Cross-reactivity and inclusivity analysis of CRISPR-based diagnostic assays of coronavirus SARS-CoV-2

Kashif Aziz Khan¹ and Marc-Olivier Duceppe²

¹ Department of Biology, York University, Toronto, ON, Canada
² Ottawa Laboratory Fallowfield, Canadian Food Inspection Agency, Ottawa, ON, Canada

ABSTRACT

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2; initially named as 2019-nCoV) is the cause of the novel coronavirus disease 2019 (COVID-19) pandemic. Its diagnosis relies on the molecular detection of the viral RNA by polymerase chain reaction (PCR) while newer rapid CRISPR-based diagnostic tools are being developed. As molecular diagnostic assays rely on the detection of unique sequences of viral nucleic acid, the target regions must be common to all coronavirus SARS-CoV-2 circulating strains, yet unique to SARS-CoV-2 with no cross-reactivity with the genome of the host and other normal or pathogenic organisms potentially present in the patient samples. This stage 1 protocol proposes in silico cross-reactivity and inclusivity analysis of the recently developed CRISPR-based diagnostic assays. Cross-reactivity will be analyzed through comparison of target regions with the genome sequence of the human, seven coronaviruses and 21 other organisms. Inclusivity analysis will be performed through the verification of the sequence variability within the target regions using publicly available SARS-CoV-2 sequences from around the world. The absence of cross-reactivity and any mutations in target regions of the assay used would provide a higher degree of confidence in the CRISPR-based diagnostic tests being developed while the presence could help guide the assay development efforts. We believe that this study would provide potentially important information for clinicians, researchers, and decision-makers.

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2; initially named as 2019-nCoV) was firstly isolated from a cluster of pneumonia patients in Wuhan, China and is the cause of novel coronavirus disease termed COVID-19 (Wu et al., 2020; Zhou et al., 2020; Zhu et al., 2020). The rapid spread of the virus has resulted in a declaration of a global pandemic by the World Health Organization (WHO) reaching more than 220 countries and territories (WHO, 2020c). SARS-CoV-2 has been classified as a member of the family Coronaviridae in the genus Betacoronavirus along with SARS-CoV and the Middle East respiratory syndrome (MERS)-CoV (Gorbalenya et al., 2020). The sequencing
of the virus from patients early in the outbreak has shown that its single-stranded RNA genome is ~30 kb in size (Chan et al., 2020; Lu et al., 2020; Wu et al., 2020). The SARS-CoV-2 genome has been predicted to encode at least 10 open reading frames (ORFs) for structural and accessory proteins, based on similarity with SARS-CoV. As per current annotation (NC_045512.2), these viral ORFs encode replicase ORF1ab, spike (S), envelope (E), membrane (M), nucleocapsid (N), and at least six accessory proteins (3a, 6, 7a, 7b, 8, and 10) (NCBI, 2020). The pandemic has serious public health and economic implications. The day-to-day life of billions of people has been affected due to different forms of social distancing measures in place in different parts of the world to mitigate the spread of the virus. Thus, the widespread availability of rapid and reliable diagnostic testing is an important tool for policymakers to make public health decisions. The current diagnosis of COVID-19 relies on the molecular detection of viral RNA from patient samples using nucleic acid amplification tests (NAAT) like polymerase chain reaction (PCR) (WHO, 2020b). However, PCR requires specialized equipment and trained staff to perform the test and interpret the results, and thus is a challenge for remote low-resource settings. One of the alternatives being explored is the CRISPR-based nucleic-acid detection methods that may be particularly useful for screening outside the laboratory setting, for example at point-of-care, airports, offices and homes.

CRISPR-Cas (clustered regularly interspaced short palindromic repeats-CRISPR-associated), a component of the bacterial immune system to infectious nucleic acid, has been widely used as a gene-editing tool. This technology exploits the ability of Cas proteins to accurately target any region in DNA in association with CRISPR RNA (crRNA) that matches the target DNA with or without the requirement of a protospacer adjacent motif (PAM) (Moon et al., 2019). Initially explored for Cas9 protein (Pardee et al., 2016), the application of CRISPR in nucleic acid detection emerged as a viable tool with the discovery of promiscuous collateral cleavage activity of Cas12a (formerly Cpf1), Cas12b (formerly C2c1) and Cas13a (formerly C2c2) after target recognition (Chen et al., 2018; Gootenberg et al., 2017). Several CRISPR-based methods have been developed for the detection of RNA and DNA viruses (Jia et al., 2020; Strich, Chertow & Kraft, 2019). With the emergence of the novel coronavirus, scientists are rapidly employing these tools for the detection of SARS-CoV-2 from patient samples as an alternative to PCR. The Cas12a has been used for the diagnosis of COVID-19 from patient samples targeting viral genes N and E (Broughton et al., 2020). Similarly, Cas12b, Cas13a and FuCas9-based assays have also been developed for the detection of SARS-CoV-2 (Ackerman et al., 2020; Azhar et al., 2021; Guo et al., 2020). Cas12a-based DETECTR and Cas13a-based SHERLOCK have been approved by the US Food and Drug Administration (FDA) under Emergency Use Authorization (Mammoth, 2020; Sherlock, 2020) and FuCas9-based FELUDA has been approved for diagnosis of COVID-19 in India (Mitra, 2020). Several other CRISPR-based methods are also under development (Petrillo et al., 2020).

The molecular diagnosis of SARS-CoV-2 may be jeopardized by potential preanalytical and analytical vulnerabilities leading to false-positive or false-negative results (Lippi, Simundic & Plebani, 2020). As molecular diagnostic assays rely on the detection of unique
sequences of viral nucleic acid, these are prone to mismatches due to genetic variability in the viral genome as well as cross-reactivity with the nucleic acid of other organisms present in the samples. The selectivity of an assay is generally validated in a laboratory using target strains, near-neighbour strains and other organisms. The use of bioinformatics tools and genome sequence databases can help to reduce wet-lab testing to a narrower focus and help to estimate more accurately the false-positive and false-negative rates of an assay (SantaLucia et al., 2020). It is known that mutations at primer/probe binding regions of the viral genome can result in potential mismatches and false-negative PCR diagnoses (Lefever et al., 2013; Stadhouders et al., 2010; Whiley & Sloots, 2005). We and others have concurrently demonstrated the genetic variability in the primer/probe binding regions of the SARS-CoV-2 genome highlighting the importance of periodic sequence verification for optimal virus detection (Farkas et al., 2020; Khan & Cheung, 2020; Osorio & Correia-Neves, 2020). Assay specificity remains a focus area in CRISPR-diagnostics as Cas proteins can result in a false-positive diagnosis due to their intrinsic capacity of mismatch tolerance. This risk has been minimized with the newer Cas proteins, Cas12a, Cas12b and Cas13, that have a lower tolerance for mismatches compared to front-runner Cas9 especially in the “seed” region (Safari et al., 2019). However, this raises the possibility that these tests may miss certain viral variants due to genetic variability in the regions targeted by these assays. The mismatch intolerant seed region of ~6 nucleotides is located in the PAM-proximal region for Cas12a (Chen et al., 2018; Kim et al., 2016) while the seed region is located in the central region of crRNA for Cas13a (Abudayyeh et al., 2017; Cox et al., 2017). Francisella novicida Cas9 (FnCas9) has been reported to have higher specificity and lower tolerance for mismatches compared to Streptococcus pyogenes Cas9 (SpCas9) tolerating only a single mismatch especially at the PAM-distal seed end (Acharya et al., 2019). It has been shown that even two mismatches would impede or even abolish the trans-cleavage activity of Alicyclobacillus acidophilus Cas12b (AaCas12b/AapCas12b) (Teng et al., 2019). It is important to design crRNA adjacent to Cas-specific PAM that is common to all SARS-CoV-2 strains, yet unique to SARS-CoV-2 coronavirus with no cross-reactivity with the genome of the host and other normal or pathogenic organisms potentially present in the patient samples.

The objective of this study is the verification of the cross-reactivity and sequence variability within the target regions of CRISPR-based COVID-19 diagnostic assays, using publicly available sequence databases. The absence of any cross-reactivity and mutations in target regions of the assay used would provide a higher degree of confidence in the alternative tests being developed while the presence of mutations could help guide assay development efforts. We believe that this study would provide important information for clinicians, researchers and policy-makers.

**METHODS**

**CRISPR-based diagnostic assays and SARS-CoV-2 sequences**

At least 15 crRNA of recently published CRISPR-based methods will be selected based on the literature review. The cross-reactivity and sequence variability within the target regions of CRISPR-based diagnostic assays will be determined using the protocol described...
below. The design planner is included in Table S1. The source code is available from the GitHub repository (https://github.com/duceppeemo/CRISPR_Assay_Tester) and is easily installable with the Conda package manager. The script will be validated (and updated if necessary) as per the method described earlier (Khan & Cheung, 2020).

SARS-CoV-2 genome sequences deposited by laboratories around the world will be obtained from the NCBI virus database (Hatcher et al., 2017). A total of 400,000 near full-length sequences will be downloaded by applying the complete filter. The RNA genome of SARS-CoV-2 is shown in DNA format as per scientific convention. The complete genome of Wuhan-Hu-1 isolate, which is 29,903 bp long, will be included as a reference (NCBI accession number: NC_045512.2).

*The number of sequences in the NCBI database is growing on a daily basis and the exact number of included sequences would be updated in the 2nd stage submission.

Cross-reactivity analysis

Each crRNA sequence along with the PAM sequence, if applicable, (Table 1) will be analyzed for reactivity with the genome sequence of the human, seven coronaviruses and 21 other species including normal or pathogenic organisms that may potentially be present in patient samples. The complete list to be tested can be found in Table S2. This step will be performed using GGGenome nucleotide sequence search online server (http://gggenome.dbcls.jp/) (Naito et al., 2015) allowing several mismatches to check for ≥80% homology according to the requirement of WHO’s Emergency Use Listing for in vitro diagnostics detecting SARS-CoV-2 nucleic acid (WHO, 2020a). The potential hits on both orientations with different numbers of mismatches for each crRNA will be returned along with a summary of number of hits for different organisms. The hits with a match in the seed and the PAM region will be discussed.

Inclusivity analysis

Multiple sequence alignment (MSA) of SARS-CoV-2 sequences will be performed using MAFFT (Multiple Alignment with Fast Fourier Transform) program v7.480 (Katoh & Standley, 2013) excluding low coverage sequences (>1% Ns) and using the Wuhan-Hu-1 genome (NC_045512.2) as reference. To evaluate the sequence variability in the target regions of each assay, referred here as the region of interest (ROI), the crRNA and PAM sequence will be extracted from all the entries in the MAFFT alignment file using the coordinates determined during cross-reactivity analysis. The MSA sequence for each ROI will be stratified by segregating sequences into discrete groups of identical sequence
variants along with their frequency. To remove extremely low prevalent variants and sequencing errors in the data, only the sequence variants occurring in more than 0.01% of all sequences will be further considered by default. Sequences with ambiguous nucleotides and stretches of Ns in ROIs will be excluded from the analysis. The summary of mutated nucleotides for each target region will be returned and results will be reported as the frequency of hits with 100% match, hits with mismatches, and sequences below threshold. The frequency of sequence variants with mismatches in the seed and PAM region will be discussed.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding
The authors received no funding for this work.

Competing Interests
The author declare that he have no competing interests.

Author Contributions
- Kashif Aziz Khan conceived and designed the experiments, performed the experiments, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Marc-Olivier Duceppe performed the experiments, authored or reviewed drafts of the paper, wrote the python script, and approved the final draft.

Data Availability
The following information was supplied regarding data availability:

  Python script is available at GitHub: https://github.com/duceppemo/CRISPR_Assay_Tester.

Supplemental Information
Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.12050#supplemental-information.

REFERENCES
Abudayyeh OO, Gootenberg JS, Esveltzichler P, Han S, Joung J, Belanto JJ, Verdine V, Cox DBT, Kellner MJ, Regev A, Lander ES, Voytas DF, Ting AY, Zhang F. 2017. RNA targeting with CRISPR-Cas13. Nature 550(7675):280–284 DOI 10.1038/nature24049.

Acharya S, Mishra A, Paul D, Ansari AH, Azhar M, Kumar M, Rauthan R, Sharma N, Aich M, Sinha D, Sharma S, Jain S, Ray A, Jain S, Ramalingam S, Maiti S, Chakraborty D. 2019. Francisella novicida Cas9 interrogates genomic DNA with very high specificity and can be used for mammalian genome editing. Proceedings of the National Academy of Sciences of the United States of America 116(42):20959–20968 DOI 10.1073/pnas.1818461116.

Ackerman CM, Myhrvold C, Thakku SG, Freije CA, Metsky HC, Yang DK, Ye SH, Boehm CK, Kosoko-Thoroddsen TF, Kehe J, Nguyen TG, Carter A, Kulesa A, Barnes JR, Dugan VG,
Hung DT, Blainey PC, Sabeti PC. 2020. Massively multiplexed nucleic acid detection with Cas13. *Nature* 582(7811):277–282 DOI 10.1038/s41586-020-2279-8.

Azhar M, Phutela R, Kumar M, Ansari AH, Rauthan R, Gulati S, Sharma N, Sinha D, Sharma S, Singh S, Acharya S, Sarkar S, Paul D, Kathpalia P, Aich M, Sehgal P, Ranjan G, Bhoyar RC, Singhal K, Lad H, Patra PK, Makharia G, Chandak GR, Pesala B, Chakraborty D, Maiti S. 2021. Rapid and accurate nucleobase detection using FnCas9 and its application in COVID-19 diagnosis. *Biosensors & Bioelectronics* 183:113207 DOI 10.1016/j.bios.2021.113207.

Broughton JP, Deng X, Yu G, Fasching CL, Servellita V, Singh J, Miao X, Streithorst JA, Granados A, Soto-Mayor-Gonzalez A, Zorn K, Gopez A, Hsu E, Gu W, Miller S, Pan C-Y, Guevara H, Wadford DA, Chen JS, Chiu CY. 2020. CRISPR–Cas12-based detection of SARS-CoV-2. *Nature Biotechnology* 38(7):870–874 DOI 10.1038/s41587-020-0513-4.

Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, Yuen KY. 2020. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerging Microbes & Infections* 9(1):221–236 DOI 10.1080/22221751.2020.1719902.

Chen JS, Ma E, Harrington LB, Da Costa M, Tian X, Palefsky JM, Doudna JA. 2018. CRISPR–Cas13a target binding unleashes indiscriminate single-stranded DNase activity. *Science* 360(6387):436–439 DOI 10.1126/science.aar6245.

Cox DBT, Gootenberg JS, Abudayyeh OO, Franklin B, Kellner MJ, Joung J, Zhang F. 2017. RNA editing with CRISPR-Cas13. *Science* 358(6366):1019–1027 DOI 10.1126/science.aap1080.

Farkas C, Fuentes-Villalobos F, Garrido JL, Haigh J, Barría MI. 2020. Insights on early mutational events in SARS-CoV-2 virus reveal founder effects across geographical regions. *PeerJ* 8:e9255 DOI 10.7717/peerj.9255.

Gootenberg JS, Abudayyeh OO, Lee JW, Essletzbichler P, Dy AJ, Joung J, Verdin V, Donghia N, Daringer NM, Freije CA, Myhrvold C, Bhattacharyya RP, Livny J, Regev A, Koonin EV, Hung DT, Sabeti PC, Collins JJ, Zhang F. 2017. Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* 356(6336):438–442 DOI 10.1126/science.aam9321.

Gorbalenya AE, Baker SC, Baric RS, RjD Groot, Drosten C, Gulyaeva AA, Haagmans BL, Lauber C, Leontovich AM, Neuman BW, Penzar D, Perlman S, Poon LLM, Samborskiy DV, Sidorov IA, Sola I, Ziebuhr J, Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. 2020. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology* 5:536–544 DOI 10.1038/s41564-020-0695-z.

Guo L, Sun X, Wang X, Liang C, Jiang H, Gao Q, Dai M, Qu B, Fang S, Mao Y, Chen Y, Feng G, Gu Q, Wang RR, Zhou Q, Li W. 2020. SARS-CoV-2 detection with CRISPR diagnostics. *Cell Discovery* 6(1):536 DOI 10.1038/s41421-020-0174-y.

Hatcher EL, Zhdanov SA, Bao Y, Blinkova O, Nawrocki EP, Ostapchuck Y, Schaffer AA, Brister JR. 2017. Virus variation resource—improved response to emergent viral outbreaks. *Nucleic Acids Research* 45(D1):D482–D490 DOI 10.1093/nar/gkv1065.

Jia F, Li X, Zhang C, Tang X. 2020. The expanded development and application of CRISPR system for sensitive nucleotide detection. *Protein & Cell* 11(9):624–629 DOI 10.1007/s13238-020-00708-8.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30(4):772–780 DOI 10.1093/molbev/mst010.
Khan KA, Cheung P. 2020. Presence of mismatches between diagnostic PCR assays and coronavirus SARS-CoV-2 genome. Royal Society Open Science 7(6):200636 DOI 10.1098/rsos.200636.

Kim D, Kim J, Hur JK, Been KW, S-H Yoon, Kim J-S. 2016. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. Nature Biotechnology 34(8):863–868 DOI 10.1038/nbt.3609.

Lefever S, Pattyn F, Hellemans J, Vandesompele J. 2013. Single-nucleotide polymorphisms and other mismatches reduce performance of quantitative PCR assays. Clinical Chemistry 59(10):1470–1480 DOI 10.1373/clinchem.2013.203653.

Lippi G, Simundic AM, Plebani M. 2020. Potential preanalytical and analytical vulnerabilities in the laboratory diagnosis of coronavirus disease 2019 (COVID-19). Clinical Chemistry and Laboratory Medicine 58(7):1070–1076 DOI 10.1515/cclm-2020-0285.

Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, Ma X, Zhan F, Wang L, Hu T, Zhou H, Hu Z, Zhou W, Zhao L, Chen J, Meng Y, Wang J, Lin Y, Yuan J, Xie Z, Ma J, Liu WJ, Wang D, Xu W, Holmes EC, Gao GF, Wu G, Chen W, Shi W, Tan W. 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 395(10224):565–574 DOI 10.1016/S0140-6736(20)30251-8.

Mammoth. 2020. SARS-CoV-2 DETECTR reagent kit. Available at https://www.fda.gov/media/141765/download (accessed 3 September 2020).

Mitra E. 2020. India’s drug authority approved paper-strip covid-19 test that could return results within hour. Atlanta: CNN.

Moon SB, Kim DY, Ko J-H, Kim Y-S. 2019. Recent advances in the CRISPR genome editing tool set. Experimental & Molecular Medicine 51(11):1–11 DOI 10.1038/s22276-019-0339-7.

Naito Y, Hino K, Bono H, Ui-Tei K. 2015. CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. Bioinformatics 31(7):1120–1123 DOI 10.1093/bioinformatics/btu743.

NCBI. 2020. Sequence viewer 3.36.0: severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome. Available at https://www.ncbi.nlm.nih.gov/projects/sviewer/?id=NC_045512.2 (accessed 15 May 2020).

Osorio NS, Correia-Neves M. 2020. Implication of SARS-CoV-2 evolution in the sensitivity of RT-qPCR diagnostic assays. Lancet Infectious Diseases 21(2):166–167 DOI 10.1016/S1473-3099(20)30435-7.

Pardee K, Green AA, Takahashi MK, Braff D, Lambert G, Lee JW, Ferrante T, Ma D, Donghia N, Fan M, Daringer NM, Bosch I, Dudley DM, O’Connor DH, Gehrke L, Collins JJ. 2016. Rapid, low-cost detection of Zika virus using programmable biomolecular components. Cell 165(5):1255–1266 DOI 10.1016/j.cell.2016.04.059.

Petrillo M, Querci M, Tkachenko O, Siska IR, Ben E, Angers-Loustau A, Bogni A, Brunetto A, Fabbri M, Garlant L, Lievens A, Munoz A, Paracchini V, Piotretti D, Puertas-Gallardo A, Raffael B, Sarno E, Tregoat V, Zaro F, Van den Eede G. 2020. The EU one-stop-shop collection of publicly available information on COVID-19 in vitro diagnostic medical devices. F1000Research 9:1296 DOI 10.12688/f1000research.27308.1.

Safari F, Zare K, Negahdaripour M, Barekat-Mowahed M, Ghasemi Y. 2019. CRISPR Cpf1 proteins: structure, function and implications for genome editing. Cell & Bioscience 9:36 DOI 10.1186/s13578-019-0298-7.

SantaLucia J, Sozhamannan S, Gans JD, Koehler JW, Soong R, Lin NJ, Xie G, Olson V, Roth K, Beck L. 2020. Appendix Q: recommendations for developing molecular assays for microbial
pathogen detection using modern in silico approaches. *Journal of AOAC International* 103(4):882–899 DOI 10.1093/jaoacint/qsaa045.

**Sherlock. 2020.** Sherlock CRISPR SARS-CoV-2 kit. Available at https://www.fda.gov/media/137746/download (accessed 3 July 2020).

**Stadhouders R, Pas SD, Anber J, Voermans J, Mes TH, Schutten M. 2010.** The effect of primer-template mismatches on the detection and quantification of nucleic acids using the 5' nuclease assay. *Journal of Molecular Diagnostics* 12(1):109–117 DOI 10.2353/jmoldx.2010.090035.

**Strich JR, Chertow DS, Kraft CS. 2019.** CRISPR-Cas biology and its application to infectious diseases. *Journal of Clinical Microbiology* 57(4):e01307-18 DOI 10.1128/jcm.01307-18.

**Teng F, Guo L, Cui T, Wang X-G, Xu K, Gao Q, Zhou Q, Li W. 2019.** CDetection: CRISPR-Cas12b-based DNA detection with sub-attomolar sensitivity and single-base specificity. *Genome Biology* 20(1):132 DOI 10.1186/s13059-019-1742-z.

**WHO. 2020a.** Instructions and requirements for Emergency Use Listing (EUL) submission: In vitro diagnostics detecting SARS-CoV-2 nucleic acid and rapid diagnostics tests detecting SARS-CoV-2 antigens. Available at https://www.who.int/publications/m/item/PQDx-347-version-4 (accessed 31 December 2020).

**WHO. 2020b.** Laboratory testing for coronavirus disease (COVID-19) in suspected human cases: Interim Guidance. Available at https://www.who.int/publications/i/item/10665-331501 (accessed 3 July 2020).

**WHO. 2020c.** WHO coronavirus disease (COVID-19) dashboard. Available at https://covid19.who.int/ (accessed 31 December 2020).

**Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ. 2020.** A new coronavirus associated with human respiratory disease in China. *Nature* 579(7798):265–269 DOI 10.1038/s41586-020-2008-3.

**Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. 2020.** A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579(7798):270–273 DOI 10.1038/s41586-020-2012-7.

**Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W. 2020.** A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine* 382:727–733 DOI 10.1056/NEJMoa2001017.