Mutations on non-structural proteins of SARS-CoV-2 are possibly responsible for adverse clinical outcomes

Takaya Ichikawa[1], Shiho Torii[2], Hikoyu Suzuki[3], Akio Takada[4], Satoshi Suzuki[1], Masahide Nakajima[5], Akihito Tampo[6], Yasutaka Kakinoki[1]

1. Division of Internal Medicine, Asahikawa City Hospital, Asahikawa, Japan

2. Department of Molecular Virology, Research Institute for Microbial Diseases, Osaka University, Suita, Japan

3. digzyme Inc., Tokyo, Japan

4. Department of Pathology, Asahikawa City Hospital, Asahikawa, Japan

5. Department of Pediatrics, Asahikawa City Hospital, Asahikawa, Japan

6. Department of Emergency Medicine, Asahikawa City Hospital, Asahikawa, Japan

Corresponding author

Takaya Ichikawa, D.V.M., M.D.

Asahikawa City Hospital, 1-1-65, Kinsei-cho, Asahikawa, Hokkaido, Japan

e-mail: ichikawa.honest2.go@gmail.com

Keywords: COVID-19, SARS-CoV-2, non-structural protein, mutation, clinical outcome

Abstract

Among a cluster of COVID-19 cases from the end of March through April 2021 in Asahikawa, we experienced the cases in which patients manifested severe clinical symptoms compared to patients who were infected before that. A hundred three patients (age range: 65 to 89 years old) enrolled in this study were divided into two groups, group A: the patients infected from November 2020 to March 2021, and group B: the patients in this cluster population. The mortality rates were 6.1% in group A and 16.2% in group B (OR: 2.97, 95%CI: 0.65-15.38). For the severity of disease, the patients in group B required higher oxygen flow rate in early course of admission (mild; p=0.892, moderate; p=0.117, severe; p=0.029). Whole viral genome sequences revealed five non-synonymous mutations by comparison of the isolates with each group. Of these, four were on non-structural proteins (NSPs) including nsp3, 6 and 15, and one was on S protein located near the C-terminus, suggesting that the mutations on NSPs could be responsible for adverse clinical outcomes in COVID-19 patients.

Introduction

The Coronavirus disease 2019 (COVID-19), which is caused by Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), initially broke out in Wuhan, China, on December 2019 and has subsequently spread around the world [1]. SARS-CoV-2 is an enveloped virus with a positive-sense single-stranded RNA genome of approximately 30 kb. Two-thirds of the viral genome at the 5' terminus contains ORF1a and ORF1b, which encodes 16 non-structural proteins (NSPs). These NSPs play crucial roles in viral replication and evasion of host immune systems [2]. The other one-third of the 3' end contains structural genes such as spike glycoprotein (S protein), envelope protein, membrane protein, nucleocapsid protein, and several accessory proteins [3]. Because replication of RNA viruses typically has a low fidelity, the viral genome of SARS-CoV-2 has accumulated mutations at an average of two nucleotides per month [4]. On 25th February 2021, World Health Organization (WHO) defined a variant of interest (VOI) as an isolate detected in several countries that changes phenotype under

46 certain conditions, and a variant of concern (VOC) as an isolate in VOIs if it has been demonstrated
47 to be associated with increase in transmissibility or virulence
48 (https://www.who.int/emergencies/diseases/novel-coronavirus-2019). For example, VOC-202012/01
49 (Pango lineage B.1.1.7), as it is called "alpha variant", has higher transmissibility and mortality
50 compared to a reference isolate [5,6]. These variants, however, were not focused on mutations in NSPs,
51 but in structural proteins, especially S protein. It remains unclear whether mutations in NSPs are
52 responsible for the virulence of COVID-19.

53 As of October 2021, COVID-19 is responsible for 1.7 million cases and 18 thousand deaths
54 in Japan (https://www.mhlw.go.jp/stf/covid-19/kokunainohasseijoukyou.html). The most remarkable
55 clinical feature of this disease is its heterogeneity in clinical manifestations, ranging from no symptoms
56 to critical illness [7]. Several risk factors for symptom severity have been known, such as age (over
57 65 years old), male sex, obesity, history of smoking, and comorbidities (including hypertension,
58 diabetes mellitus, respiratory disease, cardiovascular disease, chronic kidney disease and cancer)
59 [8,9,10,11]. In Japan, the several treatment options are currently recommended based on the symptom
60 severity: Casirivimab and Imdevimab, remdesivir, dexamethasone and baricitinib
61 (https://www.mhlw.go.jp/content/000815065.pdf). In clinical practice, methylprednisolone is also
62 used for critically ill patients when they have a poor response for the dexamethasone therapy or when
63 baricitinib is not available [12].

64 In Asahikawa city, Hokkaido prefecture located in northern part of Japan, A cluster of
65 COVID-19 cases occurred in several bars where a group of customers enjoyed "karaoke" during
66 daytime from the end of March through April 2021 (referred to bellow as cluster K), and 66 cases were
67 involved into the infection according to the city's survey
68 (https://www.city.asahikawa.hokkaido.jp/kurashi/135/136/150/index.html). We found that the
69 progression of disease among the patients in this cluster were faster compared to patients who were
70 infected before that. Therefore, in order to identify factors associated with the progression of disease,

71 we examined the viral genome sequences along with the clinical characteristics and treatment
72 strategies, and then compared those with other cases.
73
74 **Material and Method**
75 *Participants*
76 All adult patients with COVID-19 who were admitted to Asahikawa City Hospital from November
77 2020 to April 2021 were enrolled in this retrospective study. All positive results were confirmed by
78 real-time RT-PCR for the presence of SARS-CoV-2 in saliva or nasopharyngeal swab samples. The
79 recovered patients were discharged from the hospital following the discharge criteria provided by
80 Japanese Ministry of Health, Labour and Welfare when two criteria have met:10 days passed from the
81 symptom onset and 72 hours passed from the symptom resolution, or when two consecutive negative
82 PCR results are confirmed. Opt-out consent was obtained instead of written informed consent. We
83 provided the patients with information explaining the proposed research plan via the website of
84 Asahikawa City Hospital, and with the opportunity to opt out.
85
86 *Clinical data collection and definitions*
87 The epidemiological data, medical history, underlying comorbidities, symptoms and signs at
88 admission, oxygen flow rate, treatment and clinical outcomes were obtained from electronic medical
89 records. Because the date of onset was unclear for some patients who were asymptomatic, we
90 considered the date on which positive PCR result was obtained as day 0. Clinical outcomes were
91 followed up until discharge or death. The primary outcome was mortality rate, and the secondary
92 outcome was oxygen flow rate as a severity of the disease; mild: ≥1 L/min, moderate: ≥5 L/min and
93 severe: ≥10 L/min.
94
95 *Viral genome sequence*

96 The stored RNA extracts were used as clinical samples. Fist-strand cDNA was synthesized by using a
97 PrimeScript IV 1st strand cDNA Synthesis Mix (TaKaRa Bio) with random hexamer primers and
98 extracted RNA, according to the manufacture's protocols. The whole-genome sequencing of SARS-
99 CoV-2 was carried out as reported previously [13]. Briefly, a total of 10 gene fragments were amplified
100 with synthesized cDNA, specific primer sets for SARS-CoV-2 and PrimeSTAR GXL DNA
101 polymerase (TaKaRa Bio). Then, the amplified products were directly sequenced in both directions by
102 using the ABI PRISM 3130 Genetic Analyzer (Applied Biosystems) with specific primers. Our raw
103 sequence data were converted to FASTQ format by in-house Python script with BioPython module
104 [14], followed by trimming first 50 bps and the region after 900 bps by FASTX-Toolkit
105 (http://hannonlab.cshl.edu/fastx_toolkit/). Trimmed sequence reads were mapped to the SARS-CoV-2
106 reference genome (accession number MN908947.3).
107
108 *Phylogenetic analysis*
109 More than one million SARS-CoV-2 published genome sequences whose length were between 29,000
110 - 30,000 bps were downloaded from National Center for Biotechnology Information (NCBI) via
111 Application Programming Interface (API) for the Entrez Programming Utilities (E-utilities)
112 (https://www.ncbi.nlm.nih.gov/books/NBK25497/). Sequences including less than 29,000 definite
113 bases, neither 'N' nor other mixed bases, were discarded.
114 All SARS-CoV-2 genome sequence data, including our assemblies, were mapped to the
115 reference genome (accession number MN908947.3). Based on the SAM format mapping output, SNPs
116 and short INDELs were detected and categorized to synonymous, non-synonymous or intergenic
117 mutations by in-house Perl scripts, which simply picked up the differences from the reference
118 sequence for each genome. For the phylogenetic analysis, a typical genome sequence of each variant,
119 named alpha, beta, gamma, delta, epsilon, zeta, eta, theta, iota, kappa, and lambda by WHO, were
120 chosen from the published data based on the following criteria; neither 'N 'nor other mixed bases

121 were included in open reading frames (ORFs), number of substitution sites in ORFs were less than
122 100, all non-synonymous mutations which characterize each variant were included, and all other
123 substitutes observed in more than 95% of the samples that satisfy the above conditions were also
124 included. The non-synonymous mutations which characterize each variant are shown in Table S1,
125 created based on the mutation prevalence information from outbreak.info (https://outbreak.info/).
126 ORF10 was excluded from the analyses because one of our assemblies did not completely cover that
127 region. Based on the alignment, suitable nucleotide substitution models for each partition were
128 selected by ModelTest-NG [15], followed by maximum likelihood phylogenetic analyses using the
129 selected substitution models by RAxML-NG [16]. The constructed phylogenetic tree was drawn and
130 edited by MEGA X [17].
131
132 *Statistical analysis*
133 We performed group comparisons using Fisher's exact test for categorical variables, and independent
134 group t-test for continuous variables. The categorical variables were expressed as frequencies
135 (percentage), and the continuous variables as the median with interquartile range. Odds ratios (OR)
136 with 95% confidence interval were calculated for the rates of mortality and the required oxygen flow
137 rate. Cumulative probabilities of oxygen therapy were described with using the Kaplan-Meier analysis
138 and compared using log-rank test. P-values of 0.05 or less were considered statistically significant. All
139 statistical analyses were performed with EZR (Saitama Medical Center, Jichi Medical University,
140 Saitama, Japan, version 1.54) [18], which is a graphical user interface for R software (The R
141 Foundation for Statistical Computing, Vienna, Austria).
142
143 **Result**
144 A total of 191 patients diagnosed with COVID-19 were admitted to our hospital from November 2020
145 to April 2021. The patients were divided into two groups, group A: the patients infected from

146 November 2020 to March 2021, and group B: the patients whose infection was traced from the cluster
147 K. We included the patients whose ages range between 65 and 89 years old for the following reasons;
148 1) patients who are younger than 65 years old are not necessarily requested for hospitalization in our
149 city's health care system, 2) in group A, many patients over 90 years old were transferred from a
150 convalescent hospital where COVID-19 outbreak had occurred, thus the age profiles would be much
151 different from group B if the patients aged over 90 years old were included in this group. Finally, 66
152 patients were assigned to group A and 37 patients to group B (Figure 1).

153 The clinical characteristics and treatment strategies used during hospitalization are shown
154 in Table 1. Although there were no significant differences in clinical characteristics between the two
155 groups, obesity and lifestyle-related diseases such as hypertension, dyslipidemia and diabetes tended
156 to be more prevalent in group B. In treatment strategies, remdesivir tended to be used more frequently
157 in group A and methylprednisolone in group B.

158 Four patients of 66 patients (6.1%) in group A and six of 37 patients (16.2%) in group B
159 died from COVID-19 (OR: 2.97, 95%CI: 0.65-15.38) (Table 2). When limited to males, the differences
160 in number of deaths became larger (6.3% vs 29.4%, OR: 5.99, 95%CI: 0.84-71.2). The clinical
161 characteristics and the treatment of 10 deceased patients are shown in Table S2. Cumulative
162 probability of the patients who required minimal oxygen therapy (mild) were similar in both groups
163 (Table 2, Figure 2A). However, the patients in group B were likely to require more oxygen therapy in
164 earlier course of hospitalization (Figure 2B, 2C). There was no significant difference in the cumulative
165 probability between two groups for moderate oxygen therapy (p=0.117), but for severe oxygen therapy,
166 the significant difference was observed (p=0.029). As with mortality, males in group B showed worse
167 outcomes in the severity of disease (Table 2).

168 Assuming that these differences in clinical outcomes would be due to viral factors, we
169 examined the whole viral genome sequences representing two samples in each group, named
170 Asahikawa_0108 and Asahikawa_0122 isolated in January 2021, Asahikawa_0404 and

171 Asahikawa_0417 isolated in April 2021. The mutation sites which differed in these four samples are
172 summarized in Table 3. Focusing on the non-synonymous mutations, five amino acid changes were
173 identified (represented in bold). Of these, four amino acid changes were located on NSPs and one was
174 on S protein, as illustrated in the viral genome scheme of SARS-CoV-2 (Figure 3). In order to assess
175 strain relatedness, the phylogenetic analysis was carried out. The isolates in this study were closely
176 related to the strain previously isolated in Japan (accession number BS000756.1; Pango lineage
177 B.1.1.214), but not to any of the variants labeled by WHO (Figure 4).
178
179 **Discussion**
180 We here experienced the cases of a single COVID-19 cluster in which the conventional treatment was
181 unsuccessful and the hospitalized patients became more severely ill, thus performed the retrospective
182 analysis to find out the causative factors. The patients in group B had a higher mortality rate compared
183 to group A, but no significant difference was indicated. For the severity of disease, the patients in
184 group B required higher oxygen flow rate in early course of admission. Furthermore, the differences
185 became larger when limited to males, with 41.2% of male severe patients in group B against 6.3% in
186 group A.
187 In terms of treatment strategies, we were likely to use methylprednisolone monotherapy for
188 severe cases in group B. This was because we had experienced that, in the winter of 2020, patients
189 who received methylprednisolone therapy appeared to have a better response than those who were
190 treated with remdesivir after poor response for dexamethasone, although the treatment guideline for
191 COVID-19 in Japan recommends the use of remdesivir for patients with pneumonia whose oxygen
192 saturation falls below 96% (https://www.mhlw.go.jp/content/000815065.pdf). In our study, there was
193 no significant difference of remdesivir usage between two groups, thus the viral factors could have an
194 impact on the clinical outcomes.
195 Whole viral genome sequences revealed five non-synonymous mutations by comparison of

196 the isolates with each group. Interestingly, four of them were on NSPs. The sole structural mutation,
197 D1153G on S protein, was located near the C-terminus and outside the range of three-dimensional
198 structures obtained by crystallography, suggesting neither the head nor stalk of S protein. Therefore,
199 D1153G is not expected to involve in the virulence of the disease. Of the four mutation sites on NSPs,
200 two were found on nsp6 (T3598I and Y3722C), and the rest of them were on nsp3 (E2089D) and
201 nsp15 (A2143S). To our knowledge, none of these mutations are known to be the factors that determine
202 the severity of disease in clinical practice, and it is unclear whether the mutations affect the functions
203 of these NSPs. However, all of them have been reported to play roles in suppression of type 1 interferon
204 in host cells [19,20,21], and also nps6 involves in the formation of autophagosome in host cells [22].
205 The potential sites under positive selective pressure have been found on nsp6 near T3598I according
206 to an evolutionary analysis on SARS-CoV-2 genome sequences of 351 clinical samples [23]. Therefore,
207 the mutations on nsp3, 6 and 15 could be responsible for its function to interact with host immunity
208 and autophagy. Further investigations are required to reveal their functions and involvement in the
209 virulence.

210 The phylogenetic analysis revealed that the isolates in this study are derived from Pango
211 lineage B.1.1.214 which was the dominant strain in so-called "the third wave", the epidemic period
212 from October to December 2020 in Japan [24]. Considering the fact that any strains identical to the
213 viral sequence of group B cannot be found in more than 1,000,000 registered sequences in NCBI, it is
214 likely that the virus evolved locally from January to April 2021 in Hokkaido. Since May 2021, alpha
215 variant has been dominant throughout Japan including Hokkaido; 78% of the samples isolated from
216 April 26 to May 2 2021 in Hokkaido were already alpha variant
217 (https://www.mhlw.go.jp/content/10900000/000779013.pdf). As a consequence, B.1.1.214 strains in
218 Asahikawa seems to be eliminated. This is similar to the situation in influenza A virus, for example,
219 where pandemic influenza A (H1N1) pdm09 replaced Russian flu (H1N1) since 2009 [25,26].

220 There are some limitations to this retrospective study. First, the recommended therapy for

221 COVID-19 have been changing dynamically. Because the best treatment strategies are selected based
222 on the evidence at that time, it is difficult to compare clinical outcomes between different periods of
223 time in real-life practice. Second, this is a single institution study; not all cases in cluster K have been
224 assessed. The mortality and severity rate would be different when including the cases admitted to other
225 hospitals. Third, we examined only two samples from each group in this study. Due to the limited
226 human and financial resources, we were unbale to perform the viral genome analysis of much more
227 samples.

228 In conclusion, the present study demonstrated that some viral factors, four non-synonymous
229 mutations on nsp3, 6 and 15 but not on S protein, could be responsible for adverse clinical outcomes
230 in COVID-19 patients. Though scientists and physicians all over the world have focused on the S
231 protein mutations, in order to comprehensively understand the epidemiology of COVID-19, we should
232 have an insight into other type of mutations such as those on NSPs.

233

234 **Acknowledgment**

235 We thank Prof. Yoshiharu Matsuura, Osaka University, for funding the viral sequence analysis. We
236 also thank our medical team worked on COVID-19 and all of the staff at Asahikawa Health Center for
237 their support.

238

239 **Declaration of interests**

240 Authors have no conflicts of interests.

241

242 **Reference**

243 1. Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, et al. The species
244 Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it
245 SARS-CoV-2. Nat Microbiol. 2020;5: 536–544. doi:10.1038/s41564-020-0695-z

2. Maier HJ, Bickerton E, Britton P. Coronaviruses: Methods and protocols. Coronaviruses Methods Protoc. 2015;1282: 1–282. doi:10.1007/978-1-4939-2438-7

3. Hagemeijer MC, Verheije MH, Ulasli M, Shaltiël IA, de Vries LA, Reggiori F, et al. Dynamics of Coronavirus Replication-Transcription Complexes. J Virol. 2010;84: 2134–2149. doi:10.1128/jvi.01716-09

4. Duchene S, Featherstone L, Haritopoulou-Sinanidou M, Rambaut A, Lemey P, Baele G. Temporal signal and the phylodynamic threshold of SARS-CoV-2. Virus Evol. 2020;6: 1–8. doi:10.1093/ve/veaa061

5. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. Science (80-). 2021;372. doi:10.1126/science.abg3055

6. Challen R, Brooks-Pollock E, Read JM, Dyson L, Tsaneva-Atanasova K, Danon L. Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: Matched cohort study. BMJ. 2021;372: 1–10. doi:10.1136/bmj.n579

7. Grasselli G, Zangrillo A, Zanella A, Antonelli M, Cabrini L, Castelli A, et al. Baseline Characteristics and Outcomes of 1591 Patients Infected with SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. JAMA - J Am Med Assoc. 2020;323: 1574–1581. doi:10.1001/jama.2020.5394

8. Zheng Z, Peng F, Xu B, Zhao J, Liu H, Peng J. Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ' s public news and information . 2020.

9. Popkin BM, Du S, Green WD, Beck MA, Algaith T, Herbst CH, et al. Individuals with obesity and COVID-19: A global perspective on the epidemiology and biological relationships. Obes Rev. 2020;21: 1–17. doi:10.1111/obr.13128

10. Gansevoort RT, Hilbrands LB. CKD is a key risk factor for COVID-19 mortality. Nat Rev Nephrol. 2020;16: 705–706. doi:10.1038/s41581-020-00349-4
11. Liang W, Guan W, Chen R, Wang W, Li J, Xu K, et al. Cancer patients in SARS-CoV-2 infection: a nationwide analysis in China. Lancet Oncol. 2020;21: 335–337. doi:10.1016/S1470-2045(20)30096-6
12. Pinzón MA, Ortiz S, Holguín H, Betancur JF, Arango DC, Laniado H, et al. Dexamethasone vs methylprednisolone high dose for Covid-19 pneumonia. PLoS One. 2021;16: 1–13. doi:10.1371/journal.pone.0252057
13. Torii S, Ono C, Suzuki R, Morioka Y, Anzai I, Fauzyah Y, et al. Establishment of a reverse genetics system for SARS-CoV-2 using circular polymerase extension reaction. Cell Rep. 2021;35: 109014. doi:10.1016/j.celrep.2021.109014
14. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: Freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009;25: 1422–1423. doi:10.1093/bioinformatics/btp163
15. Darriba Di, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. Mol Biol Evol. 2020;37: 291–294. doi:10.1093/molbev/msz189
16. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019;35: 4453–4455. doi:10.1093/bioinformatics/btz305
17. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35: 1547–1549. doi:10.1093/molbev/msy096
18. Kanda Y. Investigation of the freely available easy-to-use software “EZR” for medical statistics. Bone Marrow Transplant. 2013;48: 452–458. doi:10.1038/bmt.2012.244

19. Shin D, Mukherjee R, Grewe D, Bojkova D, Baek K, Bhattacharya A, et al. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. Nature. 2020;587: 657–662. doi:10.1038/s41586-020-2601-5
20. Xia H, Cao Z, Xie X, Zhang X, Chen JYC, Wang H, et al. Evasion of Type I Interferon by SARS-CoV-2. Cell Rep. 2020;33: 108234. doi:10.1016/j.celrep.2020.108234
21. Yuen C-K, Lam J-Y, Wong W-M, Mak L-F, Wang X, Chu H, et al. SARS-CoV-2 nsp13, nsp14, nsp15 and orf6 function as potent interferon antagonists. Emerg Microbes Infect. 2020;9: 1418–1428. doi:10.1080/22221751.2020.1780953
22. Cottam EM, Whelband MC, Wileman T. Coronavirus NSP6 restricts autophagosome expansion. Autophagy. 2014;10: 1426–1441. doi:10.4161/auto.29309
23. Benvenuto D, Angeletti S, Giovanetti M, Bianchi M, Pascarella S, Cauda R, et al. Evolutionary analysis of SARS-CoV-2: how mutation of Non-Structural Protein 6 (NSP6) could affect viral autophagy. J Infect. 2020;81: e24–e27. doi:10.1016/j.jinf.2020.03.058
24. National Institute of Infectious Diseases, Infectious Agents Surveillance Report. 2021;42: 61-64.
25. Blyth CC, Kelso A, McPhie KA, Ratnamohan VM, Catton M, Druce JD, et al. The impact of the pandemic influenza a(H1N1) 2009 virus on seasonal influenza a viruses in the southern hemisphere, 2009. Eurosurveillance. 2010;15: 1–5. doi:10.2807/ese.15.31.19631-en
26. Broor S, Krishnan A, Roy DS, Dhakad S, Kaushik S, Mir MA, et al. Dynamic patterns of circulating seasonal and pandemic A(H1N1)pdm09 influenza viruses from 2007-2010 in and around Delhi, India. PLoS One. 2012;7: 1–9. doi:10.1371/journal.pone.0029129

Table 1. Characteristics and treatment options of COVID-19 patients

Characteristic	Group A (N=66)	Group B (N=37)	P value
Age, median (IQR)	76 (71-85)	78 (74-82)	0.75
Male sex (%)	32 (48.5)	17 (45.9)	0.84
Duration of hospitalization, median (IQR)	12.5 (9-16)	11 (59-16)	0.85
Obesity, BMI ≥30 (%)	4 (6.1)	6 (16.2)	0.16
Current or ex-smoker (%)	21 (31.8)	13 (35.1)	0.83
Comorbidities (%)			
Hypertention	38 (57.6)	28 (75.7)	0.09
Dyslipidemia	21 (31.8)	16 (43.2)	0.29
Diabetes mellitus	12 (18.2)	10 (27.0)	0.32
Chronic pulmonary disease	12 (18.2)	11 (29.7)	0.22
Cardiovascular disease	10 (15.2)	7 (18.9)	0.78
Chronic kidney disease	10 (15.2)	6 (16.2)	1.00
Cancer	9 (13.6)	2 (5.4)	0.32
Treatment options (%)			
Dexamethasone	39 (59.1)	19 (51.4)	0.54
Methylprednisolone	10 (15.2)	11 (29.7)	0.12
Remdesivir	8 (12.1)	2 (5.4)	0.32

IQR: interquartile range, BMI: body mass index

Table 2. Primary and secondary outcomes of COVID-19 patients

Outcome	Over all			Male		
	Group A (N=66)	Group B (N=37)	OR (95%CI)	Group A (N=32)	Group B (N=17)	OR (95%CI)
Death (%)	4 (6.1)	6 (16.2)	2.97 (0.65-15.38)	2 (6.3)	5 (29.4)	5.99 (0.84-71.2)
Severity* (%)						
Mild	27 (40.9)	15 (40.5)	0.98 (0.40-2.41)	15 (46.9)	9 (52.9)	1.27 (0.34-4.89)
Moderate	9 (13.6)	9 (24.3)	2.02 (0.63-6.47)	4 (12.5)	8 (47.1)	5.95 (1.25-33.9)
Severe	5 (7.6)	8 (21.6)	3.32 (0.87-14.11)	2 (6.3)	7 (41.2)	6.68 (1.24-47.8)

* Oxygen flow rate: Mild ≥1 L/min, Moderate ≥5 L/min, Severe ≥10 L/min

Table 3. Mutation sites that differed in the four samples

Nucleotide position	Gene	Group A		Group B		Amino acid change
		Asahikawa_0108	Asahikawa_0122	Asahikawa_0404	Asahikawa_0417	
4233	Orf1a	G	A	A	A	D1323G
6380		C	T	C	C	L2039F
6433		T	C	C	C	Syn
6532		G	G	T	T	**E2089D**
6997		T	C	C	C	Syn
8371		T	G	G	G	Q2702H
8917		C	T	T	T	Syn
9207		T	C	C	C	S2981F
10042		A	A	G	G	Syn
11058		C	C	T	T	**T3598I**
11430		A	A	G	G	**T3722C**
12049		C	T	T	T	Syn
13366		C	C	T	T	Syn
15240	Orf1b	T	C	C	C	Syn
15957		G	A	G	G	Syn
17502		C	T	C	C	Syn
19894		G	G	T	T	**A2143S**
20658		A	G	A	A	Syn
22502	S	C	A	C	C	Q314K
23191		C	C	T	T	Syn
23587		G	T	T	T	Q675H
23720		G	A	A	A	I720V
24748		C	C	T	T	Syn
25020		A	A	G	G	**D1153G**
29353	N	C	T	T	T	Syn

Position number based on alignment to NCBI reference sequence: NC_045512 (Wuhan-Hu-1). The different parts between the two samples in group A are shown in a gray area. Amino acid changes identified between two groups were represented in bold.

322 **Figure Legends**

323

324 **Figure 1.** Flow diagram indicating the recruitment process in this study.

325

326 **Figure 2.** Cumulative probability of the patients who required oxygen therapy with mild: ≥1 L/min,

327 moderate: ≥5 L/min and severe: ≥10 L/min. The log-rank test was used for statistical analysis.

328

329 **Figure 3.** Scheme of SARS-CoV-2 genome. The non-synonymous mutation sites are indicated by

330 arrows.

331

332 **Figure 4.** Maximum likelihood phylogenetic tree of SARS-CoV-2 variants. Black and rend characters

333 indicate synonymous and non-synonymous mutations respectively. Red bold characters indicate

334 mutations of interest. Numbers on internal nodes are bootstrap values for each divergence.

Figure 1

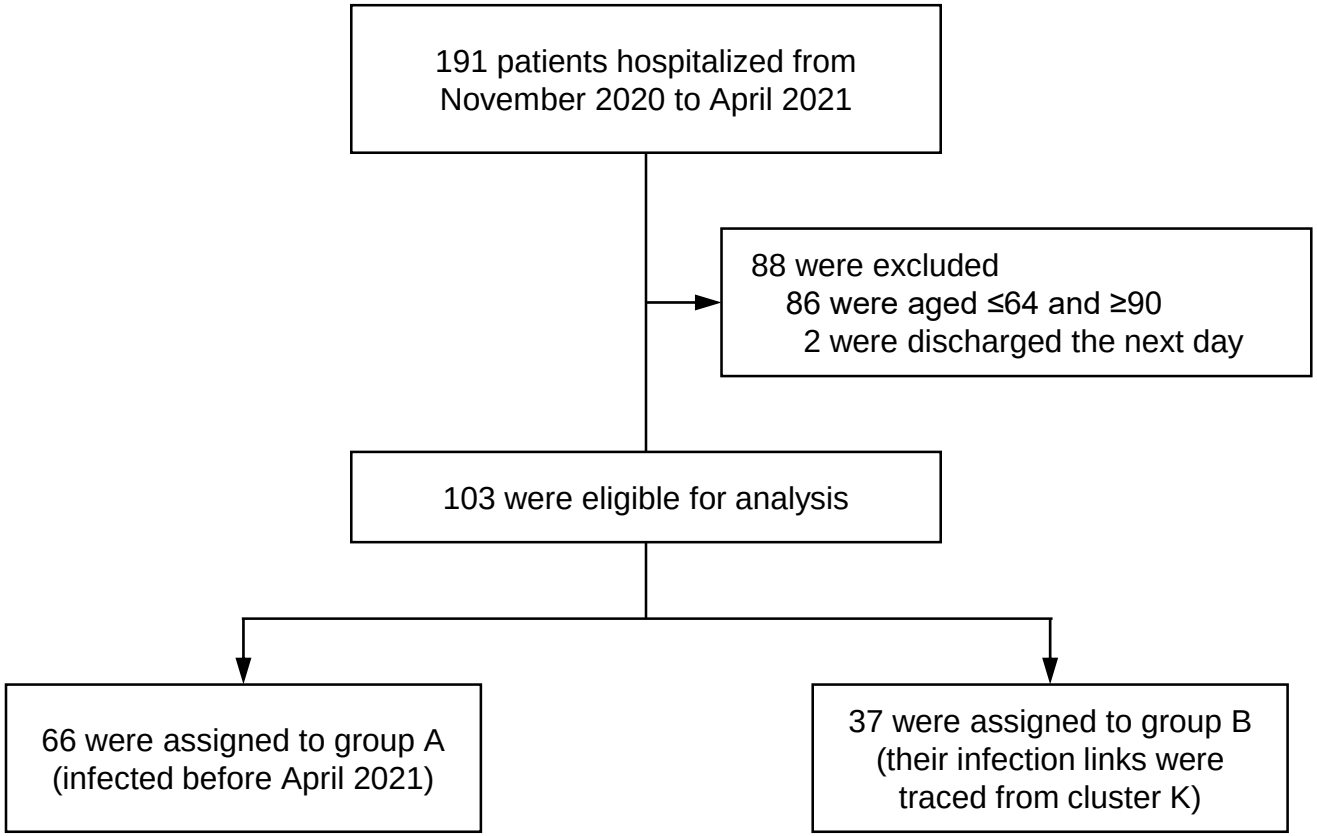

191 patients hospitalized from November 2020 to April 2021
88 were excluded
86 were aged ≤64 and ≥90
2 were discharged the next day
103 were eligible for analysis
66 were assigned to group A (infected before April 2021)
37 were assigned to group B (their infection links were traced from cluster K)

Figure 2

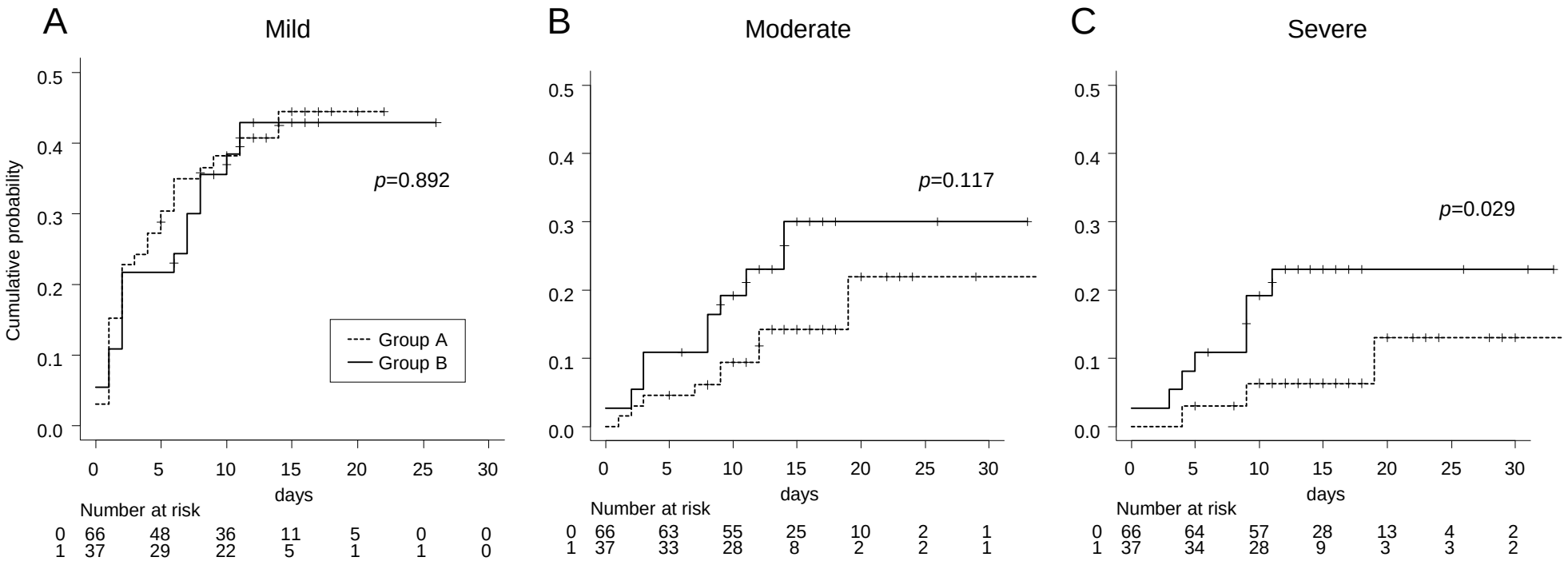

A
Mild
Cumulative probability
0.0
0.1
0.2
0.3
0.4
0.5
p=0.892
Group A
Group B
0 5 10 15 20 25 30
days
Number at risk
0 66 48 36 11 5 0 0
1 37 29 22 5 1 1 0
B
Moderate
p=0.117
0 5 10 15 20 25 30
days
Number at risk
0 66 63 55 25 10 2 1
1 37 33 28 8 2 2 1
C
Severe
p=0.029
0 5 10 15 20 25 30
days
Number at risk
0 66 64 57 28 13 4 2
1 37 34 28 9 3 3 2

Figure 3

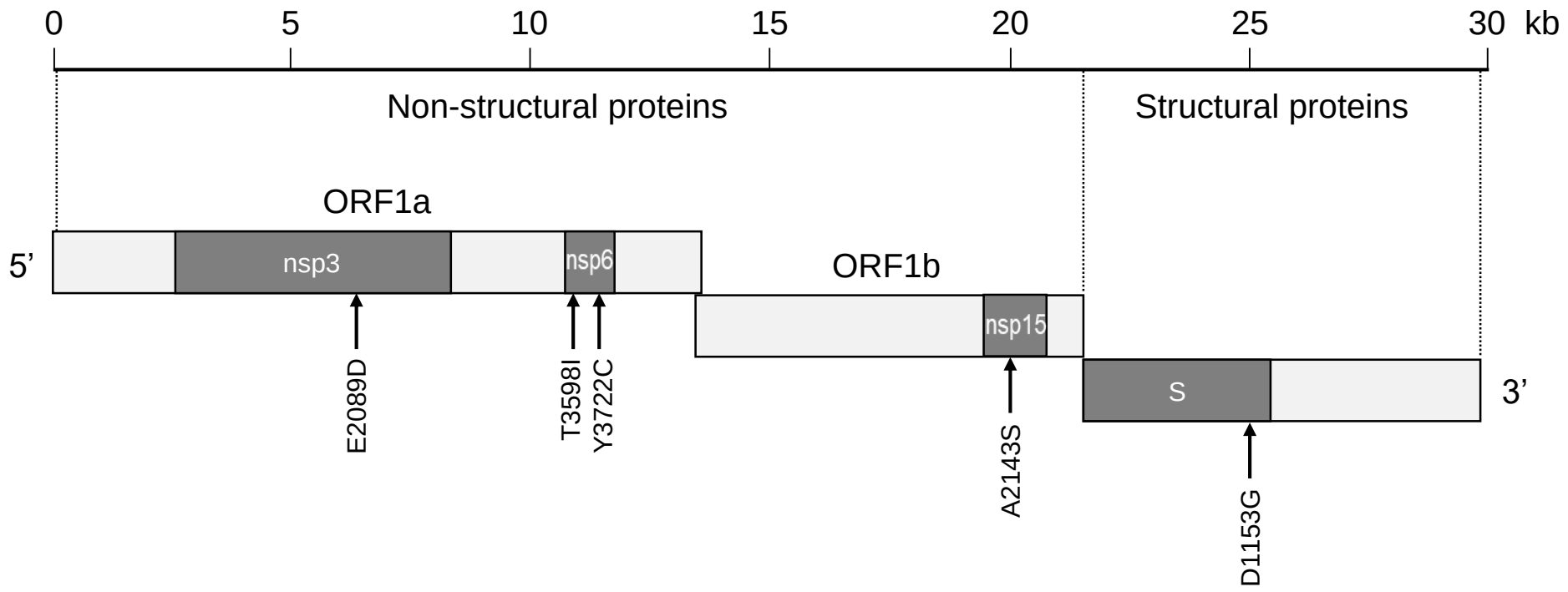
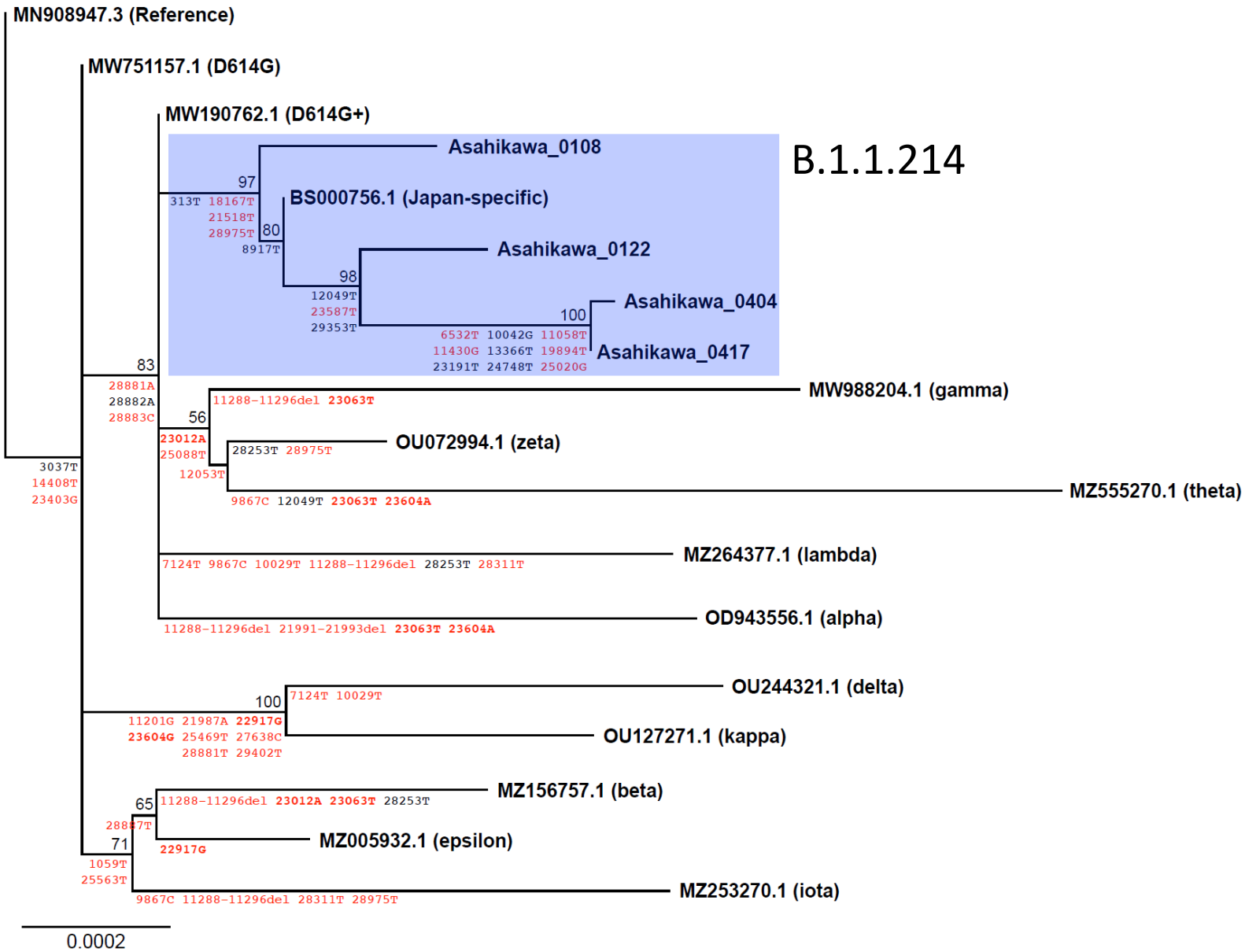

0
5
10
15
20
25
30
kb
Non-structural proteins
Structural proteins
ORF1a
5'
nsp3
nsp6
ORF1b
nsp15
S
3'
E2089D
T3598I
Y3722C
A2143S
D1153G

Figure 4


MN908947.3 (Reference)
MW751157.1 (D614G)
MW190762.1 (D614G+)
Asahikawa_0108
B.1.1.214
97
313T 18167T
21518T
28975T
BS000756.1 (Japan-specific)
80
8917T
Asahikawa_0122
98
12049T
23587T
29353T
Asahikawa_0404
100
6532T 10042G 11058T
11430G 13366T 19894T
23191T 24748T 25020G
Asahikawa_0417
83
28881A
28882A
28883C
MW988204.1 (gamma)
11288-11296del 23063T
56
23012A
25088T
12053T
28253T 28975T
OU072994.1 (zeta)
MZ555270.1 (theta)
9867C 12049T 23063T 23604A
3037T
14408T
23403G
MZ264377.1 (lambda)
7124T 9867C 10029T 11288-11296del 28253T 28311T
OD943556.1 (alpha)
11288-11296del 21991-21993del 23063T 23604A
OU244321.1 (delta)
100
7124T 10029T
11201G 21987A 22917G
23604G 25469T 27638C
28881T 29402T
OU127271.1 (kappa)
MZ156757.1 (beta)
65
11288-11296del 23012A 23063T 28253T
28887T
71
22917G
MZ005932.1 (epsilon)
1059T
25563T
MZ253270.1 (iota)
9867C 11288-11296del 28311T 28975T
0.0002