Non-SARS-CoV-2 genome sequences identified in clinical samples from COVID-19 infected patients: Evidence for co-infections

Mohamed A Abouelkhair Corresp. 1

1 Department of Biomedical and Diagnostic Sciences College of Veterinary Medicine, University of Tennessee, Knoxville, Tennessee, United States

Corresponding Author: Mohamed A Abouelkhair
Email address: mabouelk@vols.utk.edu

Background

In December 2019, an ongoing outbreak of pneumonia caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2/ 2019-nCoV) infection was initially reported in Wuhan, Hubei Province, China. Early in 2020, the World Health Organization (WHO) announced a new name for the 2019-nCoV-caused disease: coronavirus disease 2019 (COVID-19) and declared COVID-19 to be a Public Health Emergency of International Concern (PHEIC). Cellular co-infection is a critical determinant of viral fitness and infection outcomes and plays a crucial role in shaping the host immune response to infections.

Methods

In this study, 68 public next-generation sequencing data from SARS-CoV-2 infected patients were retrieved from the NCBI Sequence Read Archive database using SRA-Toolkit. Data screening was performed using an alignment-free method based on *k*-mer mapping and extension, fastv. Taxonomic classification was performed using Kraken 2 on all reads containing one or more virus sequences other than SARS-CoV-2.

Results

SARS-CoV-2 was identified in all except three patients. Influenza type A (H7N9) virus, human immunodeficiency virus, rhabdovirus, human metapneumovirus, Human adenovirus, Human herpesvirus 1, coronavirus NL63, parvovirus, simian virus 40, and hepatitis virus genomes sequences were detected in SARS-CoV-2 infected patients. Besides, a very diverse group of bacterial populations were observed in the samples.

Non-SARS-CoV-2 genome sequences identified in clinical samples from COVID-19 infected patients: Evidence for co-infections

Mohamed A. Abouelkhair [1 #]

[1]Department of Biomedical and Diagnostic Sciences, University of Tennessee College of Veterinary Medicine, Knoxville, Tennessee, USA

[#] Corresponding Author:

Mohamed A. Abouelkhair

2407 River Dr, Knoxville, TN 37996, USA

Email address: mabouelk@vols.utk.edu

Abstract

Background. In December 2019, an ongoing outbreak of pneumonia caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2/ 2019-nCoV) infection was initially reported in Wuhan, Hubei Province, China. Early in 2020, the World Health Organization (WHO) announced a new name for the 2019-nCoV-caused disease: coronavirus disease 2019 (COVID-19) and declared COVID-19 to be a Public Health Emergency of International Concern (PHEIC). Cellular co-infection is a critical determinant of viral fitness and infection outcomes and plays a crucial role in shaping the host immune response to infections.

Methods. In this study, 68 public next-generation sequencing data from SARS-CoV-2 infected patients were retrieved from the NCBI Sequence Read Archive database using SRA-Toolkit. Data screening was performed using an alignment-free method based on *k*-mer mapping and extension, fastv. Taxonomic classification was performed using Kraken 2 on all reads containing one or more virus sequences other than SARS-CoV-2.

Results. SARS-CoV-2 was identified in all except three patients. Influenza type A (H7N9) virus, human immunodeficiency virus, rhabdovirus, human metapneumovirus, Human adenovirus, Human herpesvirus 1, coronavirus NL63, parvovirus, simian virus 40, and hepatitis virus genomes sequences were detected in SARS-CoV-2 infected patients. Besides, a very diverse group of bacterial populations were observed in the samples.

Keywords: COVID-19, Co-infections, Bacteria, Influenza A virus, Human immunodeficiency virus

Introduction

In December 2019, the first cases of coronavirus disease 2019 (COVID-19) were possibly due to a zoonotic transmission in China. It was tied to a large seafood market that also traded in live wild animals (Tay et al. 2020). The causative virus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is capable of human-to-human transmission and rapidly spread to other regions of China and other countries (World Health Organization 2020). It is now a global pandemic and is a considerable concern for public health. So far, more than 16,101,367 confirmed cases were diagnosed in nearly 213 countries and territories around the world and two international conveyances, causing globally over 645,000 deaths (Worldometer 2020, July 25).

Coronaviruses are known to cause severe diseases in humans and animals. Of these, four human coronaviruses (229E, NL63, OC43, and HKU1) typically only infect the upper respiratory tract and cause relatively minor symptoms (Fehr & Perlman 2015). However, there are three coronaviruses (severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV), and SARS-CoV-2) that can replicate in the lower respiratory tract and cause pneumonia, which can be fatal. With 79% genome sequence similarity, SARS-CoV is the closest relative to SARS-CoV-2 among human coronaviruses (Gorbalenya et al. 2020). However, of all known coronavirus sequences, SARS-CoV-2 is most similar to bat coronavirus RaTG13, with a similarity of 98% (Zhou et al. 2020).

SARS-CoV-2 pathophysiology closely parallels that of SARS-CoV infection, with active inflammatory responses strongly implicated in the resulting airway damage (Wong et al. 2004). Hence the extent of the disease in patients is attributed not only to the viral infection but also to the host's response (Tay et al. 2020).

Underlying co-infections in primary infectious disease are an important variable that needs to be considered but is often undetected. Remarkable developments in next-generation sequencing

have recently made metagenomics, an unbiased shotgun method of analysis, a widely used tool in just about every field of biology, including diagnosis of infectious diseases (Kuroda et al. 2012; Lecuit & Eloit 2014). Metagenomics is powerful because it can diagnose unsuspected microbial agents (Wilson et al. 2014). It directly analyzes samples in their entirety, eliminating the need for prior knowledge to obtain comprehensive information. In this capacity, metagenomics exceeds traditional diagnostic limitations. With the microbial genomes in hand, we can now explore the possibility of using metagenomic and metatranscriptomic next-generation sequencing (mNGS) directly as a screening method of other microbes in a sample (Plyusnin et al. 2020).

A simple approach would be to first map sequencing reads from the sample to the reference microbial genome. However, the accuracy of such an alignment-based method is relatively low compared to an alignment-free approach. In the alignment-based method, genome sequences from closely related viruses can lead to false-positive results (Chen et al. 2020b). Moreover, the virus-specific reads obtained may not be adequate for unambiguous detection (degraded RNA, incompletely target-enriched sequence library by multiple-PCR (Lundberg et al. 2013) or hybrid capture (Duncavage et al. 2011), which can lead to false-negative results.
Fastv is an alignment-free, ultra-fast tool for detecting the microbial sequences in sequence data (Chen et al. 2020b). It can identify target microorganisms using unique *k*-mers. It detects SARS and other coronaviruses from sequencing data and efficiently distinguishing SARS from MERS.

In this study, public next-generation sequencing data from SARS-CoV-2 infected patients were analyzed by fastv using the pre-computed unique *k*-mer resources (Chen et al. 2020b). Taxonomic classification was performed using Kraken 2 on all reads containing more than one

99 microbial sequence (Wood et al. 2019). The present study's findings have confirmed the actual
100 existence of genome sequences of other microbes in SARS-CoV-2 infected patients.
101

102 Material and Methods

103 Computing hardware

104 Amazon Elastic Compute Cloud (EC2) instance (i.e., virtual server in the AWS cloud) was used.
105 SRA-tools package version 2.9.1 (https://github.com/ncbi/sra-tools), Kraken 2
106 (https://github.com/DerrickWood/kraken2) (Wood et al. 2019), and fastv version 0.9.0
107 (https://github.com/OpenGene/fastv) (Chen et al. 2020b) were installed within the Linux 2 EC2
108 instance.
109

110 SRA database mining

111 Next-generation sequencing technologies have enabled large-scale genomic surveillance of
112 SARS-CoV-2 as thousands of isolates are being sequenced worldwide and deposited in public
113 data repositories. The sequence data were downloaded as .sra files using the prefetch tool
114 (https://github.com/ncbi/sra-tools/tree/master/tools/prefetch), then extracted to. fastq files using
115 the NCBI fastq-dump tool (https://github.com/ncbi/sra-tools/tree/master/tools/fastq-dump).
116 Data sets for analysis were chosen through a keyword search of the SRA descriptions for
117 "COVID19" and downloaded between 27 January 2020 and 16 May 2020. Sequence data from
118 negative COVID-19 patients, experimental studies, and controlled access were excluded.
119 **(Table S1)**.
120

121 Table S1: SRA sequences used in this study with the detection result for SARS-
122 CoV-2 *k*-mer

123

Sequence data pre-processing and screening using fastv

Fastv, along with the pre-computed unique *k*-mer resources, was used as previously described (Chen et al. 2020b). Initially, fastv performed adapter trimming, quality pruning, base correction, and other pre-processing to ensure the accuracy of *k*-mer analysis on fastq input files. The fastv tool identifies a target virus from sequencing data and detects any microbial sequences for which a unique *k*-mer data is provided. A unique SARS-CoV-2 *k*-mer set and the SARS-CoV-2 reference genome were used as input files (downloaded from https://github.com/OpenGene/fastv/tree/master/data) with the *k*-mer collections for viral and microbial genomes (downloaded from http://opengene.org/microbial.kc.fasta.gz). The *k*-mer scanning results were visualized in a figure on a single HTML page by fastv.

Sequence data analysis using Kraken 2

The results of fastv were validated with Kraken 2. Kraken 2 is the latest version of Kraken, a taxonomic classification system that uses exact *k*-mer matches to achieve high accuracy and rapid classification (https://github.com/DerrickWood/kraken2) (Wood et al. 2019). Kraken constructs an index of all *k*-mers found in the reference genomes and assigns each *k*-mer to the least common ancestor (LCA) of all species that have that *k*-mer. Then Kraken matches the *k*-*mer*s contained in the reads to this index and eventually assigns the reads to the taxon with the most fitting *k*-mers by following the path from the root of the tree (Wood et al. 2019). Through lowering memory use by 85%, Kraken 2 improves upon Kraken 1, enabling higher numbers of genomic reference data to be used while retaining high accuracy and fivefold speed. A standard database containing RefSeq complete bacterial, archaeal, and viral genomes, along with the human genome and a collection of known vectors (UniVec_Core), was downloaded (https://ccb.jhu.edu/software/kraken2/downloads.shtml). The database was then constructed using 32 threads with the default parameters on an AWS EC2 h1.8xlarge storage optimized instance with 16 dual-core hyperthreaded 2.30 GHz CPUs and 132 gigabytes (GB) of RAM.

150 **Data visualization**

151 The results of Kraken 2 analysis were visualized with Pavian
152 (https://github.com/fbreitwieser/pavian) (Breitwieser & Salzberg 2016) and Krona tool
153 (https://github.com/marbl/Krona/wiki) (Ondov et al. 2011), which displays hierarchical data (like
154 taxonomic assignation) in multi-layered pie charts. The Kraken 2 outputs were converted in
155 HTML format using the program ktImportTaxonomy
156 (https://github.com/marbl/Krona/tree/master/KronaTools/scripts), which parses the information
157 relative to the query ID and the taxonomy ID.

158

159 # Results

160 **SARS-CoV-2 Identification**

161 Sequence data analysis was performed on public Illumina HiSeq/ MiSeq libraries from the NCBI
162 SRA database (Bioproject PRJNA605983) sequenced from bronchoalveolar lavage fluid from
163 five patients (WIV02, WIV04, WIV05, WIV06, and WIV07) with pneumonia at the early COVID-
164 19 outbreak in Wuhan, China. Nine libraries were downloaded as .sra files using the prefetch
165 tool; then, the fastq files were extracted using the NCBI fastq-dump tool. Prior to the *k*-mer
166 analysis, sequencing adapters and low-quality bases were removed by fastv.
167 After scanning the fastq data, fastv reported the *k*-mer coverage for each microbial genome with
168 valid hits **(Figure 1)**.

169

170 **Figure 1: SARS-CoV-2 detection by fastv.** SARS-CoV-2 was detected in WIV07 patient,
171 where twelve SARS-CoV-2 hits were included in the genome list and ordered by k-mer
172 coverage (99.4972% to 99.2308%). The number of hits (on the y-axis) of each *k*-mer key (on
173 the x-axis) were plotted. Mismatches were highlighted in red.

SARS-CoV-2 was detected in all of the 68-sequence data except SRR11772662, SRR11772663, and SRR11772664 samples were negative for SARS-CoV-2. These samples belong to one study (Bioproject PRJNA631042), where the research group used different sequencing technologies on the same sample to find a cost-effective and highly scalable method for SARS-CoV-2 sequencing. Because sequence technologies vary in reading depth and coverage thresholds, fastv could not detect SARS-CoV-2 in sequenced samples with lower coverage metrics.

Non-SARS-CoV-2 genome sequences were detected in COVID-19 infected patients

Fastv also identified influenza type A (A/Shanghai/02/2013(H7N9) and rhabdovirus genomic sequences in the data from WIV02 (SRR11092058 and SRR11092063 data), WIV04 (SRR11092057 and SRR11092062 data), WIV05 (SRR11092061 data), WIV06 (SRR11092056 and SRR11092060 data), and WIV07 (SRR11092059 and SRR11092064 data) patients. Influenza type A segment 4 hemagglutinin (HA) gene was detected with high coverage (100%, 100%, 98.57%, 95.71%, and 100%) and mean depth (5.58, 16.52, 4.8, 2.95, and 881.37) in WIV02, WIV04, WIV05, WIV06, and WIV07 patients, respectively. Genes coding for Influenza A virus polymerase, non-structural proteins, matrix proteins 1 and 2 were also identified in the previously mentioned patients but with lower coverages. The genome sequence of rhabdovirus was detected with coverage (57.36, 72.73, 57.58, 67.75, and 53.46%) and mean depth (4.82, 8.13, 4.84, 5.08, and 4.06) in WIV02, WIV04, WIV05, WIV06, and WIV07 patients, respectively. The genome sequence of the Nipah virus was detected in WIV05, WIV06, and WIV07 patients. Infection with the Nipah virus in humans causes a number of clinical manifestations that range from asymptomatic (subclinical) infection to acute respiratory infection and fatal encephalitis. In addition, human immunodeficiency virus, human herpesvirus 1, human T-lymphotropic virus 1, hepatitis virus, and simian virus 40 were found in WIV07 patient sequence data with coverage

199 below 21%. The low coverage and non-human specific pathogen hits have been overlooked
200 **(Table S2)**.
201
202 Human coronavirus 229E genome sequence was detected (coverage=14% and mean
203 depth=1.55) in SRR11772654 (Bioproject PRJNA631042). Human adenovirus 5 sequences
204 were detected in SRR11772660, SRR11772663, SRR11772666, SRR11772672,
205 SRR11772675 and SRR11772680 data. In addition, human adenovirus 1 sequences were
206 detected in SRR11772663 and SRR11772666 data. Parvovirus NIH-CQV genes coding for a
207 putative replication-associated protein (rep), and putative capsid protein (cap)
208 (coverage=14.51% and mean depth=0.41) were detected in SRR10971381 (Bioproject
209 PRJNA603194) **(Table S2).**
210 A very diverse group of bacterial populations were observed in the COVID-19 infected patients.
211 *Enterobacter hormaechei* was identified in WIV02, WIV04, WIV05, WIV06, and WIV07 patients.
212 *Enterobacter hormaechei* is an important emerging pathogen and can cause nosocomial
213 infections, and often have resistance to multiple clinically relevant antibiotics (Monahan et al.
214 2019). *Acinetobacter baumannii* sequence was detected in WIV05 and WIV07 patients. *A.*
215 *baumannii* is one of the most successful pathogens associated with hospital-acquired infections
216 worldwide (Lee et al. 2017). *Enterococcus faecalis* sequence was identified in WIV04, WIV05,
217 and WIV07 patients **(Table S2).**
218 Coliphage phi-X174 was identified in most sequence data, which might be introduced by the
219 Illumina PhiX control library (Meyer & Kircher 2010). *Stenotrophomonas* phage phiSMA7,
220 *Enterobacteria* phage phi80, DE3, Fels-2 and M13, *Proteus* virus, and *Delftia* phage RG-2014
221 were identified in the sequence data **(Table S2).**
222
223 **Kraken Taxonomic classification**

Taxonomic classification was performed using Kraken 2 on all reads containing more than one virus sequences. The genetic data for constructing the databases were retrieved from the NCBI RefSeq library. A very diverse group of viral, bacterial, and archaeal populations was observed in the samples. A taxonomic classification that was obtained from WIV04, WIV06, and WIV07 patients revealed a dominance of Bacteria (7%, 5%, and 59%, respectively), followed by Viruses (0.3%, 0.3%, and 0.3%, respectively) while Archaea was lower than 0.02% in all the patients. Among viral communities, influenza type A (8%), rather than SARS-CoV-2 (2%), was found to be dominant in the WIV07 patient, which is consistent with fastv result (**Figure 2).** SARS-CoV-2, rhabdovirus, and influenza type A dominated the sequence data from WIV06 patient (**Figure 3)** and WIV04 patient (**Figure 4)**.

A rank code, indicating unclassified (U), domain (D), kingdom (K), phylum (P), class (C), order (O), family (F), genus (G), or species (S) was used.

Figure 2: The Krona pie chart of the viral taxonomy of the WIV07 patient. Viruses identified from Kraken 2 analysis of samples collected from SRR11092059 patients were visualized using the krona tool. Among viral communities, influenza type A (8%), SARS-CoV-2 (2%), and rhabdovirus (0.6%) dominated the sequence data from WIV07 patient.

Figure 3: The Krona pie chart of the viral taxonomy of the WIV06 patient. Viruses identified from Kraken 2 analysis of samples collected from SRR11092060 patients were visualized using the krona tool. Among viral communities, SARS-CoV-2 (1%), rhabdovirus (0.6%), and influenza type A (0.4%) dominated the sequence data from WIV06 patient.

Figure 4: The Krona pie chart of the viral taxonomy of the WIV04 patient. Viruses identified from Kraken 2 analysis of samples collected from SRR11092062 patients were visualized using

the krona tool. Among viral communities, SARS-CoV-2 (1%), rhabdovirus (0.8%), and influenza type A (0.4%) dominated the sequence data from WIV04 patient.

Acinetobacter baumannii and *Enterobacter* species genomic sequences were detected in SRR11092059 (**Figure 5),** SRR11092060 (**Figure 6),** SRR11092061, SRR11092062 (**Figure 7),** and SRR11092063 sequence data. By using Kraken2, a tool for read taxonomy, many bacterial species were identified that most of them present in the normal flora and rarely cause problems. However, compared to the fastv result, *the Pasteurella multocida* sequence was detected in SRR11092062 (**Figure 7)** and SRR11092063 sequence date**.** The genome sequence of *Staphylococcus aureus* was detected in SRR11092059 data (**Figure 5).**

Figure5: Bacteria Identified in SRR11092059. Sankey diagrams of Kraken 2 report results obtained from SRR11092059. The width of the flow is proportional to the number of reads. The number above each node is the number of *k*-mer hits. A rank code, indicating domain (D), phylum (P), family (F), genus (G), or species (S) was used.

Figure 6: Bacteria Identified in SRR11092060. Sankey diagrams of Kraken 2 report results obtained from SRR11092060. The width of the flow is proportional to the number of reads. The number above each node is the number of *k*-mer hits. A rank code, indicating domain (D), phylum (P), family (F), genus (G), or species (S) was used.

Figure 7: Bacteria Identified in SRR11092062. Sankey diagrams of Kraken 2 report results obtained from SRR11092062. The width of the flow is proportional to the number of reads. The number above each node is the number of *k*-mer hits. A rank code, indicating domain (D), phylum (P), family (F), genus (G), or species (S) was used.

The most abundant bacterial phylum was Bacteroidetes, followed by Firmicutes phylum in SRR10971381 sequence data (**Figure 8)**.

Figure 8: Sankey diagrams of Kraken 2 report results obtained from SRR10971381. The width of the flow is proportional to the number of reads. The number above each node is the number of *k*-mer hits. A rank code, indicating domain (D), kingdom (K), phylum (P), family (F), genus (G), or species (S) was used.

Discussion

The importance of detecting co-infections is becoming more recognized (Griffiths et al. 2011; Li & Zhou 2013), but it remains challenging to get such information. The source of clinical samples and the sequencing technology can be inferior in co-infection detection (Birdsell et al. 2018b). A better understanding of co-infection's prevalence is required, partly because co-infecting pathogens can interact with each other directly or indirectly via the host's resources or immune system (Cox 2001; Griffiths et al. 2011). These interactions within co-infected hosts can alter the transmission, clinical progression, and control of multiple infectious diseases as compared to single pathogen species infection (Chiodini 2001; Griffiths et al. 2011; Palacios et al. 2009). Recent studies indicate that co-infection's adverse effects are more common than those with no-effects or positive impacts on human health (Pullan & Brooker 2008).
The underdiagnosis of co-infections is attributed to a lack of clinical suspicion, common symptoms, and the fact that conventional methods have little capacity to detect co-infections in the absence of a priori knowledge.

Co-infection can potentially affect the performance of laboratory testing for coronavirus disease 2019. A previous study (Lai et al. 2020) showed that reverse transcription polymerase chain reaction (**RT -PCR**) could not detect SARS-CoV-2 in two patients co-infected with influenza A

virus. Some researchers have suggested that inadequate viral specimens, the use of improperly validated assay, timing and methods of collecting specimens, the presence of mutations at the primer binding site, and co-infection with other viruses might be responsible (Arevalo-Rodriguez et al. 2020; Kucirka et al. 2020; Lai et al. 2020; Li et al. 2020). Based on a limited number of observational studies, it was found that the false-negative rate for SARS-CoV-2 RT-PCR was 20% (Lai et al. 2020; Li et al. 2020).

Given these considerations, COVID-19 might be underdiagnosed, especially during the influenza season, since typical clinical symptoms of COVID-19, including fever, cough, and dyspnea, resemble those of influenza (Chen et al. 2020a; Wu et al. 2020).
Therefore, exploring new diagnostic approaches is essential to advance understanding of co-infection contribution to disease manifestations and treatment responses (Birdsell et al. 2018a). Remarkable developments in next-generation sequencing have recently made metagenomics, an unbiased shotgun method of analysis, a widely used tool in just about every field of biology, including diagnosis of infectious diseases (Kuroda et al. 2012; Lecuit & Eloit 2014).

In this study, using *k*-mer based tools (fastv and Kraken 2), genome sequences of various microorganisms, including the SARS-CoV-2 virus, were detected. The accuracy of such an alignment-free method is relatively high compared to an alignment-based approach. Fastv is an alignment-free, ultra-fast tool for detecting the microbial sequences in sequence data using unique *k*-mers (Chen et al. 2020b). To validate the results obtained from fastv analysis, Kraken 2 was used. Kraken 2 is the latest version of Kraken, a taxonomic classification system that uses exact *k*-mer matches to achieve high accuracy and rapid classification (Wood et al. 2019). Kraken 2 is fast and requires less memory than Kraken 1. However, the standard database construction requires approximately 100 gigabytes (GB) of disk space and random-access memory (RAM) of more than 30 GB. For that purpose, an AWS EC2 h1.8xlarge storage

optimized instance with 16 dual-core hyperthreaded 2.30 GHz CPUs and 132 GB of RAM was used.

Multidrug-resistant bacteria associated with hospital-acquired infections worldwide, such as *E. hormaechei* , *S. aureus*, *P. multocida*, and *A. baumannii*, were detected in COVID-19 infected patients (Lee et al. 2017; Monahan et al. 2019). Most of the infections caused by the previously mentioned bacteria occur in critically ill and or immunocompromised patients in the intensive care unit (ICU) setting (Fournier et al. 2006).

Understanding the nature and consequences of co-infection is essential for accurate estimates of infectious disease burden. In particular, more systematic data on infectious diseases would also help measure the extent of co-infection on human health. Increased knowledge of the risk factors, the conditions in which co-infecting pathogens interact, and the mechanisms behind these interactions, particularly in experimental studies, will also help develop and evaluate infectious disease management programs.

Conclusions

In this study, screening of 68 public next-generation sequencing data from SARS-CoV-2 infected patients was performed using fastv and Kraken 2. Multiple viruses, including the SARS-CoV-2 virus, genome sequences were detected. Further large-sample studies are warranted to investigate the prevalence of COVID-19 co-infection, the impact of co-infection on the host immune system of COVID-19 patients, and their role in disease progression.

References

Arevalo-Rodriguez I, Buitrago-Garcia D, Simancas-Racines D, Zambrano-Achig P, del Campo R, Ciapponi A, Sued O, Martinez-Garcia L, Rutjes A, Low N, Bossuyt PM, Perez-Molina JA, and Zamora J. 2020. FALSE-NEGATIVE RESULTS OF INITIAL RT-PCR ASSAYS FOR COVID-

19: A SYSTEMATIC REVIEW. *medRxiv*:2020.2004.2016.20066787. 10.1101/2020.04.16.20066787

Birdsell DN, Özsürekci Y, Rawat A, Aycan AE, Mitchell CL, Sahl JW, Johansson A, Colman RE, Schupp JM, Ceyhan M, Keim PS, and Wagner DM. 2018a. Coinfections identified from metagenomic analysis of cervical lymph nodes from tularemia patients. *BMC infectious diseases* 18:319-319. 10.1186/s12879-018-3218-2

Birdsell DN, Özsürekci Y, Rawat A, Aycan AE, Mitchell CL, Sahl JW, Johansson A, Colman RE, Schupp JM, Ceyhan M, Keim PS, and Wagner DM. 2018b. Coinfections identified from metagenomic analysis of cervical lymph nodes from tularemia patients. *BMC infectious diseases* 18:319. 10.1186/s12879-018-3218-2

Breitwieser FP, and Salzberg SL. 2016. Pavian: Interactive analysis of metagenomics data for microbiomics and pathogen identification. *bioRxiv*:084715. 10.1101/084715

Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y, Xia J, Yu T, Zhang X, and Zhang L. 2020a. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 395:507-513. 10.1016/S0140-6736(20)30211-7

Chen S, He C, Li Y, Li Z, and Melançon CE. 2020b. A Computational Toolset for Rapid Identification of SARS-CoV-2, other Viruses, and Microorganisms from Sequencing Data. *bioRxiv*:2020.2005.2012.092163. 10.1101/2020.05.12.092163

Chiodini PL. 2001. Chemotherapy for patients with multiple parasitic infections. *Parasitology* 122:S83-S89.

Cox FE. 2001. Concomitant infections, parasites and immune responses. *Parasitology* 122 Suppl:S23-38. 10.1017/s003118200001698x

Duncavage EJ, Magrini V, Becker N, Armstrong JR, Demeter RT, Wylie T, Abel HJ, and Pfeifer JD. 2011. Hybrid capture and next-generation sequencing identify viral integration sites from formalin-fixed, paraffin-embedded tissue. *J Mol Diagn* 13:325-333. 10.1016/j.jmoldx.2011.01.006

Fehr AR, and Perlman S. 2015. Coronaviruses: an overview of their replication and pathogenesis. *Coronaviruses*: Springer, 1-23.

Fournier PE, Richet H, and Weinstein RA. 2006. The epidemiology and control of Acinetobacter baumannii in health care facilities. *Clinical infectious diseases* 42:692-699.

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 DV, Sidorov IA, Sola I, Ziebuhr J, and Coronaviridae Study Group of the International Committee on Taxonomy of V. 2020. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology* 5:536-544. 10.1038/s41564-020-0695-z

Griffiths EC, Pedersen AB, Fenton A, and Petchey OL. 2011. The nature and consequences of coinfection in humans. *Journal of Infection* 63:200-206. https://doi.org/10.1016/j.jinf.2011.06.005

Kucirka LM, Lauer SA, Laeyendecker O, Boon D, and Lessler J. 2020. Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction–Based SARS-CoV-2 Tests by Time Since Exposure. *Annals of Internal Medicine* 173:262-267. 10.7326/M20-1495

Kuroda M, Sekizuka T, Shinya F, Takeuchi F, Kanno T, Sata T, and Asano S. 2012. Detection of a possible bioterrorism agent, Francisella sp., in a clinical specimen by use of next-generation direct DNA sequencing. *J Clin Microbiol* 50:1810-1812. 10.1128/JCM.06715-11

Lai C-C, Wang C-Y, and Hsueh P-R. 2020. Co-infections among patients with COVID-19: The need for combination therapy with non-anti-SARS-CoV-2 agents? *Journal of Microbiology, Immunology and Infection*. https://doi.org/10.1016/j.jmii.2020.05.013

Lecuit M, and Eloit M. 2014. The diagnosis of infectious diseases by whole genome next generation sequencing: a new era is opening. *Front Cell Infect Microbiol* 4:25. 10.3389/fcimb.2014.00025

Lee C-R, Lee JH, Park M, Park KS, Bae IK, Kim YB, Cha C-J, Jeong BC, and Lee SH. 2017. Biology of Acinetobacter baumannii: Pathogenesis, Antibiotic Resistance Mechanisms, and Prospective Treatment Options. *Frontiers in Cellular and Infection Microbiology* 7:55.

Li D, Wang D, Dong J, Wang N, Huang H, Xu H, and Xia C. 2020. False-Negative Results of Real-Time Reverse-Transcriptase Polymerase Chain Reaction for Severe Acute Respiratory Syndrome Coronavirus 2: Role of Deep-Learning-Based CT Diagnosis and Insights from Two Cases. *Korean J Radiol* 21:505-508.

Li XX, and Zhou XN. 2013. Co-infection of tuberculosis and parasitic diseases in humans: a systematic review. *Parasit Vectors* 6:79. 10.1186/1756-3305-6-79

Lundberg DS, Yourstone S, Mieczkowski P, Jones CD, and Dangl JL. 2013. Practical innovations for high-throughput amplicon sequencing. *Nat Methods* 10:999-1002. 10.1038/nmeth.2634

Meyer M, and Kircher M. 2010. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor Protocols* 2010:pdb.prot5448.

Monahan LG, DeMaere MZ, Cummins ML, Djordjevic SP, Roy Chowdhury P, and Darling AE. 2019. High contiguity genome sequence of a multidrug-resistant hospital isolate of Enterobacter hormaechei. *Gut Pathogens* 11:3. 10.1186/s13099-019-0288-7

Ondov BD, Bergman NH, and Phillippy AM. 2011. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* 12:385. 10.1186/1471-2105-12-385

Palacios G, Hornig M, Cisterna D, Savji N, Bussetti AV, Kapoor V, Hui J, Tokarz R, Briese T, Baumeister E, and Lipkin WI. 2009. Streptococcus pneumoniae coinfection is correlated with the severity of H1N1 pandemic influenza. *PLoS One* 4:e8540. 10.1371/journal.pone.0008540

Plyusnin I, Kant R, Jääskeläinen AJ, Sironen T, Holm L, Vapalahti O, and Smura T. 2020. Novel NGS Pipeline for Virus Discovery from a Wide Spectrum of Hosts and Sample Types. *bioRxiv*:2020.2005.2007.082107. 10.1101/2020.05.07.082107

Pullan R, and Brooker S. 2008. The health impact of polyparasitism in humans: are we under-estimating the burden of parasitic diseases? *Parasitology* 135:783-794.

Tay MZ, Poh CM, Renia L, MacAry PA, and Ng LFP. 2020. The trinity of COVID-19: immunity, inflammation and intervention. *Nat Rev Immunol*. 10.1038/s41577-020-0311-8

Wilson MR, Naccache SN, Samayoa E, Biagtan M, Bashir H, Yu G, Salamat SM, Somasekar S, Federman S, and Miller S. 2014. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *New England Journal of Medicine* 370:2408-2417.

Wong CK, Lam CW, Wu AK, Ip WK, Lee NL, Chan IH, Lit LC, Hui DS, Chan MH, Chung SS, and Sung JJ. 2004. Plasma inflammatory cytokines and chemokines in severe acute respiratory syndrome. *Clin Exp Immunol* 136:95-103. 10.1111/j.1365-2249.2004.02415.x

Wood DE, Lu J, and Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biology* 20:257. 10.1186/s13059-019-1891-0

World Health Organization. 2020. WHO Director-General's statement on IHR Emergency Committee on Novel Coronavirus (2019-nCoV).

Worldometer. 2020, July 25. Covid-19 coronavirus pandemic.

Wu X, Cai Y, Huang X, Yu X, Zhao L, Wang F, Li Q, Gu S, Xu T, Li Y, Lu B, and Zhan Q. 2020. Co-infection with SARS-CoV-2 and Influenza A Virus in Patient with Pneumonia, China. *Emerg Infect Dis* 26:1324-1326. 10.3201/eid2606.200299

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, and Shi ZL. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579:270-273. 10.1038/s41586-020-2012-7

454

Figure 1

SARS-CoV-2 detection by fastv.

SARS-CoV-2 was detected in WIV07 patient, where twelve SARS-CoV-2 hits were included in the genome list and ordered by *k*-mer coverage (99.4972% to 99.2308%). Mismatches were highlighted in red.

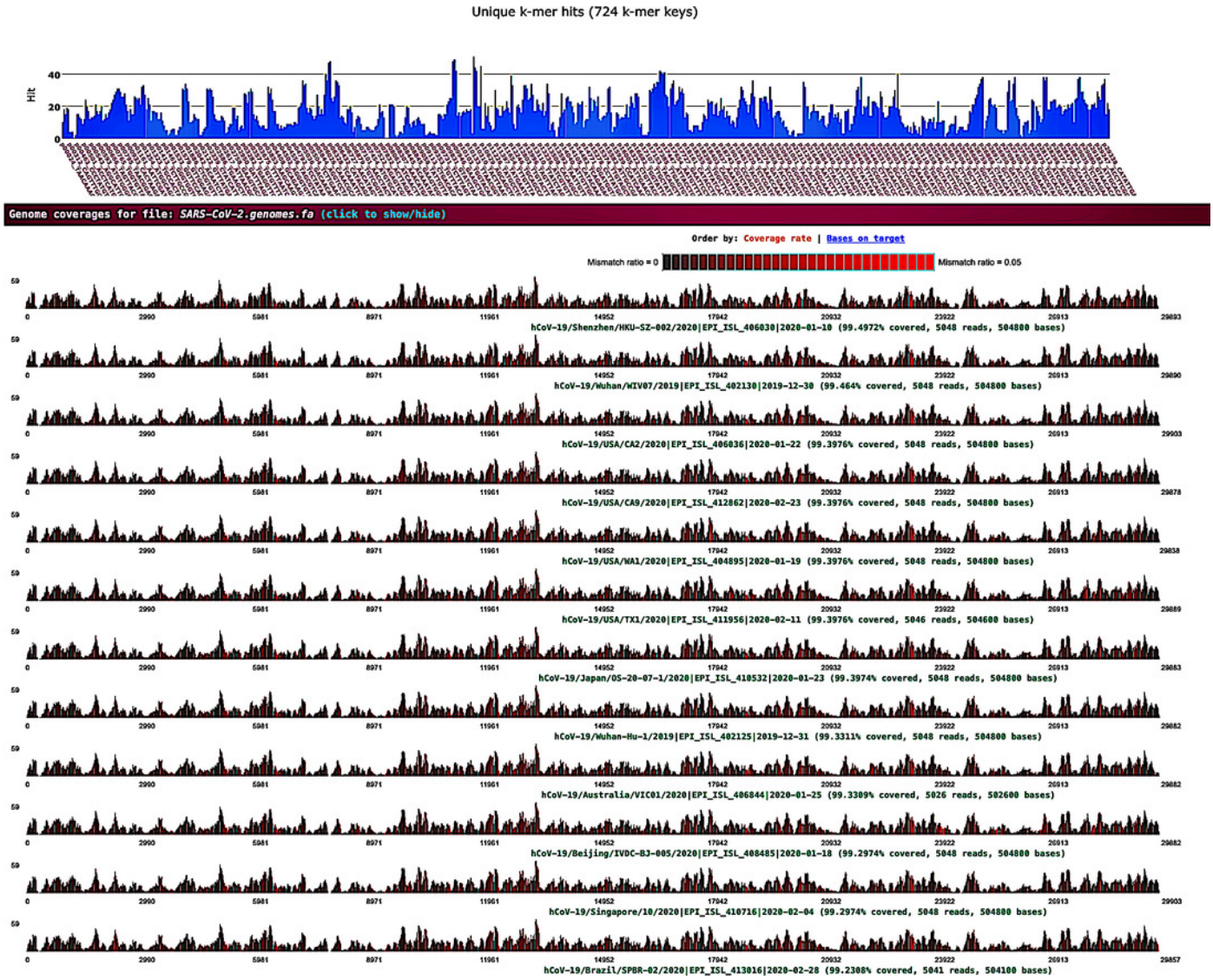


Figure 2

The Krona pie chart of the viral taxonomy of the WIV07 patient.

Viruses identified from kraken 2 analysis of samples collected from SRR11092059 patients were visualized using krona tool. Among viral communities, influenza type A (8%), SARS-CoV-2 (2%) and rhabdovirus (0.6%) dominated the sequence data from WIV07 patient.

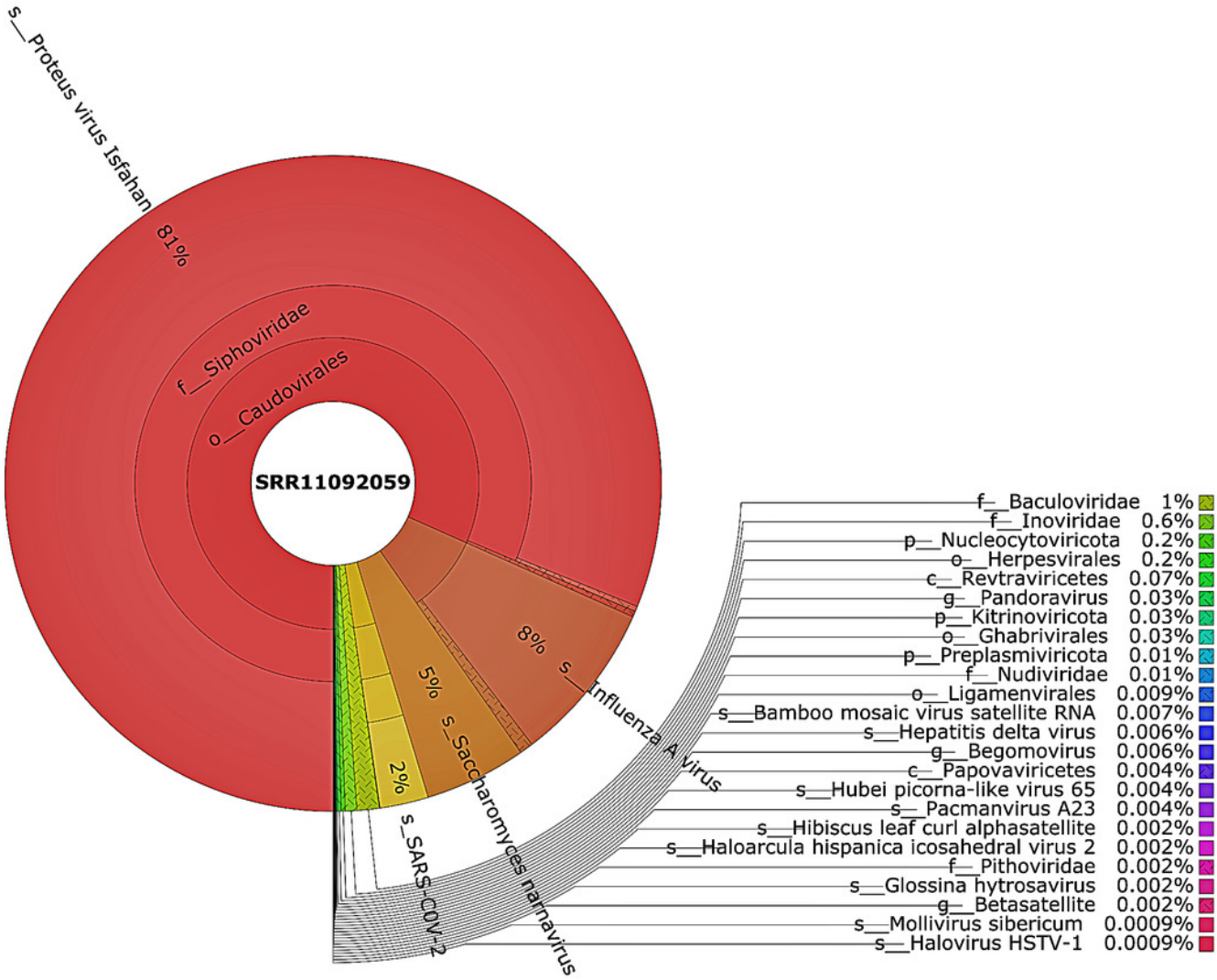


Figure 3

The Krona pie chart of the viral taxonomy of the WIV06 patient.

Viruses identified from kraken 2 analysis of samples collected from SRR11092060 patients were visualized using krona tool. Among viral communities, SARS-CoV-2 (1%), rhabdovirus (0.6%) and influenza type A (0.4%) dominated the sequence data from WIV06 patient.

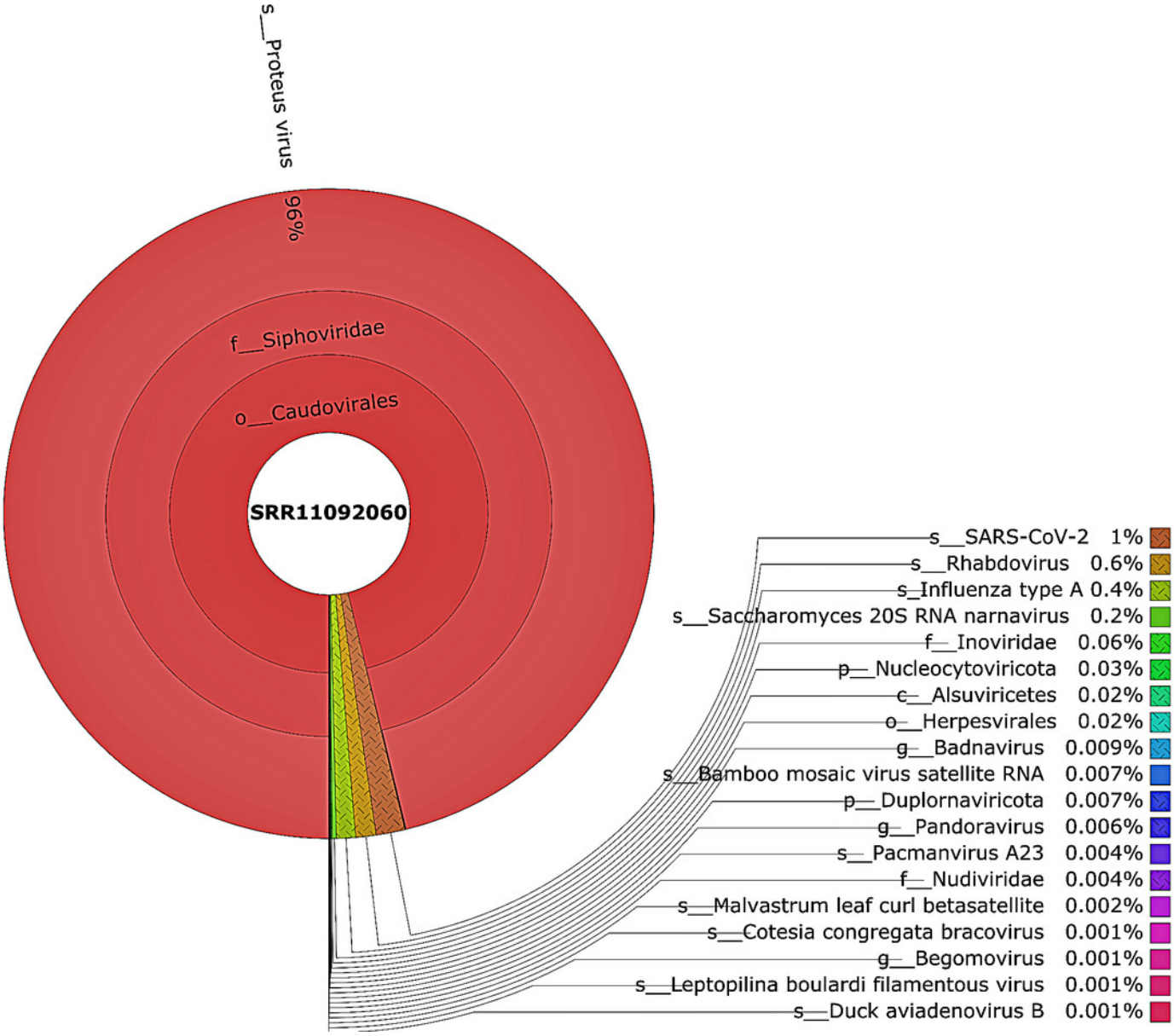


Figure 4

The Krona pie chart of the viral taxonomy of the WIV04 patient.

Viruses identified from kraken 2 analysis of samples collected from SRR11092062 patients were visualized using krona tool. Among viral communities, SARS-CoV-2 (1%), rhabdovirus (0.8%) and influenza type A (0.4%) dominated the sequence data from WIV04 patient.

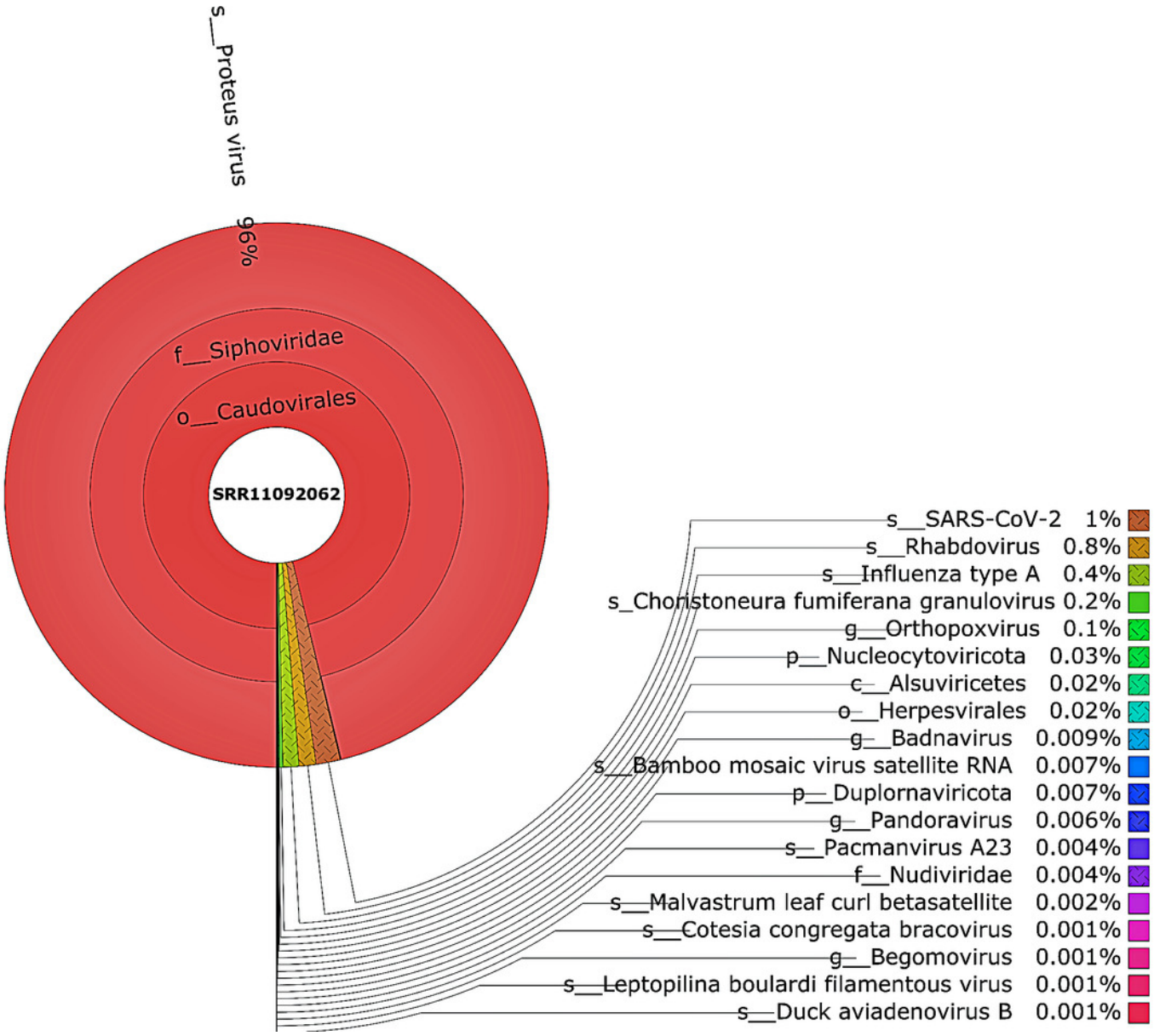


Figure 5

Bacteria Identified in SRR11092059

Sankey diagrams of Kraken 2 report results obtained from SRR11092059. The width of the flow is proportional to the number of reads. The number above each node is the number of *k*-mer hits. A rank code, indicating domain (D), phylum (P), family (F), genus (G), or species (S) was used.

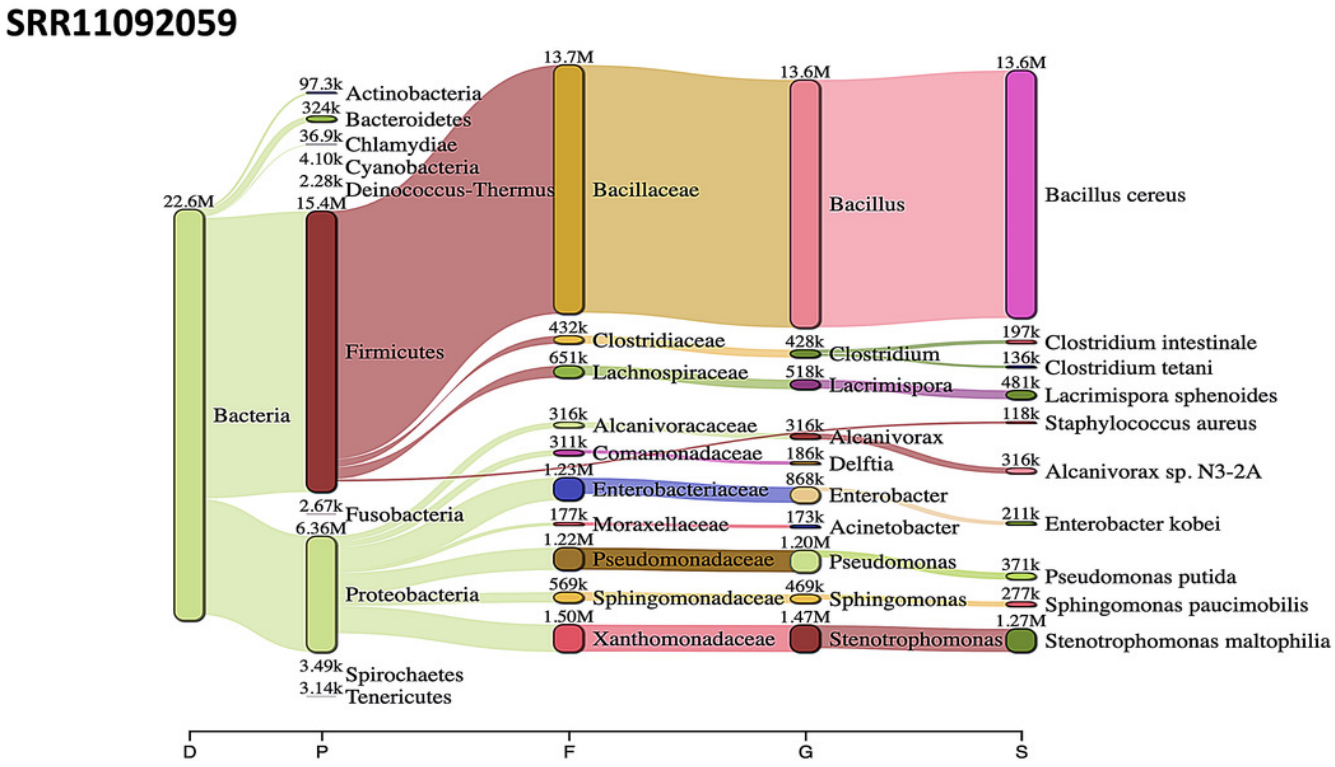


Figure 6

Bacteria Identified in SRR11092060

Sankey diagrams of Kraken 2 report results obtained from SRR11092060. The width of the flow is proportional to the number of reads. The number above each node is the number of *k*-mer hits. A rank code, indicating domain (D), phylum (P), family (F), genus (G), or species (S) was used.

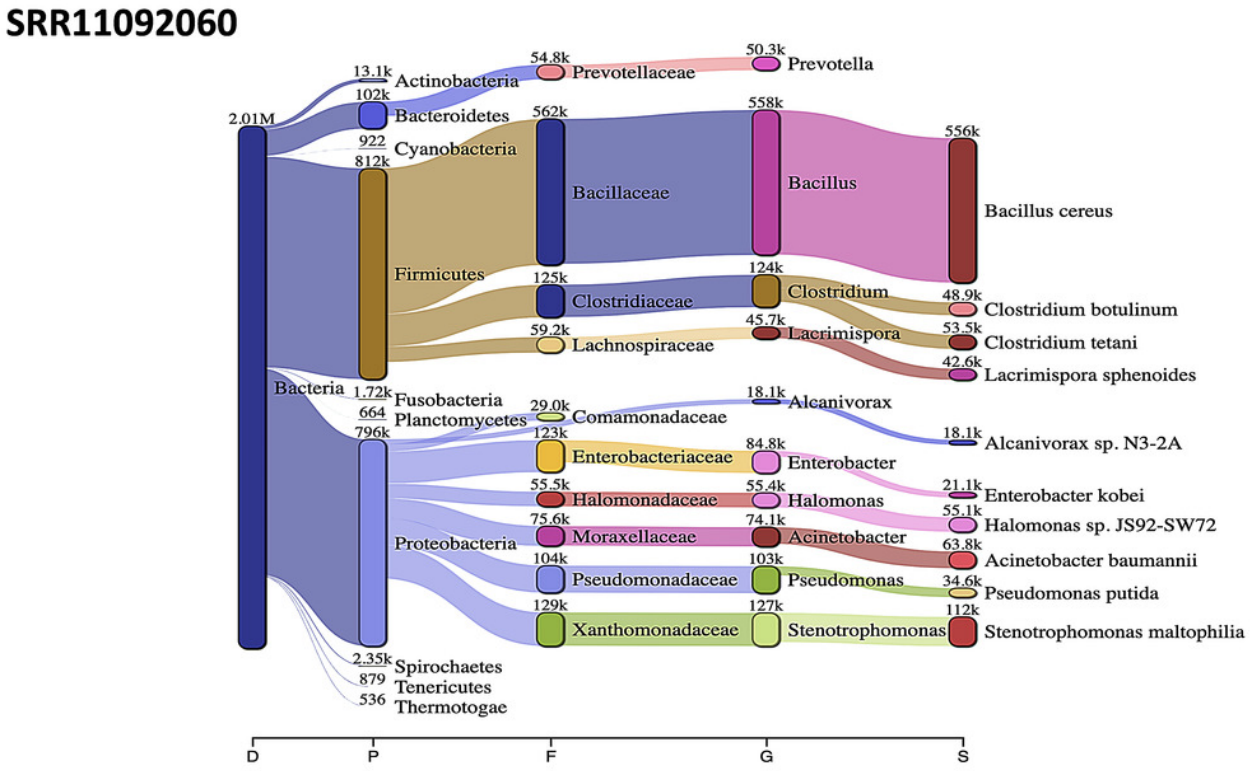


Figure 7

Bacteria Identified in SRR11092062

Sankey diagrams of Kraken 2 report results obtained from SRR11092062. The width of the flow is proportional to the number of reads. The number above each node is the number of *k*-mer hits. A rank code, indicating domain (D), phylum (P), family (F), genus (G), or species (S) was used.

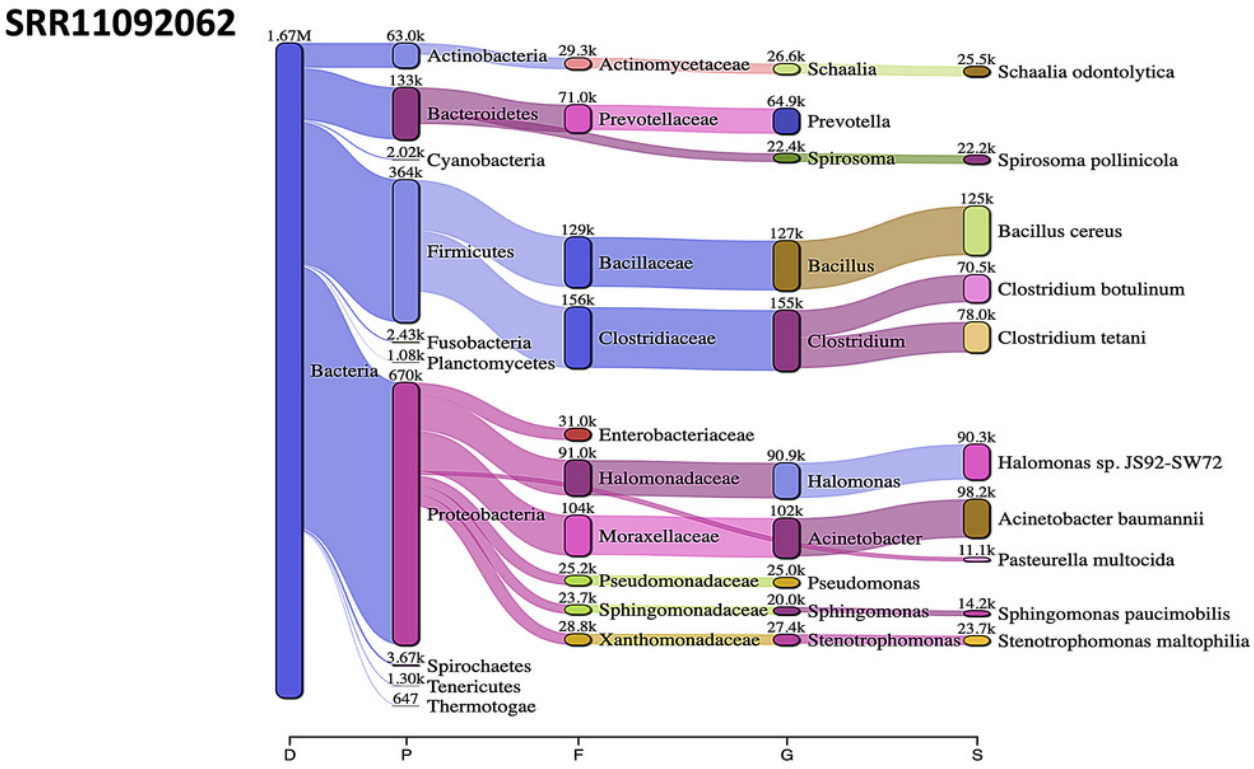
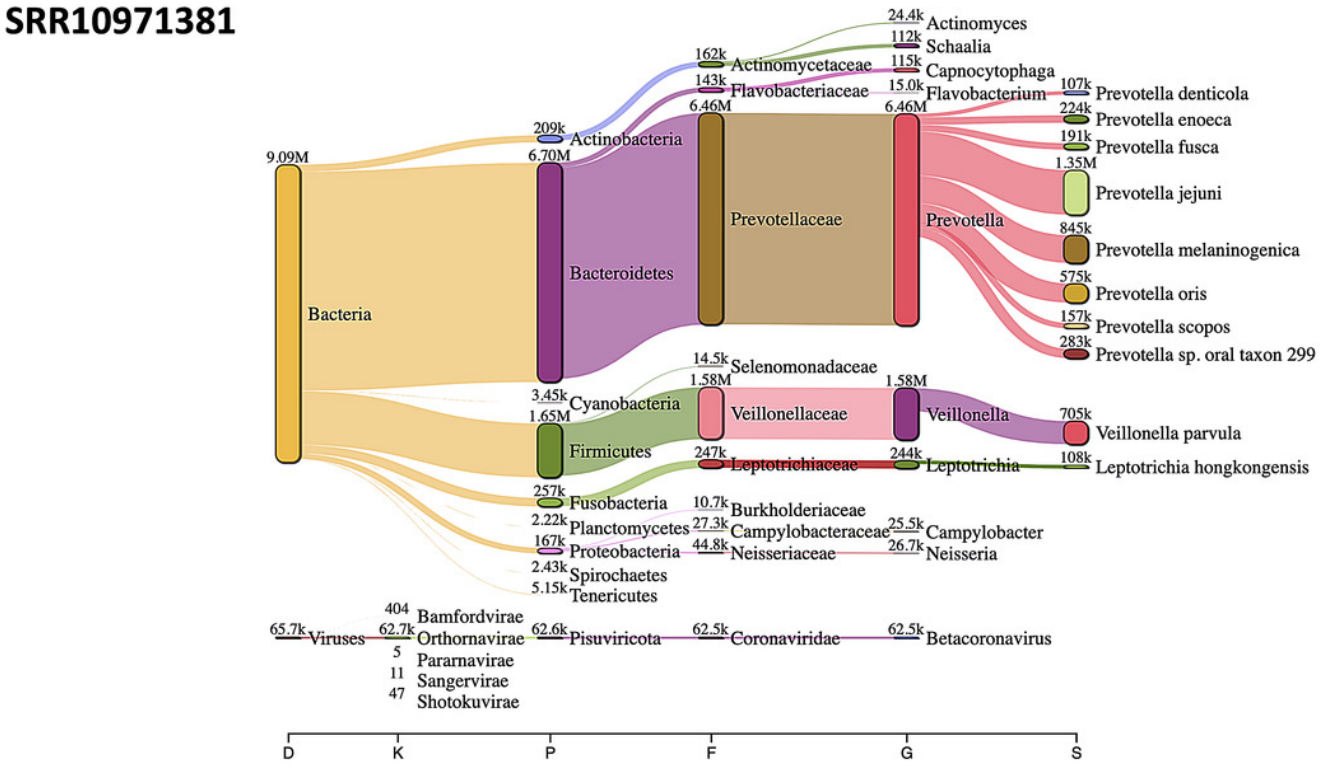


Figure 8

Sankey diagrams of Kraken 2 report results obtained from SRR10971381

The width of the flow is proportional to the number of reads. The number above each node is the number of *k*-mer hits. A rank code, indicating domain (D), kingdom (K), phylum (P), family (F), genus (G), or species (S) was used.