Title page

Myeloablation-associated deletion of ORF4 in a human coronavirus 229E infection

Alexander L. Greninger[1,2,#], Gregory Pepper[1], Ryan C. Shean[1], Anne Cent[1], Isabel Palileo[1], Jane M. Kuypers[1], Joshua T. Schiffer[2,3], Keith R. Jerome[1,2]

[1]Department of Laboratory Medicine, University of Washington, Seattle, WA

[2]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA

[3]Department of Medicine, University of Washington, Seattle, WA

[#]corresponding author, agrening@uw.edu

Abstract

We describe metagenomic next-generation sequencing (mNGS) of a human coronavirus 229E from a patient with AML and persistent upper respiratory symptoms, who underwent hematopoietic cell transplantation (HCT). mNGS revealed a 548-nucleotide deletion, which comprised the near entirety of the ORF4 gene, and no minor allele variants were detected to suggest a mixed infection. As part of her pre-HCT conditioning regimen, the patient received myeloablative treatment with cyclophosphamide and 12 Gy total body irradiation. Iterative sequencing and RT-PCR confirmation of 4 respiratory samples over the 4-week peritransplant period revealed that the pre-conditioning strain contained an intact ORF4 gene, while the deletion strain appeared just after conditioning and persisted over a 2.5-week period. This sequence represents one of the largest genomic deletions detected in a human RNA virus and describes large-scale viral mutation associated with myeloablation for HCT.

Keywords: coronavirus, 229E, human coronavirus 229E, myeloablation, genome evolution, metagenomics, mNGS, irradiation, deletion, ORF4, ORF4a, ORF4b

34 **Introduction**

35 Coronaviruses are positive-stranded RNA viruses with large genomes, ranging in
36 size from 25-35kb. Although much attention has been paid to the epidemic potential and
37 high mortality rates for MERS and SARS coronaviruses, the most common human
38 coronaviruses (HCoV) include 229E, OC43, HKU1, and NL63 viruses [1]. HCoV 229E and NL63
39 are alphacoronaviruses with genome lengths of 27-28kb, and HCoV HKU1 and OC43 are
40 betacoronaviruses with genome lengths around 30-31kb [2]. Few studies have been
41 conducted on the genome evolution of the common HCoV species [3,4].

42 HCoVs enact a major burden on hematopoietic cell transplant (HCT) patients with
43 >10% cumulative incidence of HCoV infection among HCT by day 100 of transplant [5]. HCoV
44 in HCT patients is associated with prolonged shedding, with a median duration of 3 weeks [5].
45 Conditioning regimens for HCT usually consist of chemotherapy and irradiation that may
46 induce genomic alterations [6,7]. Treatment-related myeloid leukemias are well-known to be
47 associated with radiation therapy or chemotherapeutic agents such as cyclophosphamide [8].
48 Here we describe the generation of a large deletion in an RNA virus that was temporally
49 associated with the myeloablative conditioning regimen for HCT.

50

51 **Materials and Methods**

52 *Cell culture and qRT-PCR*

53 The HCoV 229E type strain ATCC VR-740 was obtained and used as a positive
54 control for cell culture. 200ul of each patient nasal swab in viral transport media (50ul
55 229E control/150ul 2% MEM) was inoculated onto MRC-5 cells and incubated at 37C in
56 stationary rack for 30 minutes. After incubation, 1.2mL of 2% MEM was added to each tube
57 and incubated at 37C. Each day tubes were read for cytopathic effect and 50uL of
58 supernatant was stored at -80C. 50uL culture supernatants were extracted on a Roche

59 MagNA Pure LC 2.0 and RNA was eluted in 100uL water. 10uL of RNA was used for qRT-
60 PCR with primers targeting the HCoV polymerase gene [9]. qRT-PCR on patient samples was
61 performed on 200uL of nasal swab in viral transport with the same extraction and PCR
62 conditions.
63
64 *Metagenomic next-generation sequencing and confirmatory RT-PCR with deep sequencing*
65 mNGS was performed as described previously [10]. Briefly, 20uL of extracted RNA
66 was treated with DNAse I and double stranded cDNA was created using random hexamers
67 with SuperScript III and Sequenase v2.0 (Thermo Fisher). mNGS libraries were created
68 using one-third volume NexteraXT reactions with 16 cycles of PCR amplification. mNGS
69 libraries from the deletion-containing specimens were prepared twice from RNA. Dual-
70 indexed libraries were sequenced on an Illumina MiSeq over several different types of runs
71 (1x190bp, 1x188bp, 2x94bp, 2x300bp) with consistent results. Sequencing data were
72 Nextera adapter and quality (Q30) trimmed using cutadapt, iteratively mapped to the HCoV
73 229E reference genome (NC_002645) using the Geneious v9.1 read aligner, de novo
74 assembled using SPAdes v3.9, and visualized using Geneious v9.1 and manually curated [11-
75 13]. The HCoV 229E reference genome and the strains sequenced in this study have an
76 average pairwise nucleotide identity of 98%. Majority consensus genomes from mNGS
77 were called using a minimum of 3X coverage. Reads for each specimen were also mapped
78 to the day -12 consensus genome to look for minor variants that arose during treatment.
79 Due to low coverage, minor variants were called with a minimum allele frequency of 10% at
80 11X coverage, after trimming of inversions due to Nextera artifact, variants within 10
81 nucleotides of end of read, and PCR duplicates.
82 Confirmatory gap junction RT-PCR was performed on 1uL of RNA extracted as
83 above using 35 cycles following manufacturer's recommended conditions (Tm 55C) using

84 quarter-volumes of the Qiagen One-Step RT-PCR kit and primers HCoV_24034F (5´ –
85 TTGTTGTGAATCAACTAAACTTCC – 3´) and HCoV_24808R (5´ –
86 ACACACCAGAGTAGTACATTAAC – 3´). Deep sequencing of the day -9 and day -2 PCR
87 amplicons was performed using 1ng of PCR product purified with the Zymo DNA Clean and
88 Concentrator Kit-5 with 1/10th volumes of the the Kapa HyperPlus kit ligation of Y-shaped
89 adapters followed by 7 cycles of Truseq-adapter dual-indexed PCR amplification [14].
90 Sequencing was performed using a 315x310bp run on an Illumina MiSeq. From the day -9
91 amplicon 36,880 paired-end reads were recovered, while from the day -2 amplicon 231,990
92 paired-end reads were recovered. Truseq-adapter and quality-trimmed using bbduk and
93 aligned to the ORF4 intact and deleted alleles present in the day -12 and day 15 consensus
94 genomes using the Geneious read aligner with no gaps or mismatches allowed [15]. The
95 methods were performed in accordance with relevant guidelines and regulations and
96 approved by the University of Washington Institutional Review Board.
97
98 **Results**
99 The virus analyzed was from a female in her 40s with acute myelogeneous leukemia
100 presenting for hematopoetic cell transplant. The previous year she had undergone 4 cycles
101 of chemotherapy (G-CLAM, G-CLA, cytarabine, and decitabine-primed MEC). Her prior
102 month's bone marrow evaluation showed 16% blasts by flow cytometry with normal
103 cytogenetics. On day –12 prior to transplant, she was found to be infected by a coronavirus
104 229E with a cycle threshold (Ct) of 30 and persisted at the same Ct on day -9 (Figure 1A).
105 Based on the patient's risk of relapsing leukemia and relative rarity of HCoV pneumonitis,
106 the decision was made to continue with HCT despite ongoing viral shedding. On days -7
107 and -6 she received cyclophosphamide 60mg/kg, and on days -4, -3, and -2 she received a
108 cumulative dose of twice-daily 2 Gy total body irradiation for a cumulative dose of 12 Gy

109 (Figure 1A). Other significant drugs the patient received during this time period included
110 acyclovir, granisetron, micafungin, voriconazole, and sulfamethoxazole-trimethoprim. A
111 mismatched unrelated donor peripheral blood stem-cell transplant was performed on day
112 0. Throughout the HCoV infection, the patient had mild nasal congestion but never required
113 supplemental oxygen. Her course was complicated by bacteremia on day 10, thought to be
114 due to grade 2 mucositis, and she was discharged on day 20.

115 Four nasal swab samples over a four-week period were available from the patient.
116 All qRT-PCR cycle thresholds (Ct) recovered over the four-week period were comparable,
117 ranging between 31.3-33.5 (Table). The percent of total reads mapping to HCoV 229E
118 ranged from 0.01-0.75%, consistent with the low concentrations of virus found by qRT-PCR
119 (Table). mNGS of the Day 15 post-transplant specimen revealed a large deletion that
120 encompassed 548 of the 660 nucleotides of the ORF 4 gene (Figure 1B) with average
121 coverage of 68X across the viral genome. Repeat library preparations of the same specimen
122 demonstrated the same deletion. The deletion was found whether mapping to the HCoV
123 229E reference genome or de novo assembly. No read was recovered within the area of the
124 deletion, suggesting the presence of a viral population entirely containing the genomic
125 deletion. mNGS of the day -12 and -9 specimens from the same patient revealed a
126 completely intact ORF4a/4b (Figure 1B). Other than the large deletion, no consensus
127 variants were observed between the HCoV 229E consensus genomes recovered from day -
128 12, day -9, and day 15 specimens. Mapping of reads from the day –12, day -9, day 15
129 samples to the day -12 consensus genome revealed two minor allele variants that reached
130 >10% allele frequency and had consistent allele frequencies between the day -12 and day -9
131 samples. The first was a c.12332A>G, p.K4111R change in the replicase protein that went
132 from undetectable in the day -12 and -9 specimens and was at 11.9% allele frequency in the

133 day 15 specimen. The second was a c.675_676insU that maintained at 10.6%, 10.8%, and
134 8.9% allele frequency in the day -12, day -9, and day 15 specimens.
135 No junctional reads across the 548-nucleotide deletion were found in either the
136 day -12 or day -9 specimen by mNGS. mNGS of the day -2 specimen also demonstrated the
137 548-nucleotide deletion with coverage across the genome consistent with RNA degradation
138 (Figure S1). No consensus variants were observed in the genomic regions covered in the
139 day -2 specimen. Confirmatory RT-PCR across the ORF 4a/4b gene demonstrated an intact
140 gene in the pre-myeloablation specimens and a smaller PCR product consistent with the
141 548-nucleotide deletion in the day -2 and day 15 post-myeloablation specimens (Figure 1C).
142 Deep sequencing of the day -9 and day -2 PCR amplicons revealed no junction reads from
143 the day -9 specimen mapping to the day 15 deletion strain locus with 11,279X coverage at
144 the same locus for the intact ORF4 gene in the day -12 strain (Figure S2). Conversely, only
145 one paired-end read from the day -2 specimen mapped to the deleted portion of the ORF4
146 gene locus in the day -12 strain with 158,938X coverage at the ORF4 locus in the day 15
147 deletion strain (Figure S2).
148 Attempts to culture each of the patient's specimens on MRC-5 cells proved
149 unsuccessful. No CPE was visualized for any of the patient's four clinical samples and HCoV
150 229E RNA levels in the culture supernatant were undetectable by qRT-PCR after
151 inoculation, while viral growth was apparent in the positive control HCoV 229E ATCC type
152 strain (Figure S3).
153
154 **Discussion**
155 We describe metagenomic detection of a large deletion in a HCoV 229E strain
156 temporally associated with a myeloablative regimen of cyclophosphamide and total body
157 irradiation in an HCT patient. The ORF4-deleted virus maintained the same viral load over

158 a 2.5 week period and was clonal by day 15 after transplant, suggesting that it was
159 replication competent in vivo. No other majority consensus variants in the HCoV genome
160 arose during the sampled period. One minor variant arose in the day 15 specimen that was
161 not present in the earlier specimens, while another minor variant maintained a similar
162 allele frequency throughout the sampled period. This rate of single nucleotide variant
163 evolution is consistent with the molecular clock of HCoV and with a study of that found no
164 HCoV genomic variants arose during the first month of infection [16]. The ability of the virus
165 to persist in vivo despite the near complete loss of ORF4 suggests that this accessory gene is
166 completely dispensable for HCoV 229E replication and may outcompete the wild-type virus
167 under certain conditions. Indeed, HCoV 229E type strain contains a 2-nucleotide deletion
168 that splits ORF4 into ORF4a/4b, and the alpaca alphacoronavirus contains a 1-nucleotide
169 insertion that also disrupts the reading frame [17,18]. Meanwhile, all other group1b
170 coronaviruses including all known HCoV 229E clinical strains contain an intact ORF4 [17,19].
171 This is the first detection of a non-intact ORF4 in HCoV 229E associated with human
172 infection.

173 The function of ORF4 in the alphacoronaviruses is unknown. The protein sequence
174 shows no significant alignment to any protein in Protein Data Bank by HHPred [20]. ORF4a of
175 the HCoV 229E type strain has been suggested to be a viroporin that possessed ion channel
176 activity [21]. Interestingly, replacement of ORF4 in the HCoV 229E type strain with Renilla
177 luciferase resulted in viral replication in vitro comparable to that of the wild-type strain,
178 consistent with our in vivo data that ORF4 is unlikely to be required for replication [22].
179 Taken together, these data suggest that ORF4 would make a poor antiviral target for the
180 human coronaviruses.

181 Many potential mechanisms for development of the deletion exist. The
182 chemotherapy and/or irradiation could have been directly toxic to the coronavirus genome.

183 Cyclophosphamide is a well-known clastogen and has recently been associated with
184 increased development of single nucleotide variations insertions and deletions [7,23].
185 Alternatively, the myeloablative regime could have removed immune pressure, allowing for
186 the loss of ORF4. It seems that in certain circumstances, deletion of ORF4 could prove a
187 selective advantage as it was the only allele present in the day 15 specimen.
188 This study has several limitations. We were unable to culture any of the patient's
189 HCoV specimens to demonstrate ability to replicate in vitro. This is not necessarily
190 surprising as HCoV 229E are not usually cultured in the clinical virology lab and the virus
191 was present at low concentrations in the sample. We are also limited in accurately calling
192 minor variants by the relatively low coverage of each viral genome. We did not attempt
193 reverse genetics of any alphacoronaviruses to definitively show ORF4 is dispensable for in
194 vitro replication, although the ability of the type strain to replicate despite a 2-nucleotide
195 deletion in ORF4 and the luciferase gene replacement experiment described above are
196 consistent with this possibility.
197 In summary, we describe a large deletion in HCoV 229E encompassing almost an
198 entire gene that was temporally associated with myeloblation. Little is known about the
199 genome evolution of HCoV 229E as only 20 genomes are currently available in NCBI
200 Genbank. Similarly sized (500-700 nucleotide) deletions have previously been found in
201 hepatitis D virus in patients receiving antiviral therapy [24]. Our study is also the first
202 demonstration of microbial evolution associated with cyclophosphamide treatment or
203 irradiation. Further research into how chemotherapy and irradiation affect the microbiome
204 is warranted. As mNGS is increasingly performed in the clinical lab [25], much intriguing
205 biology can be unlocked from clinical specimens, providing a continuing case for the
206 sequencing clinical microbial genomes [26–28].
207

208 *Data Availability Statement*

209 These sequences are deposited in NCBI Genbank under the strain names SC677 day 15

210 (KY369909), SC379 day -12 (KY621348), and SC399 day -9 (KY674914). The non-human

211 reads are deposited in the NCBI Short Read Archive under accessions SRR5809357-

212 SRR5809372, SRR5809395, SRR5816371-SRR5816372.

213

214 *Author Contributions*

215 ALG, GP, IP performed experiments; ALG, RCS, AC, JMK analyzed data; ALG, JTS, KRJ wrote

216 the paper.

217

218 *Competing Interests*

219 The authors declare they have no competing interests.

220

221 Supplementary information is available at the *Genomic Medicine* website.

222

223 **References**
224

225 1. Lepiller, Q. *et al.* High incidence but low burden of coronaviruses and preferential
226 associations between respiratory viruses. *J. Clin. Microbiol.* **51,** 3039–3046 (2013).
227 2. Forni, D., Cagliani, R., Clerici, M. & Sironi, M. Molecular Evolution of Human Coronavirus
228 Genomes. *Trends Microbiol.* **25,** 35–48 (2017).
229 3. Dijkman, R. & Hoek, L. van der. Human Coronaviruses 229E and NL63: Close Yet Still So
230 Far. *J. Formos. Med. Assoc.* **108,** 270–279 (2009).
231 4. Posthuma, C. C., Te Welthuis, A. J. W. & Snijder, E. J. Nidovirus RNA polymerases:
232 complex enzymes handling exceptional RNA genomes. *Virus Res.* (2017).
233 doi:10.1016/j.virusres.2017.01.023
234 5. Milano, F. *et al.* Human rhinovirus and coronavirus detection among allogeneic
235 hematopoietic stem cell transplantation recipients. *Blood* **115,** 2088–2094 (2010).
236 6. Gyurkocza, B. & Sandmaier, B. M. Conditioning regimens for hematopoietic cell
237 transplantation: one size does not fit all. *Blood* **124,** 344–353 (2014).
238 7. Szikriszt, B. *et al.* A comprehensive survey of the mutagenic impact of common cancer
239 cytotoxics. *Genome Biol.* **17,** 99 (2016).
240 8. Godley, L. A. & Larson, R. A. Therapy-related Myeloid Leukemia. *Semin. Oncol.* **35,** 418–
241 429 (2008).
242 9. Kuypers, J. *et al.* Clinical disease in children associated with newly described
243 coronavirus subtypes. *Pediatrics* **119,** e70-76 (2007).
244 10. Greninger, A. L. *et al.* Rapid Metagenomic Next-Generation Sequencing during an
245 Investigation of Hospital-Acquired Human Parainfluenza Virus 3 Infections. *J. Clin.*
246 *Microbiol.* **55,** 177–182 (2017).
247 11. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing
248 reads. *EMBnet.journal* **17,** 10–12 (2011).

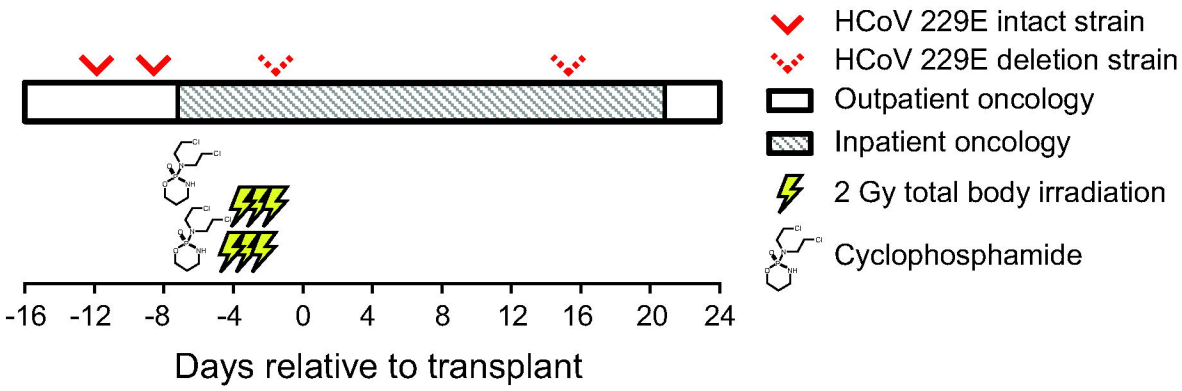
249 12. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software
250 platform for the organization and analysis of sequence data. *Bioinforma. Oxf. Engl.* **28,**
251 1647–1649 (2012).
252 13. Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to
253 Single-Cell Sequencing. *J. Comput. Biol.* **19,** 455–477 (2012).
254 14. Koelle, D. M. *et al.* Worldwide circulation of HSV-2×HSV-1 recombinant strains. *Sci.*
255 *Rep.* **7,** 44084 (2017).
256 15. BBTools. *DOE Joint Genome Institute* Available at: http://jgi.doe.gov/data-and-
257 tools/bbtools/. (Accessed: 11th July 2017)
258 16. Ogimi, C. *et al.* Prolonged shedding of human coronavirus in hematopoietic cell
259 transplant recipients: risk factors and viral genome evolution. *J. Infect. Dis.*
260 doi:10.1093/infdis/jix264
261 17. Dijkman, R. *et al.* Human coronavirus 229E encodes a single ORF4 protein between the
262 spike and the envelope genes. *Virol. J.* **3,** 106 (2006).
263 18. Crossley, B. M., Mock, R. E., Callison, S. A. & Hietala, S. K. Identification and
264 Characterization of a Novel Alpaca Respiratory Coronavirus Most Closely Related to the
265 Human Coronavirus 229E. *Viruses* **4,** 3689–3700 (2012).
266 19. Farsani, S. M. J. *et al.* The first complete genome sequences of clinical isolates of human
267 coronavirus 229E. *Virus Genes* **45,** 433–439 (2012).
268 20. Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology
269 detection and structure prediction. *Nucleic Acids Res.* **33,** W244-248 (2005).
270 21. Zhang, R. *et al.* The ORF4a protein of human coronavirus 229E functions as a viroporin
271 that regulates viral production. *Biochim. Biophys. Acta BBA - Biomembr.* **1838,** 1088–
272 1095 (2014).

273 22. Worm, S. H. E. van den *et al.* Reverse Genetics of SARS-Related Coronavirus Using
274 Vaccinia Virus-Based Recombination. *PLOS ONE* **7,** e32857 (2012).
275 23. Povirk, L. F. & Shuker, D. E. DNA damage and mutagenesis induced by nitrogen
276 mustards. *Mutat. Res.* **318,** 205–226 (1994).
277 24. Hsu, C.-W. *et al.* Detection of hepatitis D virus RNA carrying large fragment deletions in
278 patients with severe hepatitis B/D receiving oral antiviral therapy. *J. Med. Virol.* **87,**
279 634–641 (2015).
280 25. Greninger, A. L. *et al.* A Metagenomic Analysis of Pandemic Influenza A (2009 H1N1)
281 Infection in Patients from North America. *PLoS ONE* **5,** e13381 (2010).
282 26. Greninger, A. L. *et al.* Copy number heterogeneity of JC virus standards. *J. Clin. Microbiol.*
283 (2016). doi:10.1128/JCM.02337-16
284 27. Greninger, A. L. *et al.* Clinical metagenomic identification of Balamuthia mandrillaris
285 encephalitis and assembly of the draft genome: the continuing case for reference
286 genome sequencing. *Genome Med.* **7,** 113 (2015).
287 28. Canuti, M. & van der Hoek, L. Virus discovery: are we scientists or genome collectors?
288 *Trends Microbiol.* **22,** 229–231 (2014).
289

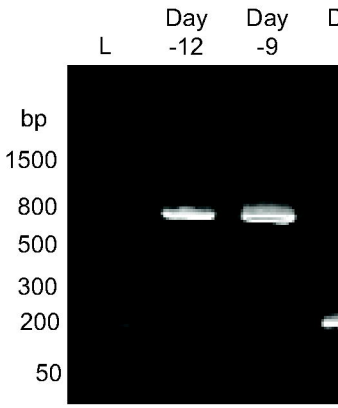
290 **Figure Legend**

291 **Figure – Detection of a 548-nucleotide deletion in a HCoV 229E strain associated with**
292 **myeloablation.** A) Depiction of the case history surrounding myeloablation and
293 coronavirus infection. A woman in her 40s with AML was found to be infected with a HCoV
294 229E just before her HCT. Due to the patient's advanced disease, the decision was made to
295 continue with the transplant. Four successive HCoV-positive nasal swab specimens were
296 available from the patient; two from before treatment with myeloablative
297 cyclophosphamide and total body irradiation and two from after. All days are reported
298 relative to HCT. B) Coverage plots of mNGS reads mapped to the reference genome for
299 HCoV 229E (NC_002645) of three samples from the patient revealed the generation of a
300 548-nucleotide deletion in the ORF4a/4b gene. In addition to deleting 83% of the coding
301 region from ORF4a/4b, the deletion resulted in a frame shift that resulted in a premature
302 stop codon 10 amino acids before the reference genome stop codon. Primer binding sites
303 are denoted in blue for confirmatory junction RT-PCR. C) Confirmatory junction RT-PCR of
304 the four available HCoV-positive nasal swabs revealed the absence of the deletion before
305 myeloablation and the presence of the deletion in specimens taken after myeloablation.
306 NTC, no template control.
307

A


HCoV 229E intact strain
HCoV 229E deletion strain
Outpatient oncology
Inpatient oncology
2 Gy total body irradiation
Cyclophosphamide
-16 -12 -8 -4 0 4 8 12 16 20 24
Days relative to transplant

C


L
Day -12
Day -9
bp
1500
800
500
300
200
50

B

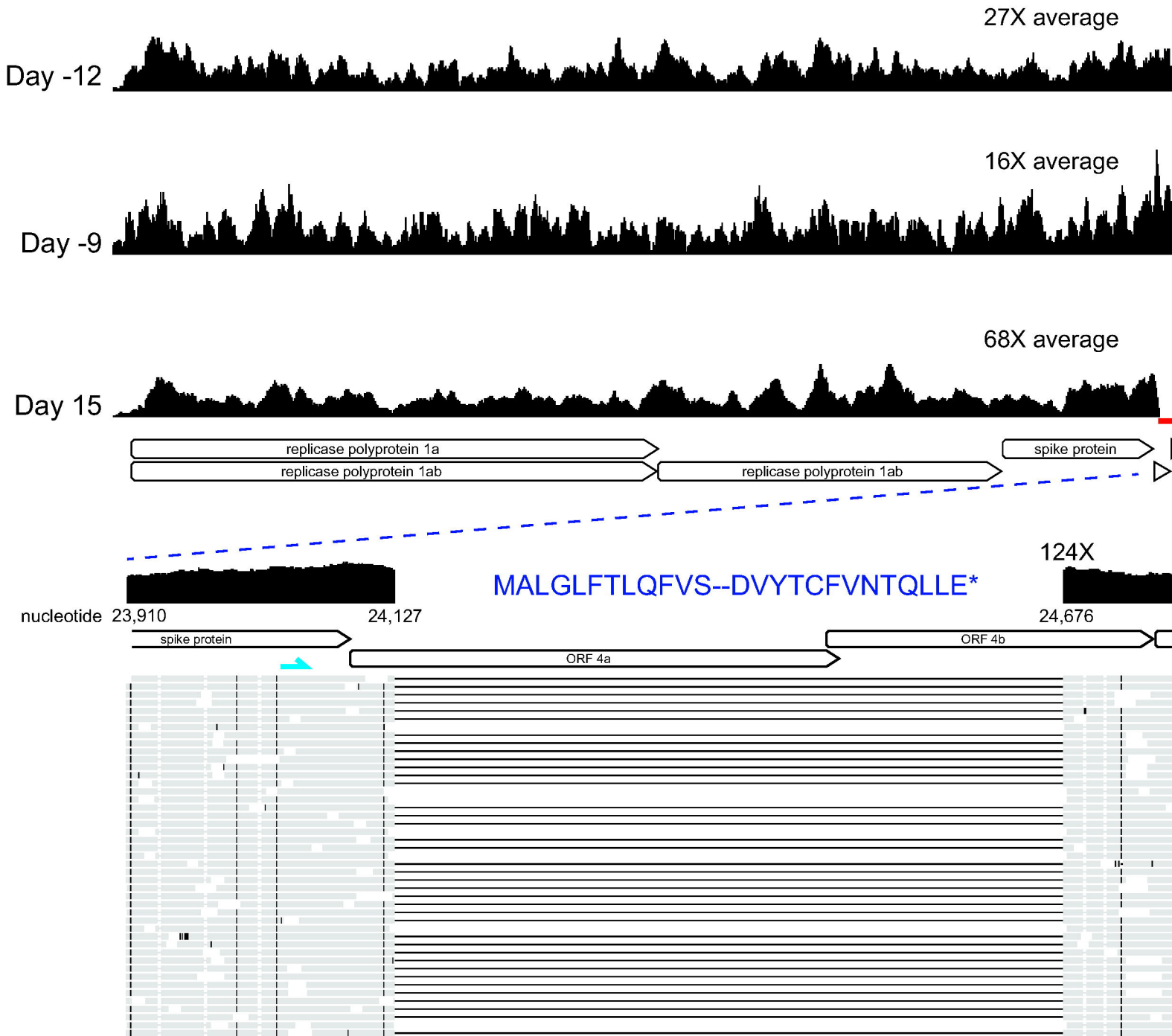

27X average
Day -12
16X average
Day -9
68X average
Day 15
replicase polyprotein 1a
replicase polyprotein 1ab
replicase polyprotein 1ab
spike protein
124X
MALGLFTLQFVS--DVYTCFVNTQLLE*
nucleotide 23,910
24,127
24,676
spike protein
ORF 4a
ORF 4b

Table 1 – Specimens sequenced in this study. mNGS sequencing reads to HCoV 229E and total reads for each specimen are indicated, along with qRT-PCR cycle threshold.

Day	HCoV Ct	HCoV reads	Total reads	Strain name
-12	31.5	7,257	4,395,609	SC379
-9	33.5	3,733	30,166,065	SC399
-2	32.7	5,064	11,185,156	SC475
15	31.3	11,990	1,594,894	SC677