

# Host genomes for the unique SARS-CoV-2 variant leaked into Antarctic soil metagenomic sequencing data

Istvan Csabai (✉ [csabai@elte.hu](mailto:csabai@elte.hu))

ELTE Eötvös Loránd University, Budapest <https://orcid.org/0000-0001-9232-9898>

Norbert Solymosi

University of Veterinary Medicine, Budapest <https://orcid.org/0000-0003-1783-2041>

---

## Research Article

**Keywords:** sars-cov-2, covid-19, origin, host genome

**Posted Date:** February 7th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1330800/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Host genomes for the unique SARS-CoV-2 variant leaked into Antarctic soil metagenomic sequencing data

István Csabai<sup>1\*</sup> and Norbert Solymosi<sup>1,2</sup>

<sup>1</sup>Department of Physics of Complex Systems, ELTE Eötvös Loránd University, Budapest, Hungary

<sup>2</sup>Centre for Bioinformatics, University of Veterinary Medicine, Budapest, Hungary

\*csabai@elte.hu

## ABSTRACT

Recently Csabai et al.<sup>1</sup> have found a most likely contaminated metagenomic sample set from Antarctica that contained traces of unique SARS-CoV-2 variants. This is a short followup of that report where we attempt to find genetic footprint of the hosts. With reasonable confidence we could identify genetic material from mitochondria of Homo sapiens, green monkey and Chinese hamster. The latter two most probably originated from cell lines Vero E6 (or COS-7) and CHO respectively, which are frequent laboratory culture media for studying viruses including SARS-CoV-2 and its closest relatives.

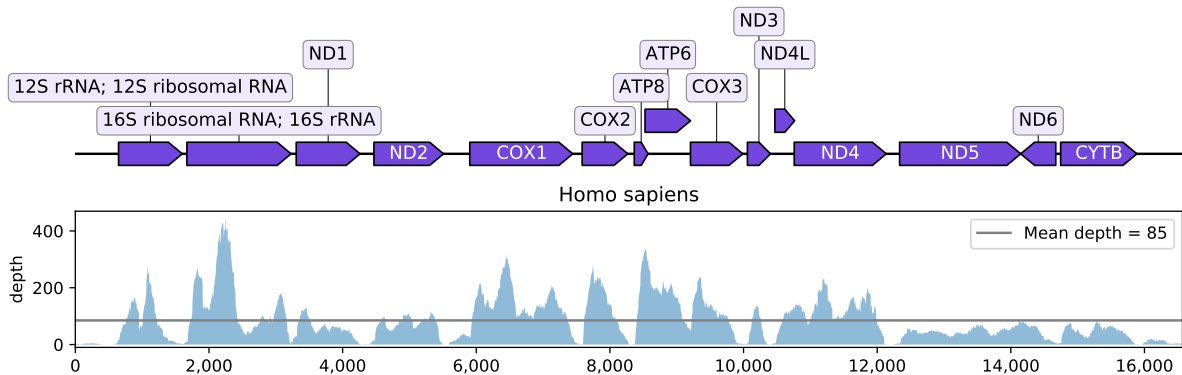
## Introduction

Despite the significant worldwide efforts we still do not know the origin of COVID-19. For this reason, detailed analysis of any sample containing early genetic material from SARS-CoV-2 is of paramount importance. Csabai et. al<sup>1</sup> have recently found samples under project ID PRJNA692319<sup>2</sup> with 12 samples, all were collected from soil at King George Island, Antarctica in a 3 week period after 2018-12-24. Three of the samples, SRR13441704, SRR13441705 and SRR13441708 contain enough sequence reads to cover the SARS-CoV-2 genome at average depth of 17x which allows to call mutations. The DNA from soil was extracted in December 2019 and sent to Sangon Biotech, Shanghai, China<sup>3</sup>, but the more exact time of the sequencing is not known. According to the phylogenetic analysis the samples seems to be very early, maybe one of the earliest - to our knowledge not published - variants most closely related to Pango lineage "A". They also contain some unique rare mutations and a long deletion starting at nucleotide 21761. After posting our preprint<sup>1</sup> the public access to the sequencing data was revoked<sup>4</sup> at SRA, but a month later the public access was restored<sup>5</sup>.

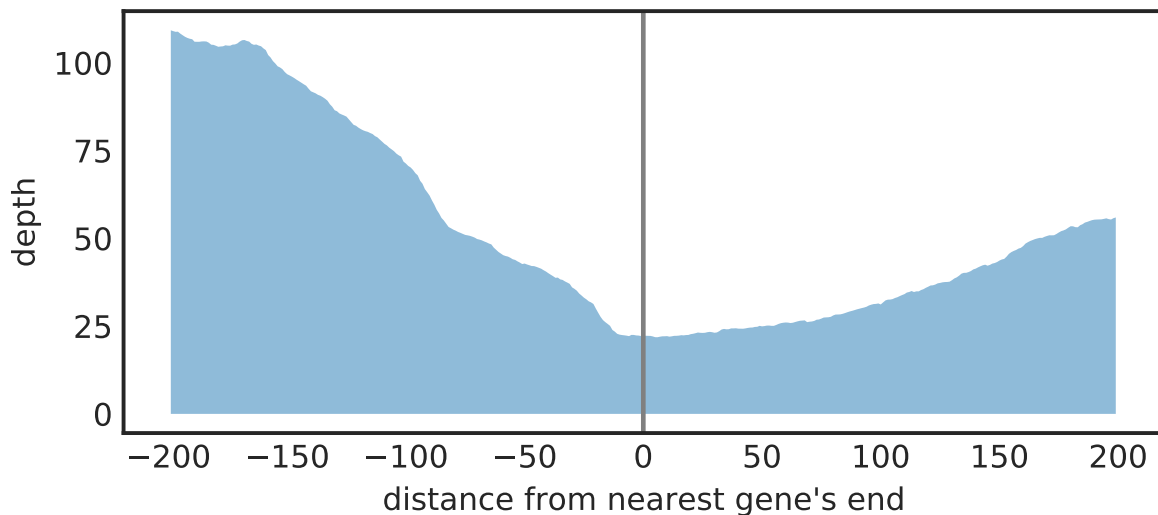
We think that the most likely scenario is that the SARS-CoV-2 sequence fragments are not from the Antarctic soil samples, but are mistakenly assigned reads from another sample(s) sequenced on the same flow cell at the Sangon Biotech sequencing facility in Shanghai. The equipment, Illumina HiSeq 4000 that was used for sequencing according to the project's metadata, is known to be prone to barcode misassignment errors<sup>6,7</sup>.

In a nutshell, as detailed in Illumina's whitepaper<sup>6</sup>, this artifact is related to the exponentially increasing density of next-generation sequencing (NGS) chips. Nowadays a single flow cell has such a large capacity that it is rarely filled up by a single sample. For example the 12 antarctic soil samples in PRJNA692319 altogether contain around 320 million paired end reads, while Illumina HiSeq 4000's capacity is up to 10 billion paired end reads. The increased capacity is utilized through multiplexing, which adds unique sequences, called indexes or barcodes, to each DNA fragment during library preparation. Before final data analysis, sequencing reads from pooled libraries need to be identified and sorted computationally in a process called demultiplexing. This allows large numbers of libraries to be pooled and sequenced simultaneously during a single sequencing run. After adapters are ligated to nucleic acid fragments, the products are cleaned up to remove any free, unligated adapters. Libraries can be cleaned up by a bead-based or gel purification step to remove free adapters or primers. Failure to remove free adapters or primers can lead to contamination of prepared libraries and may result in index hopping and misassignment. As a results, sequencing reads from other, not related samples that are sequenced on the same flow cell can get into our samples.

This putative misassignment mechanism may be related to the fact that most of the suspected contaminated paired-end sequenced samples contained mostly only the R2 mates (reverse 5' Illumina reads) of SARS-CoV-2. In the following we attempt to recover genetic sequences of the host(s) of the detected viruses.



**Figure 1.** The smoothed coverage of the human mitochondrial reference genome by the union of R2 reads from samples SRR13441704, SRR13441705 and SRR13441708. The average depth is 85 but there are large fluctuations in the coverage.



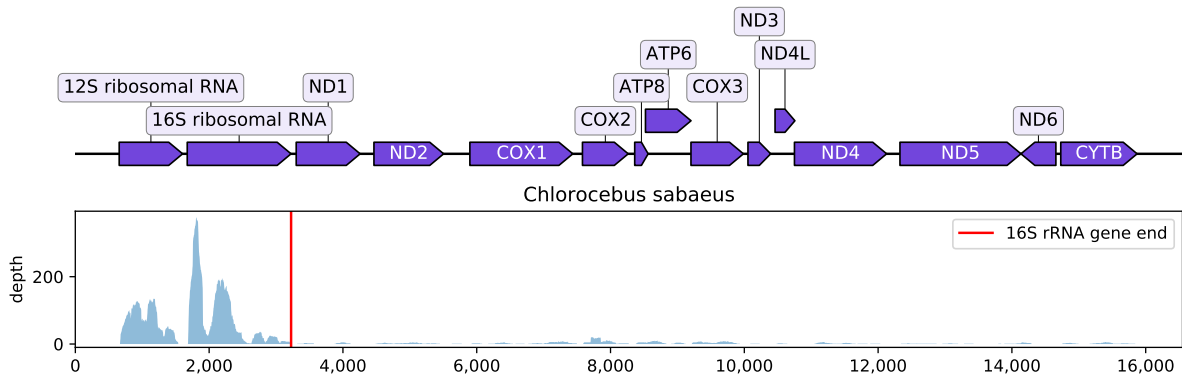
**Figure 2.** The average depth for the human mitochondrial reference genome as a function of the distance from the nearest gene's end. Union of R2 reads from samples SRR13441704, SRR13441705 and SRR13441708 was used. The depth is the lowest at the end position of the genes (distance=0), and growing towards the internal parts (larger positive or negative distances). This may indicate that not DNA, but the expressed mRNA was sequenced.

## Results

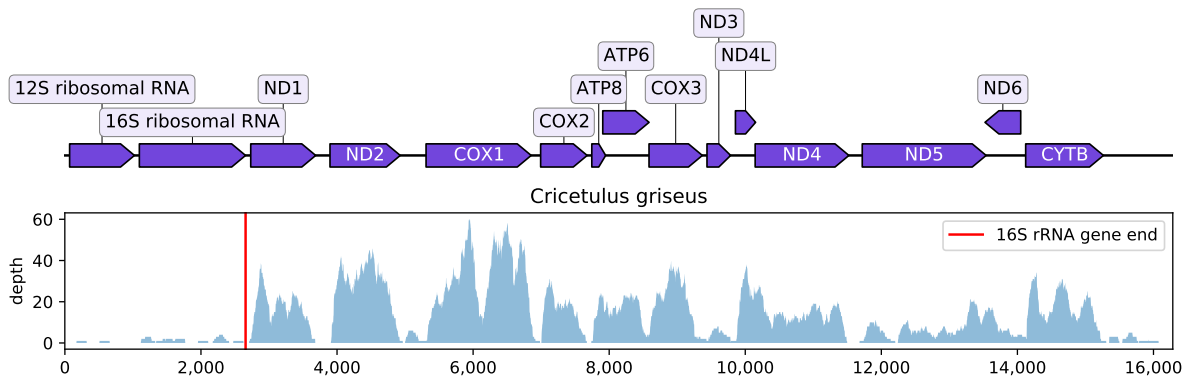
We hypothesised that, like the virus sequences, the host(s)'s genetic material may have been leaked into the sequencing data in a similar way. All known hosts of betacoronaviruses are mammals that have huge genomes so it is not likely that they will be completely captured in the studied samples. The size of mammalian mitochondrial genome (mtDNA; 16knt), on the other hand, is comparable to that of a coronavirus. Another advantage is that a single mammalian cell may contain thousands of mitochondria, each with its own genome, resulting a higher genome coverage.

We have downloaded the current database of all assembled mitochondria from NCBI and to cast our net somewhat wider than mammals only, we have extracted the ones, altogether 6158 species, that contained the term "Vertebrata" in their taxonomy annotation. We aligned separately the R1 and R2 reads of the of PRJNA692319 to this vertebrate mtDNA reference genome set.

When multiple genomes are used in parallel, the `bowtie2` alignment software assigns each read to the most perfectly matching part of the best matching genome but due to the homology between closely related species, this may sometimes



**Figure 3.** The smoothed coverage of the *Chlorocebus sabaesus* mitochondrial reference genome by the union of R2 reads from samples SRR13441704, SRR13441705 and SRR13441708. The average depth is 13.0 but there is a significantly higher coverage for the first ~ 3200nt of the genome. The vertical red line shows the end of the 16S ribosomal RNA gene’s end position.



**Figure 4.** The smoothed coverage of the *Cricetulus griseus* mitochondrial reference genome by the union of R2 reads from samples SRR13441704, SRR13441705 and SRR13441708. The average depth is 12.8 but the first ~ 2600nt part almost lacks any coverage. The vertical red line shows the end of the 16S ribosomal RNA gene’s end position.

result false alignments. Also, low complexity non-mtDNA reads produce false positive alignments, so simply the number of reads aligned to a genome may not be a good statistical measure of the true presence of a species. Instead we calculated the mitochondrial genome coverage (percentage of nucleotides covered) and average depth (average number of reads aligned to a given nucleotide) for each species. Table 1 shows the top 10 results for samples SRR13441704, SRR13441705 and SRR13441708 that were the most abundant in SARS-CoV-2 reads and as comparison two other samples, SRR13441701, SRR13441702 in which we could not identify SARS-CoV-2 reads. The SARS-CoV-2 abundant samples have in general larger mtDNA content and for certain species we have found orders of magnitudes higher coverage than in the samples where SARS-CoV-2 could not be detected. In these samples the extra counts mostly come from three species, *Homo sapiens*, *Cricetulus griseus* and *Chlorocebus sabaesus*. We can observe the same asymmetry as with the SARS-CoV-2 counts (see Table 1 in Csabai et al. 2022<sup>1</sup>) that there are more hits for the R2 mates than for the R1 ones. Since the abundance of the mtDNA content correlates with the abundance of the SARS-CoV-2 content (see also Table 1 in Csabai et al. 2022<sup>1</sup>), the mtDNA reads most probably came from the same contaminator samples and *Homo sapiens*, *Cricetulus griseus* and *Chlorocebus sabaesus* can

be considered as the hosts of the detected SARS-CoV-2. The coverage of the mitochondrial genome of this three species can be seen in Figures 1, 3 and 4.

Since these three mammals are not arctic species, this also confirms the hypothesis that SARS-CoV-2 was introduced as a contaminant in the samples. We do not examine here the other less abundant top alignments, like *Geotria australis* or pouched lamprey which is lineage of jawless fish widespread in the Southern Hemisphere. They most probably come from the original Antarctic soil samples that were collected around seal and penguin colonies or they are only chance alignments.

## Discussion

As we are investigating a small amount of genetic material from putative contaminants of unknown origin, it is difficult to make definitive statements. Accordingly, we caution the reader not to take the following as proven final scientific truth. We try to stick to the facts and refrain from making unsubstantiated hypotheses but at the same time squeeze every last drop of information out of the available data.

*Homo sapiens* is not a surprising find for a possible host. Since some of the reads aligned to other *Homo* species like *Homo sapiens neanderthalensis*, *Homo sp. Altai* and *Homo heidelbergensis* we realigned all the SRR13441704,05,08 R2 reads solely to NC\_012920.1 *Homo sapiens* mtDNA genome. As shown in Fig. 1 the mitochondrial genome is mostly covered with average depth of 85, but there are large fluctuations. Aligning the fluctuations to the genome annotation, it is very striking to the eye that the sequencing depth is lower at gene junction positions. We have calculated the average depth as a function of the distance from the nearest gene's end, as shown in Figure 2. Indeed, the depth is the lowest at the end position of the genes, and growing towards the intragenic parts. This indicates that not DNA, but the expressed mRNA was sequenced for these samples, maybe just as a by-product of the SARS-CoV-2 RNA virus sequencing.

With the caveat that the coverage is gappy and low, and there may be sequencing artifacts, we have attempted to call mutations and reconstruct the consensus human mtDNA genome. Altogether we have found 23 mutations: A263G, A750G, A862G, A866C, C870A, T903C, A918G, T921C, A929G, A939G, A1041C, T1055C, A1110G, T1119C, G1130A, A1154C, A1438G, A4769G, A8860G, G15043A, A15326G, C16223T, A16293G. We determined its haplogroup with the *mixem* tool<sup>8</sup>, that were designed for forensic analysis of low quality sequence data from complex mitochondrial mixtures. It gave a rare South-Asian haplogroup M48. We have also compared the list of mutations to the 1000 Genomes Project's<sup>9</sup> mitochondrion database. Only 11 of our 23 mutations were listed in this database. Of the total 2545 samples two, namely HG02187 (*Continental population: East Asian, Population group: CDX Chinese Dai in Xishuangbanna, China*) and HG02127 (*Continental population: EAS East Asian, Population group: KHV Kinh in Ho Chi Minh City, Vietnam*) have agreed in all of the 11 positions. It may be just an interesting coincidence that the closest known relatives of SARS-CoV-2 are coming from bats that live in caves close to Xishuangbanna in Yunnan province, and from the Indochinese peninsula<sup>10</sup>.

*Cricetulus griseus* or Chinese hamster was first used as a laboratory animal to type pneumococci obtained from human patients<sup>11</sup> and became a valuable animal model in other infectious disease and epidemiology research studies<sup>12</sup>, including MERS-CoV, SARS-CoV<sup>13</sup> and SARS-CoV-2<sup>14</sup>. There is also a Chinese hamster ovary (CHO) cell line which is widely used in biological and medical research, primarily for toxicity screening and for expression of recombinant antibodies. For example in Fukushi et al.<sup>15,16</sup> *in vitro* passage of SARS-CoV on CHO cells that expressed rat ACE2 yielded a new adapted variant that encoded two new mutations in the spike protein. The same cell line was used for SARS vaccine development<sup>17</sup> and recently also in several SARS-CoV-2 related research studies<sup>18,19</sup>.

*Chlorocebus sabeus* or green monkey is infamous for zoonosis researchers as it was associated with Marburg virus that was discovered in 1967 in a laboratory outbreak in Germany among workers exposed to cell cultures from green monkeys imported from Africa<sup>20</sup>. The 'Vero' cell line that was isolated from kidney epithelial cells of *C. sabeus*<sup>21,22</sup> has various applications in molecular and cell biology research. Vero E6 cells enable achieving high titers of SARS-CoV and are also used in the cultivation of Rabies virus, Reovirus and Japanese encephalitis virus<sup>23</sup>. The first isolation of bat coronaviruses that uses the ACE2 receptor was also done in Vero E6 cells in 2013 from Chinese horseshoe bats (Yunnan, China, RsSHC014 and Rs3367) that are known to be among the closest relatives of SARS-CoV<sup>10,24</sup>. Two years later, using also Vero cell line, genetically engineered highly pathogenic chimeric virus was reported that reproduced in human lung cells in cell culture as well as in the mouse lung with the corresponding pathogenesis in animals<sup>25</sup>. This and similar other experiments are one of the sources that feed the lab-leak hypothesis, intensively discussed in scientific and public media<sup>26</sup>. There is another SV40 transformed *C. sabeus* based fibroblast-like cell line, COS-7 that is used among other applications to study viral entry and inhibition<sup>27</sup>. The Vero E6 and COS-7 cell lines are commonly used in COVID-19 research, too. The first paper from Wuhan Institute of Virology to report isolating SARS-CoV-2<sup>28</sup> also used Vero cell line and a PubMed search for keywords "SARS-CoV-2"+"Vero"+"E6" as of January 2022 lists 322 articles<sup>29,30</sup>.

It is interesting that while the human mitochondrial genome is mostly covered from end to end, as can be seen on Figures 3 and 4, only the first ~ 3200nt of the *C. sabeus*, and almost complementarily, starting around at 2600nt the other end of the *C. griseus* genomes are covered. The position of these breakpoints may not be coincidence. Looking closer the annotations, we

noticed that the *C. saubeus* genome is covered just up to the end of its 16S ribosomal RNA gene at nucleotide 3225 and the coverage for *C. griseus* start just at the end of its respective 16S rRNA gene at nucleotide 2656. Based on this we hypothesise that the samples may contain heterokaryons from fused cells, and part of each homologous (mitochondrial) genomes are silenced. Though the details are elusive, as other viruses SARS-CoV-2 can also induce cell fusion and result syncytia formation<sup>31,32</sup>.

Combination of the above two cell lines, a binding assay between CHO cells and Vero E6 cells was also developed<sup>33</sup> for SARS-CoV research. In this assay, the SARS-CoV S protein expressed by transfection into CHO cells mediates adhesion to Vero E6 cells that possess ACE2<sup>34</sup>. Cell fusion assays or serial passage of viruses in cell lines<sup>35</sup> may be related to the fragmented coverage of the mitochondrial genomes as seen in Figures 4 and 3. Combination of green monkey and hamster cell lines are also used for SARS-CoV-2 research<sup>36–38</sup>.

All these findings support our hypothesis, that the identified species (or their cell lines) are not just chance alignments but likely hosts of the SARS-CoV-2 viruses found in the PRJNA692319 sample set. Based on the available information we cannot tell neither the exact number of the contaminating samples nor if each contained both the mammal cells and SARS-CoV-2, but we suspect that there was one or more SARS-CoV-2 infected human (or human cell line) samples and one or more green monkey (or cell line) and Chinese hamster (or cell line) or cell-fusion samples.

Since SARS-CoV-2 is an RNA virus, without reverse transcription to cDNA an infected researcher could not contaminate the samples during the lab preparation steps. Running PCR on virus containing samples in the same lab may be source of the contamination, but most probably that would not pick up host's genome fragments in comparable concentration to SARS-CoV-2, that we have found. The presence of the animal model genomes indicate that some of the samples are not just simple samples from COVID-19 patients, rather they come from research experiments using simultaneously more cell lines or animal models.

As we have shown previously<sup>1</sup> SARS-CoV-2 mutations in these samples show a unique pattern most closely related to the earliest known variants, but we could not find exact match neither in the published literature nor in the most comprehensive database of GISAID. The absence of more precise information on the contaminator samples leaves open the possibilities that either there are known samples that have escaped our attention, or that there are unpublished results that may be key to identify the origin of SARS-CoV-2.

## Materials and Methods

The analysed sequencing project can be found under project ID PRJNA692319 but after the publication of the first preprint<sup>1</sup> the authors have removed the public access to the affected three samples.

The mitochondrial genome database of the 1000 Genome Project<sup>9</sup> was downloaded from [http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/ALL.chrMT.phase3\\_callmom-v0\\_4.20130502.genotypes.vcf.gz](http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/ALL.chrMT.phase3_callmom-v0_4.20130502.genotypes.vcf.gz)

We have used The bowtie2 alignment software<sup>39</sup> with the `-local` to align the raw sequencing reads to all vertebrate mitochondrial genomes, downloaded from <https://ftp.ncbi.nlm.nih.gov/refseq/release/mitochondrion/>. The `samtools coverage` command<sup>40</sup> was used to calculate the depth and coverage for the mitochondrial genomes. For realignment to the human mitochondrial genome alone, we have used the slower but more sensitive `bowtie2 -local -very-sensitive-local` command. Mitochondrial genome coverages were visualized by the DNA Features Viewer Python library<sup>41</sup>.

The `mixemt` tool<sup>8</sup> was used to determine the possible haplogroup of the human mitochondria.

All the other analysis was done with custom scripts using `bash`, `python`<sup>42</sup> and `jupyter`<sup>43</sup>.

## Acknowledgments

The authors thanks Jesse Bloom for insightful discussion and comments on the manuscript. This work was financed by EU Horizon 2020 programs VEO No. 874735 and BY-COVID No. 101046203. Special thanks for the original researchers for collecting and analysing the PRJNA692319 samples and for all others who upload full raw sequencing data and metadata to ENA and SRA and other public archives.

## References

1. Csabai, I., Papp, K., Visontai, D., Stéger, J. & Solymosi, N. Unique sars-cov-2 variant found in public sequence data of antarctic soil samples collected in 2018-2019. *Submitted*, DOI:10.21203/rs.3.rs-1177047/v1 0, 0 (2022).
2. Prjna692319 data set. <https://www.ebi.ac.uk/ena/browser/view/PRJNA692319>. Accessed: 2021-12-10.
3. Authors of PRJNA692319, personal communication.
4. Saved sequence read archive (sra) page for the suppressed sample srr13441704. <https://web.archive.org/web/20220104013347/https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR13441704>. Accessed: 2022-01-04.

5. Saved sequence read archive (sra) page for the re-enabled sample srr13441704. <https://archive.ph/2o7WK>. Accessed: 2022-02-03.
6. Effects of index misassignment on multiplexing and downstream analysis. <https://www.illumina.com/content/dam/illumina-marketing/documents/products/whitepapers/index-hopping-white-paper-770-2017-004.pdf>. Accessed: 2022-01-13.
7. Barcode misassignment with multiplexing on the hiseq 4000. <https://rtsf.natsci.msu.edu/genomics/tech-notes/barcode-misassignment-with-multiplexing-on-the-hiseq-4000/>. Accessed: 2022-01-13.
8. Vohr, S. H. *et al.* A phylogenetic approach for haplotype analysis of sequence data from complex mitochondrial mixtures. *Forensic Science International: Genetics* **30**, 93–105 (2017).
9. Siva, N. 1000 genomes project. *Nature biotechnology* **26**, 256–257 (2008).
10. Temmam, S. *et al.* Coronaviruses with a sars-cov-2-like receptor-binding domain allowing ace2-mediated entry into human cells isolated from bats of indochinese peninsula. (2021).
11. Hsieh, E. A new laboratory animal (cricetulus griseus). *Nat. med. J. China* **5**, 20 (1919).
12. Feeney, W. P. The chinese or striped-back hamster. In *The Laboratory Rabbit, Guinea Pig, Hamster, and Other Rodents*, 907–922 (Elsevier, 2012).
13. Song, Z. *et al.* From sars to mers, thrusting coronaviruses into the spotlight. *viruses* **11**, 59 (2019).
14. Bertzbach, L. D. *et al.* Sars-cov-2 infection of chinese hamsters (cricetulus griseus) reproduces covid-19 pneumonia in a well-established small animal model. *Transboundary emerging diseases* **68**, 1075–1079 (2021).
15. Graham, R. L., Donaldson, E. F. & Baric, R. S. A decade after sars: strategies for controlling emerging coronaviruses. *Nature Reviews Microbiology* **11**, 836–848 (2013).
16. Fukushi, S. *et al.* Amino acid substitutions in the s2 region enhance severe acute respiratory syndrome coronavirus infectivity in rat angiotensin-converting enzyme 2-expressing cells. *Journal virology* **81**, 10831–10834 (2007).
17. Lin, S.-C., Leng, C.-H. & Wu, S.-C. Generating stable chinese hamster ovary cell clones to produce a truncated sars-cov spike protein for vaccine development. *Biotechnology progress* **26**, 1733–1740 (2010).
18. Sinegubova, M. V., Orlova, N. A., Kovnir, S. V., Dayanova, L. K. & Vorobiev, I. I. High-level expression of the monomeric sars-cov-2 s protein rbd 320-537 in stably transfected cho cells by the eef1a1-based plasmid vector. *PloS one* **16**, e0242890 (2021).
19. Cantuti-Castelvetri, L. *et al.* Neuropilin-1 facilitates sars-cov-2 cell entry and infectivity. *Science* **370**, 856–860 (2020).
20. Reed, K. D. Viral zoonoses. *Reference Module Biomedical Sciences* (2018).
21. Sheets, R. History and characterization of the vero cell line. *US Food Drug Administration CfBEaR, ed. US Food Drug Administration, Silver Spring, MD* (2000).
22. Yasumura, Y. The research for the sv40 by means of tissue culture technique. *Nippon Rinsho* **21**, 1201–1219 (1963).
23. Vero c1008 [vero 76, clone e6, vero e6]s. [https://www.sigmaaldrich.com/HU/en/product/sigma/cb\\_85020206](https://www.sigmaaldrich.com/HU/en/product/sigma/cb_85020206). Accessed: 2022-01-13.
24. Ge, X.-Y. *et al.* Isolation and characterization of a bat sars-like coronavirus that uses the ace2 receptor. *Nature* **503**, 535–538 (2013).
25. Menachery, V. D. *et al.* A sars-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nature medicine* **21**, 1508–1513 (2015).
26. Kaina, B. On the origin of sars-cov-2: Did cell culture experiments lead to increased virulence of the progenitor virus for humans? *vivo* **35**, 1313–1326 (2021).
27. Sha, Y. *et al.* A convenient cell fusion assay for the study of sars-cov entry and inhibition. *IUBMB life* **58**, 480–486 (2006).
28. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *nature* **579**, 270–273 (2020).
29. Ogando, N. S. *et al.* Sars-coronavirus-2 replication in vero e6 cells: replication kinetics, rapid adaptation and cytopathology. *The Journal general virology* **101**, 925 (2020).
30. Pubmed search for sars-cov-2 + vero + e6 keywords. <https://pubmed.ncbi.nlm.nih.gov/?term=sars-cov-2+vero+e6>. Accessed: 2022-01-13.

31. Lin, L., Li, Q., Wang, Y. & Shi, Y. Syncytia formation during sars-cov-2 lung infection: a disastrous unity to eliminate lymphocytes. *Cell Death & Differentiation* **28**, 2019–2021 (2021).
32. Buchrieser, J. *et al.* Syncytia formation by sars-cov-2-infected cells. *The EMBO journal* **39**, e106267 (2020).
33. Chou, C.-F. *et al.* A novel cell-based binding assay system reconstituting interaction between sars-cov s protein and its cellular receptor. *Journal virological methods* **123**, 41–48 (2005).
34. Guillon, P. *et al.* Inhibition of the interaction between the sars-cov spike protein and its cellular receptor by anti-histo-blood group antibodies. *Glycobiology* **18**, 1085–1093 (2008).
35. Yan, H. *et al.* Sars coronavirus induces apoptosis in vero e6 cells. *Journal medical virology* **73**, 323–331 (2004).
36. Nie, J. *et al.* Establishment and validation of a pseudovirus neutralization assay for sars-cov-2. *Emerging microbes & infections* **9**, 680–686 (2020).
37. Wong, L.-Y. R. *et al.* Middle east respiratory syndrome coronavirus orf8b accessory protein suppresses type i ifn expression by impeding hsp70-dependent activation of irf3 kinase ikk $\epsilon$ . *The Journal Immunology* **205**, 1564–1579 (2020).
38. Covid-19 related cell line resources. <https://cell.brc.riken.jp/en/covid-19-related-cell-lone-resources>. Accessed: 2022-01-13.
39. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with bowtie 2. *Nature methods* **9**, 357–359 (2012).
40. Danecek, P. *et al.* Twelve years of samtools and bcftools. *Gigascience* **10**, giab008 (2021).
41. Dna features viewer python visualization code. <https://edinburgh-genome-foundry.github.io/DnaFeaturesViewer/>. Accessed: 2022-01-13.
42. Van Rossum, G. & Drake, F. L. *Python 3 Reference Manual* (CreateSpace, Scotts Valley, CA, 2009).
43. Kluyver, T. *et al.* Jupyter notebooks – a publishing format for reproducible computational workflows. In Loizides, F. & Schmidt, B. (eds.) *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, 87 – 90 (IOS Press, 2016).



**Table 1. Host mtDNA coverage and depths for the samples. R1 and R2 mates were calculated separately. Only top 10 most abundant Species are shown for each Run. Green background highlights the samples that contained significant amount of SARS-CoV-2 reads. Blue background highlights coverage values above 30.**

Run	Genome ID	R1 coverage	R1 depth	R2 coverage	R2 depth	Species
SRR13441701	NC_018784.1	0.0	0.00	4.28	0.31	Mobula japonica
SRR13441701	NC_020001.1	3.78	0.26	3.85	0.24	Sooglossus thomasseti
SRR13441701	NC_029404.1	1.09	0.03	2.92	5.10	Geotria australis
SRR13441701	NC_057519.1	1.96	0.28	2.36	0.39	Didelphis pernigra
SRR13441701	NC_020599.1	1.02	0.01	2.24	0.58	Tachycineta cyaneoviridis
SRR13441701	NC_012769.1	2.25	0.07	2.2	0.08	Eulemur fulvus mayottensis
SRR13441701	NC_020605.1	0.0	0.00	2.17	0.10	Progne chalybea
SRR13441701	NC_011944.1	1.59	0.41	1.91	44.77	Thylacinus cynocephalus
SRR13441701	NC_027840.1	0.96	0.08	1.9	0.12	Amazona ochrocephala
SRR13441701	NC_012766.1	2.31	0.07	1.86	0.04	Eulemur fulvus fulvus
SRR13441702	NC_029404.1	0.91	0.01	2.68	0.31	Geotria australis
SRR13441702	NC_020001.1	3.66	0.09	2.65	0.09	Sooglossus thomasseti
SRR13441702	NC_030333.1	1.0	0.05	2.0	0.08	Telmatobius chusmisensis
SRR13441702	NC_018784.1	0.83	0.01	1.95	0.03	Mobula japonica
SRR13441702	NC_021952.1	1.51	0.03	1.69	0.03	Leontopithecus rosalia
SRR13441702	NC_035810.1	1.81	0.79	1.6	0.66	Lutra sumatrana isolate
SRR13441702	NC_057520.1	1.27	0.03	1.59	0.04	Lutreolina crassicaudata
SRR13441702	NC_057519.1	2.06	0.05	1.57	0.07	Didelphis pernigra
SRR13441702	NC_050286.1	0.81	0.01	1.56	0.12	Hirundo smithii
SRR13441702	NC_041147.1	0.62	0.02	1.53	0.03	Isopachys gyldenstolpei
<b>SRR13441704</b>	NC_012920.1	<b>61.55</b>	2.05	<b>89.17</b>	18.45	Homo sapiens
<b>SRR13441704</b>	NC_007936.1	20.89	0.26	<b>78.01</b>	4.78	Cricetulus griseus
<b>SRR13441704</b>	NC_011137.1	17.27	0.30	<b>62.31</b>	3.48	Homo sapiens neanderthalensis
<b>SRR13441704</b>	NC_008066.1	15.01	0.50	<b>48.03</b>	4.68	Chlorocebus sabaeus
<b>SRR13441704</b>	NC_013993.1	5.21	0.07	<b>35.17</b>	0.96	Homo sp. Altai
<b>SRR13441704</b>	NC_023100.1	4.61	0.05	<b>30.75</b>	0.76	Homo heidelbergensis
<b>SRR13441704</b>	NC_053822.1	0.18	0.00	19.58	0.22	Cricetulus barabensis
<b>SRR13441704</b>	NC_025654.1	3.11	0.08	8.51	0.16	Anser indicus
<b>SRR13441704</b>	NC_009748.1	0.0	0.00	7.47	0.23	Chlorocebus tantalus
<b>SRR13441704</b>	NC_034277.1	1.99	0.04	7.28	0.09	Chlorocebus djamdjamensis isolate
<b>SRR13441705</b>	NC_012920.1	<b>47.98</b>	2.24	<b>89.03</b>	17.27	Homo sapiens
<b>SRR13441705</b>	NC_007936.1	10.67	0.12	<b>73.07</b>	4.18	Cricetulus griseus
<b>SRR13441705</b>	NC_011137.1	16.76	0.66	<b>59.51</b>	3.24	Homo sapiens neanderthalensis
<b>SRR13441705</b>	NC_008066.1	11.34	0.33	<b>43.28</b>	4.33	Chlorocebus sabaeus
<b>SRR13441705</b>	NC_023100.1	4.91	0.14	29.8	0.84	Homo heidelbergensis
<b>SRR13441705</b>	NC_013993.1	4.66	0.12	25.95	0.83	Homo sp. Altai
<b>SRR13441705</b>	NC_023832.1	4.98	0.33	11.69	0.49	Anser cygnoides
<b>SRR13441705</b>	NC_053822.1	1.14	0.01	9.08	0.14	Cricetulus barabensis
<b>SRR13441705</b>	NC_011196.1	7.43	0.54	8.8	0.31	Anser anser
<b>SRR13441705</b>	NC_024933.1	1.76	0.03	7.68	0.23	Chlorocebus cynosuros
<b>SRR13441708</b>	NC_012920.1	<b>68.27</b>	3.04	<b>89.15</b>	17.16	Homo sapiens
<b>SRR13441708</b>	NC_007936.1	15.85	0.29	<b>74.71</b>	3.91	Cricetulus griseus
<b>SRR13441708</b>	NC_011137.1	24.05	0.60	<b>65.2</b>	3.40	Homo sapiens neanderthalensis
<b>SRR13441708</b>	NC_008066.1	15.08	0.52	<b>43.94</b>	4.06	Chlorocebus sabaeus
<b>SRR13441708</b>	NC_013993.1	10.54	0.16	<b>33.02</b>	1.03	Homo sp. Altai
<b>SRR13441708</b>	NC_023100.1	8.23	0.12	29.15	0.76	Homo heidelbergensis
<b>SRR13441708</b>	NC_053822.1	1.14	0.01	11.73	0.14	Cricetulus barabensis
<b>SRR13441708</b>	NC_009747.1	3.46	0.05	8.76	0.29	Chlorocebus pygerythrus
<b>SRR13441708</b>	NC_024933.1	2.06	0.02	6.28	0.17	Chlorocebus cynosuros
<b>SRR13441708</b>	NC_034276.1	1.36	0.01	5.57	0.12	Chlorocebus aethiops x