



Short Communication

Explorative assessment of coronavirus-like short sequences from host-associated and environmental metagenomes



Maximilian Mora^{a,1}, Wisnu Adi Wicaksono^{a,1}, Dilfuza Egamberdieva^b, Robert Krause^{c,d}, Jose Luis Martinez^e, Tomislav Cernava^{a,*,2}, Gabriele Berg^{a,d,*,*,2}

^a Graz University of Technology, Institute of Environmental Biotechnology, Graz, Austria

^b National University of Uzbekistan, Faculty of Biology, Tashkent, Uzbekistan

^c Medical University of Graz, Division of Infectious Diseases, Department of Internal Medicine, Graz, Austria

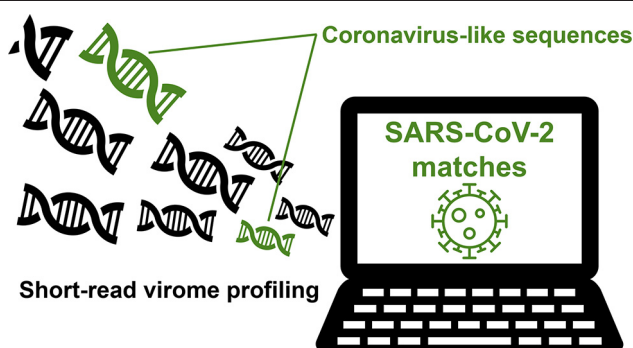
^d BioTechMed-Graz, Graz, Austria

^e Centro Nacional de Biotecnología, CSIC, Madrid, Spain

HIGHLIGHTS

- Coronavirus-like sequences can be retrieved from complex metagenomes.
- The short sequences are retrievable from DNA datasets that are free of RNA viruses.
- Their occurrence may explain incorrect detection of SARS-CoV-2.
- Processing parameters should be carefully considered for virome analyses.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 22 April 2021

Received in revised form 10 June 2021

Accepted 13 June 2021

Available online 24 June 2021

Editor: Damia Barcelo

Keywords:

Coronaviruses

COVID-19

SARS-CoV-2

ABSTRACT

The ongoing COVID-19 pandemic has not only globally caused a high number of causalities, but is also an unprecedented challenge for scientists. False-positive virus detection tests not only aggravate the situation in the healthcare sector, but also provide ground for speculations. Previous studies have highlighted the importance of software choice and data interpretation in virome studies. We aimed to further expand theoretical and practical knowledge in bioinformatics-driven virome studies by focusing on short, virus-like DNA sequences in metagenomic data. Analyses of datasets obtained from different sample types (terrestrial, animal and human related samples) and origins showed that coronavirus-like sequences have existed in host-associated and environmental samples before the current COVID-19 pandemic. In the analyzed datasets, various *Betacoronavirus*-like sequences were detected that also included SARS-CoV-2 matches. Deepening analyses indicated that the detected sequences are not of viral origin and thus should not be considered in virome profiling approaches. Our study confirms the importance of parameter selection, especially in terms of read length, for reliable virome profiling. Natural environments are an important source of coronavirus-like nucleotide sequences that should be

Abbreviations: COVID-19, CoronaVirus-Disease-2019; LCA, lowest common ancestor; SARS-CoV, severe acute respiratory syndrome coronavirus; MEM, maximum exact match; MERS, Middle East respiratory syndrome–coronavirus.

* Corresponding author.

** Correspondence to: G. Berg, BioTechMed-Graz, Graz, Austria.

E-mail addresses: tomislav.cernava@tugraz.at (T. Cernava), gabriele.berg@tugraz.at (G. Berg).

¹ These authors contributed equally.

² These authors share senior authorship.

taken into account when virome datasets are analyzed and interpreted. We therefore suggest that processing parameters are carefully selected for SARS-CoV-2 profiling in host related as well as environmental samples in order to avoid incorrect identifications.

© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The worldwide pandemic of COVID-19 (CoronaVirus-Disease-2019) due to SARS-CoV-2 started in 2019, and has since affected more than 170 million humans and caused close to 4 million deaths (WHO). Both numbers were still rising at the time point of concluding this study. It is commonly accepted that SARS-CoV-2 (also known as hCoV19 and 2019-nCoV), the pathogen behind COVID-19, is a zoonotic virus which originated from bats (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020; Zhou et al., 2020) and underwent most likely at least one more recombination event in another intermediate host, before being able to infect humans on a large scale (Ji et al., 2020). A similar mechanism has been already described for two other severe human pathogenic coronavirus outbreaks, SARS-CoV and MERS-CoV, for which civets (Childs et al., 2007) and dromedary camels (Azhar et al., 2014) were identified as intermediate hosts respectively. Regarding SARS-CoV-2, the discussion about possible intermediate hosts is still ongoing and so far, snakes (Ji et al., 2020) and pangolins (Zhang et al., 2020) have been suggested. The occurrence of a specific mutation in the receptor binding domain of the spike protein is considered crucial, which consequently allows the virus to dock on human cells as has been already excellently reviewed elsewhere (Andersen et al., 2020). In addition to the unclear origin of SARS-CoV-2, the health sector had been drastically challenged by a scarcity of reliable detection methods. Indeed, the genome of SARS-CoV-2 is very similar to those of other coronaviruses, particularly to SARS-CoV. Consequently, false-positive detections of coronaviruses might complicate epidemiological studies when One-Health approaches are implemented. Detection methods in patients are considered as accurate enough, because they rarely harbor a variety of viruses that can render false positives. However, virus detection via molecular and bioinformatic approaches is particularly problematic in natural ecosystems, containing complex microbiomes, where viruses can be abundant and diverse (Handley and Virgin, 2019). For such applications, short-read high-throughput sequencing allows to perform viral community profiling, detection of rare and even novel viruses (Cheval et al., 2011; Wylie et al., 2012; Zhang et al., 2021). In addition to homology-based annotation approaches, such as BLASTn, several bioinformatics tools including Kraken2 and Kaiju have been established to facilitate rapid microbial community profiling that also includes the virome (Menzel et al., 2016; Wood et al., 2019). On the other hand, foregoing approaches have demonstrated the importance of assembly strategies for short sequencing reads to facilitate reliable virome analyses (Sutton et al., 2019). In the context of the current pandemic, detection of coronavirus-like sequences, not necessarily belonging to SARS-CoV-2, may compromise the study of the natural distribution of this virus. In this short report, we aim to raise awareness of coronavirus-like sequences in metagenomic data. Our data supports the hypothesis that short virus-like DNA sequences of non-viral origin are an important reason for various uncertainties connected to incorrect identification of SARS-CoV-2 in different ecosystems.

2. Methods

We searched for coronavirus-like short sequences in a variety of metagenomic samples, i.e. terrestrial, animal and human related samples. Moreover, we used a combination of in-house and previously published metagenomic datasets (Table S1) that were so far not associated with any coronavirus studies. The viral taxonomy of high-quality reads

was assessed using two well-established taxonomic classifiers for metagenome and metatranscriptome studies, namely Kraken2 (Wood et al., 2019) and Kaiju (Menzel et al., 2016). Kraken2 uses a *k*-mer-based approach to classify individual metagenome reads by mapping all *k*-mers to the lowest common ancestor (LCA) of all genomes containing the given *k*-mer, whereas Kaiju searches for maximum exact matches on the protein-level using the Burrows–Wheeler transform. Kraken2 analysis was performed using default parameters against the Kraken viral database that contains RefSeq complete viral genomes/proteins (retrieved on 18.12.2020) with addition of 18 *Betacoronavirus* complete genomes that are publicly available (Table S2). Kaiju was performed using MEM (maximum exact match) mode against the NCBI RefSeq database (retrieved on 22.05.2020). Furthermore, we systematically aligned the metagenomic datasets (assembled contigs and merged paired end reads) against a list of primers and probes designed for severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1 (SARS-CoV-2) detection by different internationally renowned institutes and published on the WHO website (Table S3) and SARS-CoV-2 isolate Wuhan-Hu-1 complete genome (GenBank Accession no: MN908947.3) using BLASTn (with blastn-short for primer alignment) (Camacho et al., 2009). Degenerate primers were aligned separately for each possible nucleotide exchange. Prior to alignment, we used MEGAHIT v1.2.9 with meta-sensitive parameters (Li et al., 2015) to generate assembled metagenomic contigs. Alignment of short reads and contigs against the SARS-CoV-2 isolate Wuhan-Hu-1 complete genome were visualized with the Integrated Genomics Viewer (IGV) tool (version 2.4.10) (Thorvaldsdóttir et al., 2013). Subsequently, we performed BLASTx searches with read pairs that were tentatively identified as SARS-CoV-2 against the non-redundant protein sequence (nr) NCBI database ((Sayers et al., 2020) <https://blast.ncbi.nlm.nih.gov>; updated: 2021/01/10).

3. Results and discussion

Only a small fraction of the datasets could be assigned to viruses using Kraken2 (average: 0.40%; min: 0.02%; max: 3.30%) and Kaiju MEM (average: 1.03%; min: 0.26%; max: 4.17%) metagenomic classifiers (Fig. 1a–b). Both metagenomic classifiers detected reads that were classified as *Coronaviridae* (Fig. 1c–d). When comparatively assessed, Kraken2 detected a lower number of reads that belong to *Coronaviridae* than Kaiju (24 and 5329 reads, respectively). Interestingly, human-associated coronaviruses, namely Human coronavirus HKU1, Human coronavirus 229E, Human coronavirus NL63, Middle East respiratory syndrome–coronavirus (MERS) and severe acute respiratory syndrome-related coronavirus (SARS-CoV) were detected in various samples. In detail, SARS-CoV was detected in two terrestrial and two animal related samples according to the Kraken2 analysis and eight terrestrial, five animal related, and five human related samples according to the Kaiju analysis (Fig. 1c–d). Reads that were classified as SARS-CoV-2 were detected in five terrestrial samples (AS1, AS10, DS11, DS12, and SM), two animal related samples (COW3 and GOAT3) and one human related sample (STO3). These results indicate that coronavirus-like short sequences can be retrieved from different metagenomes including human, animal, and environmental samples. The detected reads were further annotated using BLASTx searches against the NCBI nr database (update date: 2021/01/10) in order to confirm that they did not originate from coronaviruses, which are all known to be RNA viruses (Supplementary Data 1). Short read approaches facilitate rapid

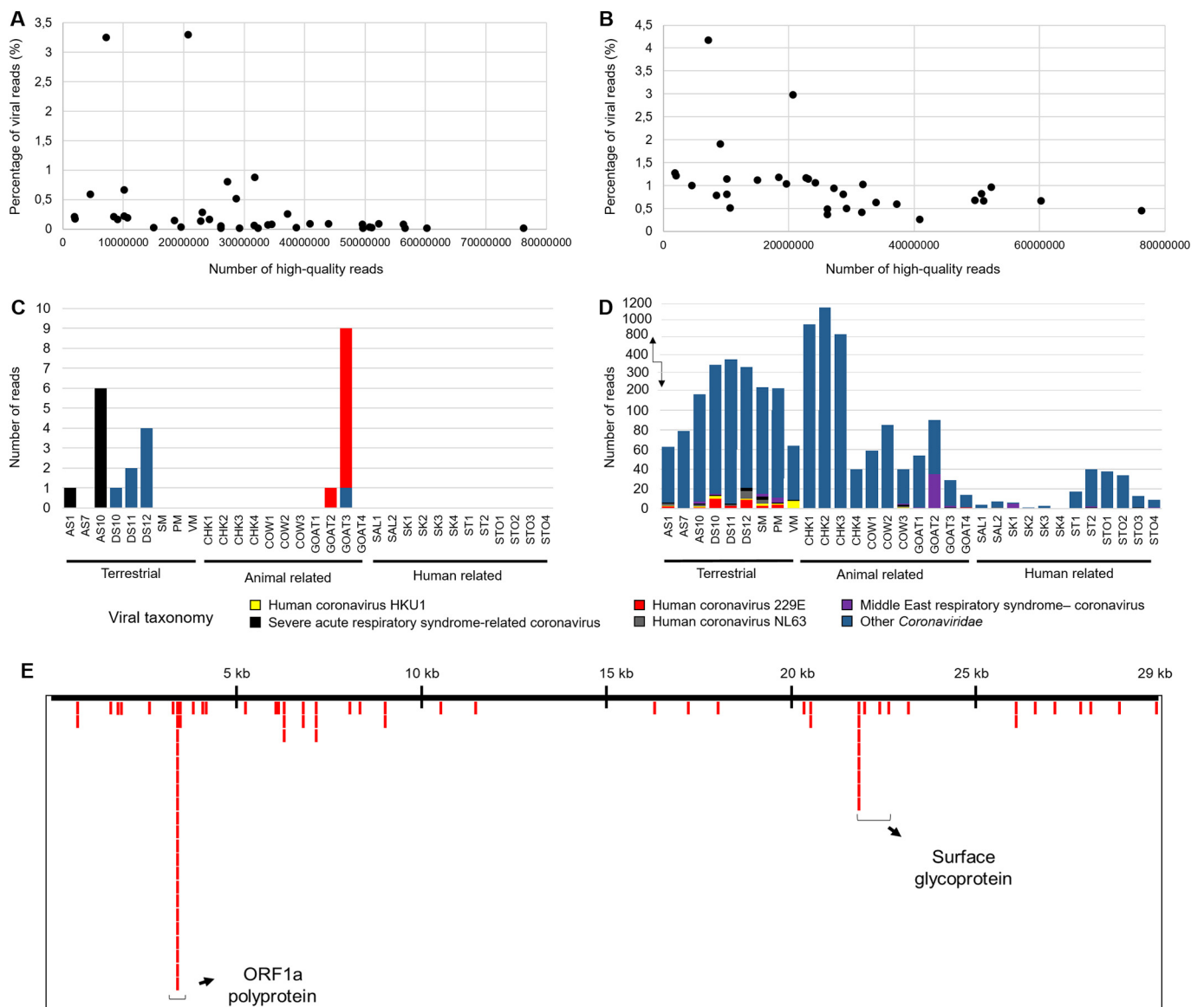


Fig. 1. Scatter plots and bar plots showing the number of viral (a and b) and *Coronaviridae* matches (c and d) in the analyzed metagenomes according to Kraken2 (a and c) and Kaiju (b and d) classifiers. Integrated Genomics Viewer (IGV) snapshot of the alignment of short reads and assembled contigs against the SARS-CoV-2 isolate Wuhan-Hu-1 complete genome (e).

but in-depth assessment of viral communities as well as their detection and surveillance (Nieuwenhuijse et al., 2020; Sabatier et al., 2020; Wylie et al., 2012; Zhang et al., 2021). Here, we want to emphasize that caution must be exercised when performing viral community profiling based on short reads due to the potential occurrence of incorrect identifications. Furthermore, we argue that while tools for short-read sequence alignments are fast, they may often not be sufficiently accurate.

By aligning short reads and contigs from the metagenomic datasets to the SARS-CoV-2 Wuhan-Hu-1 complete genome sequence (GenBank Accession no: MN908947.3), we detected a low number of matching short reads (minimum overlap of 18 bp and more than 90% identity; Fig. 1e; Table S4). The majority of the terrestrial samples included short-read alignments (8 of 9 samples) whereas no alignments were found for the human related samples. When contigs were assessed instead of short reads, only six could be aligned from the entire dataset (with low alignment coverage; 18–23 bp alignment length) to the SARS-CoV-2 Wuhan-Hu-1 complete genome (Table S5). Therefore, our analysis indicates that a minimum alignment length of 24 bp or 8 amino acids should be selected in order to avoid false identification of

coronaviruses in complex samples. The majority of the aligned reads were located at positions 3200–3500 and 21,800–22,600 of the SARS-CoV-2 Wuhan-Hu-1 complete genome, which are annotated as ORF1a polyprotein and surface glycoprotein, respectively (Fig. 1e). Following BLASTx searches with the whole contigs, it was also confirmed that they are not coronaviruses (Supplementary Data 1).

Subsequently, we analyzed if the metagenomic datasets harbor complementary regions to test primers that are currently in use for SARS-CoV-2 detection. Out of the 55 primer and probe sequences published by the WHO (World Health Organization, 2020; Table S3), eleven could be detected at least once in at least one of the two datasets (short reads and contigs) when one mismatch was allowed (Supplementary Data 2). In most cases only one primer of a primer pair was detected; the only case where forward and reverse primers of the same primer pair were detected, was with the primers nCoV_IP2-12669Fw and nCoV_IP2-12759Rv from Institute Pasteur, Paris, France and China_ORF1ab_F and China_ORF1ab_R from China CDC, China. They showed complementarity to the plant metagenomes of *Pinus mugo* (mountain pine) and *Vaccinium myrtillus* (blueberry) respectively (Supplementary Data 2). In both cases, the respective qPCR probe, which would also be required for a clear

positive test, could not be detected in any of the datasets. However, when the stringency was reduced to two mismatches, a complete set of primers and the corresponding probe (frlp_nCoV_IP4_14059Fw, frlp_nCoV_IP4_14146Rv and frlp_nCoV_IP4_14084Probe) from Institute Pasteur, Paris, France was detected in one plant metagenome, *Sphagnum magellanicum* (SM, peat moss; Supplementary Data 3). This was also the case for another primer-probe set (Jap_ORF1a_NIID_WH_1_F509, Jap_ORF1a_NIID_WH_1_R854, Jap_ORF1a_NIID_WH_1_Seq_F519 and Jap_ORF1a_NIID_WH_1_Seq_R840) from National Institute of Infectious Diseases, Japan, which was detected in one terrestrial sample obtained from the Aral Sea basin (AS10; Supplementary Data 3). In analogy to the short reads and contigs, also these reads were shown not to originate from coronaviruses (Supplementary Data 1). In previous studies, analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT-qPCR primer-probe sets were conducted (Vogels et al., 2020). The authors suggested that probe nucleotide mismatches can affect sensitivity of the PCR as viruses evolve during outbreaks, which may alter the probe binding regions. In our study, we would not expect the primer-probe hits to lead to positive test results, because these reads are not viral elements and thus would not result in the expected amplicon fragments (Supplementary Data 1). However, we highlight the importance of eliminating environmental DNA during RNA extraction for SARS-CoV-2 detection, as well as of the stringency of subsequent PCR conditions (especially annealing temperature) to avoid false-positive detections.

Considering that we found SARS-CoV-2 complete primer-probe set matches, coronavirus-like sequences in environmental RNA or contaminating DNA provide a possible explanation for the occurrence of false positive RT-PCR tests that are frequently reported and a major cause for uncertainties (Surkova et al., 2020). We assume that shotgun metagenomic/transcriptomic approaches will be increasingly conducted for early detection of SARS-CoV-2 and other virus outbreaks, in particular for wastewater surveillance (Kolarević et al., 2021; Venugopal et al., 2020) (or other environmental samples), and thus want to raise awareness of the potential occurrence of incorrect identification in complex datasets that can result from automated bioinformatic approaches based on short-read analyses.

4. Conclusions

We identified short nucleotide sequences with high similarity to coronaviruses in human, animal, and natural environments. They have a yet unknown origin and role in these ecosystems, which remains to be explored in the future. Nevertheless, we still want to stress that identification of short sequences with similarity to coronaviruses from metagenomic data is valuable. In case of insufficient removal of DNA prior to detection tests, the presence of these sequences might interfere with reliable SARS-CoV-2 detection, even if this could not be confirmed in the present study, as most primer-probe sets have so far only been validated against other known *Betacoronaviruses* and other viruses that cause respiratory diseases (e.g. Corman et al., 2020; Vogels et al., 2020). Due to the fact that SARS-CoV-2 has been detected already multiple times in wastewater samples, it has been proposed that wastewater surveillance might be a good strategy for early detection of SARS-CoV-2 outbreaks (Kolarević et al., 2021; Venugopal et al., 2020). Our results indicate that profiling of non-assembled virome datasets by standardized microbiome analysis workflows or detection of primer-probe matches for real-time PCR tests is not reliable enough to avoid false identifications of coronaviruses in complex microbiomes such as those present in soil or wastewater. RNA-centric approaches with complete exclusion of contaminating DNA in combination with standardized primers and PCR parameters will be required to elucidate the distribution of coronaviruses and to distinguish them from coronavirus-like sequences. Such ambiguities are less likely to occur when large reference databases are used (e.g. NCBI Reference Sequence (RefSeq) database and NCBI BLAST *nr* database) where potential hits are not

restricted to coronaviruses. However, when targeted searches with closed databases are required, the selection of representative sequence lengths (min. 24 bp or 8 amino acids) is crucial to avoid incorrect identifications. Results obtained from virome profiling should always be critically assessed; for example, when RNA viruses are detected in metagenomic datasets. Finally, we are confident that microbiome research can contribute to understand and predict further outbreaks and to implement required actions to avoid them, because viruses are part of the microbiome and their indigenous stability and plasticity is fundamental.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2021.148494>.

Funding

This work has not received external funding.

Authors' contributions

TC and GB conceived the idea for the study; DE organized and coordinated the sampling at the Aral Sea basin; MM performed laboratory experiments and analyzed data; MM, WAW, TC, and GB interpreted the data; MM, WAW, TC, and GB wrote and edited the manuscript with specific inputs from DE, RK and JLM. All authors read and approved the final manuscript.

CRedit authorship contribution statement

Maximilian Mora: Investigation, Writing – original draft. **Wisnu Adi Wicaksono:** Methodology, Data curation, Writing – original draft, Visualization. **Dilfuza Egamberdieva:** Investigation, Writing – original draft. **Robert Krause:** Writing – original draft. **Jose Luis Martinez:** Writing – original draft. **Tomislav Cernava:** Conceptualization, Investigation, Writing – original draft, Supervision, Visualization. **Gabriele Berg:** Conceptualization, Supervision, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We appreciate the help of Monika Schneider-Trampitsch, Barbara Fetz, and Julia Kranyecz during wet lab preparations of the samples. We also thank Darrin Higgins and users audrey and inDhika for providing icons on The Noun Project (www.thenounproject.com) that were implemented in the graphical abstract.

References

- Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., Garry, R.F., 2020. The proximal origin of SARS-CoV-2. *Nat. Med.* 26, 450–452.
- Azhar, E.I., El-Kafrawy, S.A., Farraj, S.A., Hassan, A.M., Al-Saeed, M.S., Hashem, A.M., Madani, T.A., 2014. Evidence for camel-to-human transmission of MERS coronavirus. *N. Engl. J. Med.* 370, 2499–2505.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421.
- Cheval, J., Sauvage, V., Frangeul, L., Dacheux, L., Guigon, G., Dumey, N., Pariente, K., Rousseaux, C., Dorange, F., Berthet, N., 2011. Evaluation of high-throughput sequencing for identifying known and unknown viruses in biological samples. *J. Clin. Microbiol.* 49, 3268–3275.
- Childs, J.E., Mackenzie, J.S., Richt, J.A., 2007. *Wildlife and Emerging Zoonotic Diseases: The Biology, Circumstances and Consequences of Cross-Species Transmission*. Springer Science & Business Media.
- Corman, V.M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D.K., Bleicker, T., Brünink, S., Schneider, J., Schmidt, M.L., 2020. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance* 25, 2000045.

- 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. *Nat. Microbiol.* 5, 536.
- Handley, S.A., Virgin, H.W., 2019. Drowning in viruses. *Cell* 177, 1084–1085.
- Ji, W., Wang, W., Zhao, X., Zai, J., Li, X., 2020. Cross-species transmission of the newly identified coronavirus 2019-nCoV. *J. Med. Virol.* 92, 433–440.
- Kolarević, S., Micsinai, A., Szántó-Egész, R., Lukács, A., Kračun-Kolarević, M., Lundy, L., Kirschner, A.K., Farnleitner, A.A., Djukic, A., Čolić, J., 2021. Detection of SARS-CoV-2 RNA in the Danube River in Serbia associated with the discharge of untreated wastewaters. *Sci. Total Environ.* 783, 146967.
- Li, D., Liu, C.-M., Luo, R., Sadakane, K., Lam, T.-W., 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676.
- Menzel, P., Ng, K.L., Krogh, A., 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* 7, 1–9.
- Nieuwenhuijse, D.F., Munnink, B.B.O., Phan, M.V., Munk, P., Venkatakrishnan, S., Aarestrup, F.M., Cotten, M., Koopmans, M.P., 2020. Setting a baseline for global urban virome surveillance in sewage. *Sci. Rep.* 10, 1–13.
- Sabatier, M., Bal, A., Destras, G., Regue, H., Quéromès, G., Cheynet, V., Lina, B., Bardel, C., Brengel-Pesce, K., Navratil, V., 2020. Comparison of nucleic acid extraction methods for a viral metagenomics analysis of respiratory viruses. *Microorganisms* 8, 1539.
- Sayers, E.W., Beck, J., Brister, J.R., Bolton, E.E., Canese, K., Comeau, D.C., Funk, K., Ketter, A., Kim, S., Kimchi, A., 2020. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 48, D9.
- Surkova, E., Nikolayevskyy, V., Drobniowski, F., 2020. False-positive COVID-19 results: hidden problems and costs. *Lancet Resp Med* 8, 1167–1168.
- Sutton, T.D., Clooney, A.G., Ryan, F.J., Ross, R.P., Hill, C., 2019. Choice of assembly software has a critical impact on virome characterisation. *Microbiome* 7, 12.
- Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* 14, 178–192.
- Venugopal, A., Ganesan, H., Raja, S.S.S., Govindasamy, V., Arunachalam, M., Narayanasamy, A., Sivaprakash, P., Rahman, P.K., Gopalakrishnan, A.V., Siama, Z., 2020. Novel wastewater surveillance strategy for early detection of COVID-19 hotspots. *Curr. Opin. Environ. Sci. Health.* 17, 8–13.
- Vogels, C.B., Brito, A.F., Wyllie, A.L., Fauver, J.R., Ott, I.M., Kalinich, C.C., Petrone, M.E., Casanovas-Massana, A., Muenker, M.C., Moore, A.J., 2020. Analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT-qPCR primer–probe sets. *Nat. Microbiol.* 5, 1299–1305.
- Wood, D.E., Lu, J., Langmead, B., 2019. Improved metagenomic analysis with kraken 2. *Genome Biol.* 20, 257.
- World Health Organization, 2020. *Molecular Assays to Diagnose COVID-19: Summary Table of Available Protocols.*
- Wylie, K.M., Mihindukulasuriya, K.A., Sodergren, E., Weinstock, G.M., Storch, G.A., 2012. Sequence analysis of the human virome in febrile and afebrile children. *PLoS One* 7, e27735.
- Zhang, T., Wu, Q., Zhang, Z., 2020. Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Curr. Biol.* 30, 1346–1351.
- Zhang, F., Zuo, T., Yeoh, Y.K., Cheng, F.W., Liu, Q., Tang, W., Cheung, K.C., Yang, K., Cheung, C.P., Mo, C.C., 2021. Longitudinal dynamics of gut bacteriome, mycobiome and virome after fecal microbiota transplantation in graft-versus-host disease. *Nat. Commun.* 12, 1–11.
- Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273.