

Batista, CT et al.

Artigo de Revisão

Evolução Genômica do 2019-nCoV: Revisão Sistemática da Literatura

Genomic Evolution of 2019-nCoV: Systematic Literature Review

Caio Teles Batista^{1*}, Amanda Cristina Alves da Cruz¹, João Vitor Oliveira Amorim¹ & Giovanni Monteiro Ribeiro²

1 Estudantes do Centro Universitário do Planalto Central Aparecido dos Santos – UNICEPLAC

2 Professor Doutor do Curso de Odontologia do Centro Universitário do Planalto Central Aparecido dos Santos - UNICEPLAC

*E-mail: ctelesbatista@gmail.com, Rossi Splendore apt 1501 setor industrial Gama-DF, CEP: 72445010 e Telefone: 61984567674.

Resumo:

Introdução: O SARS-COV-2 é o agente causador da doença por coronavírus (COVID-19) que infectou milhões de pessoas e é responsável por mais de 550.000 mortes em todo o mundo em um período de 7 meses. O vírus é composto por RNA de fita simples de tamanho considerável, tornando-o mais suscetível a alterações genéticas. Algumas evidências demonstram que a taxa de mortalidade do SARS-COV-2 varia de acordo com as regiões do mundo e isso pode acontecer devido a diversos fatores, sendo a mutação viral uma possível causa para esta variação. **Objetivo:** O objetivo desta revisão sistemática da literatura é reunir informações emergentes sobre a evolução genômica do SARS-COV-2. **Metodologia:** A metodologia utilizada nesta revisão sistemática de literatura foi baseada no PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. As referências foram consultadas nos bancos de dados PubMed, Scielo e Google Scholar. A estratégia de busca utilizou termos como SARS-COV-2, COVID-19 and Mutation, sem limitações quanto ao período de publicação, artigos em língua inglesa e artigos originais, relatos de caso e revisões de literatura. Resultados

e Discussão: Após o sequenciamento do primeiro genoma do SARS-CoV-2, características diferentes foram observadas durante o mapeamento de milhares de genomas coletados. Ao focarmos nas mutações genéticas no novo coronavírus, identificou-se que as mesmas ocorrem principalmente em cinco proteínas, incluindo S, N, ORF8, ORF3a e ORF1ab, sendo em grande parte mutações não-sinônimos; essas mutações podem induzir alterações conformacionais que levam a possíveis alterações de antigenicidade, modificar a afinidade de ligação a receptor de ACE2 humana ou podem determinar características quanto a patogenicidade do SARS-CoV-2. **Conclusão:** As mutações observadas são resultantes da constante interação do vírus com o genoma humano devido a adaptação contínua, mas até o presente momento não se verificou que as alterações genéticas modificaram a mortalidade ou a taxa de infecção do vírus.

Palavras-chave: COVID-19, Sars-Cov-2, Genética

Abstract:

Introduction: SARS-COV-2 is the causative agent of coronavirus disease

Batista, CT et al.

(COVID-19) that has infected millions of people and is responsible for more than 550,000 deaths worldwide in a period of 7 months. The virus is made up of single-stranded RNA of considerable size making it more susceptible to genetic changes. Some evidence shows that the mortality rate of SARS-VOC-2 varies according to the regions of the world and this can happen due to several factors, with viral mutation being a possible cause for this variation. Objective: The objective of this systematic literature review is to gather emerging information on the genomic evolution of SARS-COV-2. Methodology: The methodology used in this systematic literature review was based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyzes) guidelines. References were consulted in the PubMed, Scielo and Scholar Google databases. The search strategy used terms such as SARS-COV-2, COVID-19 and Mutation, without limitations on the period of publication, articles in English and original articles, case reports and literature reviews. Results and Discussion: After sequencing the first SARS-CoV-2 genome, different characteristics were observed when mapping thousands of collected genomes. When focusing on genetic mutations in the new coronavirus, it was identified that they occur mainly in five proteins, including S, N, ORF8, ORF3a and ORF1ab, being largely non-synonymous mutations; these mutations can induce conformational changes that lead to possible changes in antigenicity, modify the binding affinity to the human ACE2 receptor, or can determine characteristics regarding the pathogenicity of SARS-CoV-2. Conclusion: The observed mutations are the result of the constant interaction of the virus with the human genome due to continuous adaptation, but until now, it has not been verified that the genetic

alterations have changed the mortality or the infection rate of the virus.

Keywords: COVID-19, Sars-Cov-2, Genetics |

Introdução

O SARS-COV-2 é o agente causador da doença por coronavírus (COVID-19) que infectou milhões de pessoas e é responsável por mais de 550.000 mortes em todo o mundo em um período de 7 meses. O vírus tem um genoma de RNA fita simples sentido positivo com cerca de 30 kb de comprimento. O genoma codifica quatro proteínas estruturais e múltiplas não estruturais (ASTUTI; 2020). Enquanto as proteínas estruturais formam o capsídeo e o envelope do vírus, proteínas não estruturais estão envolvidas em várias etapas do ciclo viral, como replicação, tradução, montagem e liberação (LAI; 1997).

Inicialmente a infecção pelo vírus foi denominada pneumonia por coronavírus (PCN) por apresentar características clínicas semelhantes a uma pneumonia. Posteriormente, a Organização Mundial da Saúde (OMS) nomeou-a 'COVID-19' e o Comitê Internacional de Taxonomia de Vírus chamou o vírus de 'síndrome respiratória aguda grave coronavírus 2' (SARS-CoV-2) (GORBALENYA et al., 2020). O

Batista, CT et al.

sequenciamento genômico revelou que o SARS-COV-2 é um membro do gênero Betacoronavírus e pertence ao subgênero Sabecovirus, que inclui SARS-COV, enquanto o MERS-CoV pertence a um subgênero separado, Merbecovirus (LU et al., 2020). Em relação à sequência de nucleotídeos o SARS-CoV-2 e o SARS-CoV são aproximadamente 79% semelhantes, e quando comparado ao coronavírus SARS de morcego – SARSr-CoV-RaTG13 – a similaridade chegou à 96% (WANG et al., 2020).

As taxas de mutação na replicação dos vírus de RNA são muito maiores que as dos vírus de DNA e os genomas dos vírus de RNA geralmente têm menos de 10 kb de comprimento (DRAKE, 1993). O genoma do CoV é maior do que os outros vírus de RNA, com aproximadamente 30 kb de comprimento, o maior vírus de RNA conhecido (LI; YANG; REN, 2020). Como resultado de seu mecanismo único de replicação viral, os coronavírus têm uma alta frequência de recombinação (WOO et al., 2005).

O *orf1ab* é o maior gene da SARS-CoV-2, codificando uma poliproteína (PP1ab). Outro gene *orf1a* codifica para uma poliproteína (PP1a). Dois terços do RNA viral situado na primeira ORF (ORF1ab) codifica uma

poliproteína com 7096 resíduos de comprimento. Assim, *orf1a* e *orf1ab* são traduzidos para produzir poliproteínas PP1a e PP1ab, que são clivadas pelas proteases que são codificadas por ORF1a para produzir 15 proteínas não estruturais (NSPs). Os estudos de sequência observaram variações entre SARS-CoV-2 e SARS-CoV, como a falta do gene 8a ou a variação no número de pares de bases nos genes 8b e 3c no SARS-CoV-2. (ZEHRRA et al., 2020)

A principal via de mutação é dependente das enzimas que realizam o processo de replicação viral e da falha do processo de verificação e reparo pós-replicativo. Na maioria dos vírus de RNA, a polimerase não possui a capacidade de revisar o processo de duplicação, mas com algumas exceções, como a classe que engloba os coronavírus (PACHETTI et al., 2020). No caso do SARS-COV-2 muitos mecanismos são semelhantes aos encontrados no SARS-COV, como a presença do nsp12 e nsp14, ambas componentes essenciais para revisão e fidelidade na síntese de RNA (WU et al., 2020, MA et al., 2015).

Algumas evidências demonstram que a taxa de mortalidade do SARS-COV-2 varia de acordo com as regiões do mundo e isso pode acontecer

Batista, CT et al.

devido a diversos fatores, sendo a mutação viral uma possível causa para esta variação (BAUD et al., 2020). Ademais, a identificação e estudo destas mutações é muito importante para o desenvolvimento de testes diagnósticos, desenvolvimento de novas drogas, resistência às drogas, produção de vacinas (UDDIN et al., 2020). O objetivo desta revisão sistemática da literatura é reunir informações emergentes sobre a evolução genômica do SARS-COV-2.

Material e Métodos

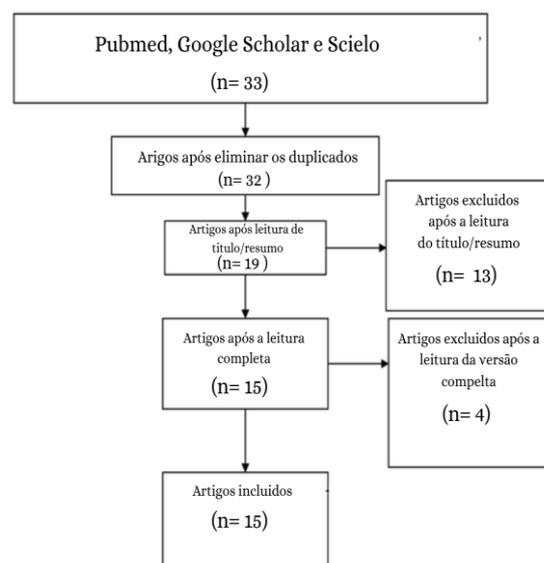
A metodologia utilizada nesta revisão sistemática de literatura foi baseada no PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. As referências foram consultadas nos bancos de dados PubMed, Scielo e Google Scholar. A estratégia de busca utilizou termos como SARS-COV-2, COVID-19 and Mutation, sem limitações quanto ao período de publicação, artigos escritos em língua inglesa e artigos originais, relatos de caso e revisões de literatura. A última busca foi realizada em julho de 2020. Uma análise inicial foi realizada baseando-se no título e resumo. Os artigos incluídos foram acessados em sua versão completa e analisados para identificar os estudos relevantes. O

processo foi realizado por todos os autores.

Resultados e Discussão

A primeira parte da seleção incluiu 33 artigos, em que 1 estava duplicado entre os bancos de dados. Após a leitura do título e resumo dos artigos, 13 artigos foram excluídos, restando 19 artigos. Os estudos restantes foram analisados em sua versão final e 4 artigos foram excluídos, restando 15 estudos para a análise desta revisão de literatura (Figura 1).

Figura 1 Prisma Flow Diagram indicando a seleção de artigos



O primeiro genoma da SARS-CoV-2 foi publicado em 24 de janeiro de 2020, apenas algumas semanas após o surto inicial na província de Wuhan, China e exibiu similaridade genômica e filogenética com a SARS-CoV, particularmente no gene S, *spike*, e no

Batista, CT et al.

domínio de ligação ao receptor (RBD), indicando a capacidade de transmissão direta de humano para humano (UDDIN et al., 2020). Um nível aumentado de diversidade viral foi encontrado em alguns pacientes infectados com SARS-CoV-2, sugerindo que o vírus começou a se adaptar ao ambiente humano e seus genomas começaram a evoluir na população (CHEN; LIU; GUO, 2020).

O genoma do SARS-CoV-2 possui uma poliproteína ORF1ab longa na extremidade 5', seguida por quatro proteínas estruturais principais, incluindo a glicoproteína da superfície do pico, a proteína do envelope pequeno, a proteína da matriz e a proteína nucleocapsídica (PHAN, 2020a). ORFs 3a, 6, 7a, 7b, 8 e 9b codificam proteínas acessórias, também presentes no coronavírus da síndrome respiratória aguda grave de 2002-2003 (SARS-CoV-1), mas menos amplamente conservadas nos coronavírus como um todo (FIRTH, 2020).

As cepas de SARS-CoV-2 foram detectadas em pacientes infectados da China (50), EUA (11), Austrália (5), Japão (5), França (4), Cingapura (3), Inglaterra (2), Inglaterra (2) e Taiwan (2), Coreia do Sul (1), Bélgica (1), Alemanha (1) e Vietnã (1) (SAITOU2; NEI, 1987). O alinhamento

da sequência nucleotídica aos pares foi realizado por ClustalX2 e a sequência da cepa China / WHU01 / 2020 / EPI_ISL_406716 foi usada como genoma de referência. Tal estudo evidenciou três deleções nos genomas de SARS-CoV-2 do Japão (Aichi), EUA (Wisconsin) e Austrália (Victoria), sendo duas deleções (três nucleotídeos e vinte e quatro nucleotídeos) na poliproteína ORF1ab e uma (dez nucleotídeos) na extremidade 3' do genoma (PHAN, 2020b). Mutações de diferentes isolados de SARS-CoV-2 ocorrem principalmente em cinco genes, incluindo S, N, ORF8, ORF3a e ORF1ab, com cerca de 42% das variações são mutações não-sinônimos (LI; YANG; REN, 2020).

Os 5' dois terços do genoma contêm dois quadros de leitura abertos longos (ORFs), ORF1a e ORF1b, que são traduzidos do RNA genômico viral (gRNA). O terço 3' do genoma contém um número de ORFs que codificam as proteínas estruturais e acessórias virais. Essas ORFs são traduzidas de uma série aninhada de mRNAs subgenômicos (sgmRNAs) produzidos durante o ciclo de infecção (FIRTH, 2020).

Um estudo recente de perfil de ribossomo de células infectadas com SARS-CoV-2 revelou 23 novas ORFs

Batista, CT et al.

traduzidas. Dez são muito curtos (≤ 15 códons). Sete dos restantes compreendem extensões de 5' ou truncamentos de 5' de ORFs previamente conhecidas (M, 6, 7a, 7b, 9b e 10). Dois são uORFs posicionados no gRNA a montante de ORF1a que podem desempenhar um papel na regulação da expressão de ORF1a / 1b, conforme proposto anteriormente para uORFs em outros coronavírus. Após excluir essas ORFs, permanecem apenas quatro novas ORFs traduzidas: ORF3c (25 457–25 582; 41 códons), outra ORF sobreposta à ORF3a (25 596–25 697; 33 códons), uma ORF sobreposta à S ORF (21 74421 863; 39 códons) e uma versão truncada do mesmo (21 768–21 863; 31 códons) (FIRTH, 2020).

Existe uma variabilidade considerável entre gêneros e subgêneros de coronavírus no complemento dos genes acessórios codificados em 3'. Mesmo dentro do subgênero sarbecovírus, existem diferenças. Por exemplo, o SARS-CoV-1 possui um ORF3b que se sobrepõe à região 3' de ORF3a, mas é truncado ou ausente no SARS-CoV-2. Além disso, em muitos isolados de SARS-CoV-1 adaptados ao ser humano, o ORF8 é dividido por uma deleção que interrompe o quadro. Aparentemente, a ORF10 está traduzida

em SARS-CoV-2 mas está truncado no SARS-CoV-1 (FIRTH, 2020).

Um estudo feito por Phan T. et al. (2020) revelou noventa e três mutações em todo o genoma de SARS-CoV-2. Dessas, quarenta e duas mutações *missense* foram identificadas em todas as principais proteínas não estruturais e estruturais, exceto a proteína do envelope, vinte e nove mutações *missense* estavam na poliproteína ORF1ab, oito na glicoproteína da superfície do pico, uma na proteína da matriz e quatro na proteína nucleocapsídica (PHAN, 2020b). Notam-se três mutações (D³⁵⁴, Y³⁶⁴ e F³⁶⁷) localizadas no domínio de ligação ao receptor da glicoproteína na superfície do pico. Essa glicoproteína desempenha um papel essencial na ligação aos receptores na célula hospedeira e determina o tropismo do hospedeiro (FUNG; LIU, 2019). É também o principal alvo de anticorpos neutralizantes (YU et al., 2020). Mutações nessa glicoproteína foi o que provavelmente levou à alteração da antigenicidade viral.(PHAN, 2020b).

A glicoproteína de pico é crítica para a infecção pelo vírus. Estudo recente sugeriu que a proteína S na SARS-CoV-2 pode sofrer um rearranjo estrutural (LU et al., 2020). Para

Batista, CT et al.

investigar esta hipótese, foram desenvolvidas duas filogenias separadas com base nas sequências full-S e RBD, respectivamente. No geral, as duas filogenias exibiram padrões de agrupamento semelhantes, separando em três grandes clados. O SARS-CoV-2 foi identificado no mesmo clado principal e estava mais estreitamente ligado ao CoVs SARS de morcego e o SARS-CoV humano. Nas duas filogenias, o SARS-CoV-2 está mais relacionado ao bat_CoV_RaTG13, sugerindo o SARS-CoV-2 pode ter se originado de morcego. Contudo, as posições evolutivas de SARS-CoV humano e morcego-SL-CoVZ45 foram trocados entre as filogenias full-S e RBD. Na filogenia full-S, bat-SL-CoVZ45 é relativamente mais semelhante ao SARS-CoV-2 humano, enquanto o SARS-CoV humano está mais próximo do SARS-CoV-2 do que o bat-SL-CoVZ45. Tomados em conjunto, esses resultados sugeriram que a RBD da SARS-CoV-2 é mais provável originada da SARS-CoV humana, enquanto a parte restante da proteína S em SARS-CoV-2 pode ter se originado de bat-SL-CoVZ45, apoiando o possível rearranjo estrutural da proteína S em SARS-CoV-2. O bat_CoV_RaTG13 é semelhante ao SARS-CoV-2, indicando a forma estrutural proposta. O rearranjo pode ter

ocorrido primeiro no bastão antes de sua transmissão ao ser humano (JIA et al., 2020).

A estrutura 3D da proteína *spike* RBD de SARS-CoV-2 (PDB: 6VW1) foi recentemente determinado em complexo com o receptor ACE2 (da sigla em inglês: angiotensin-converting enzyme 2) humano. Esse receptor é uma proteína expressa na superfície das células como o epitélio do sistema respiratório bem como músculos, rins, tireóide e outros, desempenhando um papel na regulação da função cardiovascular e renal ao converter angiotensina I e II em peptídeos Ang1-9 e Ang1-7, com efeito fisiológico oposto à angiotensina II. A ampla distribuição dessa enzima na superfície de diversos tecidos humanos levanta a possibilidade de diversos sistemas a serem alvo de vírus após disseminação por corrente sanguínea (LI, 2020). Uma única mutação de aminoácido na RBD resulta em menor afinidade de ligação a esse receptor. Uma vez que a proteína do vírus S é responsável pela primeira etapa da infecção por CoV (LAI; CAVANAGH, 1997), qualquer mutação de aminoácidos no RBD pode ter um impacto significativo na ligação ao receptor e no desenvolvimento da vacina.

Batista, CT et al.

Uma das 12 mutações de aminoácidos no RBD da proteína S (R408I) foi identificada entre os 106 SARS-Genomas de CoV-2. O alinhamento da sequência mostrou que o 408R é estritamente conservado em SARS-CoV-2, SARS-CoV e SARS-CoV do morcego. Com base na estrutura complexa CoV2_RBD-ACE2 determinada, 408R está localizado na interface entre RBD e ACE2, mas está posicionado relativamente longe do ACE2, portanto, não possui interação direta com o ACE2. Entretanto, a estrutura RBD0-ACE2 determinada mostrou que 408R forma uma ligação de hidrogênio com o glicano anexado a 90N do ACE2 (PHAN T., 2020b). A ligação de hidrogênio pode ter contribuído para o aumento da afinidade de ligação à ACE2. Por outro lado, apesar desse resíduo de arginina também ser conservado no SARS-CoV humano (correspondente a 395R no PDB: 2AJF), está posicionado relativamente distante (6,1 Å) do glicano ligado a 90N do ACE2. Curiosamente, a ligação de hidrogênio 408R-glicano parece ser interrompida pela mutação R408I em um SARS-CoV-2 acesso (ID do GeneBank: MT012098.1), que foi coletado na Índia em Janeiro de 2020 (YADAV, 2020). Além disso, ao contrário do resíduo de arginina, que é eletricamente carregado e

altamente hidrofílico, o resíduo de isoleucina mutado possui uma cadeia lateral altamente hidrofóbica sem potencial de ligação de hidrogênio. A mutação R408I identificada na Índia representa um mutante SARS-CoV-2 com afinidade de ligação à ACE2 potencialmente reduzida (WU et al., 2020).

Mutações que surgem independente várias vezes (homoplasias), são prováveis da adaptação contínua do SARS-CoV-2 ao seu novo hospedeiro humano. Uma das homoplasias mais fortes está no local 11.083 no genoma do SARS-CoV-2 em uma região de Orf1a que codifica Nsp6. Mais homoplasias menores, identificadas no Orf3a. De notar, também identificamos uma forte mutação recorrente na posição nucleotídica 21.575, correspondente à proteína *spike* SARS-CoV-2 (códon 5). Enquanto a proteína *spike* é o mediador conhecido da entrada de células hospedeiras, nossa homoplasia detectada fica fora dos domínios de ligação ao terminal N e ao receptor (VAN DORP et al., 2020). Em outro estudo foram encontradas 13 locais de variação nas regiões 1a, 1b, S, 3a, M, 8 e posição nt29095 da região N da ORF, entre os quais as posições nt28144 em ORF 8 e

Batista, CT et al.

nt8782 em ORF 1a mostraram taxa de mutação de 30,53% (29/95) e 29,47% (28/95), respectivamente (WANG et al., 2020). A maioria destas mutações observadas no SARS-CoV-2 em circulação em humanos é provavelmente neutra ou até deletéria (VAN DORP et al., 2020).

Conclusão

Percebe-se, portanto, que o SARS-COV-2 vêm sofrendo mutação a partir da interação com o genoma humano, principalmente a proteína ORF. O impacto destas mutações sobre a interação com os seres humanos não parece aumentar a taxa de infecção ou a mortalidade do vírus. Estas homoplasias dos SARS-COV-2, como as mutações *missense* na poliproteína ORF1ab e na glicoproteína da superfície do pico, reverberam a adaptação contínua que esse vírus vem sofrendo. Dessa forma, o estudo da evolução genômica do SARS-COV-2 é de suma importância para entender a origem filogenética do vírus e seus constituintes, como também para mapear suas mutações visando a busca de uma vacina para minimizar os efeitos catastróficos desse vírus.

Referências

Astuti, I. & Ysrafil. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response.

Diabetes Metab. Syndr. Clin. Res. Rev. 14: 407–412, 2020.

Baud D, Qi X, Nielsen-Saines K, Musso D, Pomar L, Favre G. Real estimates of mortality following COVID-19 infection. *Lancet Infect Dis.* 20(7): 773, 2020.

Chen Y, Liu Q, Guo D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J Med Virol.* 92(4):418-423, 2020.

Drake JW. Rates of spontaneous mutation among RNA viruses. *Proc Natl Acad Sci U S A.* 90(9): 4171-4175, 1993

Duffy S, et al. Why are RNA virus mutation rates so damn high?. *PLoS Biol.* 16(8), 2018.

Firth AE. A putative new SARS-CoV protein, 3c, encoded in an ORF overlapping ORF3a. *J Gen Virol.* Published online:jgv001469, July 15, 2020.

Fung TS, Liu DX. Human Coronavirus: Host-Pathogen Interaction. *Annu Rev Microbiol.* 73(1):529-557, 2019.

Gorbalenya AE, Baker SC, Baric RS, et al. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol.* 5(4): 536-544, 2020.

Jia Y, Shen G, Zhang Y, et al. Analysis of the mutation dynamics of SARS-CoV-2 reveals the spread history and emergence of RBD mutant with lower ACE2 binding affinity. *bioRxiv.* Published online April 11, 2020:2020.04.09.034942.

Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* 30 (4), 772-780, 2013.

Lai, M. M. C. & Cavanagh, D. The Molecular Biology of Coronaviruses. in *Advances in Virus Research.* 48: 1–100 (Elsevier, 1997).

Li C, Yang Y, Ren L. Genetic evolution analysis of 2019 novel coronavirus and coronavirus from other species. *Infect Genet Evol.* 82:104285, 2020.

Li, MY, et al. "Expression of the SARS-CoV-2 cell receptor gene ACE2 in a wide variety of human tissues." *Infectious diseases of poverty,* 2020

Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet.* 395(10224):565-574, 2020

Batista, CT et al.

Ma Y, Wu L, Shaw N, et al. Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. *Proc Natl Acad Sci U S A*.112(30): 9436-9441, 2015.

Pachetti M, Marini B, Benedetti F, et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J Transl Med*. 18(1):179, 2020.

Phan T. Genetic diversity and evolution of SARS-CoV-2. *Infect Genet Evol*. 81:104260, 2020.

Phan T. Novel coronavirus: From discovery to clinical diagnostics. *Infect Genet Evol*. 79:104211, 2020.

Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 4(4):406-425, 1987.

Uddin M, Mustafa F, Rizvi TA, et al. SARS-CoV-2/COVID-19: Viral genomics, epidemiology, vaccines, and therapeutic interventions. *Viruses*. 12(5):526, 2020.

van Dorp L, Acman M, Richard D, et al. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infect Genet Evol*. 83: 104351, 2020.

Wang C, Liu Z, Chen Z, et al. The establishment of reference sequence for SARS-CoV-2 and variation analysis. *J Med Virol*. 92(6): 667-674, 2020.

Wang L, Wang Y, Ye D, Liu Q. Review of the 2019 novel coronavirus (SARS-CoV-2) based on current evidence. *Int J Antimicrob Agents*. 55(6):105948, 2020.

Woo PCY, Lau SKP, Chu C, et al. Characterization and Complete Genome Sequence of a Novel Coronavirus, Coronavirus HKU1, from Patients with Pneumonia. *J Virol*. 79(2):884-895, 2005.

Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 579(7798): 265-269, 2020.

Yadav PD, Potdar VA, Choudhary ML, et al. Full-genome sequences of the first two SARS-CoV-2 viruses from India. *Indian J. Med. Res*. 151 (2 e 3): 200–209, 2020.

Yu F, Du L, Ojcius DM, Pan C, Jiang S. Measures for diagnosing and treating infections by a novel coronavirus responsible for a pneumonia outbreak originating in Wuhan, China. *Microbes Infect*. 22(2):74-79, 2020.

Zehra Z, Luthra M, Siddiqui SM, Shamsi A, Gaur NA, Islam A. Corona virus versus existence of human on the earth: A computational and

biophysical approach. *Int J Biol Macromol*. 161:271-281, 2020.