de saúde pública causadas por epidemias impulsionaram o desenvolvimento do sequenciamento genômico viral e da epidemiologia molecular. As sequências genômicas virais nos permitiram identificar os patógenos e entender sua origem, transmissão, diversidade genética e dinâmica de surto (Quadro 1). Esse entendimento orientou o desenvolvimento de abordagens diagnósticas, forneceu informações básicas importantes para o desenvolvimento de vacinas e projetos de medicamentos e ajudou na mitigação de doenças. *(33,*

41, 42). As análises genômicas são capazes de estimar aspectos da dinâmica epidemiológica das doenças virais que são irrecuperáveis usando apenas dados epidemiológicos *(3, 41, 43)* porque permitem tirar conclusões sobre os períodos de um surto em que não foram observados casos. Vigorosas conclusões podem ser obtidas mesmo com dados genômicos relativamente esparsos.

O SARS-CoV-2, portanto, surgiu em um contexto científico no qual as sequências genômicas podem ser geradas mais rápida e facilmente e podem ser usadas para responder a uma gama mais ampla do que nunca de questões de saúde pública.

Quadro 1. Contribuição da genômica viral para a compreensão epidemiológica em emergências de saúde pública desde a epidemia de SARS1[1]

A pandemia de gripe causada pelo vírus Influenza A (H1N1) pdm09 foi a primeira em que muitas questões epidemiológicas puderam ser investigadas por meio de análises genéticas. A avaliação da transmissibilidade do vírus a partir de sequências de genes forneceu estimativas iniciais do número de reprodução básico, R_0, que foram semelhantes àquelas produzidas por análises epidemiológicas. *(10)* A análise genômica retrospectiva confirmou que a pandemia havia começado pelo menos 2 meses antes do primeiro caso amostrado e inferiu taxas de crescimento populacional e tempos de duplicação da epidemia semelhantes aos encontrados nas análises iniciais. *(11)* No entanto, os esforços para entender as origens da epidemia de A (H1N1) pdm09 foram prejudicados pela falta de vigilância sistemática da gripe em suínos. *(12)* Um estudo retrospectivo em 2016 demonstrou grande diversidade entre os vírus da gripe no México e sugeriu que os suínos do México eram a fonte mais provável do vírus que deu origem à pandemia de 2009. *(13)*

Desde 2012, foram relatados vários surtos de síndrome respiratória do Oriente Médio (MERS) causados pelo coronavírus MERS-CoV, levantando questões sobre a origem do vírus e seu modo de transmissão. Seguindo evidências sorológicas e epidemiológicas preliminares que apoiaram o envolvimento de dromedários (camelos árabes, *Camelus dromedarius*) nesses surtos, *(14)* o sequenciamento genômico foi usado para identificar a presença do vírus em camelos *(15,16)* e para demonstrar múltiplos eventos de transmissão de vírus independentes de camelos para seres humanos. *(15, 17,18)* As análises de sequenciamento subsequentes mostraram ainda que o MERS-CoV é endêmico em camelos do Mediterrâneo Oriental e países africanos. *(19)* Em 2018, um estudo genômico abrangente confirmou que o vírus é mantido em camelos e que os seres humanos são hospedeiros terminais. *(20)* Os valores médios de R_0 estimados por meio de sequências genômicas virais foram inferiores a 0,90, sugerindo que MERS-CoV provavelmente não se tornaria endêmico em seres humanos. Isso confirmou que o foco em esforços contínuos de controle entre camelos era apropriado, ao mesmo tempo destacando a necessidade contínua de monitoramento de um possível surgimento de cepas que seriam mais facilmente transmissíveis entre seres humanos. *(20)*

[1] Ver Anexo 1 para as estratégias de amostragem empregadas nos estudos citados neste quadro.

A epidemia do vírus Ebola de 2013–2016 marcou o início de uma investigação epidemiológica genômica em grande escala num surto em andamento. As análises genômicas permitiram a vigilância epidemiológica viral durante o desenrolar da epidemia e auxiliaram na compreensão da origem, epidemiologia e evolução do vírus. As técnicas de datação por relógio molecular estimaram que o ancestral comum de todos os genomas do vírus Ebola sequenciados ocorreu no início de 2014, o que era condizente com as investigações epidemiológicas que colocaram o primeiro caso em torno do final de dezembro de 2013. *(21–24)* As análises evolutivas demonstraram que a propagação foi mantida pela transmissão de pessoa para pessoa, e não por múltiplas introduções separadas de um reservatório animal. *(21– 28)* As conclusões filodinâmicas sobre a propagação inicial da epidemia permitiram que o R_0 fosse estimado e os eventos de superdisseminação na população fossem investigados. *(29, 30)* As investigações de genética molecular apoiaram a possibilidade de transmissão sexual do vírus Ebola, resultando em recomendações da OMS para melhorar o aconselhamento de sexo seguro e testagem dos sobreviventes do Ebola. *(31, 32)* Perto do final do surto, houve no país uma mudança para o sequenciamento rápido que ajudou a resolver as cadeias de transmissão viral e a disseminação na comunidade. *(4, 33–36)*

Em 1º de fevereiro de 2016, a OMS declarou a infecção pelo vírus Zika uma emergência de saúde pública de preocupação internacional após a circulação autóctone do vírus em 33 países e fortes suspeitas de que a infecção durante a gravidez estivesse ligada a microcefalia fetal e outras anormalidades do desenvolvimento. *(37)* A reconstrução da propagação do vírus apenas a partir de dados epidemiológicos foi difícil porque os sintomas eram frequentemente leves ou ausentes e se sobrepunham aos causados por outros arbovírus co-circulantes (por exemplo, dengue, chikungunya) e também porque a vigilância diagnóstica molecular do vírus Zika era frequentemente iniciada muito tempo após o início da transmissão local. *(38)* Foram iniciados esforços colaborativos para sequenciar casos retrospectivos e novos, a fim de obter conclusões sobre a origem, as rotas de transmissão e a diversidade genética do vírus.*(38)* Análises filogenéticas e de relógio molecular preliminares mostraram que a epidemia nas Américas foi causada por um único evento de introdução de uma linhagem genotípica asiática, que foi estimada como tendo ocorrido um ano antes da detecção da doença em maio de 2015 no Brasil. *(37)* Estudos epidemiológicos genômicos subsequentemente documentaram de modo consideravelmente detalhado a disseminação do vírus Zika. *(37–40)* Por exemplo, a amostragem generalizada de sequências genômicas de pacientes infectados e mosquitos durante o surto sustentado do vírus Zika em 2016 na Flórida, EUA, permitiu que o R_0 fosse estimado em menos de 1. Isso levou à conclusão de que múltiplas introduções do vírus seriam necessárias para uma transmissão local tão extensa. *(40,41)*

2.3 Análises filogenéticas e filodinâmicas

Muitas aplicações importantes da genômica viral na orientação das respostas de saúde pública foram desenvolvidas em análises filogenéticas ou filodinâmicas. A filogenética é usada em quase todos os ramos da biologia para investigar as relações evolutivas entre diferentes organismos usando suas sequências genéticas. As árvores filogenéticas (por exemplo, ver Figura 1) são visualizações úteis dessas relações. Os padrões de ramificação e o comprimento dos ramos podem ser usados para representar a relação evolutiva. Quaisquer dois organismos, representados por nós externos ou "folhas" (pontas), terão um ancestral comum em que os ramos que levam a eles se cruzam (nós internos). Tendo em vista os dados de sequência genética homóloga de múltiplos organismos e um modelo de substituição genética de como diferentes sítios dessas sequências mudam ao longo do tempo, é possível avaliar um grande número de árvores para determinar qual é a mais provável de representar a verdadeira relação entre esses organismos.

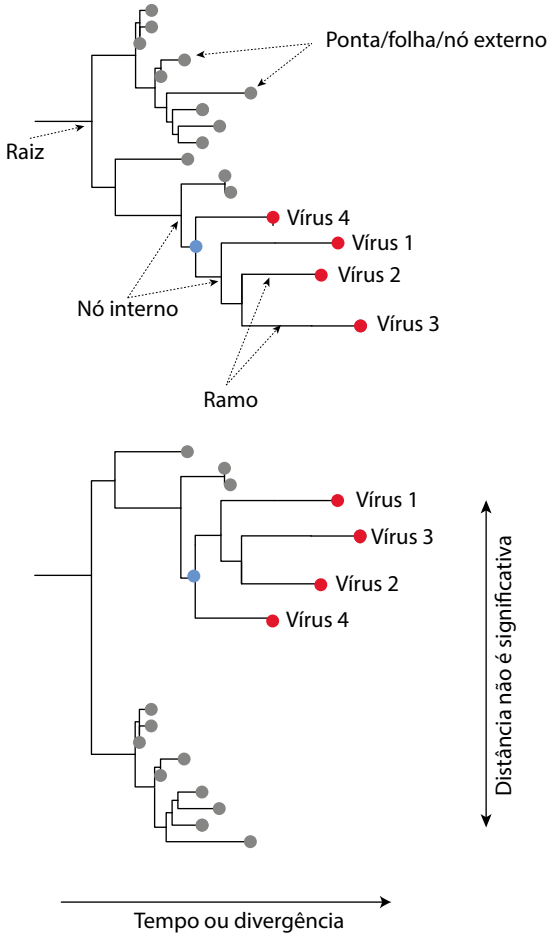

Figura 1. Árvores filogenéticas com características-chave marcadas. A distância ao longo do eixo x nas filogenias exibidas acima em formato "retangular" geralmente representa o tempo ou o volume de alterações genéticas que se acumulou. O ancestral comum mais recente dos vírus 1–4 é destacado pelo nó azul. A distância ao longo do eixo y não é significativa. Especificamente, os clados descendentes de qualquer nó podem ser girados em torno desse nó sem alterar a interpretação filogenética da árvore. As duas árvores retratadas acima são, portanto, filogeneticamente idênticas.

Ao discutir a evolução do vírus, é extremamente importante distinguir entre a taxa de mutação e a taxa evolutiva (ou taxa de substituição). A taxa de mutação é uma medida bioquímica que leva em conta o número de erros que ocorrem na cópia do RNA de um vírus parental para sua progênie e é normalmente medida em mutações por genoma por replicação. A taxa de mutação pode ser estimada experimentalmente de várias maneiras, tais como sequenciando populações inteiras de vírus para medir a diversidade genética antes e depois de um número conhecido de replicações em um ambiente laboratorial. A maioria das mutações é deletéria, *(44)* e os vírions individuais que contêm essas mutações frequentemente não conseguem se replicar.

Apenas as mutações que aumentam de frequência e se fixam em uma linhagem, depois da deriva genética ou da ação da seleção natural em uma população de vírus, contribuem para a taxa evolutiva. A taxa evolutiva é, em geral, descrita como o número de substituições de nucleotídeos por sítio, por ano (frequentemente abreviado como subs/sítio/ano). Diferentes linhagens de vírus podem ter diferentes taxas evolutivas. A taxa evolutiva pode muitas vezes ser inferida diretamente dos dados da sequência genômica viral obtidos de diferentes pacientes em diferentes datas. O intervalo de datas de coleta de amostra (ao longo de meses ou anos) necessário para permitir uma inferência robusta da taxa evolutiva varia para diferentes vírus e surtos, porque depende da taxa de substituição, da idade da linhagem viral sob investigação e do comprimento da sequência genômica sob investigação.

Para o SARS-CoV-2, a inclusão de dados genômicos coletados em intervalos de dois meses parece minimamente suficiente, (45) embora estimativas mais robustas sejam obtidas usando-se dados coletados em um período mais longo.

Os vírus de RNA normalmente têm uma alta taxa evolutiva, sendo que muitos adquirem uma alteração genética a cada poucos dias ou semanas. *(46)* Alguns vírus de RNA, portanto, adquirem substituições genéticas quase na mesma escala de tempo da transmissão entre hospedeiros. No caso do SARS-CoV-2, a taxa de eventos de transmissão entre humanos é mais alta, em média, do que a taxa na qual as linhagens virais transmissoras adquirem substituições genéticas. As linhagens do SARS-CoV-2 acumulam diversidade genética ao longo de semanas ou meses, em vez de dias, de modo que pacientes diretamente vizinhos em uma cadeia de transmissão podem ser infectados por vírus com genomas idênticos. A análise dos padrões de acúmulo da diversidade genômica viral durante um surto pode ser usada para fazer inferências sobre os processos epidemiológicos. Esse é o enfoque de um grupo de técnicos filogenéticos englobados no termo filodinâmica, que foi cunhado por Grenfell et. al. (47)

Os métodos filodinâmicos são úteis na investigação de surtos, pois podem complementar e ampliar outras análises epidemiológicas baseadas em casos confirmados que forem identificados. Em primeiro lugar, várias abordagens filodinâmicas podem ser menos afetadas – ou afetadas de forma diferente – por vieses na vigilância diagnóstica, como mudanças no trabalho de vigilância ao longo do tempo ou detecção irregular de casos.

Em segundo lugar, a filodinâmica pode revelar características da epidemia que ocorrem fora da janela de tempo de amostragem (por exemplo, antes da identificação do primeiro caso). Em

terceiro lugar, as análises filodinâmicas fornecem um meio direto de se aprender sobre a dinâmica populacional de diferentes linhagens de vírus específicas.

Os métodos filodinâmicos usam modelos probabilísticos para vincular a árvore filogenética dos genomas amostrados a parâmetros epidemiológicos de interesse. Como tal, exigem a inferência de uma árvore filogenética datada que contenha informações não apenas sobre quais sequências se agrupam, mas também quando os ancestrais comuns mais recentes sem amostragem (MRCAs) dos genomas virais amostrados existiram. Embora sejam conhecidas as datas de amostragem para vírus de amostras sequenciadas (ou seja, pontas de árvore, ver Figura 1), os MRCAs (ou seja, os nós internos) são inferidos filogeneticamente e seu tempo de existência deve ser estimado. A estimativa dessas datas requer o uso de um modelo de relógio molecular parametrizado por uma taxa de relógio – a taxa média de substituição genética ao longo dos ramos da filogenia.

Há várias famílias distintas de modelos filodinâmicos: coalescente, nascimento-morte e modelos baseados em simulação. As avaliações desses diferentes modelos estão disponíveis em outros lugares. *(48, 49)*

2.4 Características genômicas e evolutivas do SARS-CoV-2 que são importantes para as aplicações genômicas

Várias características fundamentais de qualquer vírus determinam as abordagens possíveis para a geração e uso de dados genômicos virais para orientar as autoridades de saúde pública. Essas características incluem seu material genético (RNA ou DNA), comprimento genômico, estrutura e composição do genoma e taxa evolutiva.

O SARS-CoV-2 é classificado no gênero *Betacoronavirus* (subgênero *Sarbecovirus*) na família *Coronaviridae* (subfamília *Orthocoronavirinae*), uma família de vírus RNA de fita simples. *(50)* Hoje em dia, o Comitê Internacional de Taxonomia de Vírus (ICTV) considera o SARS-CoV-2 como pertencente à espécie *Coronavírus relacionado à síndrome respiratória aguda grave*, junto com o SARS-CoV e outros vírus intimamente relacionados, amostrados em espécies não humanas. *(51)* A cepa de referência do SARS-CoV-2, Wuhan-Hu-1 (acesso do GenBank MN908947), foi amostrada de um paciente em Wuhan, China, em 26 de dezembro de 2019. *(52)* Esse genoma tem 29.903 nucleotídeos (nt) de comprimento e compreende uma ordem de gene de estrutura semelhante à observada em outros coronavírus: 5'-replicase ORF1ab-S-E-M-N-3'. O gene da replicase ORF1ab previsto do Wuhan-Hu-1 tem 21.291 nt de comprimento. Prevê-se que a poliproteína ORF1ab seja clivada em 16 proteínas não estruturais. ORF1ab é seguido por uma série de fases de leitura aberta (ORFs) a jusante. Estas incluem os genes S (espícula), ORF3a, E (envelope), M (membrana) e N (nucleocapsídeo) previstos de 3822, 828, 228, 669 e 1260 nt de comprimento, respectivamente. (52) Como o SARS-CoV, o Wuhan-Hu-1 também contém um gene ORF8 previsto (366 nt de comprimento) localizado entre os genes M e N. Por fim, as sequências terminais 5' e 3' do Wuhan-Hu-1 também são típicas dos betacoronavírus e têm um comprimento de 265 nt e 229 nt, respectivamente.

As estimativas preliminares da taxa evolutiva do SARS-CoV-2 estão próximas de uma média de 1 x 10^{-3} substituições por sítio por ano, (45, 53) que é semelhante à taxa evolutiva média observada em outros genomas virais de RNA. (46)

No momento em que este artigo foi escrito, não havia uma estimativa precisa da taxa de mutação por replicação do genoma para o SARS-CoV-2 (taxa de mutação). No entanto, espera-se que seja semelhante à de outros coronavírus. A taxa de mutação do coronavírus e de outros membros da ordem *Nidovirales* é menor do que a de outros vírus de RNA porque eles têm uma capacidade de leitura de prova intrínseca para corrigir erros replicativos que está ausente em outros vírus de RNA. *(50)*

3 Considerações práticas para a implementação de um programa de sequenciamento genômico viral

Muitos laboratórios de saúde pública agora reconhecem o impacto em potencial que as sequências genômicas virais podem ter nas decisões de saúde pública durante a atual pandemia da COVID-19 ou surtos futuros (ver também a Seção 5).

3.1 Planejamento de um programa de sequenciamento

Os laboratórios devem ter planos claros em vigor. Uma lista de verificação para auxiliar no planejamento é fornecida no Anexo 2. As principais questões a serem consideradas antes de iniciar um programa de sequenciamento incluem as seguintes.

(1) Quais são os resultados esperados do programa de sequenciamento?
(2) Quais amostras devem ser sequenciadas para atingir os resultados esperados identificados na etapa 1? Quais metadados ou fontes de dados adicionais são críticos?
(3) Quem são as principais partes interessadas e quais são suas responsabilidades? Como elas podem estar efetivamente engajadas?
(4) Como as amostras e as informações podem ser transferidas de forma rápida e adequada entre as partes interessadas, conforme necessário?
(5) O projeto foi elaborado de acordo com as leis e diretrizes éticas locais, nacionais e internacionais?
(6) Há financiamento, equipamento e recursos humanos adequados disponíveis para concluir todos os estágios de recuperação de amostras, sequenciamento em laboratório úmido, bioinformática, filodinâmica e outras análises, compartilhamento de dados e comunicação de resultados oportunos para as devidas partes interessadas?
(7) Como as metas podem ser alcançadas sem interromper outras áreas do trabalho laboratorial, como o diagnóstico clínico, e evitando a duplicação de esforços?
(8) Como o programa será avaliado em relação ao custo-efetividade e ao impacto?

3.2 Considerações éticas

Quando um programa de sequenciamento está sendo desenvolvido, é importante analisar todas as implicações éticas. Devem ser identificados os possíveis riscos de danos aos participantes da pesquisa e devem ser definidas as estratégias de mitigação. Todas as investigações propostas devem ser avaliadas e aprovadas por um comitê de revisão ética, levando em consideração o valor social e a validade científica da investigação, a seleção dos participantes, a relação risco-benefício, o consentimento informado e o respeito pelos participantes. *(54, 55)* Nos locais em que os pesquisadores tenham pouca experiência na identificação de possíveis questões éticas relacionadas ao sequenciamento de patógenos, a colaboração internacional e o envolvimento de especialistas apropriados são fortemente encorajados. A colaboração entre pesquisadores de todo o mundo ajudará a garantir parcerias de pesquisa equitativas e mutuamente benéficas. Os pesquisadores locais estão mais propensos a compreender seus

sistemas de saúde e pesquisa e a ser capazes de traduzir os resultados em políticas, sendo, portanto, muitas vezes mais adequados para assumir papéis de liderança e ativos no processo de pesquisa. *(54,55)* As considerações éticas relacionadas ao compartilhamento de dados são discutidas mais detalhadamente no Capítulo 4.

3.3 Identificação dos resultados esperados e dados necessários

Antes de embarcar em qualquer programa de sequenciamento, devem ser definidas as metas que podem ser alcançadas. Os objetivos possíveis são discutidos extensivamente na Seção 5; as metas definidas afetarão o desenho do fluxo de trabalho de sequenciamento.

Assim que os objetivos forem identificados, deve ser projetada uma estratégia de amostragem alcançável para coletar as sequências genômicas e metadados apropriados; as sequências genômicas que não contêm metadados apropriados não são úteis para a maioria dos aplicativos. Diferentes questões de saúde pública exigirão diferentes estratégias de amostragem e dados. É, portanto, de vital importância garantir que haja discussão entre as diversas partes interessadas que (a) realizam a amostragem diagnóstica, (b) escolhem amostras para sequenciamento, (c) escolhem a estratégia de sequenciamento, (d) escolhem estratégias analíticas, e (e) usam as informações geradas para a saúde pública, de modo a garantir que as estratégias de amostragem genômica e a coleta de metadados sejam corretamente direcionadas para as análises a que se destinam.

3.4 Identificação e ligação com as partes interessadas

As principais partes interessadas devem ser identificadas, consultadas e envolvidas em um estágio inicial (Quadro 2). Sua identidade e nível de envolvimento variam dependendo das circunstâncias locais e dos objetivos do programa, mas é razoável levar em consideração as partes interessadas envolvidas em todas as etapas do processo, desde a identificação de casos até a utilização dos resultados. Pode ser relevante fornecer recursos educacionais para as partes interessadas, incluindo o público em geral, para demonstrar a utilidade em potencial de um programa de sequenciamento e para explicar como as sequências serão usadas e por que metadados específicos do paciente são necessários. É essencial haver estreita colaboração e comunicação entre as partes interessadas relevantes para que as atividades de sequenciamento resolvam questões de importância para a saúde pública.

Quadro 2. As partes interessadas devem se envolver ao serem desenvolvidos programas de sequenciamento

Esta lista não é completa e outras partes interessadas devem ser levadas em consideração, dependendo das circunstâncias locais.

- **Órgãos de saúde pública.** Órgãos de saúde pública locais ou nacionais, como ministérios da saúde, frequentemente comissionam ou ajudam a estabelecer programas de sequenciamento de SARS-CoV-2. Seu envolvimento garantirá que as metas respondam às principais questões sobre políticas. Além disso, os órgãos de saúde pública muitas vezes podem ajudar a garantir a coleta ampla de amostras diagnósticas e metadados específicos.

- O ideal é que os laboratórios de diagnóstico sejam parceiros em todos os programas de sequenciamento do SARS-CoV-2. Em geral, são eles que têm melhor acesso às amostras SARS-CoV-2 e podem, com frequência, fornecer amostras positivas residuais e metadados diretamente para as instalações de sequenciamento. Em alguns ambientes, os laboratórios de diagnóstico clínico podem ter a tarefa de implementar um programa de sequenciamento interno, ao passo que em outros o sequenciamento pode ser feito por pesquisas externas ou em laboratórios nacionais de saúde pública.

- As instalações de sequenciamento podem ser públicas ou privadas; algumas instalações de sequenciamento terão a capacidade de bioinformática para gerar genomas virais de consenso, ao passo que outras fornecerão dados brutos que devem ser processados em outro lugar para gerar genomas. Nem todos os bioinformáticos terão experiência para lidar com dados produzidos por todas as técnicas e plataformas de sequenciamento de laboratório úmido possíveis. Nesse caso, é altamente recomendado o apoio de um especialista que saiba lidar com o tipo de dados que se pretenda utilizar.

- Os **grupos analíticos** que realizarão análises filogenéticas, filodinâmicas ou outras análises genômicas planejadas devem estar intimamente envolvidos na determinação de quais amostras devem ser sequenciadas, de modo que as sequências genômicas sejam apropriadas para os métodos analíticos a serem usados. Não se deve presumir automaticamente que a competência para realizar essas análises esteja presente nos laboratórios úmidos de genética molecular que realizam o sequenciamento. Onde for relevante, a estreita integração de analistas e os envolvidos na vigilância e resposta (por exemplo, equipes de saúde pública que investigam surtos locais) aumentará o impacto em potencial das análises.

- As **equipes de prevenção e controle de infecção** (por exemplo, em hospitais, lares de idosos e unidades de saúde pública) podem apoiar a identificação de *clusters* de doenças emergentes e estão bem posicionadas para identificar casos que seriam úteis para sequenciamento. Também podem atuar em relação aos achados subsequentes nos *clusters* de transmissão.

- Os **serviços de saúde ocupacional** em locais de trabalho podem ajudar a identificar possíveis *clusters* de transmissão ou rotas de transmissão que podem ser investigadas usando-se estudos genômicos virais e auxiliar a implementar atividades de prevenção e controle de infecção emergentes dos resultados desses estudos.

- Os **pacientes** devem ser envolvidos para garantir que entendam como as sequências e os metadados estão sendo usados e compartilhados e que se beneficiem dos resultados. Um programa de envolvimento da comunidade devidamente projetado e com recursos pode ajudar a identificar e abordar os possíveis obstáculos à pesquisa, relacionados, por exemplo, ao estigma, garantindo que o desenho do programa esteja ciente e responda ao ambiente sociocultural no qual o programa será implementado.

Assim que as principais partes interessadas forem identificadas, será preciso estabelecer canais apropriados de comunicação entre os vários grupos. No mínimo, os objetivos do programa devem

ser definidos em um esquema multidisciplinar que envolva representantes seniores de todas as partes interessadas.

A comunicação entre as partes interessadas deve, idealmente, ser mantida durante todo o projeto, e pode exigir reuniões diárias ou semanais entre representantes de alguns ou de todos os órgãos envolvidos, para garantir reações adequadas às mudanças de situação durante a epidemia (por exemplo, investigação de *clusters* de transmissão à medida que surgirem). As atividades com enfoque epidemiológico que integram analistas de dados genômicos diretamente nas equipes de investigação e resposta da saúde pública têm maior probabilidade de ter um impacto imediato maior do que aquelas nas quais a análise genômica viral é considerada uma atividade separada ou secundária.

Deve ser acordado desde o início como, quando e com quem os dados são compartilhados – com a comunidade científica ou entre as partes interessadas. Também deve ser acordada as responsabilidades das partes interessadas, incluindo a provisão de financiamento, se apropriado. Se forem gerados dados ou publicações, em geral é útil chegar a um acordo prévio sobre como os envolvidos serão devidamente creditados por sua contribuição na produção ou análise de dados.

Os resultados da análise de sequenciamento devem ser logo comunicados às partes interessadas em um relatório escrito padronizado e facilmente interpretável, e devem ser organizadas oportunidades para debate. A comunicação prática dos resultados e limitações analíticas deve ser transmitida em linguagem cotidiana, evitando jargão técnico. Quando uma abordagem multidisciplinar tiver sido seguida na abordagem de questões de saúde pública (por exemplo, questões que envolvam análise filogenética e modelagem matemática), os resultados do sequenciamento devem ser idealmente discutidos junto com os resultados de outras áreas.

3.5 Execução do projeto: aquisição de dados, logística e recursos humanos

As considerações técnicas referentes à adesão legal e ética, seleção de amostras, avaliação detalhada de recursos e orientação técnica são fornecidas na Seção 6.

3.6 Avaliação do projeto

Deve-se buscar regularmente um *feedback* estruturado das partes interessadas para identificar e abordar quaisquer dificuldades que possam surgir.

O potencial do sequenciamento genômico viral continua a crescer, e a comunidade científica e de saúde pública está desenvolvendo rapidamente novas estratégias para maximizar seu impacto em futuros surtos de doenças. Todos os esforços de sequenciamento devem, portanto, incluir oportunidades claras para avaliação frequente por todas as partes interessadas do que foi útil, do que estava faltando e de qual foi impacto alcançado pelo sequenciamento. É importante que haja identificação e comunicação desses achados aos pesquisadores e aos órgãos financiadores para ajudar a orientar o desenvolvimento de novas ferramentas.

4 Compartilhamento de dados

4.1 Recomendações da OMS sobre compartilhamento de dados

O rápido compartilhamento dos dados da sequência do genoma do patógeno, junto com metadados epidemiológicos e clínicos anônimos relevantes, maximizará o impacto do sequenciamento genômico na resposta da saúde pública. Esses dados, gerados durante um surto, devem ser compartilhados com a comunidade global o mais rápido possível, para garantir sua máxima utilidade na melhoria da saúde pública. Em abril de 2016, a OMS emitiu uma declaração de política sobre compartilhamento de dados no contexto de emergências de saúde pública: "A OMS advoga que as sequências do genoma de patógenos sejam disponibilizadas publicamente o mais rápido possível por meio de bancos de dados relevantes e que os benefícios decorrentes da utilização dessas sequências sejam compartilhados de forma equitativa com o país de onde a sequência do genoma do patógeno se origina". *(56)* Um dos fatores críticos para garantir o compartilhamento contínuo de dados genéticos é o devido reconhecimento dado àqueles que coletam amostras clínicas e geram sequências do genoma do vírus. As fontes de dados devem ser reconhecidas quando forem utilizados dados publicamente disponíveis, e as publicações relacionadas e os artigos preprint devem ser citados quando disponíveis. Além disso, os financiadores, os editores de periódicos e os pares revisores devem encorajar o compartilhamento contínuo de dados.

4.2 Compartilhamento de metadados apropriados

Devem ser compartilhados metadados de amostra anônimos juntamente com os dados genômicos do SARS-CoV-2 para maximizar a utilidade da sequência genômica. Os metadados compartilhados devem sempre incluir pelo menos a data e o local da coleta da amostra, mas os metadados adicionais aumentarão muito as possíveis aplicações da sequência. Sempre que possível, portanto, os metadados devem incluir dados referentes ao tipo de amostra, ao modo como a sequência foi obtida, os links para outros vírus sequenciados, a história de viagens do paciente e informações demográficas ou clínicas. Para uma descrição detalhada dos metadados, ver Seção 6, Tabela 2. Quando qualquer informação for compartilhada, é importante que o anonimato do paciente seja protegido.

4.3 Compartilhamento de sequências de consenso, sequências de consenso parciais e dados de sequência bruta

Como o SARS-CoV-2 surgiu há pouco tempo em seres humanos, a diversidade genética do vírus permanece relativamente limitada e as sequências de comprimento total são, portanto, importantes para capturar o maior número possível de sítios informativos, em termos filogenéticos. Quando não for conseguido um sequenciamento total, podem ser geradas sequências parciais. Os genomas SARS-CoV-2 com cobertura parcial ainda são importantes e devem ser compartilhados. Embora a cobertura do genoma necessária (proporção de sítios sem bases ambíguas, ou seja, Ns) varie para diferentes aplicações e para diferentes vírus, os genomas parciais muitas vezes representam fontes importantes de dados. Por exemplo, os genomas do

vírus Zika com cobertura de apenas 40% (ou seja, 60% dos sítios com Ns) foram considerados filogeneticamente informativos da estrutura do clado. *(57)*

Quanto aos genomas completos, a qualidade do genoma parcial deve ser verificada para garantir que os sítios com apoio insuficiente sejam mascarados antes que o genoma seja disponibilizado ao público. Os genomas parciais em que a cobertura ou profundidade de sequenciamento geralmente é muito baixa, mas que em algumas regiões curtas têm profundidade de sequenciamento muito alta, podem ser indicativos de contaminação com amplicons produzidos por meio da reação em cadeia da polimerase (PCR) e devem ser avaliados antes do compartilhamento, com cuidado.

O compartilhamento de leituras de sequenciamento bruto (ou seja, todos os fragmentos sequenciados individuais de um genoma viral antes de serem agrupados em um genoma de consenso) é importante porque permite que o efeito de diferentes abordagens de bioinformática para geração de genoma de consenso seja comparado diretamente e facilita a correção de erros quando necessário. Dependendo da estratégia de sequenciamento adotada e da profundidade da cobertura de sequenciamento, os dados de nível de leitura também podem ser usados para análise de variação intra-hospedeiro nos genomas virais. Os conjuntos de dados de nível de leitura do SARS-CoV-2 devem, portanto, ser disponibilizados sempre que possível. Tendo em vista que o volume de dados das bibliotecas sequenciadas pode chegar a centenas de gigabytes, o compartilhamento de dados em nível de leitura pode ser mais desafiador em locais que tenham velocidade limitada de *upload* na internet ou conexões intermitentes. Os dados brutos que contêm leituras humanas devem ser filtrados para que sejam retidos apenas os dados de sequência genética não humana (ou seja, viral) antes de serem compartilhados, a fim de garantir o anonimato do paciente (consulte a Seção 6.7.1).

4.4 Plataformas para compartilhamento

O compartilhamento de sequências por meio de plataformas pesquisáveis comumente usadas aumenta a acessibilidade dos dados. As plataformas variam no tipo de dados que hospedam, nas condições de uso que impõem aos dados e na facilidade de *upload* dos metadados. Algumas plataformas (por exemplo, o European Nucleotide Archive) oferecem modelos de planilhas para dados de sequência que podem ser preenchidos *off-line* e carregados em lotes.

Os mecanismos de compartilhamento usados para dados de sequência genômica incluem bancos de dados de domínio público e de acesso público. Os bancos de dados de domínio público fornecem acesso aos dados sem exigir a identidade de quem está acessando e utilizando os dados. Nos bancos de dados de acesso público, os usuários devem se identificar para garantir o uso transparente dos dados e permitir uma supervisão efetiva, proteger os direitos dos contribuidores de dados, envidar todos os esforços para colaborar com os provedores de dados e reconhecer sua contribuição nos resultados publicados. As sequências genéticas do SARS-CoV-2 com metadados apropriados são geralmente compartilhadas por meio de várias plataformas. Os bancos de dados de domínio público para compartilhar genomas de consenso incluem o Centro Nacional de Informações sobre Biotecnologia (NCBI), o Instituto Europeu de Bioinformática do Laboratório Europeu de Biologia Molecular (EMBL-EBI) e o Banco de Dados de DNA

do Japão (DDBJ). Os dados brutos lidos com metadados apropriados são compartilháveis por meio dos repositórios da Colaboração Internacional de Dados de Sequência de Nucleotídeos (INSDC), que inclui o Arquivo de leitura de sequências (SRA) do NCBI, o EMBL-EBI ENA e o Arquivo de leitura de sequências DDBJ. Um banco de dados de acesso público para genomas de consenso, por exemplo, é o GISAID EpiCoV ™. O portal de dados da COVID-19 tenta facilitar o compartilhamento e o acesso a todas as fontes de dados biomédicos relevantes para a COVID-19. *(58)*

Os laboratórios devem entrar em contato com as plataformas de compartilhamento de sequências para atualizar as sequências parciais enviadas anteriormente, se um erro for identificado e corrigido.

As análises preliminares de dados de sequência genética são frequentemente compartilhadas em fóruns e servidores preprint, como o medRxiv ou o bioRxiv. Isso permite que os produtores de dados forneçam para a comunidade científica em geral informações sobre os achados iniciais. Alguns fóruns, incluindo o Virological, provaram ser úteis para o compartilhamento informal e a discussão dos resultados iniciais com a comunidade de genética molecular, e as postagens podem ser atualizadas continuamente à medida que as análises progridem. Os servidores preprint são, em geral, usados para compartilhar artigos no momento do envio a um periódico revisado por pares e para comunicar claramente a intenção de publicação. A OMS encoraja fortemente o compartilhamento de dados genéticos e metadados o mais rápido possível após a verificação da qualidade dos dados, sem retenções até após a avaliação do preprint.

As análises preliminares não revisadas estão sendo usadas mais extensivamente do que nunca pelo público e pela mídia na atual pandemia. Os cientistas devem, portanto, estar cientes de como as análises podem ser interpretadas ou apresentadas na mídia e devem fornecer interpretações claras de seus achados para que os resultados não sejam facilmente mal interpretados.

5 Aplicações da genômica ao SARS-CoV-2

Esta seção analisa como o sequenciamento genômico do SARS-CoV-2 foi usado em diferentes fases da pandemia da COVID-19 e sugere possíveis aplicações futuras. Também fornece uma breve orientação sobre as limitações comuns das abordagens atuais, para auxiliar no estabelecimento de metas realistas. Para algumas das aplicações consideradas, o sequenciamento genômico viral representa apenas um pequeno componente de uma investigação maior, que pode incluir investigações laboratoriais ou clínicas essenciais.

5.1 Compreensão do surgimento do SARS-CoV-2

5.1.1 Identificação do agente causador da COVID-19

O SARS-CoV-2 foi identificado e sequenciado de forma independente no início de 2020 por Wu et al., Lu et al. e Zhou et al. *(52, 59,60)*. Várias abordagens diferentes de sequenciamento metagenômico de nova geração (mNGS) foram usadas para identificar o patógeno causador da COVID-19. O sequenciamento metagenômico permite o sequenciamento não direcionado do ácido nucleico em uma amostra e pode, portanto, identificar o RNA ou DNA viral se este estiver presente em um número suficientemente elevado de cópias em relação ao DNA ou RNA de outras fontes (ver também a Seção 6.5.1). A conclusão das sequências completas do genoma viral, incluindo os terminais do genoma, geralmente envolve o sequenciamento de Sanger e um método de amplificação rápida das extremidades 5'/3' de cDNA (RACE). Esse é um método custo-eficiente para sequenciar regiões curtas de um genoma que podem ser perdidas com métodos metagenômicos, mas depende do conhecimento prévio das informações de sequências relativamente próximas à região ausente.

5.1.2 Determinação das épocas de origem e diversificação inicial

Foi particularmente importante determinar quando o SARS-CoV-2 surgiu pela primeira vez em seres humanos, já que isso poderia indicar se houve um longo período de transmissão não detectada antes de se observarem os primeiros casos clínicos (sendo possível, portanto, haver muitos casos não detectados). Os genomas do SARS-CoV-2 de Wuhan e das áreas circundantes da província de Hubei forneceram uma série de informações importantes.

Todas as sequências estavam intimamente relacionadas, diferindo apenas por algumas variantes de nucleotídeos. Vários exercícios de datação de relógio molecular que utilizaram essas sequências forneceram estimativas do tempo de aparecimento do mais recente ancestral comum de todos os vírus SARS-CoV-2 sequenciados como sendo o período de novembro a dezembro de 2019. Essas estimativas iniciais foram confirmadas à medida que mais sequências se tornaram disponíveis. A mais recente data possível da emergência do SARS-CoV-2 em seres humanos é, portanto, novembro-dezembro de 2019. Isso está bem próximo da primeira identificação do *cluster* inicial de casos de pneumonia em Wuhan em meados de dezembro. *(59–61)*

Nos lugares em que tenha ocorrido uma única introdução em seres humanos, o momento mais precoce possível da emergência de um vírus zoonótico em seres humanos é filogeneticamente

representado pelo tempo até o ancestral comum mais recente (TMRCA) do vírus zoonótico humano e do vírus animal não humano a partir do qual ele emergiu. Uma amostragem inadequada de vírus de animais não humanos que estejam intimamente relacionados ao SARS-CoV-2 significa que o possível intervalo no qual o SARS-CoV-2 poderia ter surgido em seres humanos é relativamente amplo se os dados filogenéticos forem considerados de forma isolada. Portanto, é difícil distinguir filogeneticamente dois cenários possíveis de emergência do SARS-CoV-2. No primeiro, o SARS-CoV-2 poderia ter surgido em seres humanos no final de 2019, próximo à época da identificação da doença. Alternativamente, um progenitor do SARS-CoV-2 poderia ter surgido e circulado em seres humanos antes de adquirir alterações genômicas que permitiram que ele causasse um grande número de casos graves e iniciasse a pandemia atual. *(62)* No entanto, nenhuma amostra coletada de seres humanos antes do final de 2019 foi considerada positiva para SARS-CoV-2; o segundo cenário possível, portanto, não tem atualmente apoio de outras linhas de evidência.

Embora as sequências de Wuhan exibam diversidade genética limitada, duas linhagens filogeneticamente distintas são aparentes, indicando um evento de separação no início da emergência do vírus. Observe-se que a distinção filogenética de linhagens não implica diferenças fenotípicas na transmissibilidade ou patogenicidade entre as linhagens, porque tais distinções geralmente emergirão por meio de processos estocásticos. Essas linhagens foram recentemente classificadas como linhagens A e B (*61)* (com menos frequência denominadas linhagens S e L) (ver Seção 6.8.7 para uma discussão mais aprofundada da nomenclatura das linhagens do SARS-CoV-2). Notavelmente, embora os vírus da linhagem B tenham sido identificados e sequenciados primeiro, *(52, 59, 60)* é provável que os vírus da linhagem A sejam ancestrais porque compartilham dois nucleotídeos com os coronavírus mais estreitamente relacionados em outros animais que não são compartilhados nos vírus da linhagem B.

Apesar das fortes medidas de quarentena adotadas na província de Hubei, ambas as linhagens foram exportadas para o resto da China e semearam várias epidemias em outros países.

5.1.3 Identificação da origem zoonótica

As sequências do genoma do SARS-CoV-2 e genomas virais correlatos de outros animais foram analisados filogeneticamente na tentativa de determinar o reservatório zoonótico do qual o SARS-CoV-2 emergiu. Até o momento, houve uma amostragem relativamente limitada com o objetivo de identificar os animais envolvidos na gênese do SARS-CoV-2 e de determinar quando, onde e como o vírus surgiu em seres humanos. Embora tenham sido coletadas amostras ambientais no mercado atacadista de frutos do mar de Huanan em Wuhan no momento de seu fechamento no início de janeiro de 2020 *(63)* e estas tenham tido resultado positivo nos testes, atualmente não está claro se essas amostras eram apenas de superfícies ou de animais presentes no mercado. Se for o primeiro, isso pode simplesmente refletir contaminação humana. Além disso, nem todos os casos iniciais puderam ser associados a esse mercado. *(61)* A identificação da origem animal da qual surgiu o SARS-CoV-2 pode ajudar a combater a disseminação de teorias da conspiração relacionadas ao seu surgimento.

Pesquisas anteriores à pandemia da COVID-19 mostraram que os betacoronavírus estão presentes em várias espécies de mamíferos e exibem uma diversidade filogenética particularmente alta em morcegos.*(64– 66)* Foi confirmado que os morcegos provavelmente desempenharam um papel na história evolutiva do SARS-CoV-2 por meio da identificação de um parente próximo do SARS-CoV-2 (denominado RaTG13) em uma espécie de morcego ferradura (*Rhinolophus affinis*) amostrada na província de Yunnan, China, em 2013. *(60)* O RaTG13 e o SARS-CoV-2 têm aproximadamente 96% de similaridade de sequência no genoma como um todo, embora isso não exclua décadas de divergência evolutiva entre eles. (67)

Outro coronavírus, o RmYN02, foi identificado em uma espécie diferente de morcego-ferradura, *Rhinolophus malayanus*, na província de Yunnan em 2019. *(68)* Embora o genoma do RmYN02 tenha passado por uma série complexa de eventos de recombinação, é o parente mais próximo do SARS-CoV-2, compartilhando uma similaridade de sequência de 97% de nucleotídeos no gene ORF1ab.

Também foram encontrados parentes próximos do SARS-CoV-2 em pangolins malaios (*Manis javanica*) recuperados em atividades anti-contrabando nas províncias de Guangdong e Guangxi, no sul da China. Os coronavírus do pangolim são mais distantemente relacionados com o SARS-CoV-2 do que o RaTG13 e o RmYN02 em seus genomas como um todo, mas compartilham uma forte similaridade de sequência com o SARS-CoV-2 no domínio de ligação ao receptor chave (RBD) do gene da espícula (S) (97,4% ao nível dos aminoácidos). *(69)*

Embora esteja claro que os betacoronavírus passaram por eventos de recombinação complexos e frequentes, e que esse processo ocorreu em vírus que estão intimamente relacionados ao SARS-CoV-2, não há evidências, no momento, de que a recombinação tenha desempenhado um papel direto no surgimento desse vírus. *(67)*

Limitações Embora o SARS-CoV-2 indubitavelmente tenha origens animais, assim como o SARS-CoV e o MERS-CoV, (64) a espécie fonte só será resolvida com amostragem adicional de uma ampla gama de animais não humanos. É possível que suas origens nunca sejam totalmente resolvidas.

5.2 Compreensão da biologia do SARS-CoV

5.2.1 Uso do receptor hospedeiro

Uma vez que os vírus podem se replicar apenas dentro das células vivas de um organismo hospedeiro, a determinação do receptor celular do hospedeiro usado pelo SARS-CoV-2 é essencial para a compreensão de sua biologia básica. A ligação ao receptor é mediada pela proteína S do vírus. Semelhanças genéticas no motivo de ligação ao receptor da proteína S entre o SARS-CoV-2 e outros coronavírus previamente investigados ajudaram a identificar o receptor celular ao qual o SARS-CoV-2 se liga e, portanto, os tipos de células que ele pode infectar. Estudos iniciais indicaram que o SARS-CoV-2 provavelmente usava o mesmo receptor celular da enzima conversora de angiotensina 2 (ACE2) que o SARS-CoV de 2002-2003, e provavelmente se ligava a esse receptor com alta afinidade. *(70, 71)* A maioria dos resíduos de aminoácidos que

são conhecidos como essenciais para a ligação do ACE2 pelo SARS-CoV são conservados no SARS-CoV-2. (70) Os ensaios *in vitro* confirmam a forte especificidade para ACE2 sugerida por estudos estruturais diretos. *(72)*.

Limitações. Foram necessários experimentos *in vitro* ou *in vivo* para a confirmação completa dos resultados da sequência genética e eles são sempre necessários para investigar qualquer alteração proposta na afinidade de ligação.

5.2.2 Evolução do SARS-CoV-2: identificação de sítios genômicos candidatos que podem causar alterações fenotípicas

Todos os vírus adquirem alterações genéticas à medida que evoluem, e a maioria das alterações genéticas adquiridas não afeta substancialmente a virulência ou a transmissibilidade. Não se pode presumir que as variantes entre os genomas virais amostrados em locais diferentes causem as diferenças epidemiológicas observadas na COVID-19 nesses locais. Em vez disso, são provavelmente estocásticas. Apesar disso, é possível que ocorra uma alteração genética que cause uma alteração fenotípica correspondente no SARS-CoV-2 que seja de importância para a saúde pública.

Estudos genômicos clínicos conduzidos de forma adequada podem ser usados para propor variantes candidatas que podem causar alterações fenotípicas do vírus observadas clinicamente, mas precisariam ser realizados estudos *in vitro* ou *in vivo* depois, para avaliar essas variantes candidatas. Também seria necessário o sequenciamento genômico viral antes e depois de tais estudos experimentais para excluir a possibilidade de que a diferença fenotípica inferida não seja impulsionada por adaptações estocásticas do vírus para a replicação dentro da cultura celular. Os fenótipos observados em cultura de células e modelos animais podem não se traduzir em alterações na doença humana.

Quando os vírus associados a diferentes fenótipos têm vários sítios que diferem entre os genomas, pode ser difícil determinar quais dessas variantes genéticas, se houver, causam a diferença fenotípica observada. As variantes genômicas identificadas podem ser investigadas por genética reversa para obter uma compreensão completa de suas características fenotípicas. A genética reversa pode envolver a indução sintética sistemática de uma alteração genética em um gene viral e a investigação do efeito fenotípico que ela causa após a produção dessa proteína. Esses experimentos só devem ser realizados sob estrita conformidade com as leis e regulamentos locais e (inter)nacionais de bioproteção e biossegurança.

Se uma alteração genética com efeito fenotípico puder ser confirmada por meio desses métodos, podem ser usados estudos epidemiológicos filodinâmicos (seção 5.4) para rastrear sua disseminação global ou local.

Limitações. É extremamente desafiador identificar e fornecer evidências de alterações genômicas que podem causar alterações fenotípicas. O sequenciamento genômico viral é uma parte necessária desses estudos, que devem ser cuidadosamente planejados e controlados a fim de

validar quaisquer efeitos hipotéticos. Estudos subsequentes *in vitro* e *in vivo* com vírus mutantes podem, em alguns casos, apoiar ainda mais as avaliações dessas hipóteses.

5.3 Melhoria do diagnóstico e da terapêutica

5.3.1 Melhoria do diagnóstico molecular

Embora o SARS-CoV-2 tenha sido identificado pela primeira vez em pacientes por meio de sequenciamento metagenômico (Seção 5.1), essa abordagem é muito demorada e cara para ser usada rotineiramente para diagnosticar a infecção viral. O desenvolvimento de testes de amplificação de ácido nucleico (NAATs) rápidos, baratos e sensíveis para a detecção molecular de rotina do SARS-CoV-2 foi, portanto, priorizado no início do surto.

O rápido lançamento público dos genomas do SARS-CoV-2 foi importante para o projeto dos NAATs. De forma mais específica, esses genomas eram necessários para o projeto de primers e sondas que se ligariam efetivamente ao ácido nucleico do SARS-CoV-2 (por meio de sequências complementares exatas ou quase exatas), mas não se ligariam a outros vírus mais circulantes, como os coronavírus que causam resfriados comuns. (73) Vários NAATs SARS-CoV-2 foram projetados e validados por diferentes grupos no prazo de alguns dias desde a primeira liberação do genoma (por exemplo, *74-76*).

Como o SARS-CoV-2 continua a adquirir alterações genéticas ao longo do tempo durante esta pandemia, a geração contínua e o compartilhamento de genomas virais serão vitais para monitorar a sensibilidade esperada dos vários ensaios de diagnóstico em diferentes locais. As incompatibilidades entre os primers ou sondas e os sítios de ligação correspondentes nos genomas do SARS-CoV-2 podem reduzir a sensibilidade do NAAT ou resultar em falsos negativos. O monitoramento será muito importante se um site variante for detectado em vírus que forem filogeneticamente relacionados. A utilização de vários alvos para detecção do SARS-CoV-2, como um PCR multiplex direcionado a duas ou mais regiões do genoma do vírus, é uma abordagem custo-efetiva para reduzir a chance de falsos negativos como resultado da evolução do vírus. Uma falha constante na detecção de um alvo em várias amostras clínicas, ou o surgimento de diferenças na sensibilidade de ensaios direcionados a regiões diferentes que não foram observadas anteriormente e que ocorrem em amostras clínicas, mas não no controle positivo estabelecido, pode ser seguida por sequenciamento genômico do vírus ou gene alvo para identificar a possível causa.

Várias plataformas existentes permitem o monitoramento de incompatibilidades entre as sequências do SARS-CoV-2 enviadas pelo usuário ou disponíveis para o público, e os sítios de ligação do primer/sonda dos NAATs comumente usados. Uma série de ferramentas foram desenvolvidas para monitorar tais incompatibilidades com primers e sondas comuns, conforme descrito em outro lugar. *(77)*

Limitações. O sequenciamento genético das regiões de ligação do primer/sonda somente é suficiente para investigar o surgimento de incompatibilidades. No entanto, o sequenciamento genômico completo permite uma investigação genômica mais ampla da propagação espaço-temporal de

um vírus que contenha incompatibilidades (por exemplo, para determinar quando a variante de incompatibilidade pode ter surgido) ou o número de vezes em que a variante possa ter emergido independentemente.

5.3.2 Apoio ao desenho e monitoramento de sensibilidade de ensaios sorológicos

Os dados da sequência genômica viral podem ser importantes para ajudar a identificar proteínas do vírus que podem ser fortemente antigênicas e para indicar como esses antígenos podem ser produzidos para ensaios sorológicos. O rastreio de peptídeos indicou que é provável que as quatro proteínas estruturais do SARS-CoV-2 – S, E, M e N – sejam as mais fortemente antigênicas. *(78, 79)* Os antígenos do SARS-CoV-2 podem ser produzidos sinteticamente para uso em ensaios comerciais. Em particular, os genes de coronavírus sintéticos que codificam as quatro proteínas podem ser inseridos em sistemas de vetores de expressão, *(80, 81)* onde as proteínas são produzidas. Esse processo depende da compreensão da sequência genômica e da estrutura das proteínas do SARS-CoV-2.

Como o SARS-CoV-2 adquire substituições genômicas, é possível que possa surgir uma linhagem com propriedades antigênicas alteradas (seção 5.2.2). Isso pode significar que os testes sorológicos deixem de detectar que um indivíduo foi infectado, porque o antígeno usado no teste é diferente daquele ao qual o indivíduo foi exposto. Além disso, os ensaios de detecção de antígeno podem ser afetados por alterações virais, pois os anticorpos de captura podem não reconhecer a proteína viral adaptada que se visa detectar. A avaliação contínua da diversidade genômica, incluindo em sítios antigenicamente importantes que podem estar sob seleção, pode ajudar a identificar sítios candidatos plausíveis que podem afetar a eficácia dos ensaios sorológicos.

Limitações. As previsões *in silico* de alteração antigênica de dados de sequência genômica são inadequadas, e a possível sensibilidade de ensaios sorológicos na detecção de infecções geneticamente diversas deve sempre ser investigada por meio de validação sorológica laboratorial.

5.3.3 Apoio ao projeto da vacina

Várias vacinas candidatas contra o SARS-CoV-2 foram projetadas e várias foram avaliadas clinicamente. *(82)* As sequências do genoma do SARS-CoV-2 têm sido usadas no projeto de vacinas candidatas que dependem da inoculação com antígenos ou mRNA/DNA para estimular, direta ou indiretamente, a produção de anticorpos e as respostas celulares. Várias vacinas iniciais de mRNA foram projetadas exclusivamente com base nos genomas do SARS-CoV-2 disponíveis ao público.

Alternativamente, genomas de coronavírus sintéticos podem ser inseridos em sistemas de vetores de expressão *(80, 81)* para produzir antígenos para vacinas (Seção 5.3.2).

Limitações. Embora as sequências genômicas possam auxiliar na concepção de vacinas candidatas, estudos e ensaios clínicos *in vivo* continuam sendo de importância crítica para avaliar a eficácia da vacina.

5.3.4 Apoio ao projeto de terapia antiviral

O desenvolvimento de novos medicamentos antivirais pode ser demorado. O reaproveitamento de medicamentos existentes para o tratamento do SARS-CoV-2 poderia encurtar significativamente o tempo necessário para obter a aprovação para uso clínico. As informações genéticas e estruturais podem revelar semelhanças nas vias proteolíticas e de replicação *(78, 79)* entre o SARS-CoV-2 e outros vírus para os quais a terapia antiviral já está disponível e, portanto, ajudar a determinar quais antivirais existentes podem ser reaproveitados. Vários medicamentos candidatos que têm como alvo proteínas virais semelhantes às do SARS-CoV-2 já foram identificados *(83)* e estão atualmente em estudos pré-clínicos e clínicos.

5.3.5 Identificação de resistência antiviral ou mutações de escape de vacina

Assim que as vacinas forem implementadas e/ou antivirais se tornarem disponíveis, o sequenciamento genômico pode ser usado para apoiar a vigilância de variantes que possam conferir resistência antiviral ou permitir o escape da vacina. O sequenciamento genômico ou genético aprofundado pode ser útil na exploração do impacto da diversidade intra-hospedeiro na resistência antiviral e no escape da vacina (se ocorrer) ou na patogênese. O sequenciamento genético de regiões específicas de interesse, como o gene da espícula, pode ser suficiente para avaliar a prevalência de variantes conhecidas específicas em regiões pré-identificadas.

Limitações. Esses estudos são extremamente complexos e exigirão investigação genômica e computacional direcionada e detalhada de vírus de pacientes com histórico de vacinação e resultados clínicos conhecidos. Embora os dados de sequência de vírus cultivados sob pressão de seleção de medicamentos possam revelar possíveis marcadores de resistência antiviral, esses marcadores devem sempre ser validados pela genética reversa para determinar suas características fenotípicas.

5.4 Investigação da transmissão e disseminação do vírus

5.4.1 Apoio ou rejeição de evidências de rotas ou *clusters* de transmissão

A colocação de sequências dentro de uma árvore filogenética pode ser usada para investigar hipóteses de rotas de transmissão. O agrupamento filogenético de sequências de pacientes expostos à mesma fonte hipotética de exposição seria consistente com essa exposição (embora não seja uma evidência forte para ela). O sequenciamento de uma proporção de casos de fora de um *cluster* hipotético, e incluindo sequências de referência global que são geneticamente mais próximas das sequências do *cluster* (para representar o contexto de diversidade genômica), pode ajudar a avaliar a probabilidade de que sequências de um *cluster* filogenético identificado com uma ligação epidemiológica hipotética sejam agrupados por acaso. Quanto maior a proporção de vírus que são sequenciados no mesmo tempo e local que os vírus de interesse, mas que não são

identificados como provavelmente parte desse *cluster*, menor a chance de que essas sequências de vírus se enquadrem em um *cluster* por acaso. Em contraste, uma considerável separação filogenética nas sequências de vírus de dois pacientes (por exemplo, colocação em diferentes linhagens bem apoiadas) indicariam que os dois pacientes adquiriram infecções de fontes diferentes.

O agrupamento filogenético tem sido usado extensivamente para investigar fontes de transmissão e eventos de exposição ao SARS-CoV-2. Em um estudo anterior, foi sugerido que o agrupamento de sequências de pacientes no navio de cruzeiro Grand Princess era condizente com uma única introdução do vírus naquele navio, seguida pela transmissão entre os passageiros. *(84)* A observação de clados monofiléticos de vírus amostrados de membros da mesma família é condizente com a transmissão direta entre os membros da família ou infecção da mesma fonte (sem amostra). A análise dos *clusters* de transmissão pode orientar a decisão sobre a necessidade de medidas de controle adicionais para evitar uma transmissão futura nas situações identificadas.

Limitações. A informação filogenética não pode ser usada para confirmar a transmissão direta do vírus entre dois pacientes ou a transmissão de uma única fonte para vários pacientes, porque o envolvimento de outros indivíduos ou fontes de exposição que não foram amostradas não pode ser descartado. A taxa evolutiva do SARS-CoV-2 implica que as substituições ocorrem em uma taxa mais lenta, em média, do que a transmissão entre pacientes e, portanto, isso continua a ser verdadeiro mesmo que as sequências genômicas amostradas sejam idênticas.

5.4.2 Identificação e quantificação de períodos de transmissão

Uma vez que haja diversidade genética suficiente dentro de uma linhagem de vírus, a taxa de alteração evolutiva (taxa de substituição) pode ser estimada (Seção 2.4). Se a taxa de substituição puder ser estimada, a diversidade genética entre dois vírus amostrados com datas de amostragem conhecidas pode ser usada para estimar o TMRCA. O TMRCA de um grupo de vírus fornece uma estimativa do limite inferior da duração de sua circulação na população amostrada. É fundamental que a duração estimada da circulação pode ser anterior à primeira identificação clínica de um caso em semanas ou meses. As abordagens filogenéticas do relógio molecular são particularmente úteis na identificação de onde a circulação não detectada (ou críptica) pode ter ocorrido e na estimativa das datas possíveis de eventos não observados.

As análises iniciais sugerem que o SARS-CoV-2 já adquiriu diversidade genética suficiente para permitir que tais abordagens de relógio molecular sejam aplicadas. *(45, 59, 85)* Consequentemente, elas foram usadas para estimar que a linhagem pandêmica do SARS-CoV-2 surgiu em seres humanos, o mais tardar, entre novembro e dezembro de 2019 *(53, 59, 85)* (Seção 5.1.2). As aplicações importantes dessas abordagens para o controle da COVID-19 incluem a identificação de transmissão local não detectada em diferentes locais. A identificação de transmissão local de longa duração não detectada clinicamente em uma área pode sugerir que locais ou populações específicas devam ser alvo de programas de vigilância diagnóstica mais extensos ou adaptados.

Limitações. A resolução temporal de eventos que podem ser investigados é limitada pela razão entre a taxa evolutiva e a taxa de transmissão. As estimativas atuais da taxa evolutiva

de SARS-CoV-2 são que, em média, uma substituição ocorre aproximadamente a cada duas semanas. Isso significa que os eventos de transmissão entre indivíduos muitas vezes não serão genomicamente resolvidos, e eventos relevantes em termos epidemiológicos, que ocorrem em uma escala de tempo mais refinada, não podem ser investigados usando-se essas técnicas. No início do surto, era difícil estimar a duração da transmissão críptica porque o SARS-CoV-2 ainda não havia acumulado diversidade genômica suficiente. Assim, era difícil determinar se um determinado genoma era o resultado de transmissão local ou uma nova introdução a partir de um local com diversidade circulante semelhante. Há estudos que sugeriram que o SARS-CoV-2 pode ter circulado sem ser detectado por semanas em Seattle (EUA) e na Itália antes da detecção clínica dos primeiros casos adquiridos na comunidade. *(84, 86)* No entanto, um estudo subsequente argumentou que a duração da transmissão críptica pode ter sido superestimada em várias semanas. *(87)*

Erros no sequenciamento ou geração de consenso podem obscurecer os sinais filogenéticos quando a diversidade verdadeira é baixa. Os erros de sequenciação também podem afetar as estimativas da variação da taxa evolutiva entre as linhagens e os tempos de divergência estimados.

A duração mínima da transmissão do vírus pode ser estimada mesmo quando bem poucos (dois ou mais) casos de uma única cadeia de transmissão forem sequenciados. No entanto, a incorporação de amostras adicionais de uma ampla área geográfica e de um grande período de tempo reduzirá o risco de que os casos amostrados se agrupem próximos a uma filogenia por acaso, de modo que a duração mínima estimada seja provavelmente mais próxima da duração real.

5.4.3 Identificação de eventos de importação e circulação local

Se estiverem disponíveis metadados sobre o local de amostragem, o sequenciamento dos genomas do SARS-CoV-2 pode ajudar a determinar se as infecções resultaram de transmissão local ou foram importadas. Essa dinâmica de transmissão pode ser interpretada cautelosa e informalmente por meio do posicionamento da sequência dentro de uma filogenia (Figura 2) ou investigada por meio de análises filogeográficas ou discretas de características mais formais, nas quais seja estimada, em termos estatísticos, a localização em cada nó interno na filogenia. A incorporação de tempos de amostragem conhecidos permite que o deslocamento espaço-temporal do surto seja reconstruído.

A inferência filogeográfica formal inclui abordagens discretas e contínuas. No primeiro caso, considera-se que as linhagens de vírus estejam se movendo entre um número fixo de locais distintos. (88) As áreas exatas são definidas pelo usuário, e podem representar países, unidades administrativas, cidades, etc., dependendo das perguntas específicas que forem feitas. Na abordagem contínua, o deslocamento da linhagem do vírus é modelado com base em processos de difusão e passeio aleatório entre coordenadas geográficas. *(89)* Tanto as investigações filogeográficas discretas quanto as contínuas podem ser conduzidas sob uma série de esquemas estatísticos, que apresentam diferentes vantagens e desafios e foram extensivamente analisados em outro lugar. (90, 91)

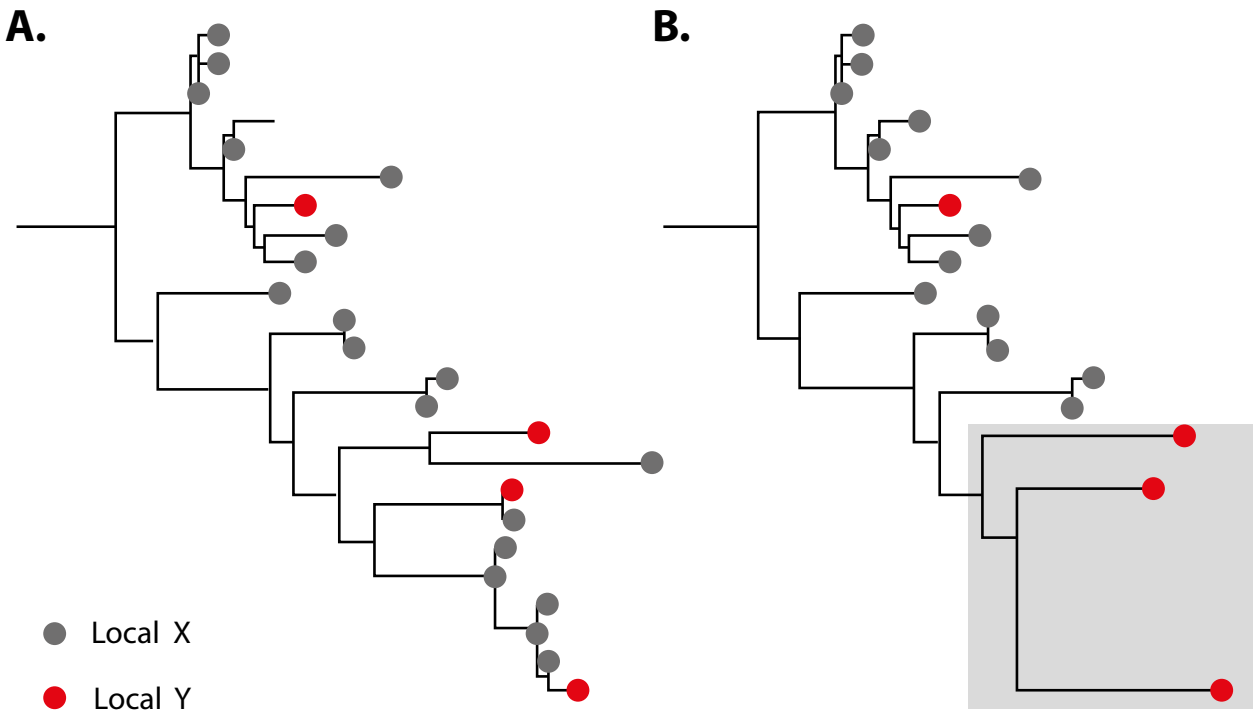

Figura 2. Efeito em potencial da amostragem genômica. As pontas em cinza representam sequências genômicas amostradas do local X, e as em vermelho representam sequências amostradas do local Y. Painel A: A árvore "verdadeira" que poderia ser reconstruída no caso de quatro eventos de introdução separados do local X ao local Y. Painel B: Amostragem insuficiente no local X do clado destacado na árvore significa que esse clado pode ser (incorretamente) inferido como transmissão local no local Y. Nesse cenário, apenas duas introduções do local X para o local Y podem ser inferidas a partir da filogenia na ausência de informações adicionais de viagens

Como a diversidade genômica do SARS-CoV-2 era baixa durante os primeiros meses da pandemia, o uso do sequenciamento genômico para rastrear sua disseminação foi amplamente limitado a introduções nacionais e regionais, em vez da transmissão comunitária. A interpretação visual informal de estruturas filogenéticas foi usada extensivamente na literatura inicial para inferir deslocamentos internacionais ou regionais. Por exemplo, a epidemiologia genômica foi usada para mostrar que muitos casos sequenciados em Connecticut (EUA) foram provavelmente importados por meio de viagens domésticas de outras partes dos EUA em vez de viagens a outros países. *(92)*

A avaliação filogeográfica da diversidade genômica pode ser usada para averiguar, por exemplo, se uma quarentena mais rigorosa de pacientes que visitaram locais específicos está efetivamente impedindo a introdução ou exportação de SARS-CoV-2 para outras regiões. Por exemplo, no Brasil, as análises filogeográficas contínuas mostraram que a disseminação do SARS-CoV-2 dentro e entre os estados brasileiros diminuiu após a implementação de intervenções não farmacêuticas. *(53)*

As abordagens filogeográficas que também incorporam tempos de amostragem permitem estimar onde e quando eventos de deslocamento de linhagem de vírus podem ter ocorrido. A duração da persistência do vírus, o número de introduções e a dimensão relativa do surto podem ser

determinados para cada local e, portanto, podem ser usados para identificar locais específicos nos quais as medidas de controle precisam ser reforçadas.

Pode ser útil estudar vírus em viajantes que retornam para ajudar a reconstruir a epidemiologia do SARS-CoV-2 no país em que a infecção foi adquirida. *(93)* Novas abordagens permitem que o histórico de viagens do paciente e sequências de locais não amostrados sejam incorporados em análises filogeográficas discretas, permitindo, assim, que padrões filogeográficos mais realistas sejam revelados e o efeito da amostragem global com viés seja avaliado. *(94)*

Limitações. As reconstruções filogeográficas são frequentemente exigentes do ponto de vista computacional. Estratégias de subamostragem cuidadosamente consideradas podem ajudar a reduzir essa carga computacional (seção 6.8.1). A dispersão de patógenos humanos nem sempre é bem capturada por esses processos. No entanto, nos lugares onde escalas geográficas e viagens de longa distância são restritas, passeios aleatórios podem capturar adequadamente o deslocamento do SARS-CoV-2. Deve-se considerar cuidadosamente a adequação de um processo contínuo para SARS-CoV-2, pois o uso de modelos de difusão inadequados pode levar a conclusões incorretas.

A forma como as sequências do genoma do vírus são amostradas pode influenciar fortemente as conclusões das análises filogeográficas. Portanto, é extremamente importante conduzir análises e interpretar os resultados com cautela, e o ideal seria envolver especialistas com experiência nesses métodos. Há várias maneiras pelas quais as análises filogeográficas podem deixar de capturar os "verdadeiros padrões" de propagação, incluindo as seguintes.

- A subamostragem de genomas virais pode levar a uma subestimação do número de introduções (e, portanto, a uma superestimativa da extensão da transmissão comunitária) (Figura 2). Isso foi bem destacado por Lu et al. *(59)*, que mostraram que um único *cluster* de sequências intimamente relacionadas, amostradas de pacientes em Guangdong, na verdade representava várias introduções independentes por meio de viagens. O histórico de viagens dos pacientes é uma informação importante que deve ser usada para apoiar os achados filogenéticos sempre que possível (o compartilhamento apropriado de dados e a proteção do anonimato do paciente são discutidos na Seção 4).
- Para análises filogeográficas discretas, a localização dos vírus ancestrais só pode ser inferida de forma confiável a partir do conjunto de localizações nas quais os vírus amostrados foram observados.*(89, 95)* Consequentemente, usando apenas os dados genômicos, em geral é impossível distinguir entre a transmissão direta entre dois locais e a transmissão indireta por meio de um local intermediário em que nenhum genoma foi produzido. A distinção entre esses cenários só é possível em raras situações nas quais as informações de viagem são conhecidas. *(94)*
- Para certas análises filogeográficas discretas (particularmente aquelas baseadas em análises de características discretas em vez de modelos coalescentes ou de nascimento-óbito), os locais que têm um número maior de sequências genômicas associadas a eles são mais propensos a serem reconstruídos como locais doadores a partir dos quais um vírus subsequentemente se espalha.*(96)* A amostragem reduzida de dados de sequência genômica disponíveis de locais super-representados pode ser útil para investigar se as conclusões têm a probabilidade de ser relativamente robustas a esse respeito. *(97)* O uso de estatísticas de apoio de fator de Bayes

ajustadas *(98)* pode fornecer ajuda adicional, para determinar se os eventos de transição são apoiados devido ao viés de amostragem geográfica.
- A amostragem de apenas certas áreas de um surto pode resultar em reconstruções imprecisas do histórico de dispersão e das estimativas de velocidade de dispersão dentro de um esquema filogeográfico contínuo. Estão atualmente sendo avaliadas formas de reduzir o impacto da amostragem com viés. *(99)*
- As informações sobre a localização do paciente costumam ser limitadas à subunidade administrativa, por exemplo, município. Frequentemente, é apropriado considerar a incerteza associada a um local de amostragem ao usar a abordagem filogeográfica contínua. Por exemplo, em vez de usar as coordenadas geográficas da cidade mais próxima, todo o polígono correspondente ao município pode ser usado para definir uma área a partir da qual as coordenadas para aquela amostra podem ser selecionadas aleatoriamente. Também é útil a repetição dessa amostragem aleatória durante a análise. *(100, 101)*

5.4.4 Avaliação dos impulsionadores de transmissão

Os métodos descritos na seção anterior também podem ser usados para investigar os fatores que impulsionaram a dispersão do vírus. *(97)* Em modelos filogeográficos discretos (incluindo aqueles implementados como modelos estruturados coalescentes e de nascimento-óbito de vários tipos), as informações sobre pares de áreas definidas são usadas como um preditor da taxa de migração da linhagem do vírus entre essas áreas. As informações podem incluir características de mobilidade humana da população, como densidade e proximidade geográfica. Os eventos de dispersão inferidos pela reconstrução filogeográfica contínua também podem ser analisados para determinar se eles são influenciados pelo "panorama" de fatores ambientais ou humanos através dos quais eles ocorrem.

No momento em que este artigo foi escrito, essas análises ainda não haviam sido aplicadas ao SARS-CoV-2. A identificação das causas de transmissão pode ajudar a formar novas estratégias para prevenir a propagação. Por exemplo, para o vírus Ebola, esse método foi usado para estabelecer que o vírus tinha maior probabilidade de se espalhar entre países que compartilham fronteiras terrestres, *(102*) sendo subsequentemente utilizado para avaliar o efeito das medidas adotadas.*(103)*

Limitações. Essas abordagens são computacionalmente exigentes, envolvendo grandes conjuntos de dados de milhares de genomas e levando dias ou semanas para serem concluídas. O uso de uma distribuição pré-estimada de árvores empíricas pode reduzir o tempo computacional necessário e é particularmente apropriado para a exploração de dados preliminares. A subamostragem de clados específicos ou uma subamostragem aleatória também pode reduzir a carga computacional (Seção 6.8.1). Também há um limite computacional para o número de áreas definidas que podem ser incluídas no modelo.

Alguns modelos têm flexibilidade relativamente limitada na identificação de fatores que possam estar resultando na transmissão do SARS-CoV-2 em momentos e lugares diferentes. Modelos de época implementados dentro de um esquema filogeográfico discreto *(104)* podem ser apropriados para investigar efeitos variáveis no tempo de diferentes fatores nos quais períodos de tempo

significativos possam ser predefinidos. No entanto, as medidas de controle estão mudando rapidamente em muitos países em épocas diferentes. Isso pode limitar a capacidade de definir épocas epidemiologicamente úteis ao aplicar essas técnicas acima da escala nacional ou regional. É necessária experiência substancial para especificar esses modelos de forma adequada e para interpretar as estimativas resultantes.

É provável que as técnicas para avaliar os efeitos das intervenções sejam aplicadas retrospectivamente, talvez meses após a intervenção. As análises do efeito das intervenções que tiveram sucesso na redução de casos podem ajudar a orientar estratégias futuras em países nos quais o surto esteja progredindo.

A amostragem com viés pode afetar os resultados (Seção 5.4.3).

5.4.5 Discernimento do envolvimento de outras espécies

Várias espécies de animais não humanos podem ser infectadas naturalmente com o SARS-CoV-2, incluindo gatos, cães e visons. *(105–107)* Nos lugares onde são observados pares epidemiologicamente ligados de ser humano infectado e animal infectado, não é possível determinar a direção da infecção entre eles. Quando vários animais estão infectados, podem ser usadas investigações filogenéticas de *clusters* para demonstrar que os animais foram infectados por diferentes rotas, como foi feito para o vison em duas fazendas da Holanda. (106) Um forte apoio (*bootstrap* ou apoio posterior elevado) para a colocação de uma sequência de genoma do SARS-CoV-2 amostrada de um ser humano em um *cluster* de múltiplas sequências amostradas de visons seria condizente com a hipótese de que seres humanos estejam se infectando a partir de animais. Se a ordem de ramificação não for fortemente apoiada, a direcionalidade não pode ser inferida de maneira robusta. Também podem ser utilizadas metodologias mais extensas que empregam a reconstrução de características ancestrais discretas formais (Seção 5.4.3).

5.4.6 Discernimento das cadeias de transmissão entre pacientes usando diversidade viral intra-hospedeiro

Já foi mencionado que as substituições de nucleotídeos parecem ocorrer aproximadamente a cada duas semanas para o SARS-CoV-2, e será um desafio responder a perguntas epidemiológicas em uma escala de tempo mais precisa. Para outros vírus, a variação genética intra-hospedeiro entre os vírions tem sido usada para aumentar a resolução na qual a transmissão pode ser inferida filogeneticamente. Variantes minoritárias de vírus intra-hospedeiro (variantes que ocorrem em baixa frequência dentro de um indivíduo) que são transmitidas entre pacientes fornecem informações que são obscurecidas pelo genoma de consenso. A análise dessas variantes tem sido usada para melhorar a compreensão das vias de transmissão de muitos vírus diferentes. *(108, 109)*

Existe variação intra-hospedeiro em coronavírus que estão intimamente relacionados ao SARS-CoV-2, como o MERS-CoV. (110) Embora os dados atuais (limitados) apoiem a existência de variação genética intra-hospedeiro no SARS-CoV-2, até o momento, há bem poucos conjuntos de dados de variação dentro do hospedeiro de *clusters* epidemiológicos

conhecidos que poderiam ser usados para determinar se essa variação é transmitida entre pacientes. *(111)* Caso contrário, o uso dessas técnicas não seria possível.

São necessárias análises bioinformáticas e filogenéticas especializadas para analisar a variação intra-hospedeiro do vírus. Dada a atual falta de compreensão da magnitude da variação intra-hospedeiro do SARS-CoV-2 ou de sua transmissibilidade, essas análises especializadas não são abordadas aqui.

Limitações. Muitos conjuntos de dados de sequência genômica viral não serão apropriados para essas análises. O sequenciamento Sanger ou o sequenciamento de nova geração que usa dispositivos com altas taxas de erro de sequenciamento por leitura sem replicação *(112)* não fornecem informações suficientes sobre as variações intra-hospedeiro. O ruído causado pela contaminação da amostra cruzada e erros de sequenciamento também pode obscurecer os sinais verdadeiros.

5.5 Parâmetros epidemiológicos inferidos

5.5.1 Número de reprodução

O número de reprodução, R_0, pode ser estimado usando modelagem genética populacional, como modelo coalescente, modelo coalescente estruturado e modelo de amostragem nascimento-morte. Essas abordagens filodinâmicas são todas baseadas no conceito de que parâmetros epidêmicos, como o R_0, afetam a forma das filogenias resolvidas no tempo. As várias abordagens são baseadas em suposições diferentes, têm requisitos de dados ligeiramente diferentes e são suscetíveis a diferentes formas de viés. Também são apropriadas em diferentes pontos da epidemia, dependendo da extensão da distribuição geográfica e da população em estudo.

Nos estágios iniciais da pandemia do SARS-CoV-2, a estrutura geográfica da população pode ser amplamente ignorada; as estimativas do R_0 se basearam em dados de sequência amostrados no mundo todo, sob a aproximação de que todos os casos estavam apenas a algumas gerações de distância da epidemia original em Hubei, China. (113) Sob essas condições de amostragem, podem ser aplicados modelos de nascimento-morte *(114),* e modelos coalescentes que tenham como premissa uma única população panmítica (mistura aleatória).

Como o SARS-CoV-2 se dispersou globalmente, tornou-se possível e apropriado estimar o R_0 em diferentes países, regiões e cidades. Uma vez que a estruturação de clados geográficos substanciais indicativos da predominância da transmissão intrarregional foi filogeneticamente aparente (Seção 5.4.3), a amostragem de nascimento-óbito e de modelos coalescentes baseados em uma população panmítica se tornou inválida. Os métodos foram então aplicados no nível dos *clusters* filogenéticos identificados individualmente que representam uma linhagem que esteja circulando na comunidade. Isso requer uma definição *a priori* dos clados filogenéticos.

É possível usar modelos genéticos populacionais mais complexos que levem em conta múltiplas importações de linhagens SARS-CoV-2 e transmissão comunitária; esses modelos não exigem uma definição *a priori* de *clusters*. Essas análises são possíveis usando modelos estruturados

coalescentes ou de nascimento-óbito de vários tipos, (85, *113)* que potencialmente fazem uso de metadados mais clínicos e demográficos que influenciam as taxas de transmissão ou padrões de transmissão. Seu desenvolvimento e implementação exigem considerável experiência e um bom domínio de modelagem epidemiológica.

Os requisitos computacionais são bem maiores do que para muitas outras aplicações filogenéticas ou filodinâmicas.

Limitações. Os profissionais devem estar cientes da robustez dos diferentes métodos no tocante às diferentes formas de viés. Todos os métodos são falíveis na presença de amostragem tendenciosa, como ocorre no sequenciamento de cadeias de transmissão identificadas por meio de rastreamento de contatos ou pequenos *clusters* identificados epidemiologicamente. A especificação incorreta do modelo é uma fonte de viés para todos os métodos. Isso é melhorado com métodos coalescentes estruturados mais complexos, mas eles exigem mais trabalho computacional. Os métodos individuais são afetados de forma diferente por diferentes fontes de viés em potencial.

- Os modelos coalescentes baseados em relações determinísticas entre o R_0 e o modelo demográfico podem fornecer uma estimativa com viés de R_0 quando a magnitude da epidemia é pequena ou o R_0 está próximo de 1 *(115)* e predominam efeitos estocásticos.
- Os modelos de amostragem de nascimento-óbito exigem uma parametrização apropriada da variação da taxa de amostragem ao longo do tempo. (116)
- Uma vez que muitos países têm testado ativamente o SARS-CoV-2 desde antes do início de seus surtos, pode ser sensato supor que a proporção da amostra seja maior do que zero para toda a duração abrangida pela análise. No entanto, se as estratégias de teste mudaram em algum ponto durante esse período, a proporção de amostragem precisará variar de maneira semelhante. Os modelos coalescentes podem fornecer estimativas mais precisas do que a amostragem de modelos de nascimento-óbito se a taxa de amostragem variar ao longo do tempo.
- As análises baseadas em *clusters* identificados *a priori* não podem ser consideradas representativas da comunidade como um todo porque negligenciam pequenas cadeias de transmissão que não são amostradas ou estão abaixo do limite de tamanho exigido para análise. Assim, os clados observados são aqueles que cresceram com mais sucesso. Os números de reprodução dos *clusters* são provavelmente maiores nesses clados do que na comunidade como um todo.
- Ao configurar qualquer análise que presuma a ausência de uma população estruturada, é fundamental garantir que haja apenas um parâmetro R_0 dentro do período de tempo abrangido pela árvore que relaciona as amostras. Se uma quarentena ou outras medidas foram introduzidas durante o período de estudo, será necessário excluir as sequências coletadas após a implantação dessas medidas ou incluir todas as sequências, mas permitir que o parâmetro R_0 mude ao longo do tempo.

Muitas abordagens, inclusive os modelos de nascimento-óbito que são implementados no pacote do software Birth Death Skyline Model (BDSKY), *(114)* exigem a incorporação explícita de informações prévias para fixar certos parâmetros a valores conhecidos e, portanto, melhorar a tratabilidade computacional. Normalmente, é comum fixar um parâmetro que possa ser verificado a partir de dados clínicos, como a taxa em que os indivíduos infectados se tornam

não infecciosos. A especificação prévia do parâmetro deve ser conduzida com cuidado para evitar possíveis fontes de viés. A realização de análises que utilizam especificações anteriores alternativas pode ajudar a determinar até que ponto os resultados filodinâmicos são sensíveis ao parâmetro anterior especificado.

5.5.2 Escala de surto ao longo do tempo e proporção de notificações de infecção por caso

Na genética populacional tradicional, o tamanho efetivo da população (o número de indivíduos de uma população que contribui com progênie para a nova geração) é estimado em vez do tamanho absoluto da população de vírus (número total de vírions) ou número de indivíduos infectados (tamanho da epidemia). O tamanho efetivo da população pode ser usado para identificar alterações relativas no tamanho da epidemia ao longo do tempo, caso sejam satisfeitas certas condições. Apenas recentemente foi tentado estimar o tamanho absoluto da epidemia a partir de dados genéticos e essa é uma área ativa do desenvolvimento metodológico filodinâmico. Uma variedade de métodos experimentais foi aplicada na atual epidemia da COVID-19. Em geral, qualquer método para reconstruir o tamanho da epidemia deve levar em conta os principais fatores que influenciam a diversidade genética dentro da estrutura de amostragem, incluindo: estrutura geográfica, variação nas taxas de transmissão, crescimento exponencial e dinâmica populacional não linear e a distribuição do tempo de geração. *(117)*

Três abordagens diferentes e suas limitações são destacadas a seguir.

- Em algumas situações, o tamanho efetivo da população estimado com modelos coalescentes pode ser traduzido em tamanho da epidemia. Por exemplo, Koelle & Rasmussen derivaram uma fórmula para fazer isso que utiliza estimativas independentes do R_0 e da variação nas taxas de transmissão em equilíbrio epidêmico. *(118)* Isso foi posteriormente ampliado para uma situação com aumento exponencial por Li, Grassly e Fraser. *(117)*

  Limitações. Esta última abordagem é limitada ao período epidêmico inicial com crescimento exponencial e ambas as abordagens podem ser inadequadas quando há uma estruturação geográfica ou demográfica substancial na transmissão do vírus. Por exemplo, nos lugares onde a transmissão do vírus ocorre separadamente em dois locais diferentes sem transmissão substancial entre os locais, dois valores de R_0 diferentes podem ser necessários.

- Sob um esquema de nascimento-morte tal como o BDSKY, a proporção de amostragem pode ser inferida e pode ser combinada com o número de sequências para produzir uma estimativa bruta do número cumulativo de casos.

  Limitações. Embora talvez seja um meio útil de obter uma estimativa rápida, essa abordagem é limitada, particularmente para tamanhos de amostra pequenos, pois ignora o efeito da estocasticidade no procedimento de amostragem. É aplicável a uma população não estruturada/panmítica, como em um único *cluster* filogenético ou no início da epidemia. Essas abordagens não levam em consideração a alta variação nas taxas de transmissão. Há abordagens menos

limitadas, como o uso de filtragem de partículas para amostrar a curva de prevalência absoluta diretamente como parte da inferência de nascimento-morte. *(119)*

- Os modelos coalescentes estruturados que são implementados no pacote PhyDyn *(120)* para o software filogenético BEAST2 foram desenvolvidos para estimar o tamanho da epidemia levando em consideração variáveis como estrutura geográfica, dinâmica não linear e alta variação nas taxas de transmissão.

  Limitações. Esses métodos exigem experiência em modelagem epidemiológica e têm elevados requisitos computacionais. Fatores como seleção natural, estrutura geográfica não modelada ou recombinação genômica ainda podem confundir as estimativas.

6 Orientação prática sobre aspectos técnicos de sequenciamento genômico e análise do SARS-CoV-2

As considerações gerais para a implementação de um programa de sequenciamento foram discutidas na Seção 3. Esta seção enfoca os diferentes aspectos técnicos dos projetos de sequenciamento genômico para COVID-19.

6.1 Estratégias de amostragem de genoma e desenho de estudo

As estratégias de amostragem do genoma dependerão das respostas buscadas. Por exemplo, a investigação da transmissão nosocomial ou a avaliação dos achados do rastreamento de contatos (Seção 5.4.1) podem exigir amostragem genômica extensa da maioria dos pacientes identificados no *cluster* epidemiológico de interesse, bem como amostras que não fazem parte do *cluster* que está sendo investigado. As amostras de fora do *cluster* são importantes para apoiar a hipótese de que as amostras de *cluster* estão epidemiologicamente mais ligadas umas às outras do que a outras infecções da comunidade. Por outro lado, as abordagens filodinâmicas (seções 5.4.2-5.5 e Tabela 1) são facilmente influenciadas pela amostragem não aleatória de todos os casos confirmados, mas, em geral, toleram uma amostragem de alguma forma esparsa, de uma proporção baixa de todos os casos. Em particular, os modelos filodinâmicos pressupõem que as sequências sejam coletadas uniformemente ao acaso de cada compartimento no modelo subjacente. Essa suposição pode ser facilmente violada se, por exemplo, as amostras forem coletadas como resultado do rastreamento de contatos.

Para abordagens filodinâmicas, os genomas virais idealmente devem, portanto, ser sequenciados em proporção à incidência de casos reais. A melhor forma de aproximar isso na prática pode variar. Nos lugares onde a cobertura diagnóstica é boa em toda uma região, pode ser sequenciado um subconjunto aleatório de amostras diagnósticas residuais positivas. No entanto, em muitos lugares, os diagnósticos clínicos são realizados de forma não aleatória, inclusive quando se utiliza um extenso rastreamento de contatos para identificar os casos. A proporção de casos a partir dos quais estão disponíveis amostras clínicas pode mudar com o tempo, conforme são implementados diferentes esquemas de amostragem. Em alguns países, as amostras positivas não refletem a verdadeira distribuição da infecção devido às disparidades de recursos ou acessibilidade entre os locais (por exemplo, há desproporcionalmente menos amostras das áreas rurais devido à dificuldade no transporte de amostras para os testes centralizados). Nesses países, pode ser mais apropriado selecionar deliberadamente um conjunto de amostras para sequenciamento que compense vieses conhecidos na amostragem. Por exemplo, se a notificação de casos suspeitos for mais representativa do que a notificação de casos confirmados, pode ser apropriado selecionar amostras de diferentes horários e locais em proporção ao número de casos suspeitos, em vez de em proporção ao número de casos confirmados. Não é possível dar recomendações universalmente apropriadas para o sequenciamento do SARS-CoV-2, pois as decisões dependerão do contexto do surto e das perguntas a serem respondidas. Os principais requisitos estão listados na Tabela 1. Além disso, o Anexo 1 destaca os tipos de estratégias de amostragem que foram usados em outros surtos virais para as aplicações filodinâmicas específicas consideradas no Quadro 1. No entanto, os números de amostra necessários para o SARS-CoV-2 serão diferentes daqueles apresentados por causa das diferenças na diversidade viral de base, no comprimento do genoma, na taxa de substituição e na dinâmica de transmissão.

Tabela 1. Considerações de amostragem genômica e dados para algumas aplicações

Aplicação	Sequência mínima de metadados	Metadados adicionais ideais	Considerações de amostragem de sequências no local de destino	Outros dados necessários
Investigação de *clusters* de transmissão (Seção 5.4.1)	• Definições de *clusters* hipotéticos (por exemplo, *cluster* familiar) • Data e local da amostragem		Amostragem densa de todos ou da maioria dos indivíduos do *cluster* de transmissão previsto **e** amostragem densa de indivíduos controle do mesmo local e tempo.	Sequências virais controle de indivíduos não vinculados em local e tempo semelhantes. As sequências apropriadas provavelmente serão geradas localmente, em vez de estarem disponíveis por meio de repositórios de compartilhamento de sequência.
Duração da transmissão (Seção 5.4.2)	Data e local da amostragem.	História de viagens nos últimos 14 dias	O ideal é que os genomas virais sejam sequenciados proporcionalmente ao número de casos COVID-19. Pode ser informativo com bem poucas sequências (> 1) do local de interesse. As estimativas geralmente são mais precisas à medida que a densidade da amostragem genômica e a diversidade aumentam.	As sequências de outros locais fora do local sob investigação às vezes podem ser obtidas por meio de repositórios de compartilhamento de sequência (Seção 4)
Eventos de importação e transmissão local (Seção 5.4.3)				
Avaliação dos impulsionadores de transmissão (Seção 5.4.4)			• Idealmente, os genomas virais devem ser sequenciados de forma proporcional ao número de casos COVID-19. • Normalmente, centenas de sequências são necessárias ao longo de vários meses.	• As sequências de outros locais fora do local sob investigação às vezes podem ser obtidas por meio de repositórios de compartilhamento de sequência (Seção 4). • São necessárias fontes adicionais de dados epidemiológicos, populacionais e/ou ambientais, as quais frequentemente estão disponíveis no setor público ou privado.
Inferência de R_0 (Seção 5.5.1)				• São necessárias sequências de outros locais fora do local sob investigação para averiguar a estrutura geográfica; às vezes elas podem ser obtidas por meio de repositórios de compartilhamento de sequência (Seção 4).
				• Tempos de geração ou intervalos de série
				• Conhecimento de medidas que possam ter alterado substancialmente o R_0 no período de tempo, por exemplo: o período de tempo de quarentena.

6.2 Metadados apropriados

Para garantir que os dados genômicos do SARS-CoV-2 sejam tão úteis quanto possível, eles devem ser acompanhados por metadados apropriados. A curadoria de metadados e seu compartilhamento local ou público pode ser demorado, mas ambos fazem parte de qualquer pipeline de sequenciamento. Os recursos necessários devem ser alocados quando o estudo estiver sendo desenhado.

Os metadados devem incluir, no mínimo, a data e o local da coleta da amostra. No entanto, a liberação de metadados adicionais aumenta muito as possíveis aplicações de uma sequência genômica. Sempre que possível, portanto, devem ser incluídas as informações sobre o tipo de amostra e sobre como a sequência foi obtida no laboratório (Tabela 2). As amostras duplicadas do mesmo indivíduo ou as sequências duplicadas da mesma amostra devem ser claramente identificadas. É incentivada a divulgação de informações demográficas e clínicas, como idade, sexo, presença de comorbidades, gravidade e desfecho da doença e links para outras sequências no banco de dados, quando essas informações não ofereçam risco de identificação do paciente.

Um consenso global sobre formatos específicos para metadados (como data) permitiria que dados de sequência genômica de muitos laboratórios diferentes fossem rapidamente compilados em conjuntos de dados maiores e reduziria ambiguidades. Alguns repositórios de genoma de consenso, incluindo o GISAID, já impõem restrições de formato em certos campos. Se os repositórios de dados ainda não impuserem formatos, são sugeridas as restrições de formato para SARS-CoV-2 mostradas na Tabela 2. A Tabela 2 também destaca exemplos de análises que exigem o fornecimento de metadados específicos.

A OMS incentiva fortemente o rápido compartilhamento público de sequências e metadados (Seção 4). No entanto, é vital proteger o anonimato do paciente. Os laboratórios devem ponderar cuidadosamente se os pacientes podem vir a ser identificados caso todos os metadados disponíveis sejam compartilhados ao mesmo tempo. Nos lugares onde poucos casos COVID-19 foram observados, há um risco maior de se comprometer o anonimato do paciente e, portanto, menos dados podem ser compartilhados. No entanto, quando for considerado impróprio compartilhar metadados detalhados por meio de repositórios disponíveis ao público, pode ser apropriado conceder acesso a um pequeno número de usuários por meio de plataformas seguras desenvolvidas localmente.

Nos lugares onde não for possível compartilhar todos os metadados sem arriscar a confidencialidade do paciente, devem ser preferencialmente compartilhados os dados que forem mais úteis para estudos globais. Por exemplo, o local de amostragem, a data e o histórico de viagens são mais úteis para estudos filodinâmicos do que a idade ou sexo do paciente (Tabela 2).

Alguns laboratórios optam por adicionar ruído às datas fornecidas para diminuir a chance de identificação dos pacientes. Isso pode ser feito por vários métodos, tais como, por exemplo, a escolha de uma data falsa dentro de um período de 5 dias antes ou depois da data da coleta da amostra ou a utilização da data do sequenciamento como a data da amostra. Essas práticas afetam negativamente a inferência filogenética com base no relógio molecular e de preferência devem ser evitadas. Se, no entanto, essa prática for seguida, as informações sobre como exatamente a nova data foi selecionada devem ser fornecidas como observação.

Tabela 2. Formato e uso de metadados[a]

Tipo de metadados	Formato recomendado se aplicável	Análises para as quais os metadados são necessários
Metadados específicos da amostra		
Data de coleta de amostra	*AAAA-MM-DD* Se a data da amostragem não estiver disponível, a data de recebimento pelo laboratório de ensaio pode ser adotada como alternativa, mas isso deve ser claramente indicado	Filogenias de relógio molecular (incluindo quaisquer modelos implementados no BEAST ou BEAST2) Estas podem fornecer estimativas das datas de introdução, alterações no tamanho do surto ao longo do tempo e na taxa evolutiva
Localização	*Continente/país/região/cidade* Para análises filogeográficas discretas (Seção 5.4.3), a resolução de localização pode ser baixa (por exemplo, informações em nível de país para consideração de deslocamento entre países), mas dados de resolução mais alta são preferíveis para permitir análises em escala mais precisa Abordagens filogeográficas contínuas em geral exigem dados de resolução relativamente alta (por exemplo, cidade ou município)	Qualquer interpretação filogenética da propagação global ou regional do vírus (incluindo modelos em BEAST ou BEAST2)
Hospedeiro	Por exemplo, ser humano ou ***Mustela lutreola***	Variação de hospedeiros e evolução do vírus
Idade do paciente	Para seres humanos, forneça a idade em anos (por exemplo, 65) ou a idade com a unidade se for inferior a 1 ano (por exemplo, 1 mês, 7 semanas) Para animais não humanos, ***jovem*** ou ***adulto***	Epidemiologia descritiva ou como uma possível característica para inferência filodinâmica discreta
Sexo	***Masculino, feminino*** ou ***desconhecido***	Epidemiologia descritiva
Informações adicionais do hospedeiro	Sem formato padrão Para animais, isso pode incluir o contexto, como "***doméstico – fazenda***", "***doméstico – casa***", "***selvagem***", etc.	Vigilância de doenças em hospedeiros humanos ou animais
Histórico de viagens	Sem formato padrão O histórico de viagens nos 14 dias anteriores ao início dos sintomas deve ser obtido dos pacientes, sempre que possível Pode ser importante a divulgação deliberada do histórico de viagens somente em uma resolução baixa (por exemplo, país) para proteger a confidencialidade do paciente	Análises filogeográficas ou filodinâmicas direcionadas à estimativa de taxas de transmissão ou rotas entre regiões
Nome do ***cluster*** ou isolado	Sem formato padrão Os formatos apropriados podem incluir "Mesmo ***cluster*** epidemiológico da amostra X", "Mesmo paciente da amostra X" ou "Amostra do paciente XYZ" (onde XYZ é um identificador anônimo que não pode ser rastreado até o paciente ou usado para acessar outros dados do paciente que possam comprometer a confidencialidade)	Amostragem filogenética para garantir a adequação dos modelos filodinâmicos Investigação de ***cluster***
Data de início dos sintomas	*AAAA-MM-DD*	Aplicações filodinâmicas especializadas que investigam ***clusters*** de transmissão
Sintomas	Sem formato padrão Grau apropriado de sintomas; pode incluir "grave", "leve" e "fora do normal"	Epidemiologia descritiva

Desfecho clínico, se conhecidos	Sem formato padrão Os formatos apropriados podem incluir "recuperado", "óbito" e "desconhecido"	Epidemiologia descritiva
Comentários	Sem formato padrão Os comentários apropriados podem incluir como as amostras foram selecionadas (por exemplo, "investigação de ***cluster***", "aleatoriamente") ou o local de armazenamento de outros arquivos de dados, como dados lidos brutos	Interpretação da qualidade ou utilidade dos dados
Metadados específicos de sequência e amostra: dados extensos devem ser compartilhados, já que o anonimato do paciente normalmente não é afetado		
Fonte da amostra, tipo de amostra	Sem formato padrão Exemplos: "Expectoração", "sangue", "soro", "saliva", "fezes", "swab nasofaríngeo"	Efeito do tropismo celular
Detalhes de passagem, história	Sem formato padrão É importante indicar que a cultura de células foi realizada (por exemplo, "Cultivada"); idealmente, essas informações devem incluir o tipo de células usadas e o número de passagens	Remoção de vírus cultivados em células (que possam ter induzido alterações genéticas)
Tecnologia de sequenciamento	Sem formato padrão Idealmente, deve incluir a abordagem de laboratório e a plataforma de sequenciamento (por exemplo, "Metagenômica em Illumina HiSeq 2500" ou "Esquema de primer ARTIC PCR em ONT MinION")	Artefatos de sequenciamento
Método de montagem, método de geração de consenso	Sem formato padrão	Artefatos de sequenciamento
Profundidade mínima de sequenciamento necessária para sítios de vocalização durante a geração da sequência de consenso	por exemplo: 20x	Artefatos de sequenciamento

[a] O compartilhamento de todas as informações listadas nesta tabela pode comprometer o anonimato do paciente. Uma revisão ética deve ser conduzida para determinar quais metadados podem ser compartilhados com segurança. Pode ser apropriado compartilhar menos dados em bancos de dados públicos do que em bancos de dados mantidos e analisados localmente.

6.3 Considerações logísticas

6.3.1 Localização

A decisão sobre onde estabelecer um laboratório de sequenciamento deve ser cuidadosamente ponderada. O sequenciamento geralmente deve ser realizado por instituições com a experiência e a infraestrutura necessárias para o sequenciamento de nova geração. Se essa infraestrutura não estiver disponível, a decisão de onde estabelecer o laboratório de sequenciamento deve levar em consideração o impacto em outros trabalhos realizados pelo laboratório. Por exemplo, a integração do sequenciamento em um laboratório de diagnóstico existente pode permitir um tempo de resposta mais curto, mas esse ganho em potencial deve ser balanceado com o risco de interrupção de outras operações do laboratório, que já podem estar em processo de ampliação de sua capacidade de diagnóstico para o SARS- CoV-2. Uma consideração cuidadosa também deve ser dada à disponibilidade de espaço e equipamentos.

Quando o manuseio de amplicons de PCR for necessário para o sequenciamento (por exemplo, métodos descritos na Seção 6.5.4), é importante reduzir a possibilidade de contaminação do amplicon por meio de uma gestão laboratorial adequada. A separação física das áreas que serão usadas para o manuseio pré e pós-PCR do material SARS-CoV-2 e um fluxo unilateral de pessoal e materiais das áreas pré e pós-PCR são fortemente recomendados. Nos lugares onde ainda não houver áreas separadas disponíveis, os laboratórios podem adotar estratégias, como a compra e uso de caixas de luvas separadas ou para atividades pré ou pós-PCR. O equipamento deve ser idealmente desenhado para uso apenas com material pré ou pós-PCR e os reagentes necessários devem ser armazenados separados (por exemplo, em congeladores diferentes ou em diferentes laboratórios) para reduzir o risco de contaminação. Como em todos os sequenciamentos, os controles negativos são importantes para detectar contaminações.

6.3.2 Bioproteção e biossegurança

Sempre devem ser realizadas avaliações de risco para averiguar a bioproteção e a biossegurança. Os resultados dessas avaliações de risco devem ser comunicados aos profissionais envolvidos nos processos relevantes.

Os laboratórios individuais devem sempre conduzir avaliações de risco locais para cada etapa de seu protocolo SARS-CoV-2. A legislação internacional, nacional e local deve ser consultada para garantir o manuseio seguro do material SARS-CoV-2. A OMS publicou diretrizes gerais de biossegurança. *(121)*

As amostras devem ser inativadas o mais cedo possível (geralmente antes da extração do RNA) usando métodos químicos que preservem a qualidade do RNA. Os métodos usados para extrair o RNA antes dos NAATs de diagnóstico são, na maioria das vezes, apropriados para o sequenciamento. Como para a maioria dos NAATs, a inativação por calor antes da extração da amostra não é recomendada por causa do risco de danificar a integridade do RNA.

6.3.3 Considerações éticas

Devem ser realizadas análises éticas para garantir que os pacientes deem consentimento apropriado para a coleta e sequenciamento das amostras e para ponderar o uso, o armazenamento e a publicação subsequentes dos dados.

Algumas abordagens de sequenciamento, como a metagenômica, geram dados genômicos humanos. Todas sequências genômicas humanas devem ser removidas do conjunto de dados virais por meio de um pipeline de análise automática o mais cedo possível, sem operação manual pela equipe (ver Seção 6.7.1), a menos que sejam obtidos a aprovação ética e o consentimento explícito do paciente para processamento de dados genéticos humanos. Se dados pessoais ou humanos tiverem que ser armazenados, é altamente recomendada a criptografia adequada de todos esses arquivos.

As revisões éticas devem determinar qual o máximo possível de metadados relevantes que podem ser compartilhados sem arriscar a confidencialidade do paciente.

6.3.4 Recursos humanos

É importante garantir que haja pessoal suficiente para apoiar todos os aspectos do programa de sequenciamento, desde a amostragem clínica até a comunicação dos resultados e compartilhamento de sequências e metadados. O financiamento de um programa de sequenciamento deve incluir custos de pessoal, bem como os custos de equipamentos de proteção individual, consumíveis, compra e manutenção de outros equipamentos e arquitetura computacional. Se vários laboratórios ou institutos estiverem envolvidos em investigações colaborativas, pode ser importante obter um acordo por escrito sobre as responsabilidades de cada laboratório (por exemplo, em relação ao financiamento, os funcionários que podem ser envolvidos e o trabalho a ser realizado) e os benefícios esperados antes do início do projeto. O conteúdo desses acordos irá variar; os acordos de colaboração institucional existentes ou os acordos de transferência de material podem fornecer modelos apropriados.

As implicações referentes aos recursos humanos de qualquer programa de sequenciamento planejado devem ser ponderadas no tocante aos padrões de trabalho esperados. Em geral, um padrão normal de trabalho deve ser incentivado para evitar o esgotamento da equipe. A probabilidade de doenças e indisponibilidade de profissionais no contexto da pandemia COVID-19 também deve ser levada em conta. As tentativas de desenvolver capacidade extra no fluxo de trabalho devem ser levadas em conta desde o início, embora seja reconhecido que a geração de genomas de patógenos a partir de amostras clínicas exige uma equipe multidisciplinar com conjuntos de habilidades altamente específicas. A intensidade e previsibilidade da carga de trabalho dependerão dos objetivos do projeto (Tabela 3).

Os laboratórios de diagnóstico costumam ser fundamentais para a identificação de casos positivos e para o processamento e armazenamento seguros de amostras de pacientes. Se um projeto de sequenciamento em grande escala for planejado, é recomendado que um representante do laboratório de diagnóstico seja designado para fazer a ligação direta com a equipe de

sequenciamento para garantir a recuperação eficiente de amostras e metadados relevantes para aplicações subsequentes.

Tabela 3. Cargas de trabalho esperadas para objetivos específicos do programa de sequenciamento

Meta	**Velocidade típica de sequenciamento necessária para impacto**	**Intensidade de trabalho**	**Carga de trabalho**
Contribuição para a filodinâmica global	Baixa (frequentemente retrospectivo)	Variável	Previsível, embora possa mudar em resposta à alteração no tamanho do surto
Identificação de eventos de importação e circulação local	Baixa (frequentemente retrospectivo)	Variável	Previsível, embora possa mudar em resposta à alteração no tamanho do surto
Investigação da especificidade do ensaio diagnóstico	Moderada	Baixa	Imprevisível se em resposta à alteração observada na especificidade do ensaio; previsível se fizer parte do monitoramento contínuo
Apoio ou rejeição de evidências de rotas ou *clusters* de transmissão	Alta	Alta	Imprevisível, em resposta à necessidade clínica

6.4 Escolha do material apropriado para sequenciamento

6.4.1 Material para sequenciamento

A aquisição de RNA de SARS-CoV-2 suficiente e de alta qualidade ajuda a maximizar o rendimento do sequenciamento e a qualidade final dos dados de sequência do genoma. A quantidade e a qualidade de uma amostra de RNA são afetadas por: escolha da amostra clínica; manuseio da amostra clínica; método de isolamento do RNA viral; e proficiência técnica da equipe.

Nos lugares onde vários tipos de amostras diferentes estiverem disponíveis, é benéfico selecionar uma que tenha alta carga viral e baixos níveis de contaminantes de material genético humano ou bacteriano (Tabela 4). Essas amostras podem ser sequenciadas usando-se tanto a metagenômica quanto técnicas direcionadas ao SARS-CoV-2 (Seção 6.5). Alguns materiais, como fezes, podem exigir centrifugação e filtração antes da extração do RNA viral, para eliminar material celular humano ou bacteriano que pode reduzir a sensibilidade do sequenciamento.

Tabela 4. Sequenciamento direto de amostra clínica e cultura de células

Material de início	Quantidade de RNA viral	Conteúdo de material não viral	Referência
Soro, sangue	Detecção muito rara	Alta em sangue total, baixa em soro	*(77, 122–126)*
Amostras respiratórias (swabs naso-orofaríngeos, expectoração, fluido de lavagem brônquica)	Detecção frequente em níveis elevados	Alta, mas pode ser reduzida por meio de filtração e centrifugação	*(122, 127–135)*
Fluidos orais e gargarejos, bochechos	Detecção altamente variável dependendo do processo de coleta e manuseio; pode ser frequente	Alta, mas pode ser reduzida por meio de filtração e centrifugação	*(133, 136–143)*
Swabs anais e fecais, fezes	Detecção variável, mas quando detectada pode estar em níveis elevados.	Alta, mas pode ser reduzida por meio de filtração e centrifugação	*(122, 144–147)*
Autópsia, amostras de tecido	Detecção possível, embora as amostras raramente sejam acessíveis	Muito alta, difícil de reduzir por meio de filtração e centrifugação	*(148–155)*
Isolado viral de amostra clínica (cultura de células, modelo animal) (necessária instalação de nível de biossegurança 3)	Níveis altos após a cultura, mas a cultura pode induzir variantes artificiais	Moderada/alta, mas às vezes pode ser reduzido por meio de filtração e centrifugação, dependendo do tipo exato de amostra	*(8, 156, 157)*

Em muitas situações, as únicas amostras rotineiramente disponíveis para sequenciamento genômico do vírus serão amostras residuais de diagnóstico. As amostras coletadas para diagnóstico de NAAT também são normalmente apropriadas para sequenciamento.*(77)* Descobriu-se que swabs nasais, swabs de garganta e saliva apresentam altas cargas virais logo após o início dos sintomas e por até 25 dias depois. *(140, 158, 159)* A carga viral do SARS-CoV-2 e a abundância de RNA viral nas amostras são, via de regra, mais elevadas na primeira semana após o início da doença. *(158, 160)*

Se possível, os isolados para sequenciamento devem ser selecionados de amostras positivas que já foram processadas por um laboratório de diagnóstico molecular (Figura 3). O compartilhamento de recursos realizado dessa forma evita a duplicação do trabalho no processamento de amostras e extração de ácido nucleico e, portanto, pode economizar recursos humanos e outros, além de custos. Alguns kits comerciais de diagnóstico molecular usam lisados virais como insumos e não permitem o armazenamento do RNA extraído. Em tais casos, nos lugares onde os componentes do tampão de lise comercial não são divulgados, pode ser extremamente difícil reutilizar lisados preparados com outros kits de extração comerciais e pode ser necessário realizar nova inativação e extração diretamente da amostra clínica original. A divulgação dos componentes dos tampões de lise comerciais ajudaria os pesquisadores no desenvolvimento de estratégias para reutilizar lisados já inativados nas atividades de sequenciamento subsequentes.

Um sistema prático e eficaz de identificação de amostras deve ser usado se as amostras forem deslocadas de um laboratório para outro; idealmente, a mesma identificação de amostra deve ser usada em todos os laboratórios de manuseio.

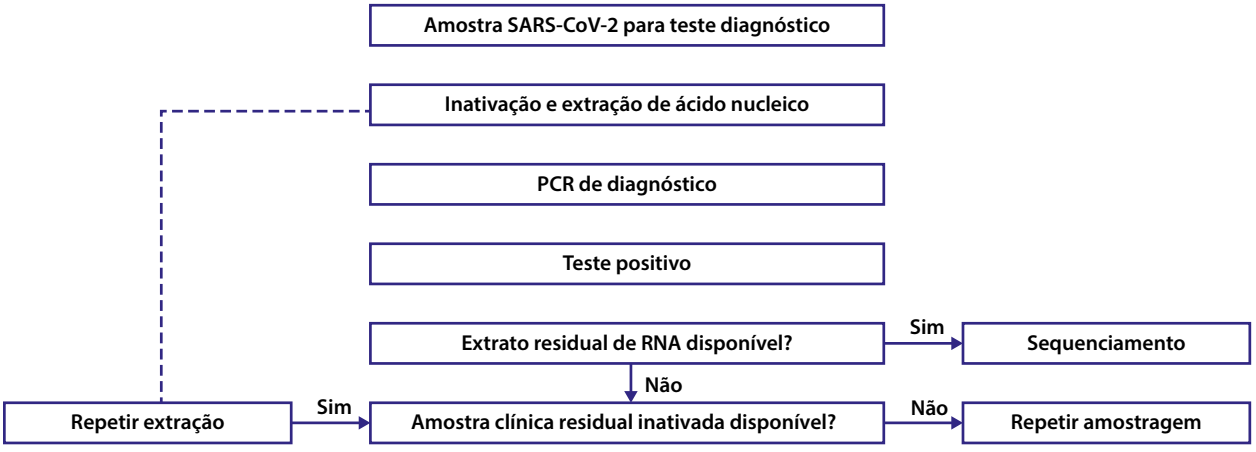

Figura 3. Exemplo de fluxo de trabalho para recuperação de amostras de um laboratório de diagnóstico.

A preservação do RNA viral é importante para a produção de dados de sequência de alta qualidade. Isso pode ser feito mantendo-se uma cadeia de frio entre a coleta e a análise da amostra, reduzindo o número de vezes em que o RNA ou as amostras são congelados e descongeladas e minimizando-se o tempo entre a coleta da amostra e o sequenciamento. É improvável que o RNA armazenado ou transportado a 4 ºC por mais de alguns dias tenha qualidade suficientemente alta para sequenciamento, a menos que tenha primeiro sido preservado em uma solução de estabilização de RNA. A qualidade será substancialmente maior se o RNA puder ser armazenado a -20 ºC ou, de preferência, a -80 ºC. Os lisados virais normalmente não

podem ser armazenados a 4 ºC por tanto tempo quanto o RNA extraído. Muitos protocolos de sequenciamento incluem etapas que melhoram a capacidade de armazenamento de uma amostra, incluindo a transcrição reversa de RNA para cDNA ou a síntese/geração de um segunda fita/geração de amplicons de PCR de DNA de fita dupla. Os amplicons de PCR podem ser armazenados a 4 ºC por muitos meses sem redução na qualidade do sequenciamento. Em alguns contextos, portanto, pode ser apropriado realizar essas etapas rapidamente após os PCRs de diagnóstico, de modo que o material possa ser armazenado ou transportado com menos restrições de temperatura antes da preparação da biblioteca.

6.4.2 Amostras de controle

As amostras de controle negativo, como tampão ou água, sempre devem ser incluídas em qualquer execução de sequenciamento que contenha várias amostras. Elas devem ser incluídas no estágio mais inicial possível e devem prosseguir com as amostras em todos os estágios do pipeline de sequenciamento. Isso é extremamente importante para descartar contaminação durante uma execução de sequenciamento que ocorra em laboratório ou durante o processamento de bioinformática.

As amostras de controle positivo com sequências genéticas conhecidas podem ser úteis para validar os pipelines de bioinformática recentemente adotados ou adaptados para a geração de uma sequência de consenso, mas não precisam ser incluídas em cada execução de sequenciamento.

6.5 Enriquecimento do material genético SARS-CoV-2 antes da preparação da biblioteca

As estratégias de sequenciamento para SARS-CoV-2 incluem abordagens metagenômicas, que não exigem conhecimento prévio da sequência genômica, e abordagens direcionadas, que dependem do conhecimento do genoma. Ambas as abordagens normalmente tentam enriquecer o material genético SARS-CoV-2 em relação a outro RNA/DNA antes do sequenciamento. Se houver RNA residual suficiente disponível e armazenado de forma adequada (Seção 6.4.1), a maioria das abordagens pode ser realizada usando RNA extraído para ensaios de diagnóstico. Muitos protocolos diferentes já foram compartilhados para o sequenciamento do SARS-CoV-2. Alguns deles são destacados abaixo; outros foram coletados pelos Centros de Controle e Prevenção de Doenças dos Estados Unidos (CDC). *(161)*

6.5.1 Análises metagenômicas de amostras clínicas não cultivadas

Os protocolos metagenômicos permitem o sequenciamento não direcionado do ácido nucleico contido em uma amostra, incluindo o material genômico viral, se presente. *(162)* Esses protocolos oferecem uma abordagem livre de hipóteses para a descoberta de patógenos, pois exigem pouco conhecimento prévio do patógeno de interesse. *(163)*

A eliminação do material genético do hospedeiro ou de outro material genético não-SARS-CoV-2 de uma amostra resulta em uma proporção maior de leituras de SARS-CoV-2 nos dados de sequência gerados e, portanto, a uma chance maior de recuperar um genoma completo.

As abordagens metagenômicas do SARS-CoV-2, portanto, normalmente incluem etapas para remoção de células hospedeiras e bacterianas, por meio de centrifugação ou filtração antes da extração do RNA, ou por remoção química ou enzimática do DNA/RNA indesejado. Isso é mais fácil nas amostras líquidas, das quais as células podem ser mais facilmente separadas, como o lavado broncoalveolar (Tabela 4). O conteúdo de RNA ribossomal (rRNA) e de DNA também é geralmente eliminado durante a preparação da biblioteca para sequenciamento do RNA viral, e o RNA transportador é muitas vezes omitido nas extrações ou substituído por poliacrilamida linear. Apesar dessas medidas, as amostras ainda podem conter grandes quantidades de DNA/RNA indesejado do hospedeiro que também podem ser sequenciados. As abordagens metagenômicas, portanto, via de regra se beneficiam com a entrada de amostras com altas cargas virais (de forma que uma proporção razoável do material genético da amostra seja de vírus). Como alternativa, geralmente é necessário gerar um grande número de leituras. Dessa forma, mesmo que o material genético do SARS-CoV-2 represente apenas uma pequena proporção das leituras, ainda será possível obter todo o genoma do vírus.

O sequenciamento metagenômico normalmente produz um grande número de leituras fora do alvo e não virais. Também é bem (embora nem sempre, dependendo da plataforma de sequenciamento e multiplexação) mais dispendioso do que as abordagens de sequenciamento baseadas em captura direcionada ou baseadas em amplicon, porque mais dados devem ser produzidos para gerar um genoma SARS-CoV-2. Além disso, as etapas de pré-tratamento que são particularmente benéficas para a metagenômica, como a centrifugação, não são, em geral, realizadas nos ensaios de diagnóstico molecular, portanto, novas extrações que incorporem etapas de pré-tratamento podem ter de ser realizadas para o sequenciamento metagenômico. As abordagens de sequenciamento direcionadas (seções 6.5.3 e 6.5.4) são frequentemente mais econômicas e exigem menos recursos; elas podem, portanto, ser mais apropriadas nos lugares onde os benefícios da abordagem metagenômica (por exemplo, descoberta de patógenos, detecção de coinfecções) não sejam necessários. O sucesso da abordagem metagenômica varia de um método para outro. Vários estudos mostraram uma redução rápida no sucesso de várias análises de sequenciamento metagenômico em amostras com limiares de ciclo (Cts) PCR em tempo real (qPCR) de aproximadamente 25-30. Para essas amostras, os métodos de PCR multiplex e baseados em captura alcançam cobertura consistentemente mais alta em todo o genoma do que o sequenciamento metagenômico. *(57, 164)* O número de leituras de sequenciamento por amostra que devem ser geradas para se obter o genoma completo dependerá do tipo de amostra, dos procedimentos de pré-tratamento para remoção do material do hospedeiro e do nível de viremia.

6.5.2 Abordagens metagenômicas após cultura de células

Para amostras com baixa carga viral, a proporção de material genético viral pode, teoricamente, ser aumentada permitindo-se que o vírus se replique em cultura de células. No entanto, os riscos de biossegurança associados à cultura de vírus são significativamente maiores do que aqueles associados a amostras clínicas não cultivadas. São necessárias instalações de nível 3 de biossegurança, com extensos procedimentos adicionais para garantir o manuseio e armazenamento seguros. Além disso, a passagem em cultura de células pode resultar em mutações artificiais nas sequências, que não estavam presentes na amostra clínica original. Isso

pode ter implicações importantes para as análises subsequentes. O uso de cultura de células apenas com a finalidade de amplificar o material genético do vírus para o sequenciamento de SARS-CoV-2 deve, portanto, ser evitado, especialmente agora que outras abordagens baseadas em amplicon e na captura com isca estão disponíveis para melhorar a sensibilidade do sequenciamento.

6.5.3 Abordagens baseadas em captura direcionada

Após a preparação de uma biblioteca de sequenciamento metagenômico, podem ser realizadas abordagens baseadas em captura que enriquecem para o material genético SARS-CoV-2 antes do sequenciamento. Essas abordagens baseiam-se na hibridização do DNA que foi transcrito reversamente de RNA viral para DNA ou iscas de RNA. Essas iscas são projetadas para serem complementares às regiões do genoma SARS-CoV-2.

O material da biblioteca fora do alvo que não se ligou com sucesso a uma isca (por exemplo, DNA do hospedeiro) pode ser removido usando-se abordagens enzimáticas ou físicas. Isso reduz a chance de detectar outras coinfecções, mas aumenta o número esperado de leituras de sequenciamento que serão mapeadas para o genoma do SARS-CoV-2, permitindo que mais amostras sejam sequenciadas ao mesmo tempo de modo efetivo em execuções multiplexadas.

Uma vantagem de se usar uma abordagem baseada em captura em vez de uma abordagem baseada em amplicon de PCR (Seção 6.5.4) é a de que as abordagens baseadas em captura podem tolerar diferenças de sequência nas sequências de sonda de 10–20%. Isso é maior do que a incompatibilidade tolerada pela PCR, na qual essa divergência das sequências do primer resultaria em alto risco de falha do amplicon. As abordagens baseadas em captura podem, portanto, ser usadas com sucesso para enriquecer sequências relativamente divergentes do SARS-CoV-2. As abordagens baseadas em captura são normalmente mais complexas de se estabelecer e mais caras do que as baseadas em amplicon de PCR.

Vários painéis de captura específicos de SARS-CoV-2 que estão comercialmente disponíveis ou que podem ser projetados sob encomenda podem resultar em um aumento de 100 a 10.000 vezes na sensibilidade. Quando várias amostras devem ser sequenciadas ao mesmo tempo em um único conjunto, é mais custo-efetivo realizar a captura em um conjunto inteiro de até 96 amostras multiplexadas após a identificação delas com códigos de barras. Vários protocolos publicados foram validados para sequenciação de SARS-CoV-2 baseada em captura (por exemplo, com base em *(165)*).

6.5.4 Abordagens direcionadas baseadas em amplicon

Os PCRs que geram amplicons em todo o genoma SARS-CoV-2 podem ser usados para amplificar o material do vírus antes da preparação da biblioteca de sequenciamento. Ao contrário das abordagens baseadas em captura, as abordagens baseadas em amplicons não toleram incompatibilidade substancial entre a sequência alvo e os primers que forem usados. A diversidade genômica alvo deve, portanto, ser relativamente baixa, e/ou a sequência alvo deve ser conhecida o bastante para permitir que os primers sejam projetados de modo a direcionar regiões

genômicas mais conservadas. Dado que o SARS-CoV-2 surgiu há pouco, em seres humanos e, portanto, mostra uma diversidade genômica global relativamente baixa, as abordagens baseadas em PCR são, hoje em dia, muito apropriadas para o seu sequenciamento. No entanto, a ocorrência de falhas do amplicon precisa ser monitorada e os primers precisam ser substituídos quando a falha ocorrer como resultado de substituições nos sítios de ligação do primer.

As abordagens otimizadas baseadas em PCR são altamente específicas e sensíveis e permitem que sejam gerados genomas inteiros do vírus SARS-CoV-2 de rotina, a partir de amostras com valores de PCR Ct de até 30. Podem ser gerados genomas parciais de rotina, a partir de amostras com valores Ct de 30–35. No entanto, esses valores são uma aproximação. O Ct não é um preditor perfeito do sucesso da amplificação, pois pode variar em diferentes métodos de diagnóstico, *(166)* e o uso de diferentes tipos e qualidade da amostra afetará a sensibilidade. Além disso, as regiões genômicas direcionadas em ensaios de diagnóstico de PCR são, via de regra, muito mais curtas do que aquelas usadas em abordagens de sequenciamento baseadas em amplicon comuns, então a degradação do RNA normalmente afetará mais o sequenciamento baseado em PCR do que os diagnósticos de PCR. Nos lugares onde a diversidade genômica direcionada é baixa, as abordagens baseadas em PCR são uma maneira barata, rápida e conveniente de aumentar a quantidade de material genético de vírus disponível em uma amostra antes do sequenciamento.

Foram descritos vários conjuntos de primers diferentes para sequenciamento genômico completo com base em amplicons. Esses amplicons têm como alvo diferentes comprimentos, normalmente 400–2000 pares de bases (pb). Amplicons mais longos exigem menos primers de PCR para estruturar todo o genoma, mas podem resultar em lacunas maiores no genoma de consenso no caso de falha de amplificação de um par de primers. Amplicons mais longos são adequados para plataformas de leitura longa, mas exigem fragmentação nas ferramentas de sequenciamento de leitura curta. O esquema mais amplamente utilizado nos tempos atuais é a abordagem baseada em amplicons projetada pela Rede ARTIC. *(167)* Embora o protocolo ARTIC se concentre amplamente no sequenciamento de nanoporos da Oxford Nanopore Technologies, vários laboratórios validaram a abordagem da ARTIC em outras plataformas de sequenciamento. *(112, 168)*

É vital adotar estratégias para evitar a contaminação de outros testes de diagnóstico ou sequenciamentos posteriores por amplicons (Seção 6.3.1).

6.6 Seleção da tecnologia de sequenciamento

Após a preparação inicial da amostra para enriquecimento do material genético SARS-CoV-2, as bibliotecas podem ser preparadas usando-se protocolos de sequenciamento padrão que sejam apropriados para qualquer vírus. O protocolo dependerá da ferramenta usada. Antes de investir na capacidade de sequenciamento pela primeira vez, ou adotar uma tecnologia alternativa, deve-se considerar o tempo de execução, custos, facilidade de uso, processamento de dados subsequente, taxa de transferência (taxa de produção de dados) e precisão de sequenciamento das várias tecnologias (Tabela 5) (ver também Seção 6.7).

O sequenciamento convencional (sequenciamento Sanger) pode ser usado para sequenciar fragmentos individuais (até 1000 pb) em reações separadas. O sequenciamento genômico completo do SARS-CoV-2 exige que pelo menos 30 amplicons individuais sejam sequenciados separadamente por amostra de paciente. O sequenciamento Sanger é, portanto, provavelmente mais útil para sequenciar fragmentos curtos de genomas, por exemplo, para preencher lacunas em montagens pós-sequenciamento de nova geração ou para investigar a diversidade de vírus em regiões curtas, como sítios de ligação de primer, após a falha de um ensaio diagnóstico.

As plataformas de sequenciamento de nova geração são mais apropriadas para o sequenciamento de genoma completo de rotina. As plataformas de sequenciamento que têm sido comumente usadas até agora para SARS-CoV-2 incluem as da Illumina, IonTorrent e da Oxford Nanopore Technologies. Ao contrário do sequenciamento Sanger, no qual todas as moléculas de DNA de uma amostra devem ter as mesmas sequências ou sequências altamente semelhantes (por exemplo, após PCR de um único amplicon), essas tecnologias permitem o sequenciamento simultâneo de vários fragmentos do genoma do SARS-CoV-2. Todas as plataformas de sequenciamento de nova geração permitem que várias amostras sejam sequenciadas ao mesmo tempo em uma única execução. As principais vantagens e limitações de cada tecnologia estão resumidas na Tabela 5. Embora todas as plataformas sejam adequadas para gerar genomas de consenso de SARS-CoV-2, algumas podem ser mais adequadas para atender aos objetivos específicos do programa de sequenciamento. Por exemplo, um tempo de resposta rápido pode ser importante para aplicações clínicas, ao passo que a precisão do nível de leitura pode ser mais importante para a investigação da diversidade intra-hospedeiro.

Tabela 5. **Plataformas comumente usadas para análise de sequência de SARS-CoV-2 e suas características**[a]

Ferramenta	Vantagens	Limitações	Tempo de execução da ferramenta	Taxa de transferência de sequenciamento	Disponibilidade e custo relativos
Sequenciamento Sanger	Amplamente acessível Fácil de usar Sequenciamento com boa relação custo-efetividade se poucas metas forem necessárias	Taxa de transferência muito baixa Amplicons (com frequência não mais do que 1000 pb) devem ser amplificados e sequenciados individualmente Dispendioso para genomas completos Impróprio para metagenômica	Normalmente algumas horas	100 kB-2 Mb por execução única	Amplamente disponível Custo relativamente baixo para alguns alvos
Illumina (por exemplo, iSeq, MiniSeq, MiSeq, NextSeq, HiSeq, NovaSeq)	Possível rendimento de sequenciamento muito alto iSeq de precisão muito alta é portátil Os métodos de tratamento de dados estão bem estabelecidos	Com exceção do Illumina iSeq, dispendioso para compra e manutenção em comparação com algumas outras plataformas Comprimento máximo de leitura 2 x 300 pb.	10–55 h, dependendo da ferramenta	1,2-6000 Gb, dependendo da ferramenta	Altos custos de manutenção e inicialização Custos de funcionamento moderados
Oxford Nanopore Technologies (Flongle, MinION, GridION, PromethION)	Dados portáteis, sequenciamento direto em tempo real Custos de inicialização e manutenção baixos Pode interromper o sequenciamento assim que dados suficientes forem obtidos Alcança comprimentos de leitura muito longos (excedendo o comprimento total do genoma SARS-CoV-2)	Dificuldades com homopolímeros A taxa de erro por leitura é de ~ 5% (células de fluxo R9.4), por isso o uso de pipelines apropriados é fundamental para obter sequências de consenso de alta precisão Atualmente inadequado para determinar a variação intra-hospedeiro, a menos que seja usado sequenciamento replicado *(112)*	Leituras disponíveis imediatamente Pode ser monitorado e executado por até vários dias, conforme necessário	Varia de <2 Gb para célula de fluxo Flongle a 220 Gb para célula de fluxo PromethION Até 48 células de fluxo podem ser usadas no PromethION	Sem manutenção e com baixo custo de inicialização Custos de funcionamento moderados.
Ion Torrent	Resposta rápida uma vez que o sequenciamento começa	Dificuldades com homopolímeros Dispendioso para compra Comprimento máximo de leitura típico em torno de 400 pb.	2h – 1 dia, dependendo do chip e dispositivo	30 Mb–50Gb dependendo do dispositivo e chips	Custos moderados.

[a] Esta lista de várias ferramentas visa fornecer uma visão geral das que são mais comumente usadas no sequenciamento genômico do SARS-CoV-2 e não implica endosso da OMS para esses produtos.

6.7 Protocolos de bioinformática

A seleção de um protocolo de bioinformática apropriado que possa processar dados lidos brutos em sequências de consenso do genoma inteiro é geralmente tão importante quanto a escolha da plataforma de sequenciamento. O uso de um protocolo de bioinformática inadequado pode produzir resultados errôneos que podem afetar gravemente as análises posteriores.

6.7.1 Visão geral das etapas típicas de bioinformática

Arquivamento de dados lidos brutos

O sequenciamento gera grandes volumes de dados (Tabela 5). Os custos da arquitetura computacional necessária para armazenar e manipular esses dados devem ser levados em consideração quando um pipeline de sequenciamento estiver sendo desenvolvido. O volume de dados brutos produzidos, geralmente armazenados como arquivos FASTQ (que armazenam sequências genéticas junto com a pontuação de qualidade de cada base na sequência), dependerá do número de amostras processadas. Os dados de leitura curta que foram enriquecidos para sequências virais, seja por captura com isca ou por amplificação por PCR, podem frequentemente abranger 1–2 milhões de leituras por amostra e exigir até 1 Gb de espaço em disco, dependendo do comprimento da leitura. As amostras não enriquecidas que forem sequenciadas metagenomicamente, em geral, exigem números de leitura 100 vezes maiores para obter uma boa cobertura genômica do SARS-CoV-2, já que a proporção de leituras virais dessas amostras pode ser inferior a 1% das leituras totais. (164)

Se a capacidade de armazenamento for limitada, o armazenamento permanente de dados brutos pode não ser viável. Embora seja preferível armazenar dados lidos brutos localmente pelo maior tempo possível, nem sempre é essencial, caso esse armazenamento se torne uma barreira para sequenciamento adicional. Uma exceção é o armazenamento de dados brutos de sequenciamento metagenômico ou metatranscriptômico, que pode conter informações sobre a coinfecção com outros vírus ou bactérias. Essas amostras representam um bem importante e devem ser feitos esforços para preservar as informações, mesmo que as leituras brutas não possam ser armazenadas em outras circunstâncias.

Uma alternativa prática recomendada para o arquivamento local permanente de dados lidos brutos é fazer o *upload* dos dados em um repositório, como o SRA (NCBI), o DDBJ ou o ENA.

A menos que a revisão ética tenha aprovado a investigação e o compartilhamento de sequências genômicas humanas, e todos os participantes tenham dado consentimento informado explícito para isso, devem primeiro ser retiradas as leituras de origem humana dos dados submetidos a repositórios públicos. Para abordagens de sequenciamento direcionado ao SARS-CoV-2, todas as leituras de sequenciamento podem ser mapeadas para o genoma do SARS-CoV-2 e as leituras mapeadas podem ser extraídas. As leituras extraídas que forem mostradas na sequência como não mapeadas para genoma humano podem normalmente ser enviadas para repositórios. Há softwares existentes que podem facilitar essa tarefa em diferentes plataformas, por exemplo, o nanostripper para dados produzidos usando dispositivos da Oxford Nanopore Technologies. *(169)*. Para

projetos metagenômicos em que um dos objetivos é identificar coinfecções, as estratégias para remoção de leituras humanas são mais complexas. Alguns repositórios, como o SRA, podem remover leituras genéticas humanas de conjuntos de dados metagenômicos se contatados diretamente. Também podem ser estabelecidos pipelines para remoção de leituras humanas usando-se softwares de classificação taxonômica, como o Kraken2 ou o CLARK, (170, 171) ou um software para remoção de mapeamento de leituras para genomas humanos, como o GSNAP. *(172)* Os processos para remoção de leituras humanas devem sempre ser avaliados como parte da revisão ética de qualquer projeto e eles devem ser amplamente testados para garantir sua eficácia. Outras abordagens de compartilhamento de dados e considerações éticas são abordadas de modo mais completo na Seção 4.

Montagem do genoma a partir de dados brutos

Foi desenvolvida uma série de pipelines de softwares gratuitamente disponíveis e ajustados para o sequenciamento do SARS-CoV-2. Muitas exigem configuração local mínima e têm instruções claras de uso. Um repositório útil (não completo) dos links para pipelines de sequenciamento, inclusive de bioinformática quando houver, é mantido pelo CDC. *(161).* Pacotes adicionais para sequenciamento viral estão disponíveis e seriam apropriados após extensa adequação para o SARS-CoV-2.

O pipeline de bioinformática dependerá dos estágios do laboratório de pré-sequenciamento (por exemplo, a amplificação por PCR exige corte bioinformático de sítios de primer) e da plataforma de sequenciamento e dos reagentes usados. Os pipelines de bioinformática geralmente incluem etapas semelhantes às mostradas na Tabela 6.

Tabela 6. Etapas comuns na construção de consenso bioinformático para as duas plataformas de sequenciamento de nova geração mais comumente usadas

Estágio	**Illumina**[b]	**Oxford Nanopore Technologies (ONT)**[b]
Chamada de bases de sinal de leitura bruta em dados de formato FASTQ	Bcl2Fastq (Illumina) As instalações de sequenciamento geralmente realizam essas etapas antes de enviar aos usuários dos dados	Guppy (ONT)
Demultiplexação de leituras em amostras diferentes		Porechop *(173)* para demultiplexação e corte de adaptador
Remoção de artefatos de sequenciamento, incluindo adaptadores de sequenciamento	Cutadapt para adaptador de corte *(174)*	
Corte de pares de base de baixa qualidade	Trimmomatic *(175)*	As leituras que são substancialmente mais longas ou mais curtas do que o comprimento de leitura esperado podem ser removidas para esquemas de PCR multiplex

Remoção de duplicatas ópticas para dados de leitura curta de protocolos que incluem enriquecimento ou amplificação	Picard Mark Duplicates	N/A
Alinhamento de leituras no alvo para um genoma de referência canônico, como a sequência de referência NCBI do genoma NC_045512 *(176)*	Bowtie2 *(177)*	Minimap2 *(178)* or BWA *(179)*
Remoção de artefatos de sequenciamento de leituras no alvo, incluindo primers para esquemas multiplex (opcional, dependendo do método de sequenciamento)	Pipelines como iVar para corte de primer *(112)*	Pipelines como a rede ARTIC *(167)*
Identificação de variantes da sequência de referência, com limites de qualidade apropriados para distinguir as variantes verdadeiras dos erros de sequenciamento. A metodologia de chamada de variantes é fortemente dependente do protocolo de biblioteca e da tecnologia de sequenciamento e, na maioria dos casos, requer um ajuste substancial de parâmetros para distinguir as variantes verdadeiras das chamadas de variantes falso positivas. Os protocolos mais simples para a filtragem de variantes seguem etapas para remoção de posições com baixa profundidade de leitura ou aquelas apoiadas por leituras com qualidade insuficiente e exigem que uma proporção significativa das chamadas de base apoie uma variante da referência	Samtools mpileup seguido por filtros e chamadas da BCFtools. *(179, 180)* As variantes podem ser mantidas nos seguintes casos, por exemplo: • uma profundidade mínima de 5 leituras em cada posição, ou maior para amostras amplificadas por PCR • uma qualidade de base média mínima de 15 • pelo menos 75% das leituras da posição apoiam a chamada • das leituras que abrangem a posição, pelo menos uma na orientação direta e pelo menos uma na direção reversa (para sequenciamento Illumina de extremidade emparelhada)	Uso de Nanopolish *(181)* ou Medaka (ONT) para melhorar as sequências de consenso É importante usar pipelines estabelecidos que foram totalmente validados. Os pipelines podem incluir várias condições, como: • uma profundidade mínima de 20 leituras para dados Oxford Nanopore para contabilizar as taxas de erro • limites nos quais os sites não são resolvidos, mas são marcados como ambíguos

[a] Com base nas sequências SARS-CoV-2 enviadas ao GISAID nos primeiros três meses da pandemia. O software mostrado é apenas para fins ilustrativos; outro software apropriado está disponível em cada estágio. Para o Ion Torrent, pode ser usado um software semelhante ao da plataforma Illumina.

[b] A menção de ferramentas e softwares específicos não implica no endosso da OMS aos produtos.

Independentemente do pipeline, as variantes de nucleotídeos não devem ser chamadas se o número de leituras de apoio exclusivas do sítio for inferior à profundidade necessária para a confiança. Em vez disso, esses sites devem ser chamados de bases ambíguas (N) no genoma de consenso final. Dependendo da precisão das leituras brutas nos métodos escolhidos, quaisquer sites com menos de 5 a 20 leituras exclusivas de apoio não podem ser chamados com precisão. O nível mínimo de contaminação esperado pode ser determinado a partir do número de leituras SARS-CoV-2 observadas no controle negativo, e os sítios só devem ser chamados se a profundidade exceder muito esse nível.

Os métodos metagenômicos e de captura são quantitativos, o que significa que a profundidade de leitura das amostras refletirá aproximadamente o número de cópias do genoma viral na biblioteca inicial. Para amostras com carga viral baixa, a chamada de variantes deve ser realizada com

cautela, pois mesmo um pequeno número de leituras contaminantes pode interferir no sinal da amostra. Os controles negativos também devem ser sequenciados para permitir a avaliação da probabilidade de contaminação.

As variantes contidas em amostras com Cts elevados que provavelmente tenham pequeno número de cópias iniciais de RNA devem ser avaliadas com cautela, porque a presença estocástica de certas variantes entre as poucas cópias presentes pode levar a erros de artefato. As variantes também devem ser ponderadas com muito cuidado se as enzimas usadas durante a transcrição reversa e/ou PCR induzirem erros frequentes. Sempre que possível devem ser usadas enzimas de alta fidelidade para proteção contra esses erros.

6.7.2 Como lidar com dados multiplexados

É custo-efetivo sequenciar várias amostras virais em uma única execução de sequenciamento. Isso geralmente é realizado pela adição de adaptadores ou códigos de barras exclusivos às leituras de sequenciamento. Quando são gerados dados brutos, eles podem ser desmultiplexados alocando-se leituras a amostras com códigos de barras correspondentes. A multiplexação introduz uma nova complexidade ao processo de controle de qualidade de saídas bioinformáticas, uma vez que é possível que os códigos de barras sejam determinados incorretamente, devido a um processo conhecido como index hopping ou index misassignment. Esses artefatos afetam particularmente as amostras com baixa carga viral, pois um pequeno número de leituras contaminantes pode ter um efeito desproporcional no consenso do genoma. Para se proteger contra isso, é recomendado que os conjuntos multiplexados contenham pelo menos um controle negativo (buffer) e, se possível, um controle não-SARS-CoV-2, e que o número de leituras atribuídas incorretamente na execução seja determinado com base em observações de leituras de controle nas amostras e no controle negativo. Os sistemas únicos com indexação dupla (por exemplo, aplicativos Illumina) ou código de barras de extremidade dupla (por exemplo, aplicativos da Oxford Nanopore Technologies e algumas preparações da Ion Torrent), devem ser usados onde for viável e deve haver controles rigorosos na demultiplexação de amostras. A demultiplexação deve ser realizada usando-se configurações rigorosas (por exemplo, dependendo da tecnologia, exigindo-se que os códigos de barras estejam presentes em ambas as extremidades de uma leitura de sequenciamento, com poucas ou nenhuma incompatibilidade com esse código de barras).

6.8 Ferramentas de análise

6.8.1 Subamostragem de dados antes da análise

Em meados de novembro de 2020, 180.000 genomas completos com boa cobertura estavam disponíveis ao público, e o número estava aumentando exponencialmente. Muitos desses genomas são provavelmente quase idênticos. Se não for necessário um genoma completo de milhares de sequências quase idênticas, podem ser empregadas estratégias de redução de amostragem para reduzir as demandas computacionais de alinhamento e análises subsequentes. As estratégias de redução de amostragem devem ser ponderadas com cuidado, pois podem afetar gravemente as análises subsequentes.

Um procedimento possível é executar uma ferramenta de agrupamento, como o cd-hit-est,*(182)* em um alto limite de agrupamento (> 99% de similaridade de sequência) e construir um alinhamento usando os genomas representativos dessa análise. Isso é computacionalmente leve e auditável, pois um relatório de agrupamento é produzido indicando quais sequências foram selecionadas para cada *cluster* e listando a associação completa do *cluster*.

Uma alternativa pode ser selecionar clados de interesse de uma árvore maior previamente calculada. Essa pode ser uma estratégia útil, particularmente nos lugares onde uma região geográfica ou outra característica é de importância primária para a análise, e a diversidade global total dos genomas virais é menos relevante. O Nextstrain *(183)* permite que os clados sejam selecionados de uma árvore global, e os metadados de sequências desses clados sejam subsequentemente extraídos e usados para ajudar a subamostragem de grandes conjuntos de dados disponíveis.

Para inferência filogeográfica em que os pesquisadores estão interessados em capturar deslocamentos de linhagens de vírus entre locais, mas não dentro de locais, pode ser apropriado realizar subamostragem com base em critérios filogenéticos. Aqui, os clados monofiléticos de sequências do mesmo local podem ser subamostrados a uma única sequência desse clado, pois sequências adicionais dentro do clado podem não adicionar mais informações de interesse em relação aos deslocamentos de linhagem viral entre locais. *(103, 184)*

6.8.2 Alinhamentos de sequência

O alinhamento de milhares de sequências do genoma do SARS-CoV-2, muitas das quais incluem regiões de ambiguidade devido a genomas parcialmente determinados, é um desafio computacional. Bem poucas ferramentas existentes podem lidar com alinhamentos desse comprimento, sendo importante notar que cada vez que uma nova sequência é gerada, ela tem o potencial de modificar o alinhamento determinado antes. É possível usar um software de alinhamento, como o MAFFT, para adicionar um pequeno número de novas sequências a um pequeno alinhamento existente com relativamente pouca sobrecarga computacional. (185) Também podem ser curados alinhamentos de até várias centenas de sequências com a ajuda de especialistas, e os autores do MAFFT *(186)* oferecem esse serviço para alinhamentos do SARS-CoV-2. No entanto, para conjuntos de amostras maiores, pode ser

necessária uma estratégia diferente. O pipeline SHIVER *(187)* produz uma versão de cada genoma montado que é alinhado para manter o posicionamento das coordenadas. Assim, cada genoma processado pode simplesmente ser adicionado em um alinhamento progressivo sem a necessidade de se realinhar todas as sequências cada vez que uma sequência é adicionada, embora seja necessário tomar cuidado para garantir que as novas inserções não sejam perdidas.

Muitas vezes, é apropriado aparar regiões não codificantes, incluindo as extremidades 5' e 3' de um alinhamento, antes de realizar análises adicionais. Pode ser um desafio analisar filogeneticamente essas regiões porque elas incorrem em inserções, deleções e substituições múltiplas no mesmo sítio com mais frequência do que regiões de codificação que estejam sob seleção mais intensa.

6.8.3 Controle de qualidade

As sequências geradas devem sempre ser submetidas a um controle de qualidade antes de serem utilizadas em qualquer análise. Os procedimentos de controle de qualidade devem ser conduzidos em diferentes estágios, para determinar vários recursos que podem estar associados a sequências de baixa qualidade.

Remoção de sequências com bases ambíguas, indels ou frameshifts com base em sequências desalinhadas/alinhadas

A maioria das ferramentas de software de construção de árvores filogenéticas, incluindo todos os métodos de máxima verossimilhança, são vulneráveis a um grande número de bases ambíguas em genomas sequenciados. São necessárias análises mais extensas para avaliar o efeito das sequências parciais nas filogenias, mas pode ser apropriado em primeiro lugar remover sequências com > 10% de Ns nas regiões de interesse.

As sequências com suspeita de erros de sequenciamento subjacentes (por exemplo, induzidos por montagens incorretas) devem ser investigadas e geralmente removidas. Erros de sequenciamento podem se manifestar como alta divergência em comparação com outras sequências ou como alto número de substituições em regiões curtas que podem indicar montagens incorretas locais. Um número elevado de bases não ACGTN pode ser indicativo de populações virais mistas como resultado de contaminação.

Há várias ferramentas úteis disponíveis para ajudar a detectar bases ambíguas, indels (inserções ou exclusões de bases) e deslocamentos de quadro, incluindo o recurso Nextclade Quality Control Metric no Nextstrain *(183),* CoV-GLUE *(188)* e Pangolin *(189).*

Remoção de sequências que formam longos ramos filogenéticos

As sequências que formam ramos com suspeita de serem longos em uma árvore filogenética (que sugerem uma divergência evolutiva incomumente alta), devem ser curadas com muito cuidado. Esses ramos podem refletir efeitos reais, como grandes indels ou eventos de recombinação, mas

no caso de genomas altamente conservados, incluindo o SARS-CoV-2, eles mais amiúde indicam uma taxa de erro substancial na sequência subjacente ou desalinhamento (Figura 4).

Remoção de sequências em que a divergência seja substancialmente maior ou menor do que o esperado.

As sequências suspeitas também podem ser identificadas usando-se uma árvore filogenética e ferramentas como o TempEst *(190)* ou o TreeTime *(191)*. De modo específico, se uma sequência for substancialmente mais ou menos divergente do que o esperado, dado o momento em que foi amostrada, ela deve ser verificada com cuidado, quanto a possíveis erros e possivelmente removida. As sequências que forem mais ou menos divergentes do que o esperado podem surgir de problemas bioinformáticos (por exemplo, chamada de variante de má qualidade ou corte inadequado) ou uma atribuição incorreta de metadados, ou seja, uma data de amostragem incorreta. Não está formalmente definido, de forma precisa, o que constitui "muito divergente", mas todas as sequências que estiverem bem fora da curva, devem ser investigadas. A inspeção manual para identificação de recursos que possam indicar erros na montagem do genoma viral costuma ser útil para conjuntos menores de dados. Essas características podem incluir inserções, substituições ou deleções que levam a códons de parada nas sequências de codificação esperadas, ou sequências curtas de bases que são altamente divergentes de todas as outras sequências do alinhamento, em particular quando sítios vizinhos têm chamadas de base ambíguas.

Figura 4. Ramo longo espúrio mostrado em uma filogenia de probabilidade máxima não enraizada construída a partir de um alinhamento de genomas do SARS-CoV-2 completos e parciais, onde um único genoma estava desalinhado em relação ao resto. Esse é um exemplo extremo. Pequenos erros de montagem ou pequenas regiões de desalinhamento podem resultar em um ramo terminal mais longo do que a média, mas não tão extremo.

6.8.4 Remoção de sequências recombinantes

Embora não haja evidências até o momento de recombinação dentro do SARS-CoV-2, sabe-se que os coronavírus se recombinam e as sequências devem ser verificadas quanto a formas recombinantes conforme a pandemia se expande. Os vírus recombinantes não podem ser colocados apropriadamente em uma árvore filogenética a partir de uma única análise de todo o genoma, pois as seções do genoma de cada vírus ancestral teriam histórias diferentes e, portanto, seriam colocadas em posições filogenéticas diferentes.

A inclusão de sequências recombinantes pode levar a estimativas incorretas da taxa evolutiva e do posicionamento filogenético. Se forem detectadas sequências recombinantes, elas podem ser removidas ou várias árvores filogenéticas podem ser estimadas a partir das subseções do alinhamento que caem em ambos os lados dos pontos de quebra recombinantes.

A detecção de uma recombinação é um desafio para muitos conjuntos de dados do SARS-CoV-2 porque as ferramentas existentes não são projetadas para uso em conjuntos de dados muito grandes, com milhares de sequências que também têm diversidade genética relativamente baixa. A detecção de múltiplas homoplasias (onde uma substituição surgiu independente em linhagens filogenéticas separadas) pode indicar a possibilidade de uma recombinação, mas deve ser investigada cuidadosamente, pois as homoplasias também podem ser causadas por mutação. O software RDP4 pode ser usado para examinar até 2.500 sequências alinhadas usando vários testes de recombinação, *(192)* embora ainda não tenha sido determinada sua sensibilidade para detecção precisa de recombinação em linhagens SARS-CoV-2. Seria benéfico haver estratégias melhoradas ou comparadas para detecção de recombinação em conjuntos de dados do SARS-CoV-2.

6.8.5 Ferramentas filogenéticas

Com um alinhamento de alta qualidade do genoma, é possível reconstruir a árvore filogenética correspondente. Os métodos filogenéticos de união de vizinhos são rápidos e podem ser úteis para a exploração inicial de grandes conjuntos de dados genéticos. No entanto, eles consideram apenas uma única árvore possível e não devem ser usados para fazer inferências sobre a relação filogenética. O FastTree também é rápido e produz uma estimativa filogenética de máxima verossimilhança aproximada, que pode ser uma alternativa apropriada aos métodos de união de vizinhos para exploração de dados. *(193)*

Muitos programas de máxima verossimilhança e bayesianos filogenéticos e filodinâmicos são apropriados para inferência filogenética. Cada um requer a especificação de um modelo de evolução do sítio. Isso pode ser escolhido com base nas informações contidas no alinhamento, usando-se um software como o ModelTest-NG. *(195)* Os softwares comumente usados para inferência de árvore de máxima verossimilhança incluem o PhyML *(195)*, o RAxML *(196)* e o IQ-TREE *(197, 198)*. O RAxML é projetado específico para velocidade de execução onde o alinhamento contém milhares de sequências, ao passo que o PhyML e o IQ-TREE são mais lentos, mas têm sido com frequência demonstrados como altamente precisos. O IQ-TREE tem a funcionalidade adicional de realizar primeiro um teste de modelo para identificar a escolha

mais apropriada do modelo de substituição a partir dos dados e também fornece um método de inicialização ultrarrápido para estimar o apoio de ramificação. O IQ-TREE também realiza uma verificação de complexidade nos dados de entrada e rejeita sequências que contenham muitas ambiguidades ou outros artefatos que possam interferir na reconstrução da filogenia. Sempre devem ser calculadas estatísticas de apoio de ramificação (por exemplo, apoio de 100 bootstraps, em que 100 árvores são reestimadas com base em alinhamentos fictícios gerados a partir de reamostragem aleatória com substituição de sítios nos sítios de alinhamento verdadeiro) para avaliar a robustez dos padrões de agrupamento. Essas abordagens filogenéticas são úteis para investigar a relação evolutiva, mas não podem ser usadas para realizar inferências filodinâmicas (seções 5.4 e 5.5).

Para pequenos conjuntos de dados (o ideal é que não sejam mais de 500-1000 para evitar problemas de convergência e conclusão de execução extremamente lenta, embora o número exato dependa da disponibilidade de computação de alto desempenho e do conjunto de dados em questão), pode ser possível usar métodos probabilísticos como os implementados no BEAST *(199)* ou no BEAST2 *(200)*. Esses métodos podem ser usados para estimar o tempo de emergência de clados específicos de interesse (por exemplo, surtos locais), a propagação geográfica de um surto e parâmetros demográficos, incluindo o tamanho da população do vírus ao longo do tempo (seções 5.4 e 5.5). Para análises focadas apenas na estimativa do tempo desde a divergência para um grupo de genomas virais, em especial quando esses conjuntos de dados são grandes, pode ser suficiente e mais tratável, em termos computacionais, usar métodos menos complexos que combinem datas de amostragem com árvores de máxima verossimilhança pré-computadas, como os programas de software de datação por mínimos quadrados (LSD) *(201)* ou o TreeTime *(191)* . Todos esses métodos exigem um "sinal temporal" suficiente dentro do conjunto de dados, de modo que a evolução das linhagens virais possa ser vista de maneira semelhante a um relógio, com a ocorrência de substituições em uma taxa relativamente previsível. A maneira exata de traçar uma linha que separe a sinalização temporal insuficiente da suficiente com respeito ao SARS-CoV-2 foi o foco de grande parte do trabalho filodinâmico inicial.*(45)* Já houve vários exemplos de análises filogenéticas e filodinâmicas do SARS-CoV-2 dimensionadas no tempo.*(59, 85)* Embora o limite filodinâmico (o ponto no tempo em que uma alteração evolutiva molecular suficiente se tenha acumulado nas amostras de genoma disponíveis de modo a obter estimativas filodinâmicas robustas) tenha sido alcançado em algumas análises, os subconjuntos de dados de sequência disponíveis correspondentes a *clusters* locais em áreas geográficas específicas devem ser tratados com cuidado e reavaliados antes do uso, a fim de determinar a aplicabilidade dos métodos filodinâmicos.

Embora os métodos baseados em rede (por exemplo, métodos de junção de haplótipos, redes de junção mediana) sejam rápidos e simples de executar e estejam presentes na literatura publicada sobre SARS-CoV-2, as redes carecem de enraizamento filogenético adequado que é importante para a compreensão das histórias evolutivas. Também carecem de um modelo apropriado de evolução do sítio, sendo baseadas, em vez disso, na similaridade das sequências do genoma isoladamente, e não avaliam ou capturam a robustez dos padrões de conectividade exibidos. A construção de uma árvore filogenética será, portanto, normalmente tão apropriada, ou mais apropriada, do que a construção de uma rede para analisar as sequências do genoma viral do SARS-CoV-2. *(202)*

6.8.6 Visualização

As árvores filogenéticas podem ser visualizadas localmente usando-se uma ampla variedade de softwares disponíveis de forma gratuita (por exemplo, o FigTree e o MEGA *(203)*) e softwares comerciais.

O aplicativo da internet Microreact fornece uma exibição interativa de uma árvore filogenética inserida pelo usuário, permitindo a estruturação filogenética por localização (longitude e latitude), categoria (por exemplo, país) e tempo a ser visualizado. *(204)* O mapeamento de localizações de pontas filogenéticas em relação à posição da árvore pode ser útil para explorar a estruturação geográfica da diversidade do SARS-CoV-2 e para confirmação rápida, quando relevante, de quaisquer dados que tenham sido geocodificados adequadamente. O Microreact requer um arquivo de entrada contendo metadados, como data e local de amostragem, e uma árvore filogenética. Os projetos carregados podem ser compartilhados publicamente ou mantidos privados e atualizados pelo usuário conforme necessário. Os projetos que estão publicamente disponíveis no momento incluem uma distribuição global de linhagens SARS-CoV-2 que está sendo atualizada pelo COVID-19 Genomics UK Consortium.

Não são mostrados arquivos de árvore filogenética com estatísticas de apoio de ramificação e, portanto, as árvores provenientes de conjuntos de dados disponíveis publicamente devem ser baixadas para inspeção local, se necessário. São fornecidas informações adicionais sobre os métodos usados para construção de filogenias a critério do autor do projeto. Isso é útil para permitir a consideração adequada dessas filogenias.

O Nextstrain *(183)* fornece uma exibição interativa da evolução e diversidade geográfica do SARS-CoV-2 e outros patógenos. Colaboradores e desenvolvedores fazem a curadoria das visualizações filogenéticas online globais e regionais que têm sido acessadas com frequência durante a pandemia da COVID-19. Os usuários podem configurar sua própria filogenia Augur local e visualização de mapas para analisar os dados com base em arquivos de entrada de sequências, filogenias e metadados. O Nextstrain é uma ferramenta poderosa e rápida para explorar padrões de ampla escala de estruturação geográfica. No entanto, toda filogenia deve sempre ser interpretada com cautela, levando em consideração os intervalos de confiança nas datas de divergência fornecidas e a incerteza na localização geográfica exibida, e dentro do contexto de “narrativas Nextstrain” explicativas, quando disponíveis. Não são exibidas árvores filogenéticas com estatísticas de apoio de ramificação, portanto, a ordem de ramificação mostrada não deve ser presumida como sendo exata nem utilizada para orientar decisões de política sem investigação adicional para confirmação dos achados. As localizações geográficas e o tempo de divergência dos ramos filogenéticos são inferidos usando-se métodos menos complexos, porém mais rápidos do que os comumente empregados no BEAST ou no BEAST2. Seria importante haver análises que comparassem o grau de concordância entre os diferentes métodos para SARS-CoV-2: não é incomum haver disparidades entre os diferentes métodos *(205)*.

Conforme descrito na Seção 5, a amostragem não aleatória de sequências pode influenciar interpretações e conclusões filogenéticas e filogeográficas. É importante estar atento a esses possíveis vieses ao interpretar quaisquer visualizações filogenéticas.

6.8.7 Classificação de linhagem

Hoje em dia não existe um sistema de nomenclatura formal universalmente aceito para as linhagens evolutivas do SARS-CoV-2. Várias nomenclaturas propostas usam os mesmos nomes (por exemplo, “A1”) para se referir a diferentes linhagens e, portanto, é importante declarar qual nomenclatura está sendo usada em toda descrição. A adoção global de um sistema único de nomenclatura facilitaria a comunicação científica sobre linhagens específicas e evitaria a confusão gerada pelo uso de múltiplos sistemas.

No momento há três sistemas de nomenclatura comumente usados para clados/linhagens SARS-CoV-2. Tanto o GISAID EpiCoV™ quanto o Nextstrain visam fornecer uma ampla categorização da diversidade em circulação global por meio da nomeação de diferentes clados filogenéticos. Rambaut et al. *(189)* propôs uma nomenclatura dinâmica para as linhagens SARS-CoV-2 que se concentra nas linhagens virais em circulação ativa e naquelas que se espalham para novos locais. Há um software disponível que permite aos usuários atribuir suas próprias sequências a essas linhagens, inclusive via Pangolin, Nextstrain e CoV-Glue. *(183, 188, 189)*

Tendo em vista que não existe hoje uma nomenclatura universalmente aceita, a melhor abordagem ao relatar linhagens é declarar a nomenclatura de clados específicos de todos os três sistemas mais usados, ou ao menos declarar, de forma explicíta, qual nomenclatura está sendo usada.

6.8.8 Enraizamento filogenético

Independentemente do software filogenético e do método usado, a escolha de um ou mais grupos externos é importante e terá um efeito em como a raiz da árvore será determinada. Isso, por sua vez, afetará as estimativas de tempo desde a divergência. Um grupo externo é uma sequência selecionada para estar tão intimamente relacionada quanto possível às sequências de interesse, mas sabendo-se que não faz parte do mesmo clado. Na prática, a mais antiga sequência de referência do SARS-CoV-2 disponível é frequentemente usada como um grupo externo ao se construir uma filogenia de genomas de uma variedade de fontes geográficas. Para investigação de *clusters* locais, pode ser apropriado escolher um genoma mais estreitamente relacionado de fora do conjunto de dados a ser analisado.

7 Conclusões e necessidades futuras

O sequenciamento rápido de genomas virais agora é possível em vários locais, e as análises das sequências genômicas do SARS-CoV-2 têm um enorme potencial para orientar os esforços de saúde pública referentes à COVID-19. A geração rápida e o compartilhamento global de sequências genômicas virais fornecem informações que contribuirão para a compreensão da transmissão e o planejamento de estratégias clínicas e epidemiológicas de mitigação.

O diálogo entre os órgãos de saúde pública, os geradores de dados e os analistas é fundamental para garantir que os dados sejam gerados e usados de forma adequada para máximo benefício da saúde pública. É necessária uma consideração prévia cuidadosa do motivo pelo qual o sequenciamento está sendo realizado, pois isso afetará a escolha das amostras, a comparação de metadados e as análises subsequentes. O sequenciamento deve ser conduzido levando-se em consideração os recursos e capacidades disponíveis e não deve desviar a capacidade de outras áreas igualmente vitais. Devem ser estabelecidos canais de comunicação claros para compartilhar resultados, amostras e dados com as partes interessadas apropriadas, de modo que as informações possam ser usadas para melhorar a saúde pública o mais rápido possível.

É complexo traduzir as sequências do genoma do SARS-CoV-2 em resultados informativos, e isso frequentemente requer substancial treinamento especializado para garantir que as violações das suposições do modelo não resultem em um entendimento incorreto da epidemiologia do vírus. Uma compreensão clara dos benefícios e limitações das análises genômicas permitirá uma avaliação confiável de onde as ferramentas genômicas podem ampliar ou apoiar as abordagens existentes e onde, por sua vez, a modelagem epidemiológica ou a experimentação laboratorial podem ser mais robustas. É importante haver uma parceria entre especialistas com diferentes conjuntos de habilidades, pois nem todos os laboratórios têm experiência local em todas as áreas. Apesar dos avanços recentes na facilidade com que podem ser geradas sequências virais, os desafios permanecem. Em muitos locais, a necessidade de importação rápida de reagentes sensíveis à temperatura foi uma barreira significativa para a adoção de abordagens de sequenciamento portátil dentro do país no início da COVID-19. Devem ser encontradas soluções para que os países desenvolvam sua capacidade de realizar atividades de sequenciamento em futuras emergências de saúde pública, bem como durante a atual pandemia. Também seria benéfico haver financiamento em apoio a atividades que validem e comparem as diversas estratégias de sequenciamento e análise publicadas para garantir uma seleção informada apropriada.

A análise e a interpretação dos dados de sequência genômica viral não são diretas. Os laboratórios que planejam adotar o sequenciamento pela primeira vez podem se beneficiar de programas que forneçam apoio para a validação formal de seus pipelines de sequenciamento. Os conjuntos de dados genômicos globais gerados para SARS-CoV-2 são muito grandes para muitas ferramentas atuais; são necessárias melhorias para permitir que conjuntos de dados cada vez maiores sejam analisados rapidamente durante emergências de saúde pública e, quando possível, aumentar o nível de automação. Também seria benéfico haver uma melhor compreensão acadêmica das necessidades das agências de saúde pública e de como os resultados podem ser melhor

apresentados para enfatizar as implicações práticas, embora levando em consideração a incerteza analítica.

Os laboratórios de saúde pública geralmente têm mais experiência em genética molecular do que em filogenética computacional e bioinformática. É necessário um investimento fortalecido de longo prazo na formação em filogenética e bioinformática para obter o máximo benefício do aumento das possibilidades de sequenciamento em laboratório nesta e nas subsequentes epidemias.

Repositórios como o GISAID encorajam e facilitam o compartilhamento de dados na COVID-19. No entanto, ainda são necessários debates mais amplos para garantir melhorias contínuas no compartilhamento de dados durante emergências de saúde pública. Atualmente, muitos pesquisadores continuam relutantes em compartilhar os dados de sequência genômica até que uma publicação preprint tenha sido preparada. As razões disso devem ser pesquisadas e devem ser propostas soluções. Também são necessários debate e acordo mais extensos sobre o credenciamento apropriado para produtores de dados em diferentes circunstâncias para encorajar o compartilhamento de dados. É necessário desenvolver novos padrões ou métricas de credenciamento de dados e haver um comprometimento por parte dos periódicos de que sejam mantidas práticas justas de uso de dados.

É importante haver envolvimento público mais amplo por parte dos cientistas para reduzir a disseminação de informações falsas durante as emergências de saúde pública atuais e futuras. Seria benéfico maior apoio e treinamento para os cientistas sobre como as mensagens científicas podem ser efetivamente compartilhadas com o público em geral. É essencial garantir que os pacientes e o público entendam o valor e as limitações dos dados de sequência genômica viral para apoiar as consultas públicas sobre o uso apropriado dos metadados dos pacientes durante emergências de saúde pública.

Referências

1. Roy S, LaFramboise WA, Nikiforov YE, Nikiforova MN, Routbort MJ, Pfeifer J et al. Next-generation sequencing informatics: challenges and strategies for implementation in a clinical environment. Arch Pathol Lab Med. 2016;140:958-75. doi: 10.5858/arpa.2015- 0507-RA.
2. Gu W, Miller S, Chiu CY. Clinical metagenomic next-generation sequencing for pathogen detection. Annu Rev Pathol. 2019;14:319-38. doi: 10.1146/annurev- pathmechdis-012418-012751.
3. Houldcroft CJ, Beale MA, Breuer J. Clinical and biological insights from viral genome sequencing. Nat Rev Microbiol. 2017;15:183-92. doi: 10.1038/nrmicro.2016.182.
4. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L et al. Real-time, portable genome sequencing for Ebola surveillance. Nature. 2016;530:228-32. doi: 10.1038/nature16996.
5. Peiris JSM, Lai ST, Poon LLM, Guan Y, Yam LYC, Lim W et al. Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet. 2003;361:1319-25. doi: 10.1016/s0140-6736(03)13077-2.
6. Drosten C, Günther S, Preiser W, van der Werf S, Brodt H-R, Becker S et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Eng J Med. 2003;348:1967-76. doi: 10.1056/NEJMoa030747.
7. Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S et al. A novel coronavirus associated with severe acute respiratory syndrome. N Eng J Med. 2003;348:1953-66. doi: 10.1056/NEJMoa030781.
8. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J et al. A novel coronavirus from patients with pneumonia in China, 2019. N Eng J Med. 2020;382:727-33. doi: 10.1056/NEJMoa2001017.
9. Organização Mundial da Saúde. Novo coronavírus (2019-nCoV): Relatório de situação 1. Genebra; 2020 (https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10_4, acessado em 2 de novembro de 2020).
10. Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, Hollingsworth TD et al. Pandemic potential of a strain of influenza A(H1N1): early findings. Science. 2009;324:1557-61. doi: 10.1126/science.1176062.
11. Rambaut A, Holmes E. The early molecular epidemiology of the swine-origin A/H1N1 human influenza pandemic. PLoS Curr. 2009;1:RRN1003. doi: 10.1371/currents.rrn1003.
12. Smith GJD, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG et al. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature. 2009;459:1122-5. doi: 10.1038/nature08182.
13. Mena I, Nelson MI, Quezada-Monroy F, Dutta J, Cortes-Fernández R, Lara-Puente JH et al. Origins of the 2009 H1N1 influenza pandemic in swine in Mexico. eLife. 2016;5:e16777. doi: 10.7554/eLife.16777.
14. WHO MERS-CoV Research Group. State of knowledge and data gaps of Middle East respiratory syndrome coronavirus (MERS-CoV) in humans. PLoS Curr. 2013;5. doi: 10.1371/currents.outbreaks.0bf719e352e7478f8ad85fa30127ddb8.
15. Haagmans BL, Al Dhahiry SHS, Reusken CBEM, Raj VS, Galiano M, Myers R et al. Middle East respiratory syndrome coronavirus in dromedary camels: an outbreak investigation. Lancet Infect Dis. 2014;14:140-5. doi: 10.1016/S1473-3099(13)70690-X.

16. Sabir JSM, Lam TTY, Ahmed MMM, Li L, Shen Y, Abo-Aba SEM et al. Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia. Science. 2016;351:81-4. doi: 10.1126/science.aac8608.

17. Azhar EI, El-Kafrawy SA, Farraj SA, Hassan AM, Al-Saeed MS, Hashem AM et al. Evidence for camel-to-human transmission of MERS coronavirus. N Eng J Med. 2014;370:2499-505. doi: 10.1056/NEJMoa1401505.

18. Memish ZA, Cotten M, Meyer B, Watson SJ, Alsahafi AJ, Al Rabeeah AA et al. Human infection with MERS coronavirus after exposure to infected camels, Saudi Arabia, 2013. Emerg Infect Dis. 2014;20:1012-5. doi: 10.3201/eid2006.140402.

19. Chu DKW, Hui KPY, Perera RAPM, Miguel E, Niemeyer D, Zhao J et al. MERS coronaviruses from camels in Africa exhibit region-dependent genetic diversity. Proc Natl Acad Sci USA. 2018;115:3144-9. doi: 10.1073/pnas.1718769115.

20. Dudas G, Carvalho LM, Rambaut A, Bedford T. MERS-CoV spillover at the camel- human interface. eLife. 2018;7:e31257. doi: 10.7554/eLife.31257.

21. Baize S, Pannetier D, Oestereich L, Rieger T, Koivogui L, Magassouba NF et al. Emergence of Zaire Ebola virus disease in Guinea. N Eng J Med. 2014;371:1418-25. doi: 10.1056/NEJMoa1404505.

22. Gire SK, Goba A, Andersen KG, Sealfon RSG, Park DJ, Kanneh L et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science. 2014;345:1369-72. doi: 10.1126/science.1259657.

23. Carroll MW, Matthews DA, Hiscox JA, Elmore MJ, Pollakis G, Rambaut A et al. Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa. Nature. 2015;524:97-101. doi: 10.1038/nature14594.

24. Dudas G, Rambaut A. Análise filogenética of Guinea 2014 ebov ebolavirus outbreak. PLoS Curr. 2014;6. doi: 10.1371/currents.outbreaks.84eefe5ce43ec9dc0bf0670f7b8b417d.

25. Park DJ, Dudas G, Wohl S, Goba A, Whitmer SLM, Andersen KG et al. Ebola virus epidemiology, transmission, and evolution during seven months in Sierra Leone. Cell. 2015;161:1516-26. doi: 10.1016/j.cell.2015.06.007.

26. Simon-Loriere E, Faye O, Faye O, Koivogui L, Magassouba N, Keita S et al. Distinct lineages of Ebola virus in Guinea during the 2014 West African epidemic. Nature. 2015;524:102-4. doi: 10.1038/nature14612.

27. Tong Y-G, Shi W-F, Liu D, Qian J, Liang L, Bo X-C et al. Genetic diversity and evolutionary dynamics of Ebola virus in Sierra Leone. Nature. 2015;524:93-6. doi: 10.1038/nature14490.

28. Ladner JT, Wiley MR, Mate S, Dudas G, Prieto K, Lovett S et al. Evolution and spread of Ebola virus in Liberia, 2014-2015. Cell Host Microbe. 2015;18:659-69. doi: 10.1016/j.chom.2015.11.008.

29. Volz E, Pond S. Phylodynamic analysis of Ebola virus in the 2014 Sierra Leone epidemic. PLOS Curr. 2014;24 doi:10.1371/currents.outbreaks.6f7025f1271821d4c815385b08f5f80e.

30. Stadler T, Kühnert D, Rasmussen DA, Plessis DL. Insights into the early epidemic spread of Ebola in Sierra Leone provided by viral sequence data. PLOS Curr. 2014. doi: 10.1371/currents.outbreaks.02bc6d927ecee7bbd33532ec8ba6a25f.

31. Mate SE, Kugelman JR, Nyenswah TG, Ladner JT, Wiley MR, Cordier-Lassalle T et al. Molecular evidence of sexual transmission of Ebola virus. N Eng J Med. 2015;373:2448- 54. doi: 10.1056/NEJMoa1509773.

32. Felsenstein J. Cases in which parsimony or compatibility methods will be positively misleading. Syst Zoology. 1978;27:401-10. doi: 10.2307/2412923.

33. Holmes EC, Dudas G, Rambaut A, Andersen KG. The evolution of Ebola virus: Insights from the 2013–2016 epidemic. Nature. 2016;538:193-200. doi: 10.1038/nature19790.

34. Arias A, Watson SJ, Asogun D, Tobin EA, Lu J, Phan MVT et al. Rapid outbreak sequencing of Ebola virus in Sierra Leone identifies transmission chains linked to sporadic cases. Vir Evol. 2016;2:vew016. doi: 10.1093/ve/vew016.

35. Hoenen T, Groseth A, Rosenke K, Fischer RJ, Hoenen A, Judson SD et al. Nanopore sequencing as a rapidly deployable Ebola outbreak tool. Emerg Infect Dis. 2016;22:331- 4. doi: 10.3201/eid2202.151796.

36. Smits SL, Pas SD, Reusken CB, Haagmans BL, Pertile P, Cancedda C et al. Genotypic anomaly in Ebola virus strains circulating in Magazine Wharf area, Freetown, Sierra Leone, 2015. Euro Surveill. 2015;20. doi: 10.2807/1560-7917.ES.2015.20.40.30035.

37. Faria NR, Azevedo RdSdS, Kraemer MUG, Souza R, Cunha MS, Hill SC et al. Zika virus in the Americas: early epidemiological and genetic findings. Science. 2016;352:345-9. doi: 10.1126/science.aaf5036.

38. Faria NR, Quick J, Claro IM, Thézé J, de Jesus JG, Giovanetti M et al. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. Nature. 2017;546:406-10. doi: 10.1038/nature22401.

39. Metsky HC, Matranga CB, Wohl S, Schaffner SF, Freije CA, Winnicki SM et al. Zika virus evolution and spread in the Americas. Nature. 2017;546:411-5. doi: 10.1038/nature22402.

40. Grubaugh ND, Ladner JT, Kraemer MUG, Dudas G, Tan AL, Gangavarapu K et al. Genomic epidemiology reveals multiple introductions of Zika virus into the United States. Nature. 2017;546:401-5. doi: 10.1038/nature22400.

41. Grubaugh ND, Ladner JT, Lemey P, Pybus OG, Rambaut A, Holmes EC et al. Tracking virus outbreaks in the twenty-first century. Nat Microbiol. 2019;4:10. doi: 10.1038/s41564-018-0296-2.

42. Gardy JL, Loman NJ. Towards a genomics-informed, real-time, global pathogen surveillance system. Nat Rev Genet. 2018;19:9-20. doi: 10.1038/nrg.2017.88.

43. Rasmussen AL, Katze MG. Genomic signatures of emerging viruses: a new era of systems epidemiology. Cell Host Microbe. 2016;19:611-8. doi: 10.1016/j.chom.2016.04.016.

44. Loewe L, Hill WG. The population genetics of mutations: Good, bad and indifferent. Philos Trans R Soc Lond B Biol Sci. 2010;365:1153-67. doi: 10.1098/rstb.2009.0317.

45. Duchene S, Featherstone L, Haritopoulou-Sinanidou M, Rambaut A, Lemey P, Baele G. Temporal signal and the phylodynamic threshold of SARS-CoV-2. Virus Evol. 2020: 19;6(2). doi:10.1093/ve/veaa061.

46. Duffy S, Shackelton LA, Holmes EC. Rates of evolutionary change in viruses: patterns and determinants. Nat Rev Genet. 2008;9:267-76. doi: 10.1038/nrg2323.

47. Grenfell BT, Pybus OG, Gog JR, Wood JLN, Daly JM, Mumford JA et al. Unifying the epidemiological and evolutionary dynamics of pathogens. Science. 2004;303:327-32. doi: 10.1126/science.1090727.

48. Volz EM, Koelle K, Bedford T. Viral phylodynamics. PLoS Comput Biol. 2013;9. doi: 10.1371/journal.pcbi.1002947.

49. Pybus OG, Rambaut A. Evolutionary analysis of the dynamics of viral infectious disease. Nat Rev Genet. 2009;10:540-50. doi: 10.1038/nrg2583.

50. Sanjuán R, Domingo-Calap P. Mechanisms of viral mutation. Cell Mol Life Sci. 2016;73:4433-48. doi: 10.1007/s00018-016-2299-6.

51. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species severe acute respiratory syndrome-related coronavirus : classifying 2019-nCov and naming it SARS-CoV-2. Nat Microbiol. 2020;5:536-44. doi: 10.1038/s41564-020- 0695-z.

52. Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579:265-9. doi: 10.1038/s41586-020-2008-3.

53. Candido DDS, Claro IM, Jesus DJG, Souza DWM, Moreira FRR, Dellicour S et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil. Science (New York, NY). 2020;369:1255-60. doi: 10.1101/2020.06.11.20128249.

54. Emanuel EJ, Wendler D, Grady C. What makes clinical research ethical? JAMA. 2000; 283(20): 2701-11.

55. Diretrizes sobre questões éticas na vigilância em saúde pública da OMS. Genebra: Organização Mundial da Saúde; 2017 (https://www.who.int/ethics/publications/public-health-surveillance/en/, acessado em 15 de novembro de 2020).

56. Organização Mundial da Saúde. Declaração de política sobre compartilhamento de dados pela Organização Mundial da Saúde no contexto de emergências de saúde pública. Genebra; 2016.

57. Thézé J, Li T, Plessis dL, Bouquet J, Kraemer MUG, Somasekar S et al. Genomic epidemiology reconstructs the introduction and spread of Zika virus in Central America and Mexico. Cell Host Microbe. 2018;23:855-64.e7. doi: 10.1016/j.chom.2018.04.017.

58. COVID-19 data portal. 2020 (https://www.covid19dataportal.org/sequences, acessado em 1º de novembro de 2020).

59. Lu J, du Plessis L, Liu Z, Hill V, Kang M, Lin H et al. Genomic epidemiology of SARS- CoV-2 in Guangdong Province, China. Cell. 2020;181:997-1003.e9. doi: 10.1016/j.cell.2020.04.023.

60. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579:270-3. doi: 10.1038/s41586-020-2012-7.

61. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med. 2020;382:1199- 207. doi: 10.1056/NEJMoa2001316.

62. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CovV2. Nature Medicine. 2020;26:450-2. doi: 10.1038/s41591-020-0820-9.

63. Relatório da missão conjunta OMS-China sobre a doença causada pelo coronavírus 2019 (COVID-19). Genebra: Organização Mundial da Saúde; 2020.(https://www.who.int/publications/i/item/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19), acessado em 28 de dezembro de 2020)

64. Cui J, Li F, Shi Z-L. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol. 2019;17:181-92. doi: 10.1038/s41579-018-0118-9.

65. Hu B, Zeng L-P, Yang X-L, Ge X-Y, Zhang W, Li B et al. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. PLoS Pathog. 2017;13:e1006698. doi: 10.1371/journal.ppat.1006698.

66. Lin X-D, Wang W, Hao Z-Y, Wang Z-X, Guo W-P, Guan X-Q et al. Extensive diversity of coronaviruses in bats from China. Virology. 2017;507:1-10. doi: 10.1016/j.virol.2017.03.019.

67. Boni MF, Lemey P, Jiang X, Lam TT-Y, Perry B, Castoe T et al. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nature Microbiology. 2020;5:1408-17. doi: 10.1101/2020.03.30.015008.

68. Zhou H, Chen X, Hu T, Li J, Song H, Liu Y et al. A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein. Curr Biol. 2020;30:2196-203.e3. doi: 10.1016/j.cub.2020.05.023.

69. Lam TT-Y, Jia N, Zhang Y-W, Shum MH-H, Jiang J-F, Zhu H-C et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. Nature. 2020:1-4. doi: 10.1038/s41586-020-2169-0.

70. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell. 2020;181:271-80.e8. doi: 10.1016/j.cell.2020.02.052.

71. Letko M, Marzi A, Munster V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nat Microbiol. 2020;5:562-9. doi: 10.1038/s41564-020-0688-y.

72. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O et al. Cryo-em structure of the 2019-nCoV spike in the prefusion conformation. Science (New York, Ny). 2020;367:1260-3. doi: 10.1126/science.abb2507.

73. Colson P, Scola BL, Esteves-Vieira V, Ninove L, Zandotti C, Jimeno M-T et al. Plenty of coronaviruses but no SARS-CoV-2. (Letter to the editor). Euro Surveill. 2020;25:2000171. doi: 10.2807/1560-7917.ES.2020.25.8.2000171.

74. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro Surveill. 2020;25. doi: 10.2807/1560-7917.ES.2020.25.3.2000045.

75. Information for laboratories about coronavirus (COVID-19). Atlanta: Centers for Disease Control and Prevention; 2020. (https://www.cdc.gov/coronavirus/2019-ncov/lab/rt-pcr-panel-primer-probes.html, acessado em 26 de junho de 2020).

76. University of Hong Kong, School of Public Health. Detection of 2019 novel coronavirus (2019-nCoV) in suspected human cases by RT-PCR. 2020 (https://www.who.int/docs/default-source/coronaviruse/peiris-protocol-16-1-20.pdf?sfvrsn=af1aac73_4.

77. Teste de diagnóstico para SARS-CoV-2. Orientação provisória. Genebra: Organização Mundial da Saúde; 2020 (https://www.who.int/publications/i/item/diagnostic-testing-for-sars- cov-2, acessado em 19 de novembro de 2020).

78. Ren Y, Zhou Z, Liu J, Lin L, Li S, Wang H et al. A strategy for searching antigenic regions in the SARS-CoV spike protein. Genomics Proteomics Bioinformatics. 2003;1:207-15. doi: 10.1016/s1672-0229(03)01026-x.

79. Kumar S, Maurya VK, Prasad AK, Bhatt MLB, Saxena SK. Structural, glycosylation and antigenic variation between 2019 novel coronavirus (2019-nCoV) and SARS coronavirus (SARS-CoV). Virusdisease. 2020:1-9. doi: 10.1007/s13337-020-00571-5.

80. Melén K, Kakkola L, He F, Airenne K, Vapalahti O, Karlberg H et al. Production, purification and immunogenicity of recombinant Ebola virus proteins – a comparison of Freund's adjuvant and adjuvant system 03. J Virol Methods. 2017;242:35-45. doi: 10.1016/j.jviromet.2016.12.014.

81. Ziegler T, Matikainen S, Rönkkö E, Österlund P, Sillanpää M, Sirén J et al. Severe acute respiratory syndrome coronavirus fails to activate cytokine-mediated innate immune responses in cultured human monocyte-derived dendritic cells. J Virol. 2005;79:13800-5. doi: 10.1128/JVI.79.21.13800-13805.2005.

82. Esboço do panorama das vacinas candidatas contra a COVID-19. Genebra: Organização Mundial da Saúde; 2020 (https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines, acessado em 26 de junho de 2020).

83. Li G, Clercq ED. Therapeutic options for the 2019 novel coronavirus (2019-nCoV). Nat Rev Drug Discov. 2020;19:149-50. doi: 10.1038/d41573-020-00016-0.

84. Bedford T, Greninger AL, Roychoudhury P, Starita LM, Famulare M, Huang M-L et al. Cryptic transmission of SARS-CoV-2 in Washington State. Science. 2020;370:571-5. doi: 10.1101/2020.04.02.20051417.

85. Volz E, Fu H, Wang H, Xi X, Chen W, Liu D et al. Genomic epidemiology of a densely sampled COVID19 outbreak in China. medRxiv. 2020:2020.03.09.20033365. doi: 10.1101/2020.03.09.20033365.

86. Zehender G, Lai A, Bergna A, Meroni L, Riva A, Balotta C et al. Genomic characterization and phylogenetic analysis of SARS-CoV-2 in Italy. J Med Virol. 2020;92(9):1637-1640. doi: 10.1002/jmv.25794.

87. Worobey M, Pekar J, Larsen BB, Nelson MI, Hill V, Joy JB et al. The emergence of SARS-CoV-2 in Europe and North America. Science. 2020;370:564-70.

88. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. PLoS Comput Biol. 2009;5:e1000520. doi: 10.1371/journal.pcbi.1000520.

89. Lemey P, Rambaut A, Welch JJ, Suchard MA. Phylogeography takes a relaxed random walk in continuous space and time. Mol Biol Evol. 2010;27:1877-85. doi: 10.1093/molbev/msq067.

90. Bloomquist EW, Lemey P, Suchard MA. Three roads diverged? Routes to phylogeographic inference. Trends Ecol Evol. 2010;25:626-32. doi: 10.1016/j.tree.2010.08.010.

91. Faria NR, Suchard MA, Rambaut A, Lemey P. Towards a quantitative understanding of viral phylogeography. Curr OpinVirol. 2011;1:423-9. doi: 10.1016/j.coviro.2011.10.003.

92. Fauver JR, Petrone ME, Hodcroft EB, Shioda K, Ehrlich HY, Watts AG et al. Coast-to- coast spread of SARS-CoV-2 during the early epidemic in the United States. Cell. 2020;181:990-6.e5. doi: 10.1016/j.cell.2020.04.021.

93. Eden J-S, Rockett R, Carter I, Rahman H, de Ligt J, Hadfield J et al. An emergent clade of SARS-CoV-2 linked to returned travellers from Iran. Virus Evol. 2020;6. doi: 10.1093/ve/veaa027.

94. Lemey P, Hong S, Hill V, Baele G, Poletto C, Colizza V et al. Accommodating individual travel history, global mobility, and unsampled diversity in phylogeography: a SARS-CoV-2 case study. bioRxiv. 2020:2020.06.22.165464. doi: 10.1101/2020.06.22.165464.

95. Ewing G, Rodrigo A. Estimating population parameters using the structured serial coalescent with Bayesian mcmc inference when some demes are hidden. Evol Bioinform Online. 2006;2:117693430600200026. doi: 10.1177/117693430600200026.

96. Maio ND, Wu C-H, O'Reilly KM, Wilson D. New routes to phylogeography: A Bayesian structured coalescent approximation. PLoS Genet. 2015;11:e1005421. doi: 10.1371/journal.pgen.1005421.

97. Lemey P, Rambaut A, Bedford T, Faria N, Bielejec F, Baele G et al. Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. PLoS Pathog. 2014;10:e1003932. doi: 10.1371/journal.ppat.1003932.

98. Chaillon A, Gianella S, Dellicour S, Rawlings SA, Schlub TE, Oliveira MFD et al. HIV persists throughout deep tissues with repopulation from multiple anatomical sources. The J Clin Invest. 2020;130:1699-712. doi: 10.1172/JCI134815.

99. Kalkauskas A, Perron U, Sun Y, Goldman N, Baele G, Guindon S et al. Sampling bias and model choice in continuous phylogeography: getting lost on a random walk. bioRxiv. 2020:2020.02.18.954057. doi: 10.1101/2020.02.18.954057.

100. Nylinder S, Lemey P, De Bruyn M, Suchard MA, Pfeil BE, Walsh N et al. On the biogeography of centipeda: a species-tree diffusion approach. Syst Biol. 2014;63:178-91. doi: 10.1093/sysbio/syt102.

101. Dellicour S, Lemey P, Artois J, Lam TT, Fusaro A, Monne I et al. Incorporating heterogeneous sampling probabilities in continuous phylogeographic inference – application to H5N1 spread in the Mekong region. Bioinformatics. 2020;36:2098-104. doi: 10.1093/bioinformatics/btz882.

102. Dudas G, Carvalho LM, Bedford T, Tatem AJ, Baele G, Faria NR et al. Virus genomes reveal factors that spread and sustained the Ebola epidemic. Nature. 2017;544:309-15. doi: 10.1038/nature22040.

103. Dellicour S, Baele G, Dudas G, Faria NR, Pybus OG, Suchard MA et al. Phylodynamic assessment of intervention strategies for the West African Ebola virus outbreak. Nature Communications. 2018;9:1-9. doi: 10.1038/s41467-018-03763-2.

104. Bielejec F, Lemey P, Baele G, Rambaut A, Suchard MA. Inferring heterogeneous evolutionary processes through time: from sequence substitution to phylogeography. Syst Biol. 2014;63:493-504. doi: 10.1093/sysbio/syu015.

105. Sit THC, Brackman CJ, Ip SM, Tam KWS, Law PYT, To EMW et al. Infection of dogs with SARS-CoV-2. Nature. 2020. doi: 10.1038/s41586-020-2334-5.

106. Oreshkova N, Molenaar RJ, Vreman S, Harders F, Munnink BBO, Honing RWH et al. SARS-CoV-2 infection in farmed minks, the Netherlands, April and May 2020. Euro Surveill. 2020;25:2001005. doi: 10.2807/1560-7917.ES.2020.25.23.2001005.

107. Segalés J, Puig M, Rodon J, Avila-Nieto C, Carrillo J, Cantero G et al. Detection of SARS-CoV-2 in a cat owned by a COVID-19-affected patient in Spain. PNAS. 2020;117(40):24790-24793. doi: 10.1073/pnas.2010817117

108. Hughes J, Allen RC, Baguelin M, Hampson K, Baillie GJ, Elton D et al. Transmission of equine influenza virus during an outbreak is characterized by frequent mixed infections and loose transmission bottlenecks. PLoS Pathog. 2012;8. doi: 10.1371/journal.ppat.1003081.

109. Worby CJ, Lipsitch M, Hanage WP. Shared genomic variants: identification of transmission routes using pathogen deep-sequence data. Am J Epidemiol. 2017;186:1209-16. doi: 10.1093/aje/kwx182.

110. Cotten M, Lam TT, Watson SJ, Palser AL, Petrova V, Grant P et al. Full-genome deep sequencing and phylogenetic analysis of novel human betacoronavirus. Emerg Infect Dis. 2013;19:736-42B. doi: 10.3201/eid1905.130057.

111. Shen Z, Xiao Y, Kang L, Ma W, Shi L, Zhang L et al. Genomic diversity of SARS-CoV- 2 in coronavirus disease 2019 patients. Clin Infect Dis. 2020; 71(15):713-720 doi: 10.1093/cid/ciaa203.

112. Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, De Jesus JG, Main BJ et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using primalseq and ivar. Genome Biol. 2019;20:8. doi: 10.1186/s13059-018-1618-7.

113. Volz E, Baguelin M, Bhatia S, Boonyasiri A, Cori A, Cucunubá Z et al. Report 5 – phylogenetic analysis of SARS-CoV-2. London: Imperial College; 2020 (http://www.imperial.ac.uk/medicine/departments/school-public-health/infectious-disease-epidemiology/mrc-global-infectious-disease-analysis/covid-19/report-5- phylogenetics-of-sars-cov-2/, acessado em 26 de junho de 2020).

114. Stadler T, Kühnert D, Bonhoeffer S, Drummond AJ. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proc Natl Acad Sci USA. 2013;110:228-33. doi: 10.1073/pnas.1207965110.

115. Boskova V, Bonhoeffer S, Stadler T. Inference of epidemiological dynamics based on simulated phylogenies using birth-death and coalescent models. PLoS Comput Biol. 2014;10:e1003913. doi: 10.1371/journal.pcbi.1003913.

116. Volz EM, Frost SDW. Sampling through time and phylodynamic inference with coalescent and birth–death models. J R Soc Interface. 2014;11. doi: 10.1098/rsif.2014.0945.

117. Li LM, Grassly NC, Fraser C. Quantifying transmission heterogeneity using both pathogen phylogenies and incidence time series. Mol Biol Evol. 2017;34:2982-95. doi: 10.1093/molbev/msx195.

118. Koelle K, Rasmussen DA. Rates of coalescence for common epidemiological models at equilibrium. J R Soc Interface. 2012;9:997-1007. doi: 10.1098/rsif.2011.0495.

119. Vaughan TG, Leventhal GE, Rasmussen DA, Drummond AJ, Welch D, Stadler T. Estimating epidemic incidence and prevalence from genomic data. Mol Biol Evol. 2019;36:1804-16. doi: 10.1093/molbev/msz106.

120. Volz EM, Siveroni I. Bayesian phylodynamic inference with complex models. PLoS Comput Biol. 2018;14:e1006546. doi: 10.1371/journal.pcbi.1006546.

121. Orientações laboratoriais de biossegurança relacionadas à doença causada pelo coronavírus (COVID-19). Genebra: Organização Mundial da Saúde; 2020 (https://apps.who.int/iris/handle/10665/332076, acessado em 21 de novembro de 2020).

122. Wang W, Xu Y, Gao R, Lu R, Han K, Wu G et al. Detection of SARS-CoV-2 in different types of clinical specimens. JAMA. 2020;323:1843-1844. doi: 10.1001/jama.2020.3786.

123. Chen W, Lan Y, Yuan X, Deng X, Li Y, Cai X et al. Detectable 2019-nCoV viral RNA in blood is a strong indicator for the further clinical severity. Emerg Microbes Infect. 2020;9:469-73. doi: 10.1080/22221751.2020.1732837.

124. Chen X, Zhao B, Qu Y, Chen Y, Xiong J, Feng Y et al. Detectable serum SARS-CoV-2 viral load (RNAaemia) is closely correlated with drastically elevated interleukin 6 (il-6) level in critically ill COVID-19 patients. Clin Infect Dis. 2020; 71(8):1937-1942. doi: 10.1093/cid/ciaa449.

125. Corman VM, Rabenau HF, Adams O, Oberle D, Funk MB, Keller-Stanislawski B et al. SARS-CoV-2 asymptomatic and symptomatic patients and risk for transfusion transmission. Transfusion. 2020; 60(6):1119-1122 doi: 10.1111/trf.15841.

126. Zhang W, Du RH, Li B, Zheng XS, Yang XL, Hu B et al. Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes. Emerg Microbes Infect. 2020;9:386-9. doi: 10.1080/22221751.2020.1729071.

127. Winichakoon P, Chaiwarith R, Liwsrisakun C, Salee P, Goonna A, Limsukon A et al. Negative nasopharyngeal and oropharyngeal swabs do not rule out COVID-19. J Clin Microbiol. 2020;58. doi: 10.1128/JCM.00297-20.

128. Ek P, Bottiger B, Dahlman D, Hansen KB, Nyman M, Nilsson AC. A combination of naso- and oropharyngeal swabs improves the diagnostic yield of respiratory viruses in adult emergency department patients. Infect Dis (Lond). 2019;51:241-8. doi: 10.1080/23744235.2018.1546055.

129. Hammitt LL, Kazungu S, Welch S, Bett A, Onyango CO, Gunson RN et al. Added value of an oropharyngeal swab in detection of viruses in children hospitalized with lower respiratory tract infection. J Clin Microbiol. 2011;49:2318-20. doi: 10.1128/JCM.02605- 10.

130. The COVID-19 Investigation Team. Clinical and virologic characteristics of the first 12 patients with coronavirus disease 2019 (COVID-19) in the United States. Nat Med. 2020;26:861-868. doi: 10.1038/s41591-020-0877-5.

131. Sutjipto HL, Yant TJ, Mendis SM, Abdad MY, Marimuthu K, Ng OT et al. The effect of sample site, illness duration and the presence of pneumonia on the detection of SARS- CoV-2 by real-time reverse-transcription pcr. Open Forum Infect Dis. 2020; 7(9):ofaa335. doi: 10.1093/ofid/ofaa335.

132. Zou L, Ruan F, Huang M, Liang L, Huang H, Hong Z et al. SARS-CoV-2 viral load in upper respiratory specimens of infected patients. N Engl J Med. 2020;382:1177-9. doi: 10.1056/NEJMc2001737.

133. Lai CKC, Chen Z, Lui G, Ling L, Li T, Wong MCS et al. Prospective study comparing deep-throat saliva with other respiratory tract specimens in the diagnosis of novel coronavirus disease (COVID-19). J Infect Dis. 2020; 222(10):1612-1619. doi: 10.1093/infdis/jiaa487.

134. Liu R, Han H, Liu F, Lv Z, Wu K, Liu Y et al. Positive rate of RT-PCR detection of SARS-CoV-2 infection in 4880 cases from one hospital in Wuhan, China, from Jan to Feb 2020. Clin Chim Acta. 2020;505:172-5. doi: 10.1016/j.cca.2020.03.009.

135. Huang Y, Chen S, Yang Z, Guan W, Liu D, Lin Z et al. SARS-CoV-2 viral load in clinical samples from critically ill patients. Am J Respir Crit Care Med. 2020;201:1435- 8. doi: 10.1164/rccm.202003-0572LE.

136. Williams E, Bond K, Zhang B, Putland M, Williamson DA. Saliva as a non-invasive specimen for detection of SARS-CoV-2. J Clin Microbiol. 2020; 24(5):422-427. doi: 10.1128/JCM.00776-20.

137. Pasomsub E, Watcharananan SP, Boonyawat K, Janchompoo P, Wongtabtim G, Suksuwan W et al. Saliva sample as a non-invasive specimen for the diagnosis of coronavirus disease-2019 (COVID-19): a cross-sectional study. Clin Microbiol Infect. 2020. doi: 10.1016/j.cmi.2020.05.001.

138. Yang JR, Deng DT, Wu N, Yang B, Li HJ, Pan XB. Persistent viral RNA positivity during the recovery period of a patient with SARS-CoV-2 infection. J Med Virol. 2020; 92(9):1681-1683. doi: 10.1002/jmv.25940.

139. Guo WL, Jiang Q, Ye F, Li SQ, Hong C, Chen LY et al. Effect of throat washings on detection of 2019 novel coronavirus. Clin Infect Dis. 2020; 71(8):1980-1981. doi: 10.1093/cid/ciaa416.

140. To KK, Tsang OT, Leung WS, Tam AR, Wu TC, Lung DC et al. Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. Lancet Infect Dis. 2020;20:565-74. doi: 10.1016/S1473-3099(20)30196-1.

141. Azzi L, Carcano G, Gianfagna F, Grossi P, Gasperina D, Genoni A et al. Saliva is a reliable tool to detect SARS-CoV-2. J Infect. 2020;81. doi: 10.1016/j.jinf.2020.04.005.

142. McCormick-Baw C, Morgan K, Gaffney D, Cazares Y, Jaworski K, Byrd A et al. Saliva as an alternate specimen source for detection of SARS.CoV-2 in symptomatic patients using cepheid xpert xpress SARS-CoV-2. J Clin Microbiol. 2020. doi: 10.1128/JCM.01109-20.

143. Wyllie AL, Fournier J, Casanovas-Massana A, Campbell M, Tokuyama M, Vijayakumar P et al. Saliva or nasopharyngeal swab specimens for detection of SARS-CoV-2. 383(13):1283-1286. N Engl J Med. 2020. doi: 10.1056/NEJMc2016359.

144. Lescure FX, Bouadma L, Nguyen D, Parisey M, Wicky PH, Behillil S et al. Clinical and virological data of the first cases of COVID-19 in Europe: a case series. Lancet Infect Dis. 2020; 20(6):697-706. doi: 10.1016/S1473-3099(20)30200-0.

145. Xing YH, Ni W, Wu Q, Li WJ, Li GJ, Wang WD et al. Prolonged viral shedding in feces of pediatric patients with coronavirus disease 2019. J Microbiol Immunol Infect. 2020; 53(3):473-480. doi: 10.1016/j.jmii.2020.03.021.

146. Zheng S, Fan J, Yu F, Feng B, Lou B, Zou Q et al. Viral load dynamics and disease severity in patients infected with SARS-CoV-2 in Zhejiang province, China, January- March 2020: retrospective cohort study. BMJ. 2020;369:1443. doi: 10.1136/bmj.m1443.

147. Wong MC, Huang J, Lai C, Ng R, Chan FKL, Chan PKS. Detection of SARS-CoV-2 RNA in fecal specimens of patients with confirmed COVID-19: a meta-analysis. J Infect. 2020;81:e31-e8. doi: 10.1016/j.jinf.2020.06.012.

148. Tang JW, To KF, Lo AW, Sung JJ, Ng HK, Chan PK. Quantitative temporal-spatial distribution of severe acute respiratory syndrome-associated coronavirus (SARS-CoV) in post-mortem tissues. J Med Virol. 2007;79:1245-53. doi: 10.1002/jmv.20873.

149. Nicholls JM, Poon LL, Lee KC, Ng WF, Lai ST, Leung CY et al. Lung pathology of fatal severe acute respiratory syndrome. Lancet. 2003;361:1773-8. doi: 10.1016/s0140- 6736(03)13413-7.

150. Pomara C, Li Volti G, Cappello F. COVID-19 deaths: are we sure it is pneumonia? Please, autopsy, autopsy, autopsy! J Clin Med. 2020;9. doi: 10.3390/jcm9051259.

151. Salerno M, Sessa F, Piscopo A, Montana A, Torrisi M, Patane F et al. No autopsies on COVID-19 deaths: a missed opportunity and the lockdown of science. J Clin Med. 2020;9. doi: 10.3390/jcm9051472.

152. Hanley B, Lucas SB, Youd E, Swift B, Osborn M. Autopsy in suspected COVID-19 cases. J Clin Pathol. 2020;73:239-42. doi: 10.1136/jclinpath-2020-206522.

153. Basso C, Calabrese F, Sbaraglia M, Del Vecchio C, Carretta G, Saieva A et al. Feasibility of postmortem examination in the era of COVID-19 pandemic: the experience of a northeast Italy university hospital. Virchows Arch. 2020 477(3):341-347. doi: 10.1007/s00428-020-02861-1.

154. Tian S, Xiong Y, Liu H, Niu L, Guo J, Liao M et al. Pathological study of the 2019 novel coronavirus disease (COVID-19) through postmortem core biopsies. Mod Pathol. 2020;33:1007-14. doi: 10.1038/s41379-020-0536-x.

155. Sekulic M, Harper H, Nezami BG, Shen DL, Sekulic SP, Koeth AT et al. Molecular detection of SARS-CoV-2 infection in FFPE samples and histopathologic findings in fatal SARS-CoV-2 cases. Am J Clin Pathol. 2020; 154(2):190-200. doi: 10.1093/ajcp/aqaa091.

156. Park WB, Kwon NJ, Choi SJ, Kang CK, Choe PG, Kim JY et al. Virus isolation from the first patient with SARS-CoV-2 in Korea. J Korean Med Sci. 2020;35:e84. doi: 10.3346/jkms.2020.35.e84.

157. Le TQM, Takemura T, Moi ML, Nabeshima T, Nguyen LKH, Hoang VMP et al. Severe acute respiratory syndrome coronavirus 2 shedding by travelers, Vietnam, 2020. Emerg Infect Dis . 2020;26:1624-6. doi: 10.3201/eid2607.200591.

158. Pan Y, Zhang D, Yang P, Poon LLM, Wang Q. Viral load of SARS-CoV-2 in clinical samples. Lancet Infect Dis. 2020;20:411-2. doi: 10.1016/S1473-3099(20)30113-4.

159. Wyllie AL, Fournier J, Casanovas-Massana A, Campbell M, Tokuyama M, Vijayakumar P et al. Saliva or nasopharyngeal swab specimens for detection of SARS-CoV-2. N Eng J Med. 2020;383:1283-6. doi: 10.1101/2020.04.16.20067835.

160. Wolfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Muller MA et al. Virological assessment of hospitalized patients with COVID-2019. Nature. 2020; 581(7809):465-469. doi: 10.1038/s41586-020-2196-x.

161. MacCannell D. SARS-CoV-2 sequencing. 2020 (https://github.com/CDCgov/SARS-CoV-2_Sequencing, acessado em 1º de novembro de 2020).

162. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol. 2017;35:833-44. doi: 10.1038/nbt.3935.

163. Bragg L, Tyson GW. Metagenomics using next-generation sequencing. Methods in Mol Biol. 2014;1096:183-201. doi: 10.1007/978-1-62703-712-9_15.

164. Xiao M, Liu X, Ji J, Li M, Li J, Yang L et al. Multiple approaches for massively parallel sequencing of SARS-CoV-2 genomes directly from clinical samples. Genome Med. 2020;12:57. doi: 10.1186/s13073-020-00751-4.

165. Cesare MD. Probe-based target enrichment of SARS-CoV-2. Protocolsio. 2020; 66(11):1450-1458. doi: 10.17504/protocols.io.bd5di826.

166. Vogels CBF, Brito AF, Wyllie AL, Fauver JR, Ott IM, Kalinich CC et al. Analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT–qPCR primer–probe sets. Nat Microbiol. 2020:1-7. doi: 10.1038/s41564-020-0761-6.

167. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K et al. Multiplex PCR method for Minion and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nature Protoc. 2017;12:1261-76. doi: 10.1038/nprot.2017.066.

168. Matteson N. Primalseq: generation of tiled virus amplicons for miseq sequencing. Protocolsio. 2020. doi: 10.17504/protocols.io.bez7jf9n.

169. Gordon P, Mabon P. Nanostripper2020 (https://github.com/nodrogluap/nanostripper, acessado em 15 de julho de 2020).

170. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019;20:257. doi: 10.1186/s13059-019-1891-0.

171. Ounit R, Wanamaker S, Close TJ, Lonardi S. Clark. Fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers. BMC Genomics. 2015;16:236. doi: 10.1186/s12864-015-1419-2.

172. Wu TD, Reeder J, Lawrence M, Becker G, Brauer MJ. Gmap and gsnap for genomic sequence alignment: enhancements to speed, accuracy, and functionality. Methods Mol Biol. 2016;1418:283-334. doi: 10.1007/978-1-4939-3578-9_15.

173. Wick RR, Judd LM, Gorrie CL, Holt KE. Completing bacterial genome assemblies with multiplex minion sequencing. Microb Genom. 2017;3:e000132. doi: 10.1099/mgen.0.000132.

174. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal. 2011;17:10-12.

175. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114-20. doi: 10.1093/bioinformatics/btu170.

176. NCBI. Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome. 2020 (https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2, acessado em 1º de novembro de 2020).

177. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods. 2012;9:357-9. doi: 10.1038/nmeth.1923.

178. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094-100. doi: 10.1093/bioinformatics/bty191.

179. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754-60. doi: 10.1093/bioinformatics/btp324.

180. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27:2987-93. doi: 10.1093/bioinformatics/btr509.

181. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat Methods. 2015;12:733-5. doi: 10.1038/nmeth.3444.

182. Li W, Jaroszewski L, Godzik A. Clustering of highly homologous sequences to reduce the size of large protein databases. Bioinformatics. 2001;17:282-3. doi: 10.1093/bioinformatics/17.3.282.

183. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C et al. Nextstrain: real- time tracking of pathogen evolution. Bioinformatics. 2018;34:4121-3. doi: 10.1093/bioinformatics/bty407.

184. Hong SL, Dellicour S, Vrancken B, Suchard MA, Pyne MT, Hillyard DR et al. In search of covariates of HIV-1 subtype B spread in the United States – a cautionary tale of large- scale Bayesian phylogeography. Viruses. 2020;12:182. doi: 10.3390/v12020182.

185. Katoh K, Standley DM. Mafft multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772-80. doi: 10.1093/molbev/mst010.

186. Katoh K, Rozewicki J, Yamada KD. Mafft online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2019;20:1160-6. doi: 10.1093/bib/bbx108.

187. Wymant C, Blanquart F, Golubchik T, Gall A, Bakker M, Bezemer D et al. Easy and accurate reconstruction of whole HIV genomes from short-read sequence data with shiver. Virus Evol. 2018;4. doi: 10.1093/ve/vey007.

188. Singer J, Gifford R, Cotten M, Robertson D. CoV-glue: a web application for tracking SARS-CoV-2 genomic variation. Preprints 2020; 2020060225. doi: 10.20944/preprints202006.0225.v1.

189. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol. 2020:1-5. doi: 10.1038/s41564-020-0770-5.

190. Rambaut A, Lam TT, Carvalho LM, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-o-gen). Virus Evol. 2016;2. doi: 10.1093/ve/vew007.

191. Sagulenko P, Puller V, Neher RA. Treetime: maximum-likelihood phylodynamic analysis. Virus Evol. 2018;4. doi: 10.1093/ve/vex042.

192. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. Rdp4: detection and analysis of recombination patterns in virus genomes. Virus Evol. 2015;1. doi: 10.1093/ve/vev003.

193. Price MN, Dehal PS, Arkin AP. Fasttree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 2009;26:1641-50. doi: 10.1093/molbev/msp077.

194. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. Modeltest-ng: a new and scalable tool for the selection of DNA and protein evolutionary models. Mol Biol Evol. 2019; 37(1):291-294. doi: 10.1093/molbev/msz189.

195. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59:307-21. doi: 10.1093/sysbio/syq010.

196. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019;35:4453-5. doi: 10.1093/bioinformatics/btz305.

197. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268-74. doi: 10.1093/molbev/msu300.

198. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530-4. doi: 10.1093/molbev/msaa015.

199. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 2018;4:vey016. doi: 10.1093/ve/vey016.

200. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A et al. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2019;15:e1006650. doi: 10.1371/journal.pcbi.1006650.

201. To T-H, Jung M, Lycett S, Gascuel O. Fast dating using least-squares criteria and algorithms. Syst Biol. 2016;65:82-97. doi: 10.1093/sysbio/syv068.

202. Kong S, Sánchez-Pacheco SJ, Murphy RW. On the use of median-joining networks in evolutionary biology. Cladistics. 2016;32:691-9. doi: 10.1111/cla.12147.

203. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. Mega x: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35:1547-9. doi: 10.1093/molbev/msy096.

204. Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microb Genom. 2016;2. doi: 10.1099/mgen.0.000093.

205. Nadeau SA, Vaughan TG, Sciré J, Huisman JS, Stadler T. The origin and early spread of SARS-CoV-2 in Europe. 2020 (http://medrxiv.org/lookup/doi/10.1101/2020.06.10.20127738, acessado em 17 de julho de 2020).

Anexo 1. Exemplos de estudos de sequenciamento para epidemiologia molecular

Estudo (citado na Seção 2.2, Quadro 1)	Tipo de análise	Nº de sequências	Características das amostras
Vírus da gripe A(H1N1)pdm09			
Fraser et al. *(1)*	Análise filogenética medida pelo tempo, taxa evolutiva e estimativa de R_0	11	30 de março de 2009 a 25 de abril de 2009
		23	30 de março de 2009 a 29 de abril de 2009
Mena et al. *(2)*	Análise filogenética medida pelo tempo	58 x 8 segmentos	2010 a 2014
	Análise filogeográfica	422	1º de março de 2009 a 31 de maio de 2009 Suínos amostrados em 20 países, e humanos amostrados globalmente
Rambaut & Holmes *(3)*	Análise filogenética medida pelo tempo, taxa evolutiva e estimativa da taxa de crescimento	242	23 países
Smith et al. *(4)*	Análise filogenética medida pelo tempo e estimativa da taxa evolutiva	168	30 de março de 2009 a 2 de maio de 2009
Coronavírus MERS-CoV			
Azhar et al. *(5)*	Análise filogenética	27 (gene da espícula) 34 (genoma completo)	2012 a 2013
Dudas et al. *(6)*	Análise filogenética medida pelo tempo e análise coalescente estruturada pelo hospedeiro	274	5 de fevereiro de 2013 a 17 de agosto de 2015
Haagmans et al. *(7)*	Análise filogenética	20	NA
Memish et al. *(8)*	Análise filogenética medida pelo tempo	69	2012 a 2013
Sabir et al. *(9)*	Análise filogenética e análise filogenética medida pelo tempo	173	de maio de 2014 a abril de 2015
Vírus Ebola			
Arias et al. *(10)*	Análise filogenética e análise filogenética medida pelo tempo	1573 1058	2014 a 2015
Baize S et al. *(11)*	Análise filogenética medida pelo tempo	51	1976 a 2014 República Democrática do Congo, Gabão e Guiné
Carroll et al. *(12)*	Análise filogenética medida pelo tempo e análise filogenética	179 262	27 de março de 2014 a 31 janeiro de 2015 1976 a 2015
Dudas & Rambaut *(13)*	Análise filogenética medida pelo tempo	49	1976 a 2014

Gire et al. *(14)*	Análise filogenética medida pelo tempo	81	17 de março de 2014 a 18 de junho de 2014
		123	1976 a 2014
Hoenen et al. *(15)*	Time-measured phylogenetic and taxa evolutiva analysis	296	de novembro de 2014 a janeiro de 2015
Ladner et al. *(16)*	Análise filogenética medida pelo tempo	922	março de 2014 a fevereiro de 2015
Park et al. *(17)*	Análise filogenética medida pelo tempo	318	17 de março de 2014 a 12 de março de 2015
Quick et al. *(18)*	Análise filogenética medida pelo tempo e estimativa da taxa evolutiva	728	17 de março de 2014 a 24 de outubro de 2015
Simon-Loriere et al. *(19)*	Análise filogenética medida pelo tempo	195	janeiro de 2014 a outubro de 2015
Stadler et al. *(20)*	Análise filogenética medida pelo tempo e análise filodinâmica	72	de maio a junho de 2014
Tong et al. *(21)*	Análise filogenética medida pelo tempo	256	17 de março de 2014 a 11 de novembro de 2014
Volz et al. *(22)*	Análise filogenética medida pelo tempo e análise filodinâmica	78	de maio de 2014 a junho de 2015
Vírus Zika			
Faria et al. *(23)*	Análise filogenética medida pelo tempo	23	19 de fevereiro de 2013 a 15 de dezembro de 2015
Faria et al. *(24)*	Análise filogenética medida pelo tempo e análise filogeográfica	254 328	23 de fevereiro de 2015 a 12 de outubro de 2015 Brasil e Américas
Grubaugh et al. *(25)*	Análise filogenética medida pelo tempo e estimativa da taxa evolutiva	104	28 de novembro de 2013 a 27 de abril de 2016
Metsky et al. *(26)*	Análise filogenética medida pelo tempo	174	12 de dezembro de 2014 a 12 de outubro de 2016

NA, não aplicável

Referências

1. Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, Hollingsworth TD et al. Pandemic potential of a strain of influenza A(H1N1): early findings. Science. 2009;324:1557-61. doi: 10.1126/science.1176062.
2. Mena I, Nelson MI, Quezada-Monroy F, Dutta J, Cortes-Fernández R, Lara-Puente JH et al. Origins of the 2009 H1N1 influenza pandemic in swine in Mexico. eLife. 2016;5:e16777. doi: 10.7554/eLife.16777.
3. Rambaut A, Holmes E. The early molecular epidemiology of the swine-origin A/H1N1 human influenza pandemic. PLoS Curr. 2009;1:RRN1003. doi: 10.1371/currents.rrn1003.
4. Smith GJD, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG et al. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature. 2009;459:1122-5. doi: 10.1038/nature08182.

5. Azhar EI, El-Kafrawy SA, Farraj SA, Hassan AM, Al-Saeed MS, Hashem AM et al. Evidence for camel-to-human transmission of MERS coronavirus. N Eng J Med. 2014;370:2499-505. doi: 10.1056/NEJMoa1401505.

6. Dudas G, Carvalho LM, Rambaut A, Bedford T. MERS-CoV spillover at the camel- human interface. eLife. 2018;7:e31257. doi: 10.7554/eLife.31257.

7. Haagmans BL, Al Dhahiry SHS, Reusken CBEM, Raj VS, Galiano M, Myers R et al. Middle East respiratory syndrome coronavirus in dromedary camels: an outbreak investigation. Lancet Infect Dis. 2014;14:140-5. doi: 10.1016/S1473-3099(13)70690-X.

8. Memish ZA, Cotten M, Meyer B, Watson SJ, Alsahafi AJ, Al Rabeeah AA et al. Human infection with MERS coronavirus after exposure to infected camels, Saudi Arabia, 2013. Emerg Infect Dis. 2014;20:1012-5. doi: 10.3201/eid2006.140402

9. Sabir JSM, Lam TTY, Ahmed MMM, Li L, Shen Y, Abo-Aba SEM et al. Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia. Science. 2016;351:81-4. doi: 10.1126/science.aac8608.

10. Arias A, Watson SJ, Asogun D, Tobin EA, Lu J, Phan MVT et al. Rapid outbreak sequencing of Ebola virus in Sierra Leone identifies transmission chains linked to sporadic cases. Virus Evol. 2016;2:vew016. doi: 10.1093/ve/vew016.

11. Baize S, Pannetier D, Oestereich L, Rieger T, Koivogui L, Magassouba NF et al. Emergence of Zaire Ebola virus disease in Guinea. N Eng J Med. 2014;371:1418-25. doi: 10.1056/NEJMoa1404505.

12. Carroll MW, Matthews DA, Hiscox JA, Elmore MJ, Pollakis G, Rambaut A et al. Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa. Nature. 2015;524:97-101. doi: 10.1038/nature14594.

13. Dudas G, Rambaut A. Phylogenetic analysis of Guinea 2014 ebov ebolavirus outbreak. PLoS Curr. 2014;6. doi: 10.1371/currents.outbreaks.84eefe5ce43ec9dc0bf0670f7b8b417d.

14. Gire SK, Goba A, Andersen KG, Sealfon RSG, Park DJ, Kanneh L et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science. 2014;345:1369-72. doi: 10.1126/science.1259657.

15. Hoenen T, Groseth A, Rosenke K, Fischer RJ, Hoenen A, Judson SD et al. Nanopore sequencing as a rapidly deployable Ebola outbreak tool. Emerg Infect Dis . 2016;22:331- 4. doi: 10.3201/eid2202.151796.

16. Ladner JT, Wiley MR, Mate S, Dudas G, Prieto K, Lovett S et al. Evolution and spread of Ebola virus in Liberia, 2014-2015. Cell Host Microbe. 2015;18:659-69. doi: 10.1016/j.chom.2015.11.008.

17. Park DJ, Dudas G, Wohl S, Goba A, Whitmer SLM, Andersen KG et al. Ebola virus epidemiology, transmission, and evolution during seven months in Sierra Leone. Cell. 2015;161:1516-26. doi: 10.1016/j.cell.2015.06.007.

18. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L et al. Real-time, portable genome sequencing for Ebola surveillance. Nature. 2016;530:228-32. doi: 10.1038/nature16996.

19. Simon-Loriere E, Faye O, Faye O, Koivogui L, Magassouba N, Keita S et al. Distinct lineages of Ebola virus in Guinea during the 2014 West African epidemic. Nature. 2015;524:102-4. doi: 10.1038/nature14612.

20. Stadler T, Kühnert D, Rasmussen DA, Plessis DL. Insights into the early epidemic spread of Ebola in Sierra Leone provided by viral sequence data. PLoS Curr. 2014. doi: 10.1371/currents.outbreaks.02bc6d927ecee7bbd33532ec8ba6a25f.

21. Tong Y-G, Shi W-F, Liu D, Qian J, Liang L, Bo X-C et al. Genetic diversity and evolutionary dynamics of Ebola virus in Sierra Leone. Nature. 2015;524:93-6. doi: 10.1038/nature14490.

22. Volz E, Pond S. Phylodynamic analysis of Ebola virus in the 2014 Sierra Leone epidemic. PLoS Curr. 2014;24:ecurrents.outbreaks.6f7025f1271821d4c815385b08f5f80e.

23. Faria NR, Quick J, Claro IM, Thézé J, de Jesus JG, Giovanetti M et al. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. Nature. 2017;546:406-10. doi: 10.1038/nature22401.

24. Faria NR, Azevedo RdSdS, Kraemer MUG, Souza R, Cunha MS, Hill SC et al. Zika virus in the Americas: early epidemiological and genetic findings. Science. 2016;352:345-9. doi: 10.1126/science.aaf5036.

25. Grubaugh ND, Ladner JT, Kraemer MUG, Dudas G, Tan AL, Gangavarapu K et al. Genomic epidemiology reveals multiple introductions of Zika virus into the United States. Nature. 2017;546:401-5. doi: 10.1038/nature22400.

26. Metsky HC, Matranga CB, Wohl S, Schaffner SF, Freije CA, Winnicki SM et al. Zika virus evolution and spread in the Americas. Nature. 2017;546:411-5. doi: 10.1038/nature22402.

Anexo 2. Lista de verificação para o estabelecimento de um programa de sequenciamento

Meta

- [ ] Definir os objetivos esperados do programa de sequenciamento; quais informações o sequenciamento provavelmente fornecerá que serão adicionais ou mais custo-efetivas do que as abordagens existentes?

Identificação e envolvimento das partes interessadas

- [ ] Identifique as principais partes interessadas.
- [ ] Discuta os objetivos do programa com representantes seniores de grupos de partes interessadas e defina as responsabilidades de cada grupo.
- [ ] Considere a possibilidade de compartilhar materiais educacionais sobre o potencial e os requisitos do sequenciamento do SARS-CoV-2 com as partes interessadas.
- [ ] Identifique os vínculos necessários entre as principais partes interessadas para permitir o deslocamento rápido de amostras, a solicitação de informações e o uso dos resultados.
- [ ] Certifique-se de que sejam estabelecidos vínculos claros e apropriados entre as partes interessadas.

Considerações técnicas

- [ ] Determine o nível de amostragem genômica necessária para atingir os objetivos desejados, em discussão com membros seniores de identificação de casos e equipes analíticas.
- [ ] Identifique os metadados necessários para atingir os objetivos desejados, em discussão com membros seniores das equipes analíticas e de identificação de casos.
- [ ] Escolha os protocolos apropriados de preparação de amostras e biblioteca.
- [ ] Escolha protocolos bioinformáticos apropriados.
- [ ] Escolha protocolos analíticos apropriados.

Considerações logísticas

- [ ] Pondere onde o sequenciamento e a análise serão realizados (por exemplo um laboratório de diagnóstico existente ou um laboratório comercial ou acadêmico externo).
- [ ] Identifique fontes apropriadas de financiamento que serão adequadas para apoiar o sequenciamento laboratorial, o armazenamento de dados e a análise de dados.
- [ ] Certifique-se de que reagentes e recursos computacionais suficientes estejam disponíveis e possam ser obtidos de forma sustentável conforme necessário.
- [ ] Certifique-se de que haja recursos humanos suficientes e apropriados para concluir o programa em todas as fases.
- [ ] Certifique-se de que a integridade da amostra possa ser mantida em todas as etapas ao longo do pipeline por meio da cadeia de frio ou outras medidas.

- [ ] Assegure a coleta e armazenamento adequados de metadados e a associação correta com amostras biológicas.
- [ ] Considere a possível pressão adicional que o sequenciamento exercerá sobre os grupos existentes de resposta de saúde pública e procure maneiras de aliviar essa pressão.
- [ ] Para programas de sequenciamento em grande escala, identifique como agilizar o compartilhamento de dados e amostras entre os grupos participantes (por exemplo, a viabilidade de usar uma identificação de amostra única e formatos de metadados idênticos).

Garantir um ambiente seguro e ético

- [ ] Conduza análises éticas adequadas para a geração, uso e armazenamento de dados de sequência e metadados associados.
- [ ] Realize avaliações de risco das atividades de sequenciamento para garantir a biossegurança apropriada em todos os estágios.
- [ ] Realize avaliações de risco das atividades de sequenciamento para garantir biossegurança apropriada, se relevante de acordo com a legislação nacional e regional.
- [ ] Pondere o impacto sobre os recursos humanos, incluindo a realocação de pessoal ou a contratação de pessoal adicional para manter a carga de trabalho individual em níveis razoáveis.
- [ ] Certifique-se de que os profissionais possam se deslocar para o trabalho e estar no local de trabalho com segurança e de acordo com as diretrizes nacionais de prevenção da transmissão durante o surto da COVID-19.
- [ ] Defina estratégias para manter o programa de sequenciamento se os principais membros da equipe ficarem doentes ou precisarem se isolar.

Compartilhamento de dados

- [ ] Certifique-se de que todas as partes interessadas estejam de acordo sobre quais sequências e metadados serão compartilhados publicamente, por meio de quais plataformas e quando.
- [ ] Certifique-se de que todas as partes interessadas estejam de acordo quanto à possibilidade de qualquer metadado ser restrito a um número limitado de usuários locais e elabore estratégias para compartilhar esses dados com segurança.
- [ ] Garanta que o compartilhamento de dados esteja em conformidade com os esquemas regulatórios nacionais e internacionais.

Avaliação

- [ ] Garanta oportunidades regulares para avaliar o programa de sequenciamento, incluindo sucessos e dificuldades contínuas.
- [ ] Garanta que seja implementado um esquema de monitoramento e avaliação para medir o desempenho do programa de sequenciamento e o sucesso no cumprimento de seus objetivos

OPAS

Organização Pan-Americana da Saúde

Organização Mundial da Saúde

ESCRITÓRIO REGIONAL PARA AS Américas

9 789275 723890 >