 doi:10.20944/preprints202002.0385.v1

Rapid design of a bait capture platform for culture- and amplification-free next-generation sequencing of SARS-CoV-2

Jalees A. Nasir[1,2], David J. Speicher[1,2], Rob A. Kozak[3], Hendrik N. Poinar[1,4], Matthew S. Miller[1,2,5], Andrew G. McArthur[1,2,#]

[1] Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, Ontario, Canada

[2] Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario, Canada

[3] Division of Microbiology, Department of Laboratory Medicine and Molecular Diagnostics, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada

[4] McMaster Ancient DNA Centre, Department of Anthropology and Biochemistry, McMaster University, Hamilton, Ontario, Canada

[5] McMaster Immunology Research Centre, McMaster University, Hamilton, Ontario, Canada

Address correspondence to Andrew G. McArthur, mcarthua@mcmaster.ca

Keywords: Coronavirus, SARS-CoV-2, enrichment, next-generation sequencing

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2020 doi:10.20944/preprints202002.0385.v1

Abstract

SARS-CoV-2 is a novel betacoronavirus and the aetiological agent of the current COVID-19 outbreak that originated in Hubei Province, China. While polymerase chain reaction is the front-line tool for SARS-CoV-2 surveillance, application of amplification-free and culture-free methods for isolation of SARS-CoV-2 RNA, partnered with next-generation sequencing, would provide a useful tool for both surveillance and research of SARS-CoV-2. We here release into the public domain a set of bait capture hybridization probe sequences for enrichment of SARS-CoV-2 RNA from complex biological samples. These probe sequences have been designed using rigorous bioinformatics methods to provide sensitivity, accuracy, and minimal off-target hybridization. Probe design was based on existing, validated approaches for detecting antimicrobial resistance genes in complex samples and it is our hope that this SARS-CoV-2 bait capture platform, once validated by those with samples in hand, will be of aid in combating the current outbreak.

Introduction

In late 2019, a novel coronavirus subsequently coined “SARS-CoV-2” (synonymous with “2019 novel coronavirus” and “2019-nCoV”) was identified as the aetiological agent of an outbreak of febrile respiratory illness, COVID-19, possibly associated with the Huanan South China Seafood Market in Wuhan, Hubei Province, China (1, 2). Since the initial outbreak, clinical symptoms have ranged from mild to severe pneumonia with the disease spreading through human-to-human transmission (1). Coronaviruses (CoVs) infect humans and other animals and are large (120-160 nm), roughly spherical, enveloped viruses, which carry a non-segmented positive-sense-strand RNA genome ~30 kb in length. There are four human coronaviruses (HCoVs) that circulate seasonally and cause mild to acute upper and lower respiratory infections, namely HCoV-OC43, HCoV-229E, HCoV-NL63, and HCoV-HKU1(3). However, in the past decade two zoonotic CoVs have emerged from bats, infected an intermediate host, and then caused severe disease in humans: Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) and Middle East Respiratory Syndrome Coronavirus (MERS-CoV) (4). SARS-CoV emerged in Guangdong, China in 2002 and infected 8,098 people in 29 countries with a case fatality of 9.6%. MERS-CoV originated in Saudi Arabia and transmitted from camels to 2,494 humans with case fatality of 34%. While cases of MERS-CoV continue to appear sporadically, there has not been a case of SARS-CoV since the

end of the pandemic in 2004. SARS-CoV-2, like SARS-CoV, is a betacoronavirus of group 2B that forms a clade within the subgenus Sarbecovirus (5). As SARS-CoV-2 is 89% identical to two bat SARS-like CoVs (ZC45 and ZC21), it is presumed that SARS-CoV-2 is another zoonotic emergence with the intermediate host currently unknown (6). However, as SARS-CoV-2 is an RNA virus, mutation rates are high (approximately 10^{-4} nucleotide substitutions per site per year) and whole-genome surveillance is critical for proper molecular epidemiology (i.e., tracking the outbreak) (6). In order to perform phylogenetic (forensic) analysis, the virus needs to be first isolated from a sample, but as viremia levels can vary dramatically, depending on what stage of the infection, sequencing the virus from complex metagenomic pools can be expensive and time consuming. Currently the method of choice is a bronchoalveolar lavage using cell culture, either human airway epithelial cell culture, Huh7 or VeroE6, requiring both time and expertise (2, 6). In order to minimize the risk of handling infectious viral cultures, preserve critical sample volumes, reduce labour and turn-around time, perform quick phylogenetic analysis, and to enable new scientific enquiry, a reliable method for both culture- and amplification-free enrichment of SARS-CoV-2 RNA from complex samples containing a mixture of host and bacterial DNA and RNA would be of value for next-generation sequencing (NGS) workflows. In recent years, hybridization bait capture methods for enrichment of viral targets in metagenomics samples have been designed and validated for a range of applications (7-11). Our own team members have used bait capture to isolate and sequence the aetiological agent of the 'Black Death', *Yersinia pestis*, (12), and have recently designed and validated an accurate and sensitive enrichment platform for NGS detection of antimicrobial resistance genes in complex biological samples based on our Comprehensive Antibiotic Resistance Database (13, 14). In the last year, we have been combining our expertise in bioinformatics (Nasir, McArthur), viral molecular & clinical epidemiology (Speicher, Kozak, Miller), and bait capture (Poinar) to design a sensitive and accurate bait capture platform for all known viruses involved in human respiratory diseases, with particular emphasis on suppression of off-target hybridization. While our algorithms were naive about SARS-CoV-2 itself, our design workflow included all other known coronavirus genome sequences and *in silico* analysis illustrated it could be effective for enrichment of SARS-CoV-2 RNA. As such, we here describe and release into the public domain a set of bait capture probe sequences for enrichment of SARS-CoV-2 RNA from complex biological samples. Given the unknown severity of the current SARS-CoV-2 outbreak and associated illness and mortality of COVID-19, we have opted to release our SARS-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2020 doi:10.20944/preprints202002.0385.v1

CoV-2 bait capture platform without experimental validation, in part relying on past success of similar bait capture design approaches and in part hopeful that it can be rapidly validated by those with samples in hand. We hope our platform proves useful for those combatting and researching the current SARS-CoV-2 outbreak.

Methods

Design

Using BaitsTools (v1.6.2) software (15), we designed 80 nucleotides (nt) hybridization probes by tiling with an offset of 20 nt across all SARS-CoV-2 sequences available at the National Center for Biotechnology Information (NCBI) prior to February 5, 2020 (n=37; complete genomes = 21, partial coding sequence = 16). Built-in BaitsTools functions were used to remove probe sequences that were incomplete or contained incorrect nucleotides. We also used BaitsTools to remove probes with GC content <25% or >55%, leaving 30,853 probes remaining. Melting temperature (T_m) was predicted using the OligoArrayAux function melt.pl (settings, -n RNA -t 65 -C $1.89e^{-9}$) and used to remove probes with a T_m <55°C or >105°C (16). To prevent off-target hybridization between the probes and any non-viral DNA or RNA, the candidate set of probes was compared against GenBank's nucleotide database using high-throughput BLASTN (default settings) (13, 17). Probes with high-scoring segment pairs (HSPs) >50 nt and high sequence similarity (>80%) to non-viral targets were discarded. Finally, the candidate set of hybridization probes was compared against itself through BLASTN analysis and a custom filter applied to mitigate the number of redundant probes sharing overlapping sequence space (discarding one member of non-identical pairs with >60 nt alignment and >80% sequence similarity), resulting in a candidate list of 1,310 bait capture hybridization probes for SARS-CoV-2.

Assessment

To predict the efficacy of capture by the candidate probes a Bowtie2 alignment (settings, bowtie2 --end-to-end -N 1 '-L 32' -a) (18) was performed to align the set of 1,310 probes to 21 complete SARS-CoV-2 genome sequences, with the resulting alignment file analyzed using SAMtools (19). Statistics to determine the number of instances that a probe mapped to a section of a SARS-CoV-2 genome, the length coverage of probes across known genomes, and the depth

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2020 doi:10.20944/preprints202002.0385.v1

of coverage by probes for each genome were calculated by adapting the Next Generation Sequencing Capture Assessment Tool (ngsCAT) in Python 3.6.8 (20). The GC content for each probe was calculated using the GCcontent.py Python3 script available at https://gist.github.com/wdecoster and the melting temperature predicted using melt.pl, as described above. An individual Bowtie2 alignment was performed comparing the candidate probe set against a single SARS-CoV-2 genome (www.ncbi.nlm.nih.gov/nuccore/MN908947.3), analyzed using SAMtools, and visualized using JBrowse (21). Plots were generated using R (v3.5.1) (22) or using the adapted ngsCAT software. To further predict efficacy of capture, a Bowtie2 alignment was performed to compare the candidate set of probes against all SARS-CoV-2 genomes uploaded to the Global Initiative on Sharing All Influenza Data (GISAID) database (n=96 complete genomes) (23) prior to February 13, 2020, as well as all other coronavirus genomes, particularly SARS-CoV and MERS-CoV, with analysis of alignments performed as described above.

Availability

Software, high resolution copies of images, and the complete set of SARS-CoV-2 hybridization probe sequences are available at https://github.com/jaleezyy/covid-19-baits.

Results

Our proposed SARS-CoV-2 bait capture platform of 1,310 80-mer nucleotide bait capture hybridization probes was custom designed and analyzed *in silico* through alignment analyses. The probes span the assembled complete genomes for SARS-CoV-2 with physical properties restricted to 25-55% GC content and T_m between 55°C-105°C (Figure 1). The number of aligned probes per SARS-CoV-2 genome averaged 1,284. However, one genome (GenBank accession LR757997.1) was an outlier with only ~60% of probes aligning. With this outlier removed, the number of aligned probes per SARS-CoV-2 genome averaged 1,309 (Figure 2). Although we did not investigate the poor mapping for the outlier genome LR757997, it was the only genome exhibiting poor mapping amongst the NCBI (n=37) and GISAID genomes (n=96) (plots available as Supplementary data at the GitHub repository).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2020 doi:10.20944/preprints202002.0385.v1

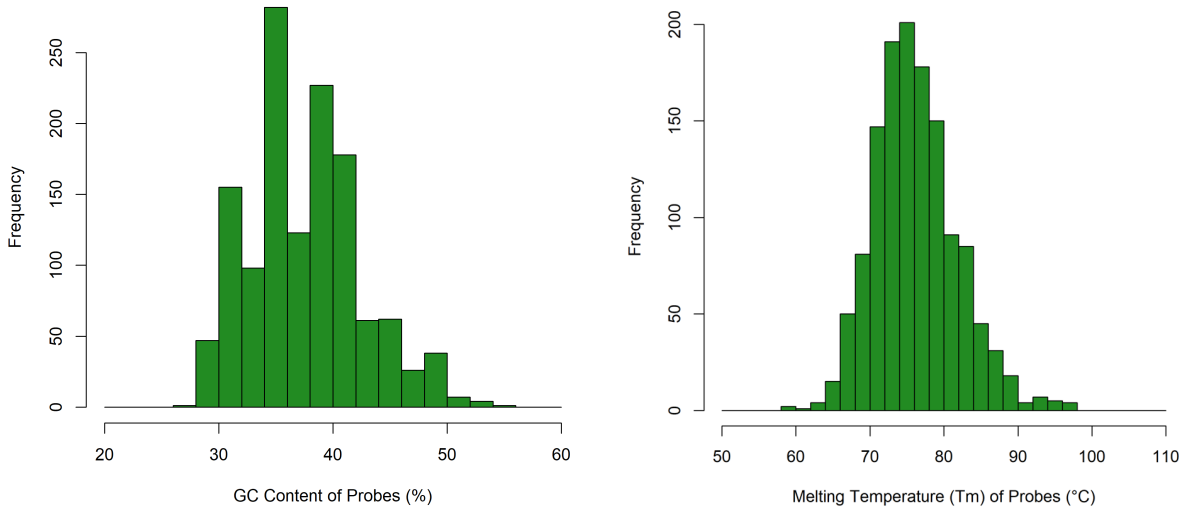

Figure 1: Preliminary Candidate Probe Set Statistics. *The GC content (left) and predicted melting temperatures (right) of the final list of probes. The average GC content of the reference SARS-CoV-2 genomes used (NCBI, n=37) was 37.5%.*

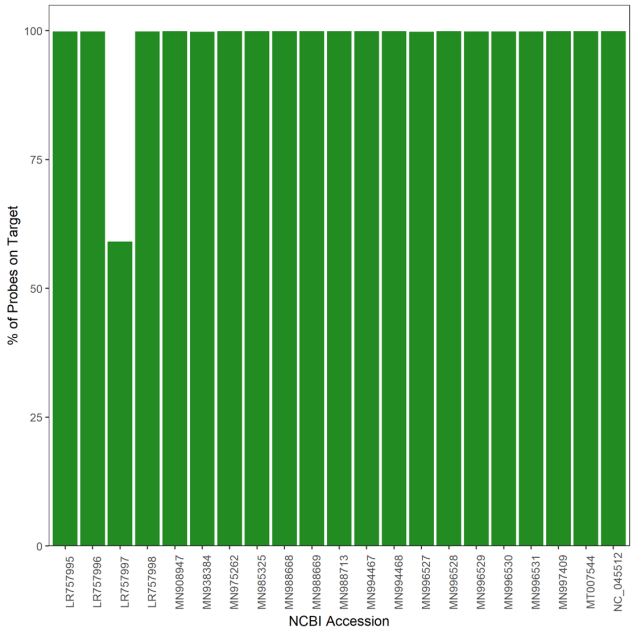

Figure 2: Summary of alignment of candidate probes to SARS-CoV-2 genomes.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2020 doi:10.20944/preprints202002.0385.v1

Across possible target regions within 21 SARS-CoV-2 genomes, 64% of all targeted regions are covered by 4 or more probes (Figure 3). Overall, 406,169 out of 626,874 possible nucleotides are covered by 4 probes or more (Figure 4).

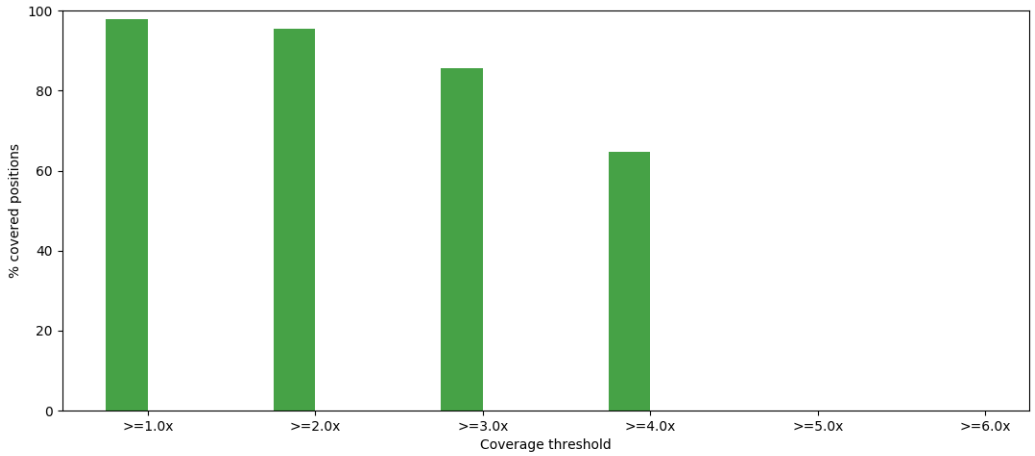

Figure 3: Coverage plot of probes aligned to SARS-CoV-2 genomes*. Percentage of target nucleotide positions covered as a function of coverage threshold. Over 60% of nucleotides in the SARS-CoV-2 genomes were covered by 4 or more probes.*

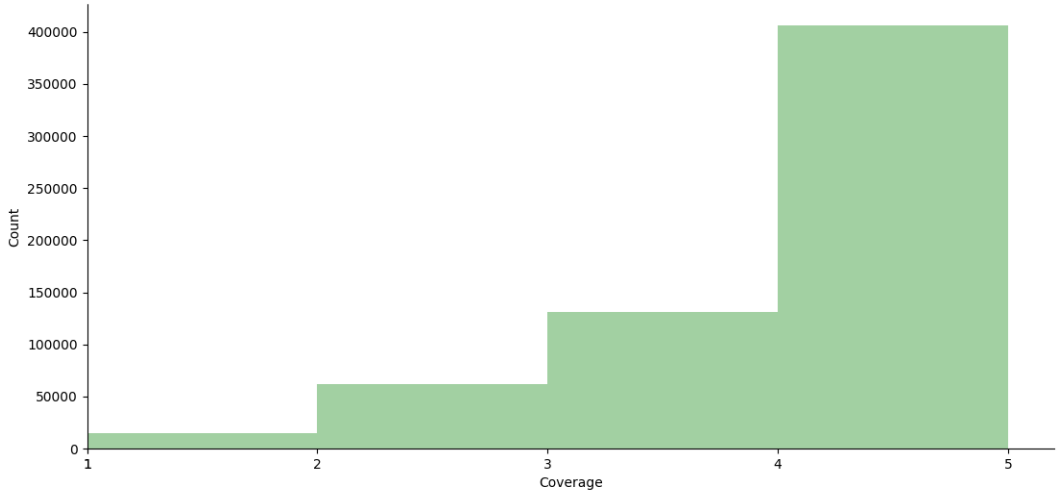

Figure 4: Uniformity plot denoting depth of coverage of candidate probes across all possible nucleotide targets from the NCBI collection of reference SARS-CoV-2 genomes* (n=37, summing 630,282 nt). The y-axis gives the number of nucleotides for each coverage category and the majority of target bases have >4x coverage by probes.*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2020 doi:10.20944/preprints202002.0385.v1

Our probe set consistently displayed 3.4x coverage across all genomes (Figure 5). A drop in coverage occurs at approximately the 15 kB region of ORF1ab, but retains ~3x coverage. Candidate probes were visualized against one SARS-CoV-2 genome, MN908947.3, highlighting uniform coverage across the genome and visualizing the regions where drops in coverage reside. (Figure 6). Additional SARS-CoV-2 genomes were downloaded from GISAID and 3.5x average coverage was observed across all 96 genomes, with the lowest depth of coverage seen at the 3'-UTR (plots available as Supplementary data at the GitHub repository).

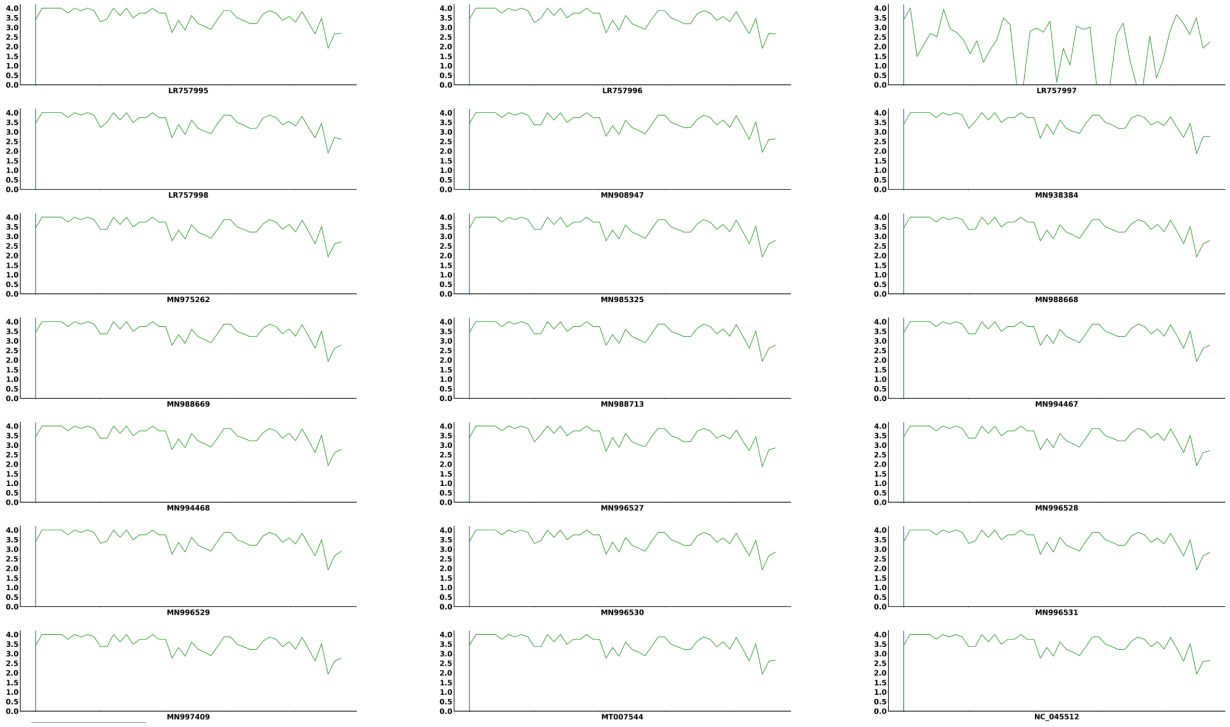

Figure 5: Coverage plots of candidate probes across known SARS-CoV-2 complete genomes*. The x-axis represents the base position on the genome denoted by the accession. The y-axis ranges from 0 to 4 fold coverage (high resolution versions available at the GitHub repository). As mentioned above, LR757997 (top right) was an outlier.*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2020 doi:10.20944/preprints202002.0385.v1

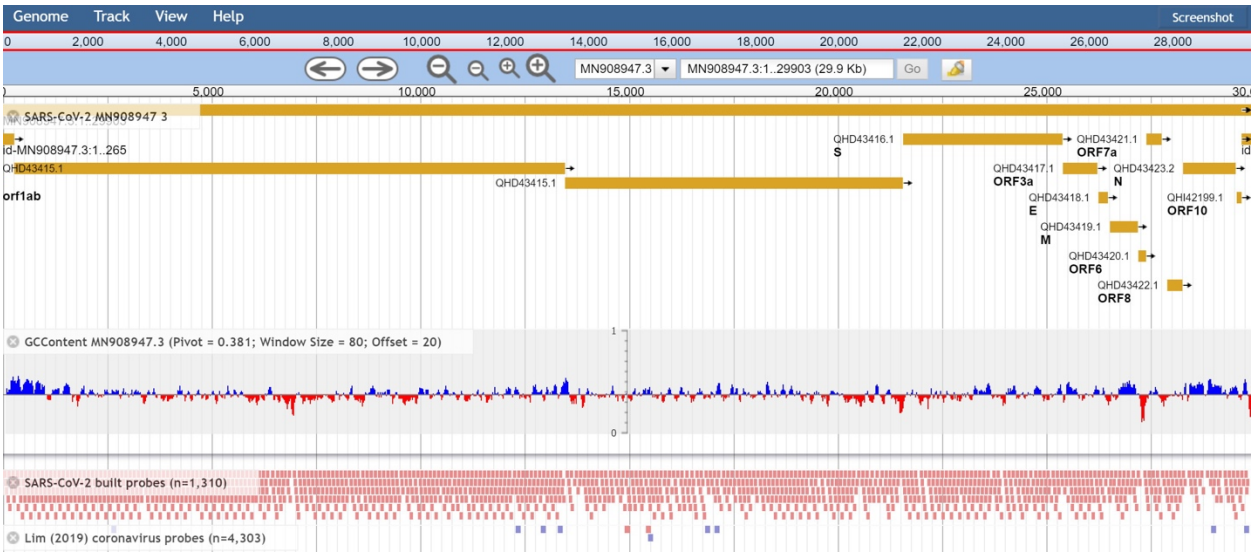

Figure 6: Alignment of candidate probes against a single SARS-CoV-2 genome (MN908947.3). *Top track: MN908947.3 GenBank annotation. Second track: MN908947.3 GC content between 0 and 100%. Third track: Alignment of candidate probes against MN908947.3. Bottom track: Alignment of CoV bait capture probes from Lim et al. 2019 (Journal of General Virology 100:1363–1374) against MN908947.3 - these probes were designed prior to knowledge of SARS-CoV-2 genome sequence.*

As the above workflow did not explicitly examine cross-hybridization of SARS-CoV-2 probes with other coronavirus, particularly SARS-CoV and MERS-CoV, we additionally aligned the 1,310 candidate probes against all other available coronavirus sequences using Bowtie2, as described above. Alignment of the candidate probe set against all known coronavirus genome sequences revealed 117 probe sequences aligning to SARS-CoV (with the most coverage at the 15kb region encoding ORF1ab), none aligning to MERS-CoV, and 62 and 1 aligning to bat coronavirus BM48-31 (NC_014470.1) and bat Hp-betacoronavirus (NC_025217.1), respectively. Additionally, no alignments were found against seasonal human coronaviruses HCoV-OC43, HCoV-229E, HCoV-NL63, and HCoV-HKU1.

Discussion

The growing SARS-CoV-2 outbreak will require continued sequencing efforts to collect information pertaining to this novel coronavirus. At present, in order to sequence SARS-CoV-2, there is a culturing step required that increases the risk to laboratory staff, making the analysis of

 doi:10.20944/preprints202002.0385.v1

SARS-CoV-2 difficult in resource-constrained countries without the facilities to culture the virus. Bait capture of SARS-CoV-2 followed by NGS using small Nanopore sequencers provides a simpler alternative (24). Targeted SARS-CoV-2 RNA enrichment promotes enrichment through subtraction, by physically separating target RNA from a complex patient sample, resulting in most sequenced fragments being on-target and thus reducing required metagenomics sequencing effort. This added benefit of lowered sequencing volume in turn also reduces overall turn-around time and cost.

While probe sets for targeted enrichment of a broad range of viruses already exist (7-10), we propose a probe set designed specifically in response to the current SARS-CoV-2 outbreak. For probe sets based on a broad diversity of viruses there is concern for off-target hybridization, resulting in off-target enrichment and added time and cost to the workflow (7, 8). Probe design improvements in detecting rare DNA pertaining to antimicrobial resistance highlight the importance considering off-target hybridization and balancing length and depth of coverage when designing probes for specific targets (13). Our probe design aims to maximize specificity and sensitivity to SARS-CoV-2 by removing candidate probes that could hybridize to human or other eukaryotic, bacterial, or archaeal RNA or DNA. Yet, *in silico* alignment tools do not accurately reflect hybridization in solution and it is entirely possible that our proposed probe set may cross-hybridize with other coronaviruses, such as SARS-CoV and MERS-CoV, or may have unanticipated off-target hybridization with unrelated viruses, bacterial members of the microbiome, or mammalian host DNA or RNA. While co-infection with multiple coronaviruses are rare, possible off-target hybridization with non-coronavirus DNA or RNA is an important concern (3). While our design reflects optimal methods, validated for bait capture of antimicrobial resistance (AMR) genes (13), experimental validation of our proposed SARS-CoV-2 bait capture platform is required. To this aim, we note that for our antimicrobial resistance gene bait capture platform we had 80 nt biotinylated ssRNA probes synthesized by Arbor Biosciences (Ann Arbor, MI) using the custom myBaitsR kit and that our protocol for capture of AMR DNA is described by Guitor *et al*. and is available at the Comprehensive Antibiotic Resistance Database, https://card.mcmaster.ca/download.

Acknowledgements

J.A.N. was supported by funds from the Comprehensive Antibiotic Resistance Database. D.J.S. was supported by McMaster University's Michael G. DeGroote Initiative for Innovation in Healthcare. H.N.P. was supported by funding from the Natural Sciences and Engineering Research Council of Canada (NSERC), Social Sciences and Humanities Research Council of Canada (SSHRC), and a Boris Family Fund. M.S.M. was supported in part by a New Investigator Award from the Canadian Institutes for Health Research and an Early Researcher Award from the Ontario Ministry of Research, Innovation and Science. Computer resources were supplied by the McMaster Service Lab and Repository computing cluster, funded in part by grants to A.G.M. from the Canadian Foundation for Innovation. Additional cloud computing needs were funded by the Comprehensive Antibiotic Resistance Database. We gratefully acknowledge the authors, the originating and submitting laboratories for their sequence data shared through GISAID, upon which this research is based.

References

1. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B. 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* doi:10.1016/S0140-6736(20)30183-5.
2. 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, China Novel Coronavirus I, Research T. 2020. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* doi:10.1056/NEJMoa2001017.
3. Mackay IM, Arden KE, Speicher DJ, O'Neil NT, McErlean PK, Greer RM, Nissen MD, Sloots TP. 2012. Co-circulation of four human coronaviruses (HCoVs) in Queensland children with acute respiratory tract illnesses in 2004. *Viruses* 4:637-53.
4. Gralinski LE, Menachery VD. 2020. Return of the Coronavirus: 2019-nCoV. *Viruses* 12.

5. Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, Haagmans BL, Lauber C, Leontovich AM, Neuman BW, Penzar D, Perlman S, Poon LLM, Samborskiy D, Sidorov IA, Sola I, Ziebuhr J. 2020. Severe acute respiratory syndrome-related coronavirus: The species and its viruses – a statement of the Coronavirus Study Group. *bioRxiv* doi:10.1101/2020.02.07.937862:2020.02.07.937862.

6. 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* doi:10.1016/S0140-6736(20)30251-8.

7. Li B, Si HR, Zhu Y, Yang XL, Anderson DE, Shi ZL, Wang LF, Zhou P. 2020. Discovery of bat coronaviruses through surveillance and probe capture-based next-generation gequencing. *mSphere* 5: e00807-19.

8. Briese T, Kapoor A, Mishra N, Jain K, Kumar A, Jabado OJ, Lipkin WI. 2015. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *mBio* 6:e01491-15.

9. Chalkias S, Gorham JM, Mazaika E, Parfenov M, Dang X, DePalma S, McKean D, Seidman CE, Seidman JG, Koralnik IJ. 2018. ViroFind: A novel target-enrichment deep-sequencing platform reveals a complex JC virus population in the brain of PML patients. *PLoS One* 13:e0186945.

10. Metsky HC, Siddle KJ, Gladden-Young A, Qu J, Yang DK, Brehio P, Goldfarb A, Piantadosi A, Wohl S, Carter A, Lin AE, Barnes KG, Tully DC, Corleis B, Hennigan S, Barbosa-Lima G, Vieira YR, Paul LM, Tan AL, Garcia KF, Parham LA, Odia I, Eromon P, Folarin OA, Goba A, Viral Hemorrhagic Fever C, Simon-Loriere E, Hensley L, Balmaseda A, Harris E, Kwon DS, Allen TM, Runstadler JA, Smole S, Bozza FA, Souza TML, Isern S, Michael SF, Lorenzana I, Gehrke L, Bosch I, Ebel G, Grant DS, Happi CT, Park DJ, Gnirke A, Sabeti PC, Matranga CB. 2019. Capturing sequence diversity in metagenomes with comprehensive and scalable probe design. *Nat Biotechnol* 37:160-168.

11. Bonsall D, Ansari MA, Ip C, Trebes A, Brown A, Klenerman P, Buck D, Consortium S-H, Piazza P, Barnes E, Bowden R. 2015. ve-SEQ: Robust, unbiased enrichment for streamlined detection and whole-genome sequencing of HCV and other highly diverse pathogens. *F1000Res* 4:1062.
12. Bos KI, Schuenemann VJ, Golding GB, Burbano HA, Waglechner N, Coombes BK, McPhee JB, DeWitte SN, Meyer M, Schmedes S, Wood J, Earn DJ, Herring DA, Bauer P, Poinar HN, Krause J. 2011. A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* 478:506-10.
13. Guitor AK, Raphenya AR, Klunk J, Kuch M, Alcock B, Surette MG, McArthur AG, Poinar HN, Wright GD. 2019. Capturing the resistome: a targeted capture method to reveal antibiotic resistance determinants in metagenomes. *Antimicrob Agents Chemother* 64:e01324-19.
14. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen AV, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran HK, Werfalli RE, Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-Koutoucheva A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar GV, McArthur AG. 2020. CARD 2020: Antibiotic resistome surveillance with the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res* 48:D517-D525.
15. Campana MG. 2018. BaitsTools: Software for hybridization capture bait design. *Mol Ecol Resour* 18:356-361.
16. Rouillard JM, Zuker M, Gulari E. 2003. OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Res* 31:3057-62.
17. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.
18. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357-9.
19. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078-9.

20. Lopez-Domingo FJ, Florido JP, Rueda A, Dopazo J, Santoyo-Lopez J. 2014. ngsCAT: a tool to assess the efficiency of targeted enrichment sequencing. *Bioinformatics* 30:1767-8.
21. Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH. 2009. JBrowse: a next-generation genome browser. *Genome Res* 19:1630-8.
22. R Core Team. 2018. R: A language and environment for statistical computing., R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
23. Shu Y, McCauley J. 2017. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill* 22:30494.
24. Oxford Nanopore Technologies. 2020. ARTIC Network provides protocol for rapid, accurate sequencing of novel coronavirus (nCoV-2019): first genomes released. https://nanoporetech.com/about-us/news/artic-network-provides-protocol-rapid-accurate-sequencing-novel-coronavirus-ncov-2019. Accessed 2020/02/14.