Running title: An autoencoder-classified cluster of SARS-CoV-2 strain with two mutations in helicase

Clusters consisting only of virus types with two mutations in the helicase found by Autoencoder analysis in Washington State, USA

Jun Miyake*[1,2,3†], Mitsuaki Yoshino[4†], Takaaki Sato[1], Hirohiko Niioka[5], Yasushi Sakata[3], Yoshihisa Nakazawa[2]

[1] Department of Material and Life Science, Graduate School of Engineering, Osaka University, Suita, Osaka, Japan.
[2] Hitz Research Alliance Laboratory, Graduate School of Engineering, Osaka University, Suita, Osaka, Japan.
[3] Global Center for Medical Engineering and Informatics, Osaka University, Suita, Osaka, Japan.
[4] National Institute of Biomedical Innovation, Health and Nutrition, Ibaraki, Osaka, Japan.
[5] Institute for Datability Science, Osaka University, Suita, Osaka, Japan.

* Corresponding Author (email-address: jun_miyake@bpe.es.osaka-u.ac.jp)
[†]These authors are equally contributed.

Keywords: COVID-19, SARS-CoV-2 (nCoV-2019), Genome, Helicase, Mutant, Autoencoder, Washington State, Classification, Cluster

Abstract

Using an autoencoder-based analysis to classify genomes of SARS-CoV-2 coronaviruses, we found a cluster consisting only of a specific genotype with two mutations in the helicase. This virus genotype, called C-type SARS-CoV-2, was almost exclusively prevalent in the United States from March to July 2020. This type of virus, characterized by a pair of the C17747T (P504L) and A17858G (Y541C) mutations on the nsp13 gene, had never been highly prevalent at any other time or in any other part of the world. In the U.S., Washington State was the center of the epidemic, and

the C-type viruses, along with the viruses with wild-type helicase, seemed to have aroused the pandemic. In Washington State, USA, the CoViD-19 epidemic during the first two months of the year, starting at the end of February 2020, was mainly caused by the type-C virus. During this period, the infection spread rapidly; from May onwards, the number of viruses with wild-type helicases became higher than that of type-C viruses, and no type-C viruses have been collected since early July. The involvement of the helicase in this COVID-19 disease was discussed.

Introduction

Coronaviruses have had a major impact on human society since the 2002–2003 SARS epidemic (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. 2020). With the exception of influenza, there are few other viral infections that have been so prevalent, and it is important to elucidate the characteristics of viruses (Hua *et al.* 2020). There are questions about how host cells attack coronaviruses and whether there are other ways to defend against them besides antibodies. The coronavirus–host cell interactions such as the viral RNA-synthesizing machinery and the evasion of host innate immune responses are detailed in a review (de Welde *et al.* 2018).

Host organisms, including humans, have the ability to detect viral infection as "non-self-nucleic acid invasion," which is expected to trigger antiviral innate immunity and suppress viral replication. New coronaviruses have continued to mutate at a high rate (Gómez-Carballa *et al.* 2020; Jones and Manrique 2020; Nie *et al.* 2020; Rochman *et al.* 2020). Since its emergence, new variants have appeared one after another, and there are now more than eight in total (Miyake et al. 2021). The types, viral infections, disease patterns, and mutations of the encoded proteins have not yet been comprehensively understood. We believe that significant advances in our understanding of this complex system will be made in the near future.

The viral helicase (nsp13 in ORF1ab), which plays a central role in the immune response of host cells, is stably maintained in various strains, and a mutant form was first isolated in Washington State, USA in February 2020 (Chan et al. 2020). This variant is present only in a specific genotype (Zhao et al. 2020), which Chan et al. define as

coevolutionary variant group 4 (CEVg4). In our previous paper (Miyake et al. 2021), we presented a method to analyze clusters of viral genomes using an autoencoder. Among the clusters classified by this method, those consisting only of genomes with mutations in the nsp13 gene on ORF-1ab (C-type SARS -CoV-2 virus genome). Since the other clusters did not show the said mutation in nsp13, it was considered that the genomes constituting the clusters had some aggregate characteristics. A closer look at the genomic variation revealed that the helicase translated and synthesized from nsp13 had double mutations, P504L and Y541C.

In silico analysis to evaluate protein-drug interactions suggests that helicase, the nsp13 product with the double mutation of P504L and Y541C, has an altered shape of the ATP-binding site (Ugurel et al.). In this paper, along with the analysis of the clusters, we also examined their relationship to disease during the pandemic. The ration of wild type (WT)/C SARS-CoV-2 virus ratio based on the weekly number of viruses classified by the autoencoder and compared the relationship with the number of infected people and deaths. In the case of the pandemic in Washington State, USA, we found that the type C virus was dominant in the first two months, and that the virus with wild-type helicase became overwhelmingly dominant after May 2020. We will also discuss the involvement of helicases in the current COVID-19 disease.

Methods

We analyzed the coronavirus epidemic in Washington State, U.S.A. from 2020/2/29 to 2020/7/31. Genome sequence data of novel coronaviruses were collected from the NCBI Virus database (downloaded on February 26, 2021). Only complete genome sequences were analyzed (genes with more than two consecutive bases were excluded from the analysis). Cluster analysis of gene sequences using autoencoders was performed using the Tensor-Flow library (downgraded from V2.0 to V1.0). computers equipped with GPUs (NVIDIA Quadoro P-6000) were used. The autoencoder previously classified the genotypes of the human leukocyte antigen A (HLA-A) gene (Miyake et al. 2018): a histogram of the frequency of occurrence of 1,024 5-mer words in the SARS-CoV-2 ORF1ab gene was generated using a four-layer autoencoder (1,000,000 epochs with a batch size of 96) was compressed into a three-dimensional array, which were plotted in

three-dimensional (3D) space as (x, y, z) coordinates. As a result, 31,050 SARS-CoV-2 ORF1ab genes were classified into eight categories (A1, A2, B1, B2, C, D, E1, and E2 types) (Miyake et al. 2021). Epidemiological data were obtained from the Centers for Disease Control and Prevention (CDC) database.

Based on these autoencoder-classified SARS-CoV-2 ORF1ab genes, we examined time courses of weekly ratios of SARS-CoV-2 viral genomes with wild-type (WT) helicase (WT helicase genomes belonging to clusters A1, A2, B1, B2, D, E1 and E2) to those of C-type helicase (C-type genomes belonging to cluster C) that were collected in Washington State. Note that the C genome was collected only during weeks 4-20 starting from February 1, 2020; the period starting from week 21 was excluded from the analysis.

Results

The ORF1ab gene was extracted from the SARS-CoV-2 genome and classified into eight clusters in three-dimensional space by autoencoder. Among them, C-type was located at the far end of the cluster spread. It was a small cluster with relatively clear boundaries. ORF1ab genes of cluster C which have C-type helicase were expanded in March (Fig. 1a) and disappeared after July 3, 2020 (Fig. 1b).

Helicase is an essential enzyme for viral replication, but a special type of mutant was found in the United States in February to July, 2020. This mutant is characterized by the presence of mutations in the nsp13 (helicase) gene of ORF1ab that differ by two bases from the wild type (C17747T (P504L) and A17858G (Y541C)) (Chan *et al*., 2020). C-type of ORF1ab was consisted only with the mutant genome. The mutant was not found in any other clusters. We will refer to this helicase as helicase-C in this paper. Note that there is also a helicase in which only one of the two mutations was different, but due to the very small number, it was excluded from this analysis. This cluster (category) corresponds to the CEVg4 reported by Chan *et al*. in their combinatorial analysis of genomic variation (Chan *et al*., 2020).

The classification of the clusters as in Fig.1 were plotted against the time axis and the distance of the 3D data from the center for nine states in the United States (Fig. 2).

Each cluster was plotted in a different color (A1, red; A2, pink; B1, green; B2, light green; C, purple; D, orange; E1, blue; E2, light blue). The vertical axis is the distance of the 3D data used in the autoencoder classification (The nucleotide sequence (29903 nt in the reference sequence) was compressed into 3 dimensions by autoencoder analysis. The genomes in the cluster were easily identifies by their positions. Weekly numbers of helicase-C genomes in the United States, Washington State, and Australia were shown in Fig. 3.

The usefulness of the autoencoder classification method is that it can automatically classify the genomes into clusters without knowing the differences in gene sequences beforehand, it can handle a huge number of genomes, and it is easy to understand the temporal variation. In other words, it is a way to spatially represent the "artificial intelligence recognition" of the full-length sequence of genes. The classification of clusters depends on where the boundaries are drawn, so there can be some differences in the number of genes included in some cases.

The country subcategories of helicase-C SARS-CoV-2 genomes were 1,460 (97.6%) in the United States, 32 (2.1%) in Australia, 2 in the United Kingdom, 1 in Peru, and 1 in Puerto Rico (country data were taken from NCBI Virus Database). Australia and less countries were not examined in this study due to small sample size.

In the U.S., helicase-C has been detected in areas where the SARS-CoV-2 outbreak has caused infections, such as California, New York, and Washington State. In the U.S., the state of Washington had the highest number of database-registered helicase-C genomes at 68.9% (67.4% worldwide) throughout the period of February to June. This is followed by California (10.5%). Other states were Minnesota (2.8%), Utah (2.8%), Florida (1.5%), and Massachusetts (1.4%). The portion of the infections and the high mortality rate in Washington State is remarkable. Washington State was thought to be the single source of the outbreak and that it spread to other states.

In Washington State, helicase-C SARS-CoV-2 genomes were initially predominant in the four–eight weeks from February 1, 2020. The WT/C-type helicase ratio logarithmically increased with a kink at nine weeks when the dominance of WT helicase genomes was initiated (Fig. 4a). For comparison, weekly numbers of cumulative

158 COVID-19 cases and deaths in Washington State (obtained from the Center for Disease
159 Control and Prevention (CDC) webpage) were shown (Fig. 4b). Weekly numbers of
160 cumulative cases also logarithmically increased with a kink at nine weeks when the
161 growth rate markedly decreased. The cumulative number of deaths showed a one-week
162 lagged kink (week 10) in the logarithmic increase, with a marked decrease in the late
163 phase. The coincidence of sinks in WT/C helicase ratios, cases, and deaths in Washington
164 State with a 1-2 week lag suggests that the change in dominance of the SARS-CoV-2
165 helicase-C genome to the WT helicase-C genome may be responsible for the decrease in
166 cases and deaths.
167

168 ## Discussion

169 In the cluster analysis using the autoencoder, the genome of SARS-CoV-2 with helicase
170 C could be extracted as one cluster. The other genomes with helicase WT were distributed
171 in several clusters. These results indicate that the autoencoder can identify the ORF1ab
172 gene of SARS-CoV with only two mutations in the helicase. Although the genomic
173 sequence of the helicase is highly conserved in nature, mutations in such a highly
174 conserved sequence may form specific clusters that are clearly separated by autoencoder
175 analysis. The ability of autoencoders to extract genomes with significant mutations in
176 specific gene regions as separate clusters will be beneficial not only for helicases but also
177 for a variety of other application.

178 Type-C SARS-CoV-2 virus (hereafter referred to as "type-C virus") was first
179 discovered in Washington State on February 20 (Chan et al. 2020). in June (after week
180 21), the number of type-C virus collections has decreased significantly. The last virus was
181 obtained in Florida on July 2. In the early stages of the pandemic (4-8 weeks counting
182 from February 1, 2020), type-C virus initially predominated and may have contributed to
183 the number of COVID-19 cases. By weeks 10-20, the WT helicase genome predominated.
184 Looking at the semilog plot in Fig. 4, it is understood that the ratio of deaths/infected
185 cases was considerably larger in the early phase than in the later phase. The inflection
186 point of the number of infected and deaths was off by almost a week. Assuming a one-
187 week gap between infection and outcome/death, the ratio of deaths/infected may have
188 been temporarily more than twice as high in the early phase (4-8 weeks) as in the late

phase. In the late phase (10-20 weeks), this ratio was consistently less than 6%. In the later stages, this ratio was consistently less than 6%. If this large difference is dependent on the type of virus, it should be fully considered as a preparation for future viral diseases. The emergence, disappearance, and pathogenesis of type-C virus A detailed analysis of the mechanisms, mortality dependence, and other factors may be useful as a key to developing effective treatments and antiviral measures.

All C-type viruses contain a variant of the helicase, helicase-C, which is highly suspected as a cause of serious disease. Helicase is a protein that unwinds nucleic acids and is essential for viral genome replication. Adedeji *et al*. found that helicase activity was enhanced in the presence of nsp12 in the SARS-CoV epidemic of 2002-2003 (Adedeji *et al*. .2012). Kindler *et al*. (Kindler *et al*. 2017; Ancar *et al*. 2020) and Deng *et al*. (Deng *et al*. 2017; Deng and Baker 2018) considered the following scheme. The viral RNA forms a double-stranded structure at the poly-U sequence. If the virus can use RNA-degrading enzymes (EndoU) to cleave its own RNA-poly U sequence, the immune response of the host cell is inactivated. The structure and interactions of the proteins discussed here are being studied. Pillon *et al*. published the four-dimensional protein structure of SARS-CoV-2 nsp15 using Cryo-EM (Pillon *et al*. 2021). Perry *et al*. also studied that nsp 15 protein is surrounded by nsp14, 16, nsp7, 8, and 12 proteins, as well as nsp13 (helicase) using Cryo-EM (Perry *et al*. 2021).

Several researchers have pointed out that helicases are important targets for viral clearance (Gurung 2020; Habtemariam *et al*. 2020; White *et al*. 2020). Possible strategies for drug discovery include the application of existing drugs (Li and De Clercq 2020; Pandey *et al*. 2020) and RNAi (Liu *et al*. 2020). Existing compounds that have the effect of inhibiting the RNA or protein molecules of helicases could also be used. Computer modeling studies are also underway (Gurung 2020; White *et al*. 2020). It is expected that various efforts will be made in drug development.

Various vaccines have been developed and used now. The biochemical reactions involved in viral replication in host cells may be used as a countermeasure. If it is possible to inhibit the intracellular replication mechanism common to various viruses, it may be an effective countermeasure.

Acknowledgements

We should like to express our sincere thanks to Dr. Tomonori Kimura of NIBIOHN for support and discussions. Thanks are to Hayao Nakamura, Yuta Nitada and Shunsuke Baba for programming and computer operation. This work was supported partially by Global Center for Medical Engineering and Informatics, Osaka University, Hitz Research Alliance Laboratory, Nihon Unisys, Ltd. and Japan Agency for Medical Research and Development (Grant Number 20bm0804008h0004.PI. Prof. S. Miyagawa).

References

Adedeji AO, Marchand B, Te Velthuis AJ, Snijder EJ, Weiss S, Eoff RL, Singh K, Sarafianos SG. Mechanism of nucleic acid unwinding by SARS-CoV helicase. *PLoS One.* 7:e36521 (2012). doi: 10.1371/journal.pone.0036521.

Ancar R, Li Y, Kindler E, Cooper DA, Ransom M, Thiel V, Weiss SR, Hesselberth JR, Barton DJ. Physiologic RNA targets and refined sequence specificity of coronavirus EndoU. *RNA.* 26:1976–1999 (2020). doi: 10.1261/rna.076604.120.

Centers for Disease Control and Prevention (CDC, Official Updates Coronavirus - COVID-19 in United States): https://coronavirus.dc.gov/?gclid=Cj0KCQiA5vb-BRCRARIsAJBKc6L-50jCtxFFaFZe05Os84e-Mb6dLiWHBepIx9bJdHXF2rABJmLpyyIaAmxBEALw_wcB.

Chan AP, Choi Y, Schork NJ. Conserved genomic terminals of SARS-CoV-2 as coevolving functional elements and potential therapeutic targets. *mSphere* 5: e00754-20 (2020). doi: 10.1128/mSphere.00754-20.

Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* 5:536–544 (2020). doi: 10.1038/s41564-020-0695-z.

de Wilde AH, Snijder EJ, Kikkert M, van Hemert MJ. Host factors in coronavirus replication, *Curr Top Microbiol Immunol.* 419:1–42 (2018). doi: 10.1007/82_2017_25.

Deng X, Hackbart M, Mettelman RC, O'Brien A, Mielech AM, Yi G, Kao CC, Baker SC. Coronavirus nonstructural protein 15 mediates evasion of dsRNA sensors and limits apoptosis in macrophages. *Proc Natl Acad Sci U S A.* 114:E4251–E4260 (2017). doi: 10.1073/pnas.1618310114.

Deng X, Baker SC. An "Old" protein with a new story: Coronavirus endoribonuclease is important for evading host antiviral defenses. *Virology.* 517:157–163 (2018). doi: 10.1016/j.virol.2017.12.024.

252 Gómez-Carballa A, Bello X, Pardo-Seco J, Martinón-Torres F and Salas A. Mapping genome
253 variation of SARS-CoV-2 worldwide highlights the impact of COVID-19 super-spreaders.
254 *Genome Res.* 30:1434–1448 (2020). .doi: 10.1101/gr.266221.120.
255 Gurung AB. *In silico* structure modelling of SARS-CoV-2 Nsp13 helicase and Nsp14 and
256 repurposing of FDA approved antiviral drugs as dual inhibitors. *Gene Rep.* 21:100860 (2020).
257 doi: 10.1016/j.genrep.2020.100860.
258 Habtemariam S, Nabavi SF, Banach M, Berindan-Neagoe I, Sarkar K, Sil PC, Nabavi SM. Should
259 we try SARS-CoV-2 helicase inhibitors for COVID-19 therapy? *Arch Med Res.* 51:733–735
260 (2020). doi: 10.1016/j.arcmed.2020.05.024.
261 Hua Li, Zhe Liu, Junbo Ge, Scientific research progress of COVID-19/SARS-CoV-2 in the first
262 five months, *J Cell Mol Med.* 24:6558–6570 (2020). doi: 10.1111/jcmm.15364.
263 Jones LR, Manrique JM. Quantitative phylogenomic evidence reveals a spatially structured
264 SARS-CoV-2 diversity. *Virology* 550:70–77 (2020). doi: 10.1016/j.virol.2020.08.010.
265 Kindler E, Gil-Cruz C, Spanier J, Li Y, Wilhelm J, Rabouw HH, Züst R, Hwang M, V'kovski P,
266 Stalder H, Marti S, Habjan M, Cervantes-Barragan L, Elliot R, Karl N, Gaughan C, van
267 Kuppeveld FJ, Silverman RH, Keller M, Ludewig B, Bergmann CC, Ziebuhr J, Weiss SR,
268 Kalinke U, Thiel V. Early endonuclease-mediated evasion of RNA sensing ensures efficient
269 coronavirus replication. *PLoS Pathog.* 13:e1006195 (2017). doi:
270 10.1371/journal.ppat.1006195.
271 Li G, De Clercq E. Therapeutic options for the 2019 novel coronavirus (2019-nCoV). *Nat Rev*
272 *Drug Discov.* 19:149–150 (2020). doi: 10.1038/d41573-020-00016-0.
273 Liu C, Zhou Q, Li Y, Garner LV, Watkins SP, Carter LJ, Smoot J, Gregg AC, Daniels AD, Jervey
274 S, Albaiu D. Research and Development on Therapeutic Agents and Vaccines for COVID-19
275 and Related Human Coronavirus Diseases. *ACS Cent Sci.* 6:315–331 (2020). doi:
276 10.1021/acscentsci.0c00272.
277 Miyake J, Kaneshita Y, Asatani S, Tagawa S, Niioka H, Hirano T. Graphical classification of
278 DNA sequences of HLA alleles by Deep learning. *Human Cell* 31:102–105 (2018). doi:
279 10.1007/s13577-017-0194-6.
280 Miyake J, Sato T, Baba S, Nakamura H, Niioka H, Hirano T, Nakazawa Y. Cluster analysis of
281 SARS-CoV-2 gene using deep learning autoencoder: Gene profiling for mutations and
282 transitions. *bioRxiv* 2021.03.16.435601 (2021). doi: 10.1101/2021.03.16.435601. Preprint.
283 NCBI Virus Database: https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/find-data/virus
284 Nie Q, Li X, Chen W, Liu D, Chen Y, Li H, Li D, Tian M, Tan W, Zai J. Phylogenetic and
285 analyses of SARS-CoV-2. *Virus Res.* 287:198098 (2020). doi: 10.1016/j.virusres.2020.198098.

Pandey SC, Pande V, Sati D, Upreti S, Samant M. Vaccination strategies to combat novel corona virus SARS-CoV-2. *Life Sci.* 256:117956 (2020). doi: 10.1016/j.lfs.2020.117956.

Perry JK, Appleby TC, Bilello JP, Feng JY, Schmitz U, Campbell EA. An atomistic model of the coronavirus replication-transcription complex as a hexamer assembled around nsp15. *bioRxiv* 2021.06.08.447516 (2021). doi: https://doi.org/10.1101/2021.06.08.447516. Preprint.

Pillon MC, Frazier MN, Dillard LB, Williams JG, Kocaman S, Krahn JM, Perera L, Hayne CK, Gordon J, Stewart ZD, Sobhany M, Deterding LJ, Hsu AL, Dandey VP, Borgnia MJ, Stanley RE. Cryo-EM structures of the SARS-CoV-2 endoribonuclease Nsp15 reveal insight into nuclease specificity and dynamics, *Nature Communications* 12, Article number: 636 (2021) https://doi.org/10.1038/s41467-020-20608-z

Rochman ND, Wolf YI, Faure G, Zhang F, Koonin EV. Ongoing adaptive evolution and globalization of Sars-Cov-2. *bioRxiv.* 2020.10.12.336644 (2020). doi: 10.1101/2020.10.12.336644. Preprint.

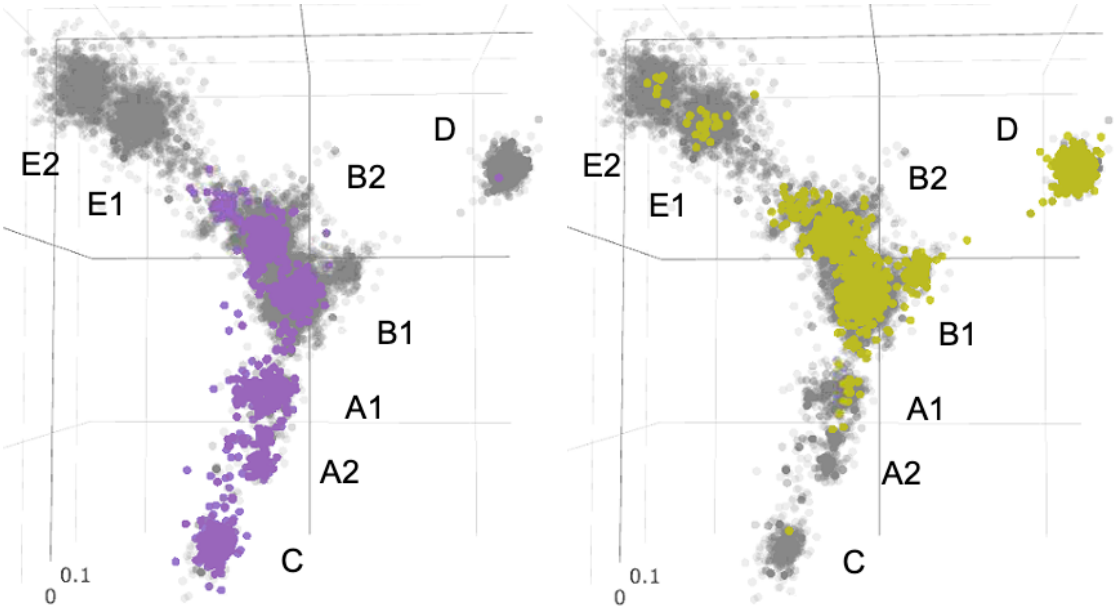
Ugurel OM, Mutlu O, Sariyer E, Kocer S, Ugurel E, Inci TG, Ata O, Turgut-Balik D. Evaluation of the potency of FDA-approved drugs on wild type and mutant SARS-CoV-2 helicase (Nsp13). *Int J Biol Macromol.* 163:1687–1696 (2020). doi: 10.1016/j.ijbiomac.2020.09.138.

White MA, Lin W, Cheng X. Discovery of COVID-19 inhibitors targeting the SARS-CoV-2 Nsp13 helicase. *J Phys Chem Lett.* 11:9144–9151 (2020). doi: 10.1021/acs.jpclett.0c02421.

Zhao J, Zhai X, Zhou J. Snapshot of the evolution and mutation patterns of SARS-CoV-2. *bioRxiv.* 2020.07.04.187435 (2020). doi: 10.1101/2020.07.04.187435. Preprint.

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

Figures



(a) (b)

Fig.1. Autoencoder clustered 3D view of ORF1ab gene of SARS-CoV-2 (March and July 2020).

The number of C-type ORF1ab genes expanded rapidly in March, and new appearances stopped as of July 3, 2020. (a) ORF1ab gene that appeared in March 2020 (purple). (b) ORF1ab gene that appeared in July 2020 (yellow-green). Shown in gray as background is the cluster of ORF1ab genes for the period 19/12/2019-2021/2/16.

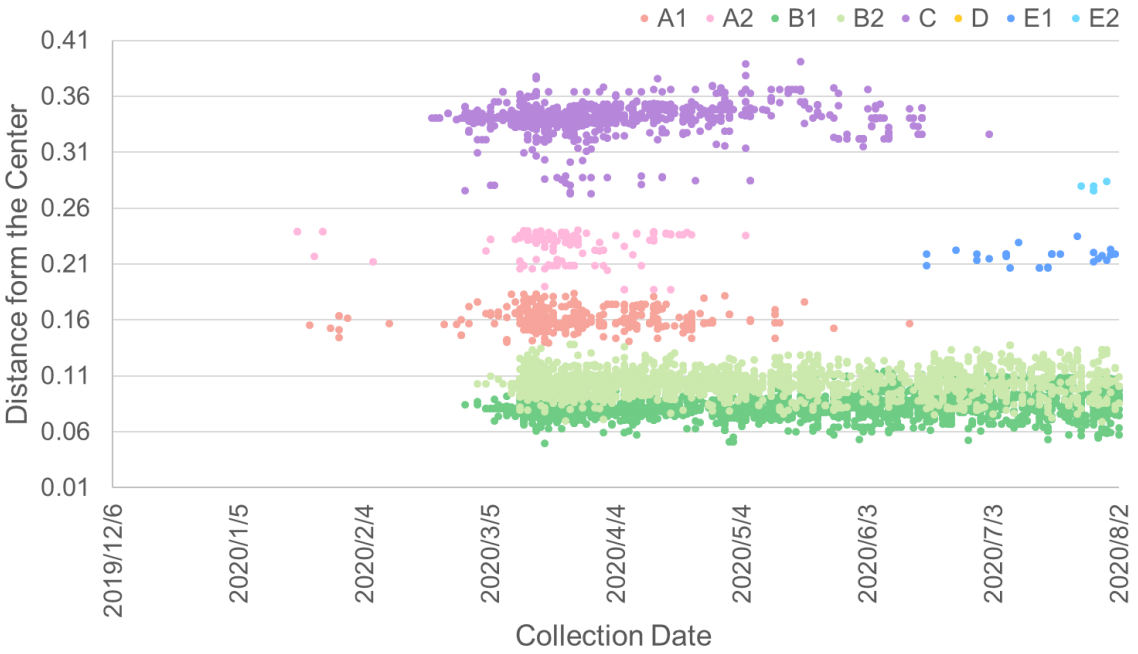
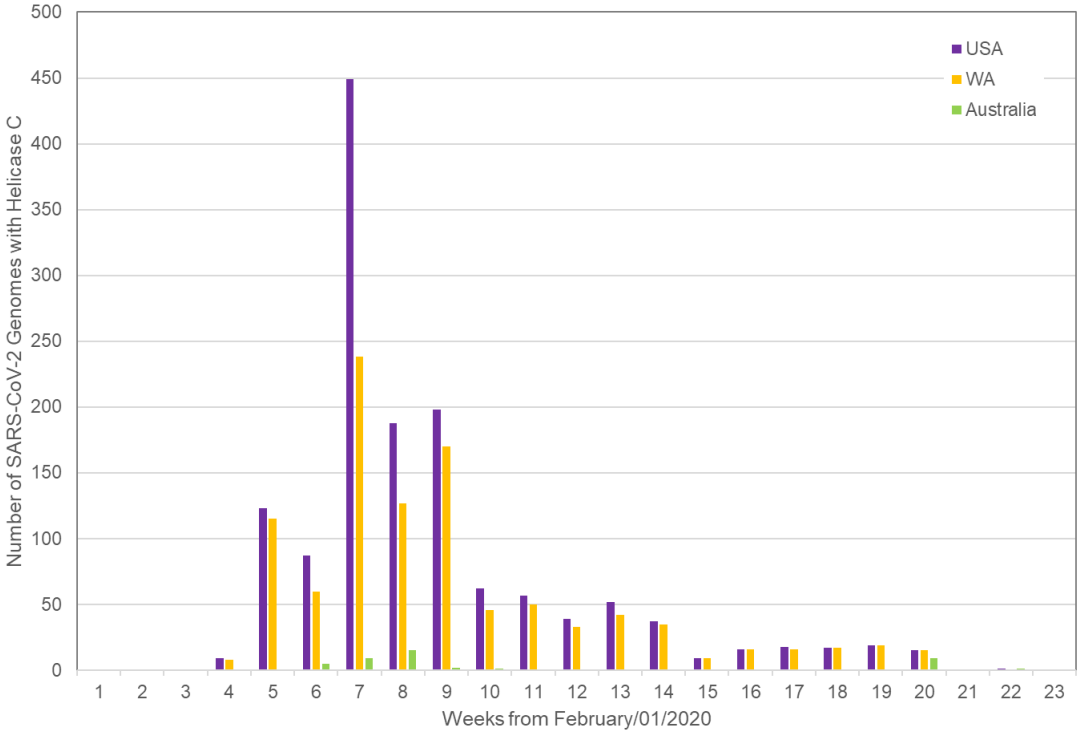


Fig.2. Evolution of clusters in nine US states.

Eight clusters classified by the autoencoder of the SARS-CoV-2 ORF1ab gene in nine U.S. states (WA, CA, UT, FL, MA, NY, MN, WI, ME) plotted in different colors by collection date (A1, red; A2, pink; B1, green; B2, light green; C, purple; D, orange; E1, blue; E2, light blue). Cluster C with helicase C appeared in February and disappeared after July 3, 2020.

340



341

342 **Fig. 3. Transition of C-type SARS-CoV-2 in USA, Washington State and Australia.**

343 Weekly numbers of autoencoder-classified SARS-CoV-2 ORF1ab genes in USA,

344 Washington State, and Australia are shown by number of weeks from

345 February/01/2020.

346

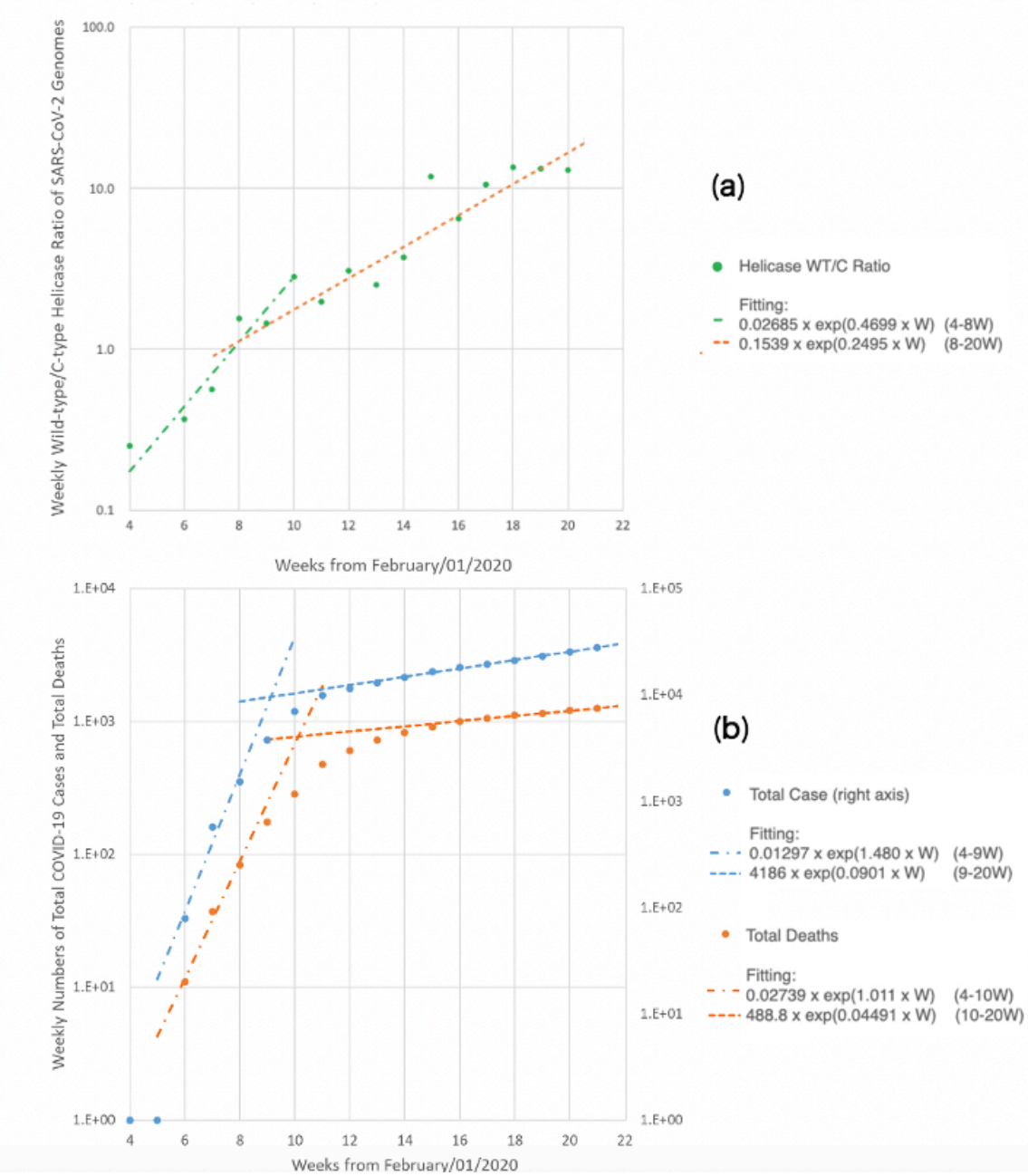


Fig. 4. Ratio of the numbers of wild-type and C-type helicase SARS-CoV-2 viruses collected in Washington State in USA.

(a) The ratio of wild-type/C-type helicase SARS-CoV-2 genomes (green) showed logarithmic increases with a kink around nine weeks from February 1, 2020 in Washington State in USA. The first half corresponds to the period when the C-type helicase viruses rapidly increased, and the second half to the period when the wild-type

356 helicase viruses became predominant. Total deaths (orange) and total cases (blue) also
357 showed logarithmic increases with kinks around ten and nine weeks, respectively.