

ARTIGO DE PESQUISA

Um método computacional para identificação de espécies e genótipos dos vírus Dengue, Zika e Chikungunya

Vagner Faria^{1,6}, T. Nunes⁷, Maria I. Restovic⁸, Bruna Maria Azevedo⁹, Maria Clara Abete¹⁰, Silke Koppers¹¹, Pieter JK Libin^{4,5}, Kristof Theys⁵, Nuno R. Faria^{1,6}, Márcio R. Faria^{1,6}, Isadora C. de Siqueira⁸, Emmanuel J. San2 de Filippis¹, Gilberto A. Santiago¹¹, Luiz CJ, Kaliane CB Machado⁸, Vasco Azevedo³, Ana Maria Bispo-Rivaldo Venâncio da Cunha¹², Oliver G. Pybus⁶, Anne-Mieke Vandamme^{5,9}, Alcântara^{1,3}*, Túlio de Oliveira²*



Citação: Fonseca V, Libin PJK, Theys K, Faria NR, Nunes MRT, Restovic MI, et al. (2019) Um método computacional para identificação de espécies e genótipos dos vírus Dengue, Zika e Chikungunya. *PLoS Negl Trop Dis* 13(5): e0007231. <https://doi.org/10.1371/journal.pntd.0007231>

Editora: Isabel Rodriguez-Barraquer, Universidade de Califórnia São Francisco, ESTADOS UNIDOS

Recebido: 14 de agosto de 2018

Aceito: 11 de fevereiro de 2019

Publicado: 8 de maio de 2019

Direitos autorais: © 2019 Fonseca et al. Este é um artigo de acesso aberto distribuído sob os termos da [Licença Creative Commons Attribution](#), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor e a fonte originais sejam creditados.

Declaração de disponibilidade de dados: Todas as informações relevantes dos dados estão no documento e os detalhes foram incluídos como arquivos de informações de apoio.

Financiamento: Apoiado por uma bolsa emblemática de pesquisa do Conselho Sul-Africano de Pesquisa Médica (MRC-RFA-UFSP-01-2013/UKZN HIVEPI), um Real Society Newton Advanced Fellowship (TdO), a O projecto VIROGENESIS recebe financiamento do Programa de Investigação e Desenvolvimento do Horizonte 2020 da União Europeia Programa de Inovação (sob Acordo de Subvenção

1 Laboratório de Flavivírus, IOC, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil, **2** KwaZulu-Natal Plataforma de Inovação e Sequenciamento de Pesquisa (KRISP), Faculdade de Ciências da Saúde, Universidade de KwaZulu-Natal, Durban, África do Sul, **3** Laboratório de Genética Celular e Molecular, ICB, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil, **4** Laboratório de Inteligência Artificial, Departamento de Ciência da Computação, Vrije Universiteit Brussel, Bruxelas, Bélgica, **5** KU Leuven — Universidade de Leuven, Departamento de Microbiologia e Imunologia, Instituto Rega de Pesquisa Médica, Clínica e Virologia Epidemiológica, Leuven, Bélgica, **6** Departamento de Zoologia, Universidade de Oxford, Oxford, Estados Unidos Reino, **7** Instituto Evandro Chagas, Ministério da Saúde, Ananindeua, Brasil, **8** Laboratório de Patologia Experimental, Fundação Oswaldo Cruz, Salvador, Brasil, **9** Centro de Saúde Global e Medicina Tropical, Unidade de Microbiologia, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Lisboa, Portugal, **10** EMWEB (empresa privada), Herent, Bélgica, **11** Divisão de Doenças Transmitidas por Vetores, Centros de Controle e Prevenção de Doenças, San Juan, Porto Rico, Estados Unidos da América, **12** Coordenação de Vigilância em Saúde e Laboratórios de Referências, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil

† Esses autores contribuíram igualmente para este trabalho. *

alcantara13@gmail.com (LCJA); tuliodna@gmail.com (TDO)

Abstrato

Nos últimos anos, um número crescente de surtos de vírus Dengue, Chikungunya e Zika foi relatado na Ásia e nas Américas. A monitorização da diversidade genotípica do vírus é crucial para compreender o surgimento e a propagação de surtos, aspectos vitais para o desenvolvimento de estratégias eficazes de prevenção e tratamento. Assim, desenvolvemos um método eficiente para classificar sequências de vírus em relação às suas espécies e subespécies (isto é, sorotipo e/ou genótipo). Esta ferramenta fornece uma implementação de software fácil de usar deste novo método e foi validada em um grande conjunto de dados que avalia o desempenho da classificação em relação a sequências do genoma completo e sequências do genoma parcial. Disponível on-line: <http://krisp.org.za/tools.php>.

Resumo do autor

Dengue (DENV), Chikungunya (CHIKV) e Zika (ZIKV) são considerados grandes desafios de saúde pública. Além da epidemia causada pelo DENV, descrita em muitos países tropicais, a introdução do CHIKV e do ZIKV nesses países é uma grande preocupação de saúde pública. Esses arbovírus são transmitidos principalmente por mosquitos da espécie *Ae. Aegypti* e suas doenças relacionadas resultam em aumento dos custos financeiros associados

não. 634650) e o Fundo Comum dos Institutos Nacionais de Saúde, número de concessão U24HG006941. Pieter Libin foi apoiado por uma bolsa de doutoramento da FWO (Fonds Wetenschappelijk Onderzoek - Vlaanderen). Este trabalho contou com apoio do Decit/SCTIE/MS e CNPq (440685/2016-8 e 440856/2016-7); pela CAPES (88887.130716/2016-00, 88881.130825/2016-00 e 88887.130823/2016-00); e pelo Programa Horizonte 2020 da UE através da ZIKAlliance (PRES-005-FEX-17-4-2-33). Ana Abecasis foi apoiada pela Fundação para a Ciência e Tecnologia (FCT) através de fundos para GHTM-UID/Multi/04413/2013. Os financiadores não tinham papel no desenho do estudo, coleta e análise de dados, decisão de publicação ou preparação do manuscrito.

Conflitos de interesses: Dr. Koen Deforche é um dos proprietários da empresa comercial EMWEB.

diagnóstico e tratamento. Para apoiar a concepção de estratégias eficientes de diagnóstico, prevenção e tratamento, foi desenvolvida uma ferramenta de bioinformática para a genotipagem destes vírus com base em modelos evolutivos apropriados de forma automática, precisa e rápida. Um conjunto de sequências de referência de vírus foi obtido do GenBank e utilizado para o desenvolvimento das ferramentas. Este processo envolveu o alinhamento das sequências de referência seguido de reconstruções de árvores filogenéticas. Para atribuir os genótipos carregados pelo usuário, a ferramenta analisa as sequências uma a uma, genótipos por meio de identificação, alinhamento e reconstrução filogenética. Este método computacional permite a classificação de alto rendimento dessas espécies e genótipos de vírus em segundos. Conforme mostrado experimentalmente, os genótipos são classificados com mais segurança usando o gene do envelope ou sequências completas do genoma.

Introdução

Nos últimos anos, um número crescente de surtos de vírus Dengue (DENV), Chikungunya (CHIKV) e Zika (ZIKV) foi relatado na Ásia e nas Américas [1–3]. As espécies predominantes de mosquitos que transmitem DENV, CHIKV e ZIKV são *Aedes aegypti* e *Aedes Albopictus*, que estão amplamente distribuídos em regiões tropicais e subtropicais [4]. Nos últimos anos, vários estudos relataram surtos simultâneos de DENV, CHIKV e ZIKV na mesma área geográfica [5, 6]. Atualmente, surtos sem precedentes de DENV, CHIKV e ZIKV estão ocorrendo simultaneamente no Brasil. Em 2017, o Ministério da Saúde brasileiro estimou que aproximadamente 251.000 casos suspeitos de DENV, 185.000 casos suspeitos de CHIKV e cerca de 18.000 casos suspeitos de ZIKV ocorreram no Brasil [7].

A monitorização da diversidade genotípica do vírus é crucial para compreender o surgimento e a propagação de surtos, aspectos vitais para o desenvolvimento de estratégias eficazes de prevenção e tratamento. Tanto as epidemias de DENV como de CHIKV estão associadas a uma mortalidade e morbidade que impõe um fardo económico significativo às regiões afetadas [8,9]. Embora as infecções pelo ZIKV raramente sejam fatais, como afirmado anteriormente, as infecções pelo ZIKV podem resultar na síndrome de Guillain-Barré e malformações congênitas [10,11]. A vigilância genômica de epidemias na resolução apropriada e a classificação consistente das sequências genéticas relacionadas também permitem a identificação de cepas associadas a maior potencial epidêmico [12] ou gravidade da doença [13].

No entanto, actualmente faltam métodos que classifiquem consistentemente as sequências de arbovírus ao nível de espécies e subespécies (ou seja, serotipo e/ou genótipo). Além disso, as sequências completas do genoma muitas vezes não estão disponíveis em ambientes clínicos de rotina, forçando o uso de sequências genéticas mais curtas para classificar em nível de espécie ou subespécie viral. No entanto, não foi suficientemente explorado quais regiões genômicas são mais adequadas para uma classificação precisa.

É apresentado um novo método computacional para a identificação de sequências de DENV/CHIKV/ZIKV, com relação a espécies e subespécies (ou seja, serotipo e/ou genótipo). O método de classificação é implementado na ferramenta de software Genome Detective, que foi validada em um grande conjunto de dados avaliando o desempenho de classificação de sequências do genoma completo, sequências do genoma parcial e produtos de métodos de sequenciamento de próxima geração. Além disso, foi avaliada a adequação de diferentes regiões genômicas para classificação de vírus.

Materiais e métodos

Conjuntos de dados

Conjunto de dados global de sequência do genoma completo (Global-WG). Um conjunto de dados de sequências do genoma completo publicadas anteriormente no GenBank [14] foi compilado. Este conjunto de dados consiste em 4.118 sequências de DENV, 653 sequências de CHIKV e 413 sequências de ZIKV e contém DENV

sequências para cada um dos quatro sorotipos conhecidos: DENV-sero1 (n = 1688), DENV-sero2 (n = 1317), DENV-sero3 (n = 897) e DENV-sero4 (n = 216). A lista de números de acesso do GenBank para este conjunto de dados global do genoma completo está disponível na seção Informações de Apoio

([Arquivo S1](#)). No restante deste manuscrito, este conjunto de dados será referido como Global-WG.

Conjunto de dados de sequência de envelope global (Global-ENV). Um conjunto de dados publicado anteriormente sequências de envelope do GenBank [14] foram compiladas. Este conjunto de dados consiste em 4.118 sequências de DENV, 2.531 sequências de CHIKV e 413 sequências de ZIKV e contém DENV sequências para cada um dos quatro sorotipos conhecidos: DENV-sero1 (n = 1688), DENV-sero2 (n = 1317), DENV-sero3 (n = 897) e DENV-sero4 (n = 216). A lista de adesão ao GenBank números para este conjunto de dados de envelope global estão disponíveis na seção Informações de Apoio ([S2 Arquivo](#)). No restante deste manuscrito, este conjunto de dados será referido como **Global-ENV**.

Identificação de genótipos e seleção de sequências de referência. Para identificar o vírus genótipos, um alinhamento de múltiplas sequências foi construído com o software de alinhamento MAFFT [15] por espécie de vírus, usando o conjunto de dados **Global-WG**. Cada alinhamento foi editado manualmente até que um alinhamento correto de códons fosse alcançado em todos os genes. O próximo passo nesta exploração envolveu uma análise filogenética usando PhyML (ou seja, máxima verossimilhança, 1000 réplicas de bootstrap) e MrBayes (ou seja, Bayesiano) [16,17]. Com esta abordagem, quatro clados principais de DENV (ou seja, sorotipo 1 a 4) e 19 genótipos (ou seja, 1I, 1II, 1III, 1IV, 1V, 2I, 2II, 2III, 2IV, 2V, 2VI, 3I, 3II, 3III, 3V, 4I, 4II, 4III e 4IV) foram identificados. Essas descobertas estão de acordo com o atual consenso na classificação DENV [18–21]. Para o CHIKV, três clados filogenéticos podem ser distinguidos: o genótipo Leste-Centro-Sul-africano (ECSA), o genótipo Asiático-Caribenho e o genótipo da África Ocidental. Sendo o genótipo da África Ocidental mais divergente e menos difundido do que o genótipo ECSA e o genótipo Asiático-Caribenho [22,23]. ZIKV, como bem, podem ser classificados em dois genótipos. O genótipo africano, originalmente identificado em Uganda em 1947 [24], é encontrada em muitos países africanos [25]. O genótipo asiático foi identificado na Malásia em 1966 [26], este genótipo causou recentemente a epidemia mundial em Ásia e Pacífico [27,28], e é responsável pela epidemia nas Américas [5].

A precisão e consistência com que um método identifica espécies e genótipos virais clados depende da seleção de um conjunto de sequências de referência representativas [29–31].

O passo inicial na seleção de cepas de referência para o nosso método envolveu a identificação de sequências de genoma completo altamente divergentes, mas equidistantes, que são representativas para a diversidade dentro dos diferentes genótipos de DENV, CHIKV e ZIKV, através da triagem de todas as sequências completas do genoma publicadas em nosso conjunto de dados **Global-WG**. Por exemplo, normalmente começamos selecionando 5 a 10 sequências que representam a diversidade de cada genótipo de vírus. Sequências que atenderam a esses critérios de seleção foram controlados de qualidade quanto à presença de inserções, exclusões, mudanças de quadro e caracteres não IUPAC usando VIRULIGN [32]. Para DENV, usamos o sequências de referência incluídas no software VIRULIGN, para ZIKV, usamos a sequência de referência apresentada em [33], e para CHIKV construímos uma nova sequência de referência de NC_004162 que adicionamos ao repositório VIRULIGN. Sequências que passam na qualidade controle foram alinhados usando MAFFT [15] e foram submetidos à análise filogenética usando PAUPy (ou seja, adesão de vizinho), MrBayes (ou seja, bayesiano) e PhyML (ou seja, probabilidade máxima) [16,17,34,35] utilizando GTR+G+I. Sequências que forneceram topologias consistentes usando todas as três árvores os métodos de inferência foram mantidos como possíveis sequências de referência (ver [Informações de Apoio, Tabela S1](#)) e usados na próxima etapa do processo de avaliação.

Estabelecemos que nenhuma das cepas de referência selecionadas eram recombinantes ([S2 Fig](#)) usando o programa de detecção de recombinação RDP4 [36]

Adequação de regiões subgenômicas para fins de genotipagem. O conjunto de dados de cepas de referência ([Tabela S1](#)) foi então explorado para estabelecer a adequação de regiões subgenômicas para genotipagem automatizada. Dois métodos diferentes foram usados.

O primeiro foi um método de verificação de inicialização, usando uma abordagem de janela deslizante explorando o variam entre 200 e 2.000 nucleotídeos. Todas as janelas do genoma foram usadas para a construção de árvores vizinhas unindo árvores com 1.000 réplicas de bootstrap. O objetivo era encontrar o tamanho e segmentos do genoma que classificariam corretamente uma sequência de consulta com suporte de boot-strap de > 70%.

O segundo método envolveu o cálculo do sinal filogenético presente em cada um dos os genes DENV, CHIKV e ZIKV, utilizando o mesmo conjunto de sequências de referência. Para calcular Para o sinal filogenético, foi utilizada a implementação do software TreePuzzle [37] do método de mapeamento de verossimilhança [38]. Apenas quartetos entre genótipos foram avaliados. Quarteto intrigante essencialmente é um procedimento de três etapas, primeiro reconstruindo todas as árvores de máxima verossimilhança possíveis do quarteto (etapa de máxima verossimilhança) e, em seguida, combinando repetidamente as árvores do quarteto para um árvore geral (etapa intrigante) e, finalmente, calcular o consenso da regra majoritária de todas as árvores intermediárias, resultando na árvore intrigante do quarteto (etapa de consenso).

Método de classificação e implementação

Método de classificação. Nosso método envolve um pipeline de classificação viral, inspirando-se naquele descrito anteriormente para classificar sequências de HIV, vírus da hepatite C e vírus linfotrópico T humano [29,30]. O pipeline de classificação apresentado aqui consiste em dois componentes de classificação. O primeiro componente de classificação permite espécies e subespécies atribuições. A análise de classificação submete uma sequência de consulta a uma análise BLAST contra um conjunto de sequências de referência [39]. Uma consulta é atribuída a um tipo específico quando o BLAST reporta uma tarefa com uma pontuação que excede um limite predefinido.

O segundo componente de classificação envolve a construção de uma árvore filogenética Neighbor Joining. Este componente permite atribuições em nível de genótipo e/ou subtipo. Primeiro, a sequência de consulta está alinhada com um conjunto de sequências de referência.

O alinhamento é produzido utilizando a opção de alinhamento de perfil no software ClustalW [40], de modo que a sequência de consulta seja adicionada ao alinhamento existente de sequências de referência. Após o alinhamento, uma árvore filogenética Neighbor Joining, com 100 réplicas de bootstrap, é construída. A árvore é construída usando a métrica de distância HKY com gama variação da taxa entre locais, conforme implementado no software PAUP® [34]. A sequência de consulta é atribuído a um genótipo específico se ele se agrupar monofileticamente com aquele clado genotípico com suporte de inicialização > 70%. Se o suporte de bootstrap for <70%, o genótipo é relatado como não atribuído.

Implementação de software. Embora o método de classificação tenha sido inspirado naquele anterior apresentado [29], uma nova estrutura de software foi desenvolvida para ser facilmente adaptável aos procedimentos de classificação para vários patógenos virais. Todo o código-fonte é escrito na linguagem de programação Java (Fig. 1). A estrutura do software faz parte da cadeia de ferramentas do Genome Detective [41]

Método de classificação e implementação do ArboTyping. Em primeiro lugar, a espécie viral é determinado usando BLAST, classificando a sequência como DENV, CHIKV ou ZIKV.

Caso a sequência enviada tenha sido atribuída como ZIKV ou CHIKV, um vizinho que se junte árvore é inferida para determinar o respectivo genótipo ZIKV ou CHIKV. Somente para DENV, outro procedimento BLAST é invocado para atribuir primeiro o sorotipo. Com base no sorotipo inferido, uma árvore de junção de vizinhos específica do sorotipo é construída para determinar a Dengue. genótipo.

Para cada uma dessas etapas, foram utilizadas as cepas de referência discutidas anteriormente, no que diz respeito ao nível de tipagem apropriado (ou seja, espécie de vírus, sorótipo ou genótipo). Este processo é resumido em uma árvore de decisão na Fig 2.

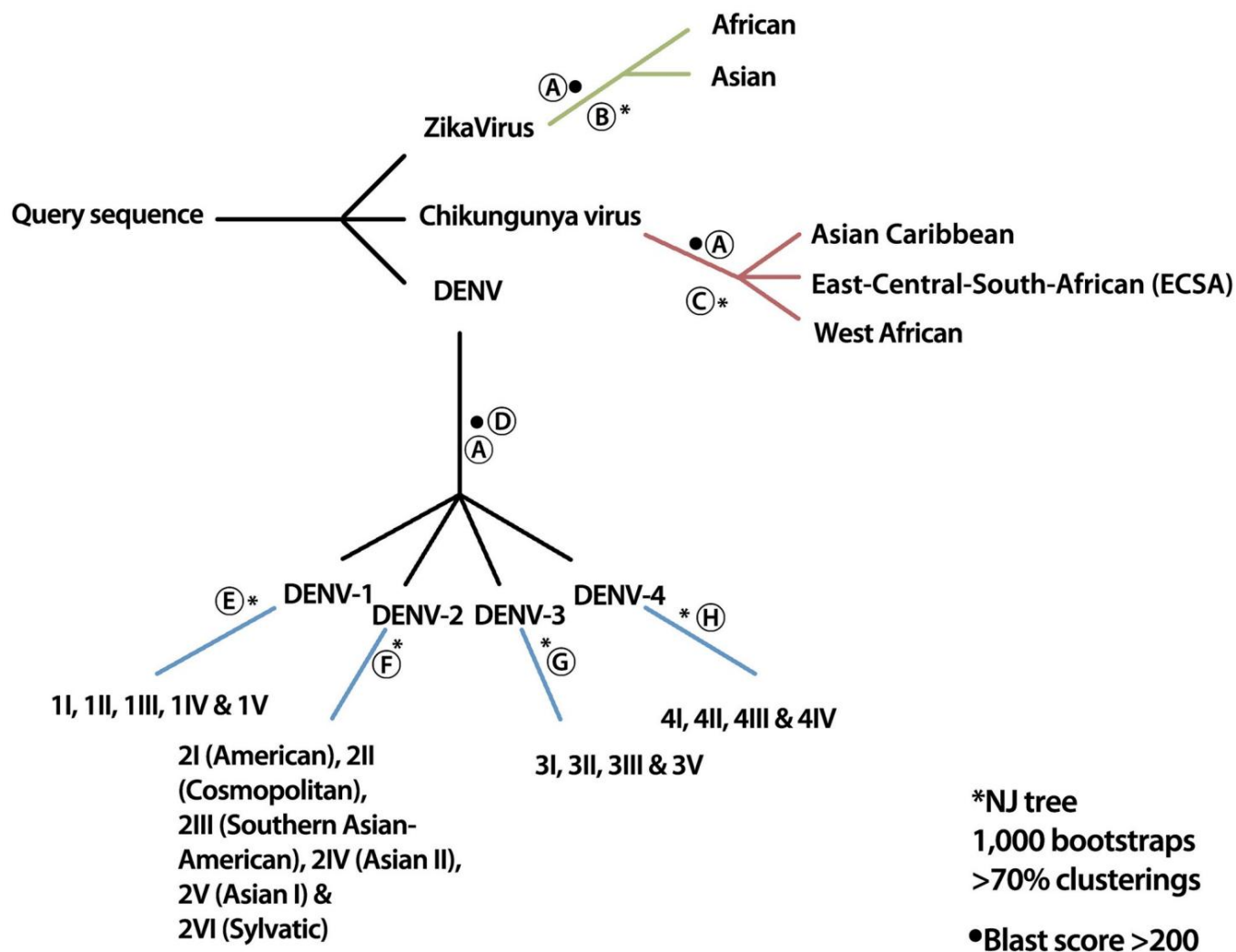


Fig 2. Esboço do procedimento de classificação. Primeiramente (A), a espécie viral é determinada usando BLAST. Quando a sequência submetida é um vírus Zika, uma árvore de junção de vizinhos é construída para determinar o genótipo do Zika (B). Quando a sequência submetida é um vírus Chikungunya, uma árvore de junção de vizinhos é construída para determinar o genótipo Chikungunya (C). Quando a sequência submetida é um vírus da Dengue, o sorotipo é determinado usando outra invocação do BLAST (D). Com base no sorotipo inferido, uma árvore de união de vizinhos específica do sorotipo é construída para determinar o genótipo da Dengue (E, F, G, H).

<https://doi.org/10.1371/journal.pntd.0007231.g002>

Desempenho de classificação para genomas completos e regiões subgenômicas. Para determinar a precisão do método automatizado para sequências do genoma completo, o método foi avaliado em um conjunto de dados de sequência do genoma completo (ou seja, conjunto de dados **Global-WG**).

Como as sequências de regiões subgenômicas estão mais comumente disponíveis do que as do genoma completo, a precisão do método também foi avaliada neste contexto. Para este propósito, as sequências de envelope no conjunto de dados **Global-ENV** foram utilizadas para avaliação.

Cada uma das sequências consideradas para avaliação foi atribuída usando tanto o padrão ouro quanto o método automatizado aqui descrito. O padrão ouro, uma classificação manual consiste em realizar uma atribuição usando tanto Bayesiano (ou seja, MrBayes, atribuição com posterior > 90% [17]) e Máxima verossimilhança (ou seja, PhyML, 1000 réplicas de bootstrap, atribuição com > 70% de réplicas [17]). [16]) análise filogenética. Quando as atribuições geradas pelas técnicas bayesianas e de máxima verossimilhança coincidem, a classificação é confirmada [31].

A sensibilidade, especificidade e precisão do nosso método foram calculadas para ambas as espécies atribuição e genotipagem. A sensibilidade foi calculada pela fórmula $\frac{TP}{TP+FN}$ e especificidade pelo $\frac{TN}{TN+FP}$. A precisão foi calculada pela fórmula $\frac{TP}{TP+FP}$ [42]. Nestas fórmulas: TP = Verdadeiros Positivos, FP = Falsos Positivos, TN = Verdadeiros Negativos e FN = Falsos Negativos.

Resultados

Método de classificação e implementação de ArboTyping

Um método eficiente para classificar sequências de vírus em relação às suas espécies e subespécies (ou seja, sorotipo e/ou genótipo) foi desenvolvido. Este método foi implementado em Java e esta implementação foi integrada em uma interface web fácil de usar. Uma descrição detalhada do método e sua implementação podem ser encontrados na subseção Métodos 'Método de classificação e implementação'.

Adequação de regiões subgenômicas para fins de genotipagem

Dois métodos diferentes foram usados para verificar a adequação de regiões subgenômicas para fins de genotipagem: um método de inicialização e um método de mapeamento de verossimilhança (ver [Métodos](#)).

Para DENV, a única região subgenômica que suporta atribuição confiável de genótipos em os quatro sorotipos diferentes foi o gene do envelope. Para o CHIKV, a região do envelope E1 foi a única região que permitiu atribuição consistente. A análise de inicialização mostrou que para o ZIKV, segmentos de cerca de 1.200–1.500 pares de bases apoiam a atribuição de genótipo com bootstrap > 70% ([Fig. 3](#)). Este foi o caso em todo o genoma, com exceção do final do genoma (ou seja, a região não codificadora) e perto da região NS3, onde os bootstraps caíram abaixo de 60%.

Nossas análises de mapeamento de verossimilhança mostram que para DENV, o envelope, NS1, NS3 e NS5 teve bom sinal filogenético em todos os quatro sorotipos. Para CHIKV, o gene do envelope E2 tinha o melhor sinal, mas esta região não forneceu um bom suporte de inicialização para a classificação do genótipo ECSA ([Fig. 3](#)). Para ZIKV, o envelope, NS1, NS2A, NS3, NS4A, NS4B e as regiões NS5 apresentaram bom sinal filogenético. Uma visão geral detalhada dos resultados da análise do mapeamento de probabilidade pode ser encontrada na Tabela S2 das Informações de Apoio.

Em resumo, estas análises mostram que os genes do envelope dos conjuntos de dados de referência de três patógenos (DENV, 1.485 nucleotídeos; CHIKV, 1.317 nucleotídeos; ZIKV, 1.525 nucleotídeos) são os alvos mais adequados para uma classificação genotípica confiável.

Desempenho de classificação para sequências do genoma completo

Nosso método automatizado proporcionou especificidade, sensibilidade e precisão de 100% para a identificação de genomas completos para todas as espécies e genótipos virais em comparação ao padrão ouro, uma classificação manual. Para uma visão geral detalhada da atribuição de DENV, CHIKV e ZIKV desempenho, nos referimos à [Tabela S3 de informações de suporte](#).

Apenas dez dos 4.118 genomas completos do DENV não puderam ser classificados em nível genotípico por análise filogenética manual ou pelo nosso método automatizado. Notavelmente, as sete sequências (AF298807, KF864667, EU179860, JQ922546, KF184975, KF289073, EF457905 de DENV--Sero1 eram discrepantes na árvore filogenética (ver [Informações de Apoio, S1 Fig](#)). Testamos tudo dez sequências para recombinação usando varredura de inicialização (consulte [Informações de suporte, S2 Fig](#)) e o programa de detecção de recombinação RDP4 [36]. Encontramos apenas a sequência AY496879 para ser um recombinante claro do genótipo 3I e 3II do DENV. As outras duas sequências (DENV-Sero2 KF744408 e DENV-Sero3 JF262783) também foram identificados como valores discrepantes divergentes.

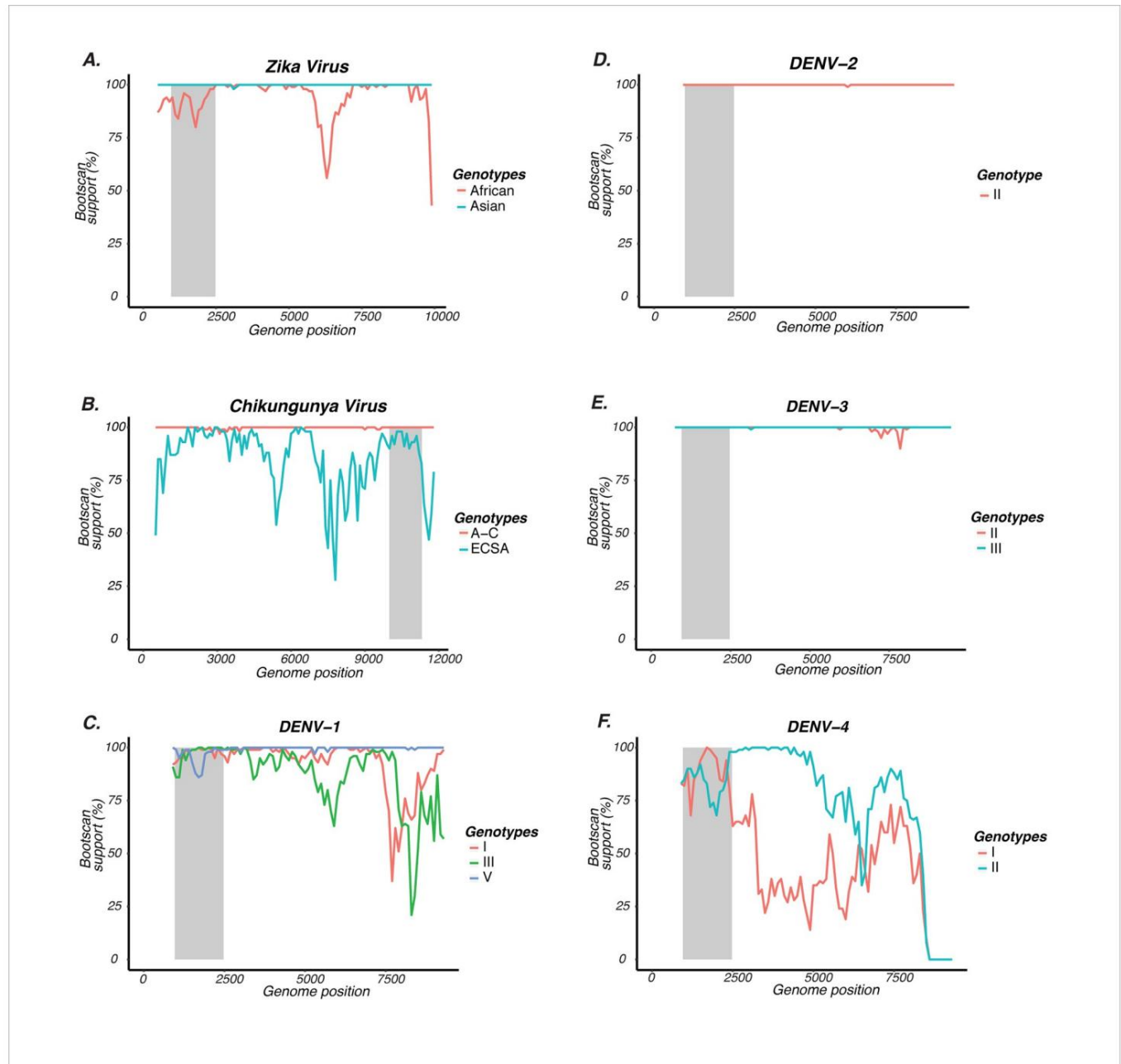


Fig 3. Investigando a adequação de regiões subgenômicas para genotipagem por meio de boot scan. Este gráfico foi construído usando resultados de bootstrap de árvores Neighbor Joining (1000 réplicas de bootstrap), realizadas no conjunto de dados das cepas de referência indicadas. O método de inicialização usa uma janela deslizante de um segmento de 1.500 pares de bases que se move em etapas de 100 pares de bases ao longo do genoma. O eixo X representa a posição do nucleotídeo no genoma e o eixo Y representa o suporte de bootstrap em percentagens. A área retangular cinza claro marca a localização do gene do envelope em cada gráfico. O código de cores inserido mostra os genótipos definidos em cada espécie de vírus. Para o ZIKV (A), são africanos e asiáticos. Para o CHIKV (B), trata-se da Ásia-Caribe (AC), da África Ocidental (WA) e da África Centro-Leste-Sul (ECSA). Para DENV (CF), o genótipo é visualizado por um algarismo romano. Apenas os genótipos que mostraram menos de 100% de suporte de bootstrap em todo o genoma são mostrados.

<https://doi.org/10.1371/journal.pntd.0007231.g003>

Desempenho de classificação para regiões subgenômicas Nossa

análise mostra que os resultados de classificação para a região subgenômica do envelope em nível de espécie e genótipo foram semelhantes aos obtidos usando sequências do genoma completo e em grande parte de acordo com o padrão-ouro, uma classificação manual.

Para DENV, a maioria dos genótipos foi classificada com grande precisão (ou seja, especificidade e sensibilidade > 99%) utilizando o gene do envelope. A exceção foi o genótipo IV do DENV-sero2, do qual estavam disponíveis 41 sequências de envelope e para as quais 33 foram corretamente identificadas (ou seja, sensibilidade 80,49%, especificidade 100%). As sequências do CHIKV cobrindo a região E1 foram classificadas com precisão para todos os três genótipos (ou seja, 100% de sensibilidade e especificidade). Todas as sequências do envelope do ZIKV foram classificadas com 100% de sensibilidade e especificidade. Para obter uma visão geral detalhada do desempenho da atribuição de DENV, CHIKV e ZIKV, consulte a [Tabela S4 de informações de suporte](#).

Como foi relatado um bom sinal filogenético para a região DENV e ZIKV NS5 e para a região Região E2 do CHIKV, uma análise de classificação também foi realizada para essas regiões. Para a região DENV NS5 foi observada sensibilidade de 57,48% e especificidade de 31,35%. Quase todas as sequências NS5 do ZIKV foram corretamente atribuídas ao genótipo africano (ou seja, sensibilidade de 97,72% e especificidade de 100%). Isto indica que a região NS5 do ZIKV também pode ser usada para classificação genotípica. Para o CHIKV, a região E2 apresentou acurácia perfeita, semelhante à região E1 (ou seja, especificidade e sensibilidade de 100%). No entanto, o nosso apoio anterior ao bootscanning mostrou que a região E2 geneticamente variável pode causar problemas para que algumas estirpes sejam corretamente identificadas como genótipo ECSA.

Em resumo, nossos resultados sugerem que a região do envelope do DENV e ZIKV e a região do envelope E1 do CHIKV são adequadas para fins de genotipagem. Além disso, estas regiões contêm o maior número de sequências em bases de dados públicas, o que permite facilmente uma ampla gama de análises comparativas e experiências de validação.

Discussão

As doenças infecciosas emergentes causadas por agentes patogênicos virais ainda representam uma grande ameaça para a saúde pública em todo o mundo, como recentemente demonstrado pelos surtos de Ébola, Zika, Síndrome Respiratória do Médio Oriente (MERS) e vírus da Febre Amarela. A monitorização rápida e precisa em tempo real dos surtos e a vigilância das epidemias em curso são cruciais para antecipar a propagação viral e para conceber estratégias eficazes de prevenção ou tratamento. Para tanto, foi desenvolvido um método preciso e confiável para classificação dos arbovírus ZIKV/DENV/CHIKV: a ferramenta ArboTyping.

A ferramenta ArboTyping implementa um pipeline de classificação que consiste em uma atribuição de espécies baseada no BLAST e avaliação filogenética para identificar subespécies (ou seja, genótipos) em relação a um conjunto de cepas de referência, como exemplificado para outras espécies de vírus em trabalhos anteriores [29–31]. Para permitir uma classificação precisa, um conjunto de sequências de referência que cobrem a extensão da diversidade dentro de espécies e subespécies foi cuidadosamente selecionado.

O desempenho de classificação da ferramenta ArboTyping foi avaliado em um conjunto de dados completos. sequências do genoma. Todas as sequências do genoma completo neste conjunto de dados que puderam ser atribuídas com segurança a uma espécie e genótipo com o padrão ouro, um procedimento de classificação manual, foram concordantes com a ferramenta de tipagem.

Houve, no entanto, 10 sequências que não puderam ser classificadas usando a classificação manual. procedimento de seleção: análises adicionais mostram que essas 10 sequências consistem em 3 sequências atípicas, 2 clados de sequências atípicas (3 sequências em cada clado atípico) e 1 sequência recombinante. Como esses valores discrepantes foram identificados anteriormente [43], esses resultados precisam ser investigados mais detalhadamente para avaliar se esses valores discrepantes formam novos genótipos [44].

No entanto, as sequências do genoma completo não estão actualmente disponíveis de forma rotineira e a adequação das diferentes regiões genômicas foi avaliada no que diz respeito à sua utilização para classificação. Como o gene do envelope é um alvo popular para classificação filogenética, há uma grande disponibilidade de sequências de envelope em bancos de dados públicos. Portanto, o desempenho da ferramenta ArboTyping foi avaliado em um grande conjunto de dados de sequências de envelope (ou seja, conjunto de dados **Global-ENV**). Para essas sequências de envelope, foi relatado um desempenho de classificação próximo ao desempenho da ferramenta em sequências do genoma completo.

Embora a disponibilidade de produtos sequenciais originários de outras regiões genômicas seja atualmente baixa, pode-se esperar que essas regiões aumentem em relevância dado o interesse no desenvolvimento de agentes antivirais que tenham como alvo proteínas não estruturais. Portanto, estudos mais detalhados para avaliar o desempenho da classificação de outras regiões genômicas são necessários [44].

Neste manuscrito nos concentramos na classificação de sequências de consenso nas espécies e nível de subespécie. No entanto, Genome Detective, a estrutura na qual nossas ferramentas estão integradas, também é uma cadeia de ferramentas de descoberta de vírus [41]. A interface de usuário do Genome Detective permite que os usuários forneçam leituras brutas de sequência de próxima geração que podem ser montadas automaticamente em um consenso e passadas para a ferramenta ArboTyping. Detalhes sobre os métodos usados para montar leituras no Genome Detective e uma extensa validação usando leituras NGS brutas podem ser encontrados em [41].

Em conclusão, o novo método aqui apresentado permite a classificação rápida, precisa e de alto rendimento de espécies e genótipos de DENV, CHIKV e ZIKV. As espécies podem ser classificadas usando diferentes produtos de sequenciamento (ou seja, sequências de genoma completo, sequências de envelope e leituras individuais de sequenciamento de próxima geração) e os genótipos podem ser classificados com mais segurança ao usar sequências de envelope ou sequências de genoma completo. Este método acomoda a necessidade de classificar de forma consistente e precisa as sequências de DENV/CHIKV/ZIKV, o que é essencial para implementar o rastreamento de epidemias e para apoiar os esforços de vigilância de surtos. Além disso, apresentamos uma estrutura sólida que tem potencial para servir de base para muitas outras ferramentas de classificação de arbovírus. Essas ferramentas também são úteis para serem integradas em ambientes de gerenciamento de dados [45].

Nosso método é implementado na estrutura do software Genome Detective, adequado para muitas ferramentas de digitação de vírus. A aplicação web que disponibiliza nossa ferramenta através de uma interface web fácil de usar está disponível online através de um servidor dedicado hospedado em <http://www.krisp.org.za/tools.php>.

Informações de apoio S1 Fig. Árvore

filogenética de máxima verossimilhança dos outliers DENV-sero1. Todas as sequências completas do genoma DENV-sero1 foram atribuídas ao nível do genótipo usando análise filogenética manual e classificação pela ferramenta de digitação automatizada. No total, sete genomas completos de DENV-sero1 não puderam ser classificados ao nível do genótipo por nenhum dos métodos de classificação. Essas sete sequências são visualizadas em uma árvore filogenética dos conjuntos de dados WGS, coloridas de acordo com o genótipo. (1I em azul, 1II em verde, 1III em vermelho, 1IV em amarelo, 1V em rosa) Pode-se observar que um aglomerado divergente de seis genomas (AF298807, KF864667, KF184975, EU179860, KF289073 e JQ922546 em preto) forma um outlier clado e um genoma (EF457905 em preto) podem ser considerados discrepantes.

No entanto, note que estes sete genomas poderiam ser adequadamente atribuídos ao serótipo 1. (TIF)

S2 Fig. Análise de recombinação para as sequências completas do genoma do DENV. São mostrados os resultados do bootscan para os dez genomas completos do DENV que não puderam ser classificados em nível genotípico. A análise de inicialização foi realizada usando um comprimento de janela de 1.500 pares de bases e um tamanho de passo de 100 pares de bases. As diferentes cores representam os genótipos de cada sorotipo. O eixo X representa a posição do nucleotídeo no genoma e o eixo Y representa os resultados de bootstrap em

porcentagens. No total, foram analisadas 7 sequências de DENV-sero1 e 1 sequência para cada um dos outros sorotipos, ou seja, DENV-sero2, DENV-sero3 e DENV-sero4. Encontramos apenas a sequência AY496879 é um recombinante do genótipo 3I e 3II do DENV. As outras sequências são outliers (ou seja, JF262783, KF744408, EF457905) ou clados de outliers (ou seja: AF298807, KF864667 e KF184975 forma um clado discrepante; EU179860, KF289073 e JQ922546 formam um clado discrepante). (TIF)

Tabela S1. Cepas de referência selecionadas para os genótipos DENV, CHIKV, ZIKV. Essas sequências de referência foram selecionadas para serem representativas da diversidade dentro dos diferentes DENV, Genótipos de CHIKV e ZIKV que circulam dentro dessas espécies de vírus. (DOCX)

Tabela S2. Sinal filogenético estimado por mapeamento de verossimilhança para DENV (DENV-sero1 para DENV-sero4), regiões subgenômicas de CHIKV e ZIKV. O sinal filogenético foi calculado separadamente por proteína pelo método de mapeamento de verossimilhança implementado no software Tree-Puzzle. A análise de mapeamento de probabilidade calcula a probabilidade das três árvores possíveis que pode ser construído a partir de todos os possíveis quartetos intergenotípicos de táxons. Os resultados para o quartetos resolvidos e quartetos não resolvidos são mostrados na tabela, enquanto os parcialmente resolvidos quartetos não estão listados (podem ser obtidos por quartetos 100%—(não)resolvidos). Parcialmente resolvido quartetos representam os quartetos para os quais estão presentes sinais filogenéticos conflitantes ou recombinação potencial. Regiões genômicas para as quais a porcentagem de quartetos resolvidos é maior mais de 90% estão sombreados em laranja e são considerados caracterizados por sinal filogenético suficiente. (DOCX)

Tabela S3. Avaliação do método filogenético automatizado para classificação de DENV, CHIKV e genomas do genoma completo do ZIKV. O novo método de classificação consiste em 2 partes: determinação da espécie (e para DENV também do sorotipo) usando um procedimento BLAST, seguido de determinação do genótipo usando um método filogenético automatizado. Nosso método foi capaz de atribuir todas as sequências no conjunto de dados de validação do genoma completo às espécies corretas e DENV sorotipo. Portanto, nesta tabela, focamos no desempenho da classificação em relação a atribuição de genótipo, com base no resultado da etapa BLAST (ou seja, um conjunto de dados da espécie e do sorotipo adequados). Os resultados da classificação foram comparados com a análise filogenética manual. Nomes das colunas: TP = total de positivos, TN = total de negativos, FP = falso positivo, FN = falso negativo, SENS = sensibilidade, SPEC = especificidade, ACC = precisão. (DOCX)

Tabela S4. Avaliação do método filogenético automatizado para classificação de DENV, CHIKV e genomas de envelope ZIKV. O novo método de classificação consiste em 2 partes: determinação a espécie (e para DENV também o sorotipo) usando um procedimento BLAST, seguido pela determinação do genótipo usando um método filogenético automatizado. Nosso método foi capaz de atribuir todos sequências no conjunto de dados de validação de envelope para a espécie correta e o sorotipo DENV. Portanto, nesta tabela, nos concentramos no desempenho da classificação em relação à atribuição de genótipos, com base no resultado da etapa BLAST (ou seja, um conjunto de dados da espécie e do sorotipo adequados). Os resultados da classificação foram comparados com a análise filogenética manual. Nomes das colunas: TP = total de positivos, TN = total de negativos, FP = falso positivo, FN = falso negativo, SENS = sensibilidade, SPEC = especificidade, ACC = precisão. (DOCX)

Arquivo S1. Número de acesso das sequências coletadas de DENV, ZIKV e CHIKV genomas do genoma completo. Uma mineração de sequências do GenBank foi realizada contra

genomas desses vírus que tiveram o genótipo reportado para testes de sensibilidade, especificidade e precisão da ferramenta.

(XLSX)

Arquivo S2. Número de acesso das sequências coletadas dos genomas do envelope DENV, ZIKV e CHIKV. Uma mineração de sequências no GenBank foi realizada contra genomas de envelope desses vírus que tiveram o genótipo relatado para testes de sensibilidade, especificidade e precisão da ferramenta. (XLSX)

Contribuições do autor

Conceituação: Vagner Fonseca.

Curadoria de dados: Vagner Fonseca, Pieter JK Libin, Lize Cuypers.

Análise formal: Vagner Fonseca, Pieter JK Libin, Kristof Theys, Maria I. Restovic, Murilo Freire, Lize Cuypers.

Aquisição de financiamento: Luiz CJ Alcantara, Tulio de Oliveira.

Investigação: Vagner Fonseca, Pieter JK Libin, Lize Cuypers.

Metodologia: Vagner Fonseca, Maria I. Restovic, Murilo Freire.

Administração do projeto: Luiz CJ Alcantara, Tulio de Oliveira.

Recursos: Luiz CJ Alcântara, Túlio de Oliveira.

Programas: Vagner Fonseca, Kristof Theys, Maria I. Restovic, Murilo Freire, Koen Deforche, Emmanuel J. San, Kaliane CB Machado.

Supervisão: Luiz CJ Alcântara, Túlio de Oliveira.

Validação: Vagner Fonseca, Pieter JK Libin, Kristof Theys, Maria I. Restovic, Murilo Freire, Lize Cuypers, Koen Deforche, Anne-Mieke Vandamme.

Redação – rascunho original: Vagner Fonseca, Nuno R. Faria, Marcio RT Nunes, Marta Giova-netti, Ann Nowe', Ana Abecasis, Koen Deforche, Gilberto A. Santiago, Isadora C. de Siqueira, Vasco Azevedo, Ana Maria Bispo- de Filippis, Rivaldo Venâncio da Cunha, Oliver G. Pybus, Anne-Mieke Vandamme, Luiz CJ Alcantara, Tulio de Oliveira.

Redação – revisão e edição: Vagner Fonseca, Nuno R. Faria, Marcio RT Nunes, Marta Gio-vanetti, Ann Nowe', Ana Abecasis, Koen Deforche, Gilberto A. Santiago, Isadora C. de Siqueira, Emmanuel J. San, Vasco Azevedo, Ana Maria Bispo-de Filippis, Rivaldo Venâncio da Cunha, Oliver G. Pybus, Anne-Mieke Vandamme, Luiz CJ Alcantara, Tulio de Oliveira.

Referências

1. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. A distribuição global e a carga da dengue. *Natureza*. 2013; 496:504–507. <https://doi.org/10.1038/nature12060> PMID: 23563266
2. Weaver SC, vírus Lecuit M. Chikungunya e a propagação global de uma doença transmitida por mosquitos. *Jornal de Medicina da Nova Inglaterra*. 2015; 372(13):1231–1239. <https://doi.org/10.1056/NEJMr1406035> PMID: 25806915
3. Fauci AS, Morens DM. Vírus Zika nas Américas: mais uma ameaça de arbovírus. *New England Journal of Medicine*. 2016; 374(7):601–604. <https://doi.org/10.1056/NEJMp1600297> PMID: 26761185
4. Kraemer MUG, Sinka ME, Duda KA, Mylne AQN, Shearer FM, Barker CM, et al. A distribuição global dos vetores de arbovírus *Aedes aegypti* e *Ae. Albopictus*. *eVida*. 2015; 4(08347). <https://doi.org/10.7554/eLife.08347> PMID: 26126267

5. Cardoso CW, Paploski IAD, Kikuti M, Rodrigues MS, Silva MMO, Campos GS, et al. Surto de doença exantemática associada aos vírus Zika, Chikungunya e Dengue, Salvador, Brasil. *Doenças Infecciosas Emergentes*. 2015; 21(12):2274–2276. <https://doi.org/10.3201/eid2112.151167> PMID: 26584464
6. Roth A, Mercier A, Leprosos C, Hoy D, Duituturaga S, Benyon E, et al. Surtos simultâneos de infecções por dengue, chikungunya e vírus Zika – uma onda epidêmica sem precedentes de vírus transmitidos por mosquitos no Pacífico 2012–2014. *Fiscalização do Euro*. 2014; 19(41):20929. <http://dx.doi.org/10.2807/1560-7917.ES2014.19.41.20929> PMID: 25345518
7. Ministério da Saúde B. Boletim Epidemiológico Secretaria de Vigilância em Saúde; 2018. v.
8. Shepard DS, Undurraga EA, Halasa YA, Stanaway JD. O fardo econômico global da dengue: uma análise sistemática. *Lanceta Infect Dis*. 2016; 16(8):935–941. [https://doi.org/10.1016/S1473-3099\(16\)00146-8](https://doi.org/10.1016/S1473-3099(16)00146-8) PMID: 27091092
9. Morens DM, Fauci AS. Enfrentando o Desafio da Epidemia de Chikungunya. *Jornal de Doenças Infecciosas*. 2016; 214(suplemento 5):S434–S435. <https://doi.org/10.1093/infdis/jiw291> PMID: 27920168
10. Rasmussen SA, Jamieson DJ, Honein MA, Petersen LR. Vírus Zika e defeitos congênitos – revisando as evidências de causalidade. *Jornal de Medicina da Nova Inglaterra*. 2016; páginas 1–7. <https://doi.org/10.1056/NEJMs1604338> PMID: 27074377
11. Brasil P, Sequeira PC, Freitas AD, Zogbi HE, Calvet GA, de Souza RV, et al. Síndrome de Guillain-Barré associada à infecção pelo vírus Zika. *A Lanceta*. 2016; 1482. [https://doi.org/10.1016/S0140-6736\(16\)30058-7](https://doi.org/10.1016/S0140-6736(16)30058-7)
12. Manokaran G, Finol E, Wang C, Gunaratne J, Bahl J, Ong EZ, et al. RNA subgenômico da dengue se liga TRIM25 para inibir a expressão de interferon para aptidão epidemiológica. *Ciência*. 2015; 350(6257):217–221. <https://doi.org/10.1126/science.aab3369> PMID: 26138103
13. Katzelnick LC, Fonville JM, Gromowski GD, Bustos Arriaga J, Green A, James SL, et al. Os vírus da dengue agrupam-se antigenicamente, mas não como sorotipos distintos. *Ciência (Nova York, NY)*. 2015; 349(6254):1338–43. <https://doi.org/10.1126/science.aac5017> PMID: 26383952
14. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. Pesquisa de ácidos nucleicos. 2013; 41(D1):D36–D42. <https://doi.org/10.1093/nar/gkr1202>
15. Katoh K, Standley DM. Software de alinhamento de sequências múltiplas MAFFT versão 7: Melhorias no desempenho e usabilidade. *Biologia Molecular e Evolução*. 2013; 30(4):772–780. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690
16. Guindon S, Gascuel O. Um algoritmo simples, rápido e preciso para estimar grandes filogenias por máxima verossimilhança. *Biologia Sistemática*. 2003; 52(5):696–704. <https://doi.org/10.1080/10635150390235520> PMID: 14530136
17. Ronquist F, Huelsenbeck JP. MrBayes 3: Inferência filogenética bayesiana sob modelos mistos. *Bioinformática*. 2003; 19(12):1572–1574. <https://doi.org/10.1093/bioinformatics/btg180> PMID: 12912839
18. Rico-Hesse R. Evolução molecular e distribuição dos vírus da dengue tipo 1 e 2 na natureza. *Virologia*. 1990; 174(2):479–493. [https://doi.org/10.1016/0042-6822\(90\)90102-W](https://doi.org/10.1016/0042-6822(90)90102-W) PMID: 2129562
19. Twiddy SS, Farrar JJ, Vinh Chau N, Wills B, Gould EA, Gritsun T, et al. Relações filogenéticas e pressões de seleção diferencial entre genótipos do vírus dengue-2. *Virologia*. 2002; 298(1):63–72. <https://doi.org/10.1006/viro.2002.1447> PMID: 12093174
20. Chungue E, Deubel V, Cassar O, Laille M, Martin PMV. Epidemiologia molecular dos vírus da dengue 3 e relação genética entre cepas de dengue 3 isoladas de pacientes com forma leve ou grave de dengue na Polinésia Francesa. *Jornal de virologia geral*. 1993; 74(12):2765–2770. <https://doi.org/10.1099/0022-1317-74-12-2765>
21. Klungthong C, Zhang C, Mammen MP, Ubol S, Holmes EC. A epidemiologia molecular da dengue sorotipo 4 do vírus em Bangkok, Tailândia. *Virologia*. 2004; 329(1):168–179. <https://doi.org/10.1016/j.virol.2004.08.003> PMID: 15476884
22. Nunes MRT, Faria NR, de Vasconcelos JM, Golding N, Kraemer MU, de Oliveira LF, et al. Emergência e potencial de disseminação do vírus Chikungunya no Brasil. *Medicina BMC*. 2015; 13(1):102. <https://doi.org/10.1186/s12916-015-0348-x> PMID: 25976325
23. Volk SM, Chen R, Tsetsarkin KA, Adams AP, Garcia TI, Sall AA, et al. Filogenética em escala genômica análises do vírus chikungunya revelam emergências independentes de epidemias recentes e diversas taxas evolutivas. *Jornal de virologia*. 2010; 84(13):6497–6504. <https://doi.org/10.1128/JVI.01603-09> PMID: 20410280
24. Dick GWA, Cozina SF, Haddow AJ. Vírus Zika (I). Isolamentos e especificidade serológica. *Transações da Royal Society of Tropical Medicine and Hygiene*. 1952; 46(5):509–520. [https://doi.org/10.1016/0035-9203\(52\)90042-4](https://doi.org/10.1016/0035-9203(52)90042-4) PMID: 12995440

25. Kindhauser MK, Allen T, Frank V, Santhana RS, Dye C. Zika: a origem e propagação de um vírus transmitido por mosquitos. *Órgão Mundial de Saúde Bull.* 2016;171082.
26. Marchette NJ, Garcia R, Rudnick A. Isolamento do vírus Zika de mosquitos *Aedes aegypti* na Malásia. *Jornal Americano de Medicina Tropical e Higiene.* 1969; 18(3):411–415. <https://doi.org/10.1056/NEJMp1002530> PMID: 4976739
27. Cao-Lormeau VM, Roche C, Teissier A, Robin E, Berry AL, Mallet HP, et al. Zika Vírus, Polinésia Francesa, Pacífico Sul, 2013. *Doenças Infecciosas Emergentes.* 2014; 20(6):1085–1086. <https://doi.org/10.3201/eid2006.140138> PMID: 24856001
28. Hayes EB, outros. Vírus Zika fora da África. *Emerg Infect Dis.* 2009; 15(9):1347–1350. <https://doi.org/10.3201/eid1509.090442> PMID: 19788800
29. Alcantara LCJ, Cassol S, Libin P, Deforche K, Pybus OG, Van Ranst M, et al. Um quadro padronizado trabalham para genotipagem precisa e de alto rendimento de sequências virais recombinantes e não recombinantes. *Pesquisa de ácidos nucleicos.* 2009; 37(Suplemento 2):634–642. <https://doi.org/10.1093/nar/gkp455> PMID: 19483099
30. de Oliveira T, Deforche K, Cassol S, Salminen M, Paraskevis D, Seebregts C, et al. Um sistema automatizado de genotipagem para análise de HIV-1 e outras sequências microbianas. *Bioinformática.* 2005; 21 (19):3797–3800. <https://doi.org/10.1093/bioinformatics/bti607> PMID: 16076886
31. Pineda-Peña AC, Faria NR, Imbrechts S, Libin P, Abecasis AB, Deforche K, et al. Subtipagem automatizada de sequências genéticas do HIV-1 para fins clínicos e de vigilância: Avaliação de desempenho da nova versão 3 do REGA e sete outras ferramentas. *Infeção, Genética e Evolução.* 2013; 19:337–348. <https://doi.org/10.1016/j.meegid.2013.04.032> PMID: 23660484
32. Libin P., Deforche K., Abecasis AB, & Theys K. (2018). VIRULIGN: alinhamento rápido com correção de códons e anotação de genomas virais. *Bioinformática (Oxford, Inglaterra).*
33. Theys K., Libin P., Dallmeier K., Pineda-Peña AC, Vandamme AM, Cuypers L., & Abecasis AB (2017). A genômica do zika precisa urgentemente de sequências de referência padronizadas e selecionadas. *Patógenos PLoS*, 13(9), e1006528. <https://doi.org/10.1371/journal.ppat.1006528> PMID: 28880955
- Salemi M, Vandamme AM. *O manual filogenético: uma abordagem prática à filogenia do DNA e das proteínas.* Cambridge University Press; 2003.
35. Nylander JAA, Wilgenbusch JC, Warren DL, Swofford DL. AWTY (já chegamos?): um sistema para exploração gráfica da convergência MCMC na filogenética Bayesiana. *Bioinformática.* 2008; 24 (4):581–583. <https://doi.org/10.1093/bioinformatics/btm388> PMID: 17766271
36. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuve P. RDP3: Um programa de computador flexível e rápido para analisar recombinação. *Bioinformática.* 2010; 26(19):2462–2463. <https://doi.org/10.1093/bioinformatics/btq467> PMID: 20798170
37. Ha Schmidt, Strimmer K, Vingron M, von Haeseler A. TREE-PUZZLE: análise filogenética de máxima verossimilhança usando quartetos e computação paralela. *Bioinformática (Oxford, Inglaterra).* 2002; 18(3):502–504. <https://doi.org/10.1093/bioinformatics/18.3.502>
38. Strimmer K, von Haeseler A. Mapeamento de probabilidade: um método simples para visualizar o conteúdo filogenético de um alinhamento de sequência. *Anais da Academia Nacional de Ciências.* 1997; 94(13):6815–6819. <https://doi.org/10.1073/pnas.94.13.6815> PMID: 9192648
39. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Ferramenta básica de busca de alinhamento local; 1990.
40. Larkin MA, Blackshields G, Brown NP, Chenna R, Mcgettigan PA, McWilliam H, et al. Clustal W e Clustal X versão 2.0. *Bioinformática.* 2007; 23(21):2947–2948. <https://doi.org/10.1093/bioinformatics/btm404> PMID: 17846036
41. Vilsker M, Moosa Y, Nooij S, Fonseca V, Ghysens Y, Dumon K, Pauwels R, Alcantara LC, Vanden Eyn-den E, Vandamme AM, Deforche K, de Oliveira T, *Bioinformática* (2019), <https://doi.org/10.1093/bioinformática/bty695> PMID: 30124794
42. Banoo S, Bell D, Bossuyt P, Herring A, Mabey D, Poole F, et al. Avaliação de testes diagnósticos de doenças infecciosas: Princípios gerais. *Nature Review Microbiologia.* 2006; 4(9 SUPPL.) S21–S31. <https://doi.org/10.1038/nrmicro1523> PMID: 17034069
43. Libin P., Vanden Eynden E., Incardona F., Nowe´ A., Bezenchek A.; Grupo de Estudo EucoHIV, Sonnerborg A., Vandamme A.-M., Theys K., Baele G.(2017). PhyloGeoTool: explorando interativamente grandes filogenias em um contexto epidemiológico. *Bioinformática*, 33(24):3993–3995. <https://doi.org/10.1093/bioinformatics/btx535> PMID: 28961923
44. Cuypers L., Libin PJK, Simmonds P., Nowe´ A., Muñoz-Jorda´n J., Alcantara LCJ, Vandamme A.-M., Santiago GA, Theys K. (2018). É hora de harmonizar a nomenclatura e classificação da dengue. *Vírus*, 10(10), pii: E569. <https://doi.org/10.3390/v10100569> PMID: 30340326
45. Libin P., Beheydt G., Deforche K., Imbrechts S., Ferreira F., Van Laethem K., Theys K., Carvalho AP, Cavaco-Silva J., Lapadula G., Torti C., Assel M., Wesner S., Snoeck J., Ruelle J., De Bel A., Lacor P.,

De Munter P., Van Wijngaerden, E., Zazzi M., Kaiser R., Ayoub A., Peeters M., de Oliveira T., Alcantara LC, Grossman Z., Sloot P., Otelea D., Paraschiv S., Boucher C., Camacho RJ, Vandamme A.-M. (2013). RegaDB: gerenciamento e análise de dados voltados para a comunidade para doenças infecciosas. *Bioinformática*. 2013, 29(11):1477–80. <https://doi.org/10.1093/bioinformatics/btt162> PMID: [23645815](#)