bioRxiv preprint doi: https://doi.org/10.1101/2020.03.29.014290. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

1

2 **Sequence variation among SARS-CoV-2 isolates in Taiwan**

3 Yu-Nong Gong[1,2,†], Kuo-Chien Tsao[1,2,3,†], Mei-Jen Hsiao[2], Chung-Guei Huang[2,3], Peng-Nien
4 Huang[1], Po-Wei Huang[2], Kuo-Ming Lee[1], Yi-Chun Liu[2], Shu-Li Yang[2,3], Rei-Lin Kuo[1,3,4,5],
5 Ming-Tsan Liu[6], Ji-Rong Yang[6], Cheng-Hsun Chiu[7,8], Cheng-Ta Yang[9,10], Shin-Ru Shih[1,2,3,11,*],
6 Guang-Wu Chen[1,2,12,*]

7

8 [1] Research Center for Emerging Viral Infections, College of Medicine, Chang Gung University,
9 Taoyuan, Taiwan

10 [2] Department of Laboratory Medicine, Linkou Chang Gung Memorial Hospital, Taoyuan,
11 Taiwan

12 [3] Department of Medical Biotechnology and Laboratory Science, College of Medicine, Chang
13 Gung University, Taoyuan, Taiwan

14 [4] Division of Asthma, Allergy, and Rheumatology, Department of Pediatrics, Linkou Chang
15 Gung Memorial Hospital, Taoyuan, Taiwan.

16 [5] Graduate Institute of Biomedical Sciences, College of Medicine, Chang Gung University,
17 Taoyuan, Taiwan.

18 [6] Centers for Disease Control, Taipei, Taiwan

19 [7] Division of Pediatric Infectious Diseases, Department of Pediatrics, Chang Gung Memorial
20 Hospital, Chang Gung University College of Medicine, Taoyuan, Taiwan

21 [8] Molecular Infectious Disease Research Center, Chang Gung Memorial Hospital, Chang Gung
22 University College of Medicine, Taoyuan, Taiwan

23 [9] Department of Respiratory Therapy, College of Medicine, Chang Gung University, Taoyuan,
24 Taiwan

25 [10] Department of Thoracic Medicine, Linkou Chang Gung Memorial Hospital, Taoyuan, Taiwan

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.29.014290. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

26 [11] Research Center for Chinese Herbal Medicine, Research Center for Food and Cosmetic Safety,
27 and Graduate Institute of Health Industry Technology, College of Human Ecology, Chang Gung
28 University of Science and Technology, Taoyuan, Taiwan

29 [12] Department of Computer Science and Information Engineering, School of Electrical and
30 Computer Engineering, College of Engineering, Chang Gung University, Taoyuan, Taiwan

31

32 † These authors contributed equally to this work.

33

34 * Corresponding author:

35 Guang-Wu Chen, Ph.D.

36 E-mail: gwchen@mail.cgu.edu.tw

37 Shin-Ru Shih, Ph.D.

38 E-mail: srshih@mail.cgu.edu.tw

39

40 **Keywords:**

41 SARS-CoV-2, Illumina Sequencing, Phylogeny, ORF8 Deletion

42

43

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.29.014290. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

44 **Abstract:**

45 Taiwan experienced two waves of imported cases of coronavirus disease 2019 (COVID-19), first
46 from China in January to late February, followed by those from other countries starting in early
47 March. Additionally, several cases could not be traced to any imported cases and were suspected
48 as sporadic local transmission. Twelve full viral genomes were determined in this study by
49 Illumina sequencing either from virus isolates or directly from specimens, among which 5
50 originated from clustered infections. Phylogenetic tree analysis revealed that these sequences
51 were in different clades, indicating that no major strain has been circulating in Taiwan. A
52 deletion in open reading frame 8 was found in one isolate. Only a 4-nucleotide difference was
53 observed among the 5 genomes from clustered infections.

54

55

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.29.014290. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

56 ***Introduction***

57 A novel coronavirus emerged from Wuhan, Hubei province in China in December 2019
58 (1). This virus has been designated as severe acute respiratory syndrome coronavirus 2 (SARS-
59 CoV-2), and the disease is named as coronavirus disease 2019 (COVID-19). The World Health
60 Organization declared this disease a Public Health Emergency of International Concern on
61 January 30, 2020. As of March 26, 2020, the outbreak of COVID-19 has resulted in 462,684
62 confirmed cases and 20,834 deaths worldwide (2), and 252 confirmed cases and two deaths were
63 reported in Taiwan (3).

64 There have been two waves of COVID-19 cases in Taiwan. The first occurred from late
65 January to the end of February, with most cases imported from China, either by Chinese tourists
66 or Taiwanese businessmen returning for Chinese New Year. This wave was smaller than the
67 second wave. The second wave started in early March, during which the disease occurred largely
68 in Taiwanese tourists, business travelers, or students returning from other countries. Although
69 most of these cases were traced to their foreign origins, some small and clustered infections were
70 suspected to have been acquired by local transmission.

71 In this study, we performed virus culture and full-genome sequencing of isolates or
72 clinical specimens of SARS-CoV-2. We compared the genomes obtained from Taiwanese
73 samples to those of other strains in a database to understand their evolutionary trajectory. An
74 open reading frame 8 (ORF8) deletion was found in one strain. Moreover, we assessed the
75 number of nucleotide substitutions that may have accumulated in clustered infections during a
76 short period of time.

77 ***Methods***

78 ***Specimen Collection***

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.29.014290. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

79 Infection of patients by COVID-19 was confirmed by real-time reverse-transcriptase
80 polymerase chain-reaction (RT-PCR) according to the guidelines of the Taiwan Centers for
81 Disease Control (CDC; https://www.cdc.gov.tw/En), and all nasopharyngeal (NP), throat (TH)
82 swab, and sputum (SP) samples were maintained in universal transport medium for further
83 analysis.

84 ***Cell Culture and Virus Isolation***

85 Vero-E6 (ATCC, Manassas, VA, USA) and MK-2 (ATCC) cells were maintained in
86 Modified Eagle Medium (MEM, Thermo Fisher Scientific, Waltham, MA, USA) supplemented
87 with 10% fetal bovine serum and 1x penicillin-streptomycin at 37°C in the presence of 5% CO_2.
88 To isolate the virus, all procedures following the laboratory biosafety guidelines of the Taiwan
89 CDC were conducted in a biosafety Level-3 facility. Cells grown to 80–90% confluency in a T-
90 25 flask were inoculated with 500 μL of virus solution, which was prepared by diluting 100 μL
91 of specimen samples with 1.5 mL of sample pretreatment medium consisting of MEM and 2x
92 penicillin-streptomycin solution, followed by incubation at 37°C for 1 h. The absorption was
93 performed at 37°C for 1 h, then cells were refreshed with 5 mL virus culture medium composed
94 of MEM, 2% fetal bovine serum, and 1x penicillin-streptomycin solution and maintained at
95 37°C. Infected cells were observed daily to determine their cytopathic effect. Additionally, RT-
96 PCR analysis using the RNA extracted from part of the culture supernatant every two days after
97 inoculation was performed to monitor viral growth. We continuously observed the infected cells
98 until cytopathic effects occurred in more than 75% of the cells, after which the culture
99 supernatant was harvested.

100 ***Whole-Genome Sequencing***

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.29.014290. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

101 RNA was extracted either from the culture supernatant or directly from the specimens
102 using a QIAmp viral RNA mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's
103 instructions, except that the carrier RNA was replaced with linear acrylamide (Thermo Fisher
104 Scientific) as the co-precipitant. The amount of viral RNA was evaluated by quantitative RT-
105 PCR to examine the Ct value of the viral E gene. For RNAs showing a high Ct value, we used
106 the Ovation RNA-Seq System V2 (Nugen Technologies, San Carlos, CA, USA) to synthesize
107 cDNA which was further processed into a library using the Celero DNA-Seq System (Nugen
108 Technologies). Other samples with lower Ct values were used for library preparation by using
109 the Trio RNA-Seq kit (Nugen Technologies). Sequencing was performed on an Illumina MiSeq
110 System (San Diego, CA, USA) with paired-end reads. More than 0.75 and 2.5 Gb of raw data
111 were generated for samples from viral isolates and clinical specimens, respectively.

112 ***Next-generation Sequencing Data Analysis Pipeline***

113 We first trimmed the raw data by removing low-quality and short reads using
114 Trimmomatic (version 0.39) (4). Next, quality reads were mapped to the human reference
115 genome to remove host sequences using HISAT2 (version 2.1.0) (5). SPAdes (version 3.14.0) (6)
116 was used to perform *de novo* assembly for constructing contig sequences. Fourth, the BLASTN
117 tool was used to search the assembled contigs against the nucleotide sequence (NT) database of
118 the National Center for Biotechnology Information (NCBI). Viral candidates were identified
119 using the reported top BLASTN hits for each of the queried contig sequences. Finally, we used
120 an iterative mapping approach (7) to increase the read depth and coverage of quality contigs to
121 obtain the whole genome.

122 ***Phylogenetic and Sequence Analysis***

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.29.014290. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

123 Twelve whole genomes were assembled by using our pipeline, including three genomes
124 from specimens and nine genomes from isolates, which were deposited in the Global Initiative
125 on Sharing All Influenza Data (GISAID, https://www.gisaid.org/) with accession numbers
126 EPI_ISL_411915, EPI_ISL_417518, EPI_ISL_415741–3, and EPI_ISL_417519–25, according
127 to CGMH-CGU No. 1–12. We further downloaded all complete and high-coverage genomes
128 from GISAID as of March 14, 2020, and obtained 335 sequences after removing those with
129 sequences gaps or ambiguous nucleotides. One reference strain (accession number MN908947.3)
130 was downloaded from GenBank (NCBI). In total 348 sequences were aligned using MAFFT
131 (version 7.427) (8) for further analyses. The phylogenetic tree was inferred using RAxML
132 (version 8.2.12) (9) under the GTRGAMMA model with a bootstrap value of 1000 to investigate
133 the genomic relationships.

134 ***Results***

135 ***Phylogenetic Tree of Taiwanese and Global Strains***

136 Twelve complete genomes from three specimens (CGMH-CGU No. 1, 7, and 8) and nine
137 isolates (No. 2–6 and 9–12) were uploaded to GISAID. Table 1 shows their next-generation
138 sequencing (NGS) coverage and depth. All average depths were greater than 10,000, except for
139 CGMH-CGU-04 and -08 which showed values of 446.0 and 53.0, respectively. Table 1 also
140 includes two earlier strains, hCoV-19/Taiwan/2/2020 and hCoV-19/Taiwan/3/2020, previously
141 submitted by Taiwan CDC.

142 The phylogenetic tree revealed that the SARS-CoV-2 viral genomes from Taiwan
143 (highlighted) were in different clades (Figure 1). Viral genomes of No. 3–7 were from clustered
144 infections, together with No. 8 (a case originating from the United Kingdom), and some from

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.29.014290. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

145 Australia (AUS) and New Zealand (NZ) in the yellow clade. Three patients with AUS/NZ
146 infections had a travel history to Iran. This figure also shows eight additional Taiwanese isolates
147 (highlighted), which appeared in distinct lineages, indicating that no single dominant strain has
148 been circulating in Taiwan.

149 The two earliest sequences in this yellow clade were dated to mid-January from Wuhan
150 and Shangdong, which may have been the origin of the yellow clade. CGMH-CGU-03 had no
151 travel history and the specimen was collected nearly 6 weeks after the two Chinese isolates were
152 collected. All other viruses in this clade were also dated after February 26. Separated by this long
153 duration from the two Chinese strains in mid-January, it is unlikely the later strains were directly
154 linked to the Wuhan strains. Although some AUS/NZ cases in this clade had a travel history to
155 Iran, the transmission route of these five Taiwanese cases remains unclear.

156 ***ORF8 Deletion Revealed by NGS Data Analysis***

157 Figure 2A shows the NGS coverage and depth of CGMH-CGU-01. This strain was
158 identical to the WuHan-1 strain (accession number MN908947.3). The most divergent strain
159 among the 14 Taiwanese sequences was CGMH-CGU-04 which showed nine nucleotide changes
160 (resulting in five amino acid changes) in the coding region compared to CGMH-CGU-01.
161 Notably, we detected a deletion in a 382-nucleotide (nt) sequence at genomic positions 27,848–
162 28,229 in CGMH-CGU-02. Figure 2B shows the coverage and depth of this strain. According to
163 the reference strain (WuHan-1), the genomic position of ORF8 was 27,894–28,259 (Figure 2C).
164 This 382-nt deletion begins upstream of ORF8 to nearly the end of ORF8. We further performed
165 NGS using a specimen isolated from the same patient. Reads yielding this 382-nt deletion were
166 confirmed in original specimen, although only the partial genome was assembled (Table 1).

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.29.014290. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

167 ***Within Four Nucleotide Changes among Virus Isolates from Clustered Patients***

168 COVID-19 has been reported to be transmitted through close contact among confirmed
169 cases. Regardless of whether individuals are symptomatic, their family members and co-workers
170 are at risk of becoming infected. Viral genomes No. 4–7 were from patients who had contact
171 with an index patient (CGMH-CGU-03). To identify the number of nucleotides changed in the
172 viral genome during clustered infections, we determined the viral full genomes either from viral
173 isolates (No. 3–6) or specimens (No. 7) of these 5 cases. Although the genomes of samples No.
174 3, 5, and 6 were identical, they differed from that of No. 4 at 3 ORF1ab nucleotide positions
175 A4788G, C10809T, and G21055A; the third position showed a synonymous change with a
176 G7019S amino acid substitution (Figure 3). Number 7 showed only one nucleotide difference
177 from No. 3, 5, and 6. These results suggest that only 4 nucleotide changes occurred in the viral
178 genome among cases in clustered infections.

179 ***Discussion***

180 Twelve full viral genomes were resolved in this study either from virus isolates or
181 directly from specimens. Phylogenetic tree analysis showed that these sequences were in
182 different clades, indicating that no major strain is currently circulating in Taiwan. A deletion in
183 ORF8 was found in one isolate, which has also been detected in patients in Singapore (10). Four
184 or fewer nucleotide differences were observed in the 5 genomes from clustered infections.

185 We detected a 382-nt deletion covering nearly the entire open reading frame 8 of the
186 CGMH-CGU-2 isolate obtained from a patient who returned from Wuhan in January. A similar
187 observation was reported for eight hospitalized patients in Singapore. During the SARS-CoV

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.29.014290. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

188 outbreak in 2003, deletions in ORF8 were observed, which were associated with a reduced
189 ability for virus replication in human cells (11).

190 RNA viruses show variations in their genomes due to nucleotide substitutions generated
191 by the low fidelity of RNA-dependent RNA polymerase during replication. The genome
192 variation of these viruses is thought to facilitate successful adaption to the environment of
193 various hosts. However, previous studies showed that the mutation rates of RNA viruses vary in
194 different viruses and depend on the viral transmission modes (12). Sequence analysis of SARS-
195 CoV-2 isolated from 5 patients from February 26 to March 9, 2020 in CGMH Taiwan revealed
196 only 4 mutations in their 29,903-nt genomic RNA. This suggests that the nucleotide substitution
197 rate is controlled during viral RNA replication. The nsp14 exoribonuclease encoded by several
198 coronaviruses plays a role in proofreading during genome replication (13, 14); further studies are
199 required to investigate the function of SARS-CoV-2 nsp14 in replication fidelity.

200 Timely sharing full genomes of SASR-CoV-2 from different locations is important for
201 monitoring genetic changes in the virus which may be associated with viral spreading and
202 clinical manifestations. We determined the sequences of SARS-CoV-2 in Taiwan in different
203 clades. Moreover, four or fewer nucleotide changes in viral genomes from five cases in clustered
204 infections indicated that sequencing is a useful tool for tracing the source of infection for this
205 type of RNA virus.

206

207 ***Acknowledgments***

208 This work was financially supported by the Research Center for Emerging Viral
209 Infections from The Featured Areas Research Center Program within the framework of the

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.29.014290. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

210 Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan, the Ministry of
211 Science and Technology (MOST), Taiwan (MOST 108-3017-F-182-001, MOST 107-2221-E-
212 182-064-MY2, and MOST 106-2320-B-182A-013-MY3), and Linkou Chang Gung Memorial
213 Hospital, Taiwan (No. CLRPG3B0048 and CMRPD1H0231–3).

214

215

216

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.29.014290. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

217 ***References***

218 1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients
219 with Pneumonia in China, 2019. N Engl J Med. 2020 Feb 20;382(8):727-33.

220 2. Coronavirus disease (COVID-2019) situation reports. [cited 2020 March 26];
221 https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

222 3. Number of Confirmed Cases of COVID-19 in Taiwan. [cited 2020 March 26];
223 https://sites.google.com/cdc.gov.tw/2019-ncov/taiwan

224 4. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
225 Bioinformatics. 2014 Aug 1;30(15):2114-20.

226 5. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory
227 requirements. Nat Methods. 2015 Apr;12(4):357-60.

228 6. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a
229 new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol.
230 2012 May;19(5):455-77.

231 7. Gong YN, Chen GW, Yang SL, Lee CJ, Shih SR, Tsao KC. A Next-Generation Sequencing
232 Data Analysis Pipeline for Detecting Unknown Pathogens from Mixed Clinical Samples and
233 Revealing Their Genetic Diversity. PLoS One. 2016;11(3):e0151495.

234 8. Kuraku S, Zmasek CM, Nishimura O, Katoh K. aLeaves facilitates on-demand exploration of
235 metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity.
236 Nucleic Acids Res. 2013 Jul;41(Web Server issue):W22-8.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.29.014290. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

237 9. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
238 phylogenies. Bioinformatics. 2014 May 1;30(9):1312-3.

239 10. Su YC AD, Young BE, Zhu F, Linster M, Kalimuddin S. Discovery of a 382-nt deletion
240 during the early evolution of SARS-CoV-2. biorxiv. 2020.

241 11. Chinese SMEC. Molecular evolution of the SARS coronavirus during the course of the
242 SARS epidemic in China. Science. 2004 Mar 12;303(5664):1666-9.

243 12. Hanada K, Suzuki Y, Gojobori T. A large variation in the rates of synonymous substitution
244 for RNA viruses and its relationship to a diversity of viral infection and transmission modes. Mol
245 Biol Evol. 2004 Jun;21(6):1074-80.

246 13. Eckerle LD, Lu X, Sperry SM, Choi L, Denison MR. High fidelity of murine hepatitis virus
247 replication is decreased in nsp14 exoribonuclease mutants. J Virol. 2007 Nov;81(22):12135-44.

248 14. Denison MR, Graham RL, Donaldson EF, Eckerle LD, Baric RS. Coronaviruses: an RNA
249 proofreading machine regulates replication fidelity and diversity. RNA Biol. 2011 Mar-
250 Apr;8(2):270-9.

251

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.29.014290. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

252 ***Table 1. Specimen collection, culture, and sequencing***

CGMH-CGU ID/Strain name**	Collection date	Viral culture (day)	Source* (Ct value of E gene)	Coverage and avg. depth of SARS-CoV-2
1	2020-01-25	-	SP (17.01)	99.9%; 1157.4
2	2020-02-04	14	MK2 (10.0)	100.0%; 4735.8
2	2020-02-04	-	NP (29.07)	80.0%; 3.3
3	2020-02-26	10	MK2 (14.25)	100.0%; 18,299.0
4	2020-02-27	4	Vero E6 (26.15)	99.2%; 446.0
5	2020-02-27	4	MK2 (12.78)	100.0%; 26,521.5
6	2020-03-05	5	MK2 (12.82)	100.0%; 13,029.9
7	2020-03-09	-	SP (22.98)	99.9%; 53.0
8	2020-03-10	-	NP (23.18)	100.0%; 10,412.2
9	2020-03-13	3	MK2 (10.89)	100.0%; 30,044.7
10	2020-03-13	3	MK2 (10.45)	100.0%; 29,614.0
11	2020-03-14	3	Vero E6 (11.08)	100.0%; 24,326.9
12	2020-03-14	3	MK2 (10.11)	100.0%; 34,422.0
TW/2	2020-01-23	-	-	-
TW/3	2020-01-24	-	-	-

253

254 * Sources from sputum (SP), nasopharyngeal swab (NP), and throat swab (TH) specimens, or
255 supernatant on MK2 and Vero E6 cells

256 ** Twelve GISAID accession numbers of CGMH-CGU No. 1–12 are EPI_ISL_411915,
257 EPI_ISL_417518, EPI_ISL_415741–3, and EPI_ISL_417519 –25. The other two Taiwanese

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.29.014290. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

258 strains (TW/2 and TW/3) were previously submitted to GISAID by Taiwan CDC, with accession

259 numbers (EPI_ISL_406031 and EPI_ISL_411926, respectively).

260

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.29.014290. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

261 ***Figure legends***

262 **Figure 1. Phylogenetic tree of Taiwanese and global strains.** Phylogeny was inferred using a
263 maximum likelihood approach. Taiwanese strains are highlighted. Strains isolated from different
264 locations and clades with specific variations are marked in different colors. Significant bootstrap
265 support values greater than 70% are shown.

266 **Figure 2. ORF8 deletion in SARS-CoV-2 genome.** A and B) NGS depths of CGMH-CGU-01
267 and CGMH-CGU-02 and C) genomic regions of ORF8 and ORF8 deletion according to the
268 reference strain are shown.

269 **Figure 3. Nucleotide and amino acid variations in SARS-CoV-2 genomes.** Compared to
270 CGMH-CGU-01 (identical to the reference strain), nucleotide and amino acid variations in the
271 SARS-CoV-2 genomes from Taiwanese strains are shown. Synonymous and nonsynonymous
272 mutations were marked by blue and red color, respectively. Amino acid changes were annotated
273 in parentheses. ORF8 deletion was marked in gray.

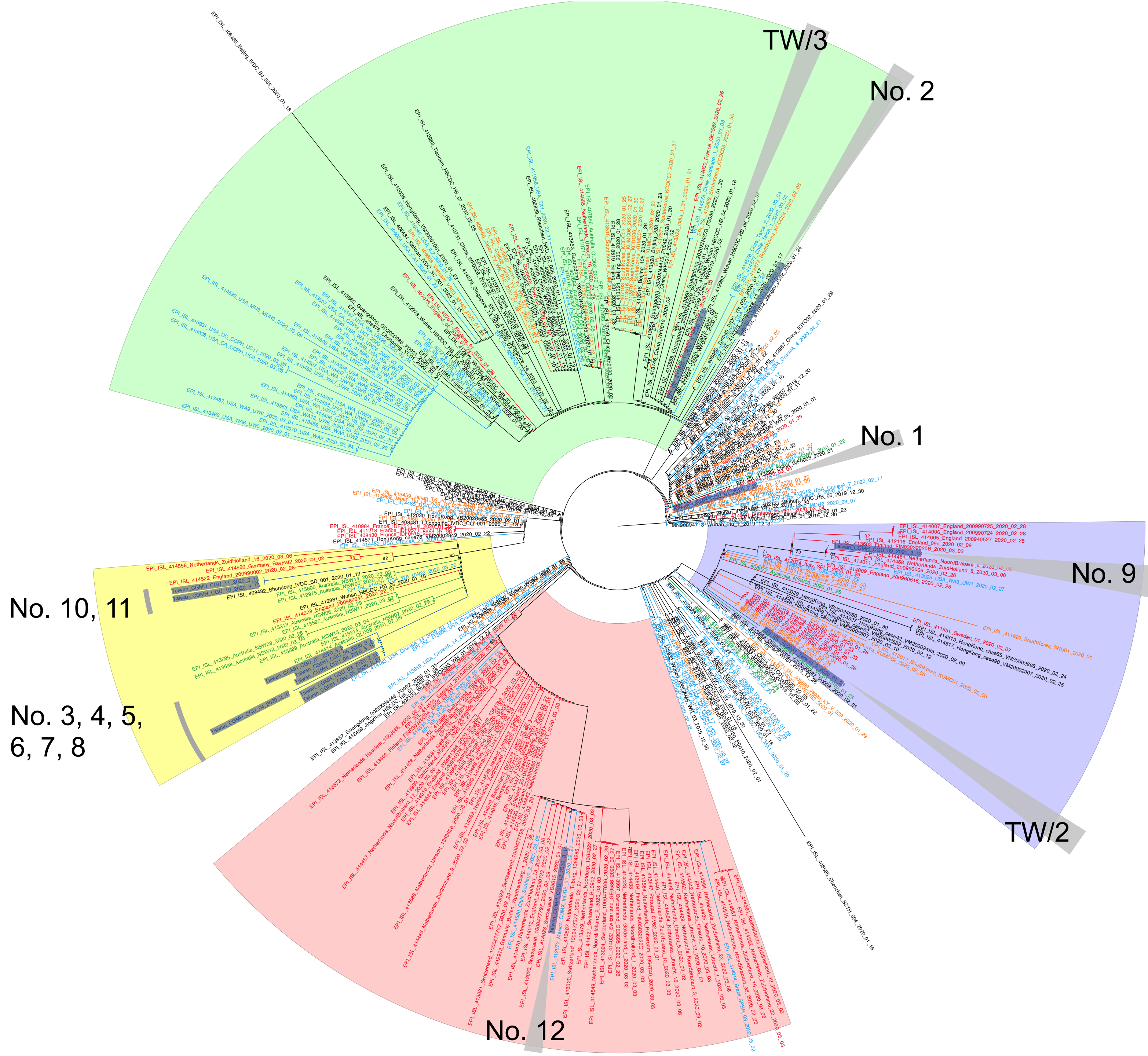
274

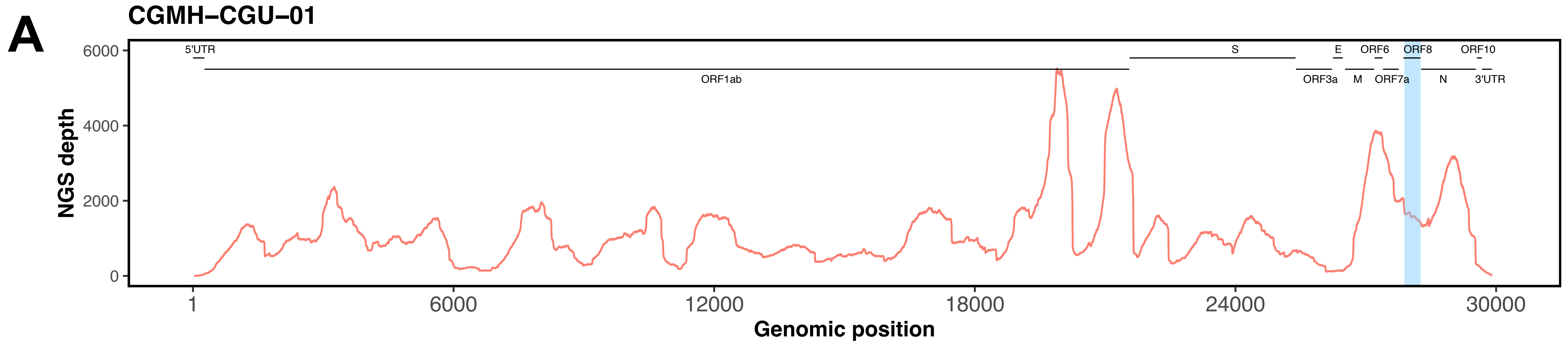
Location

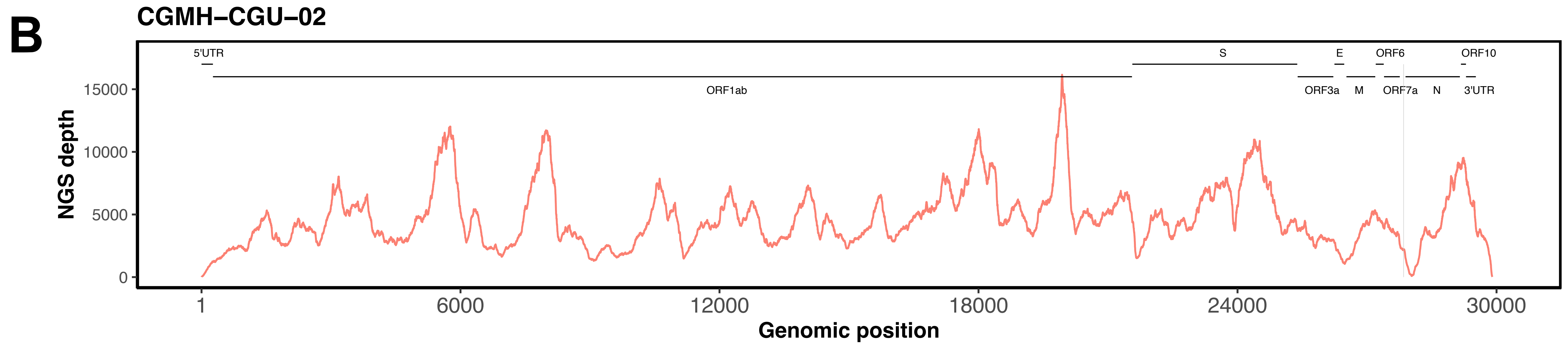
- Taiwan
- China
- Other Asian countries
- Europe
- Americas
- Oceania
- Africa

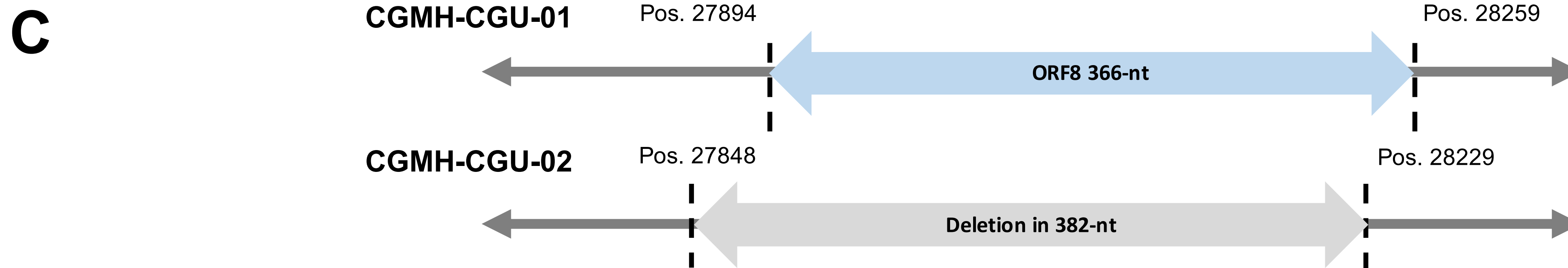
Clade

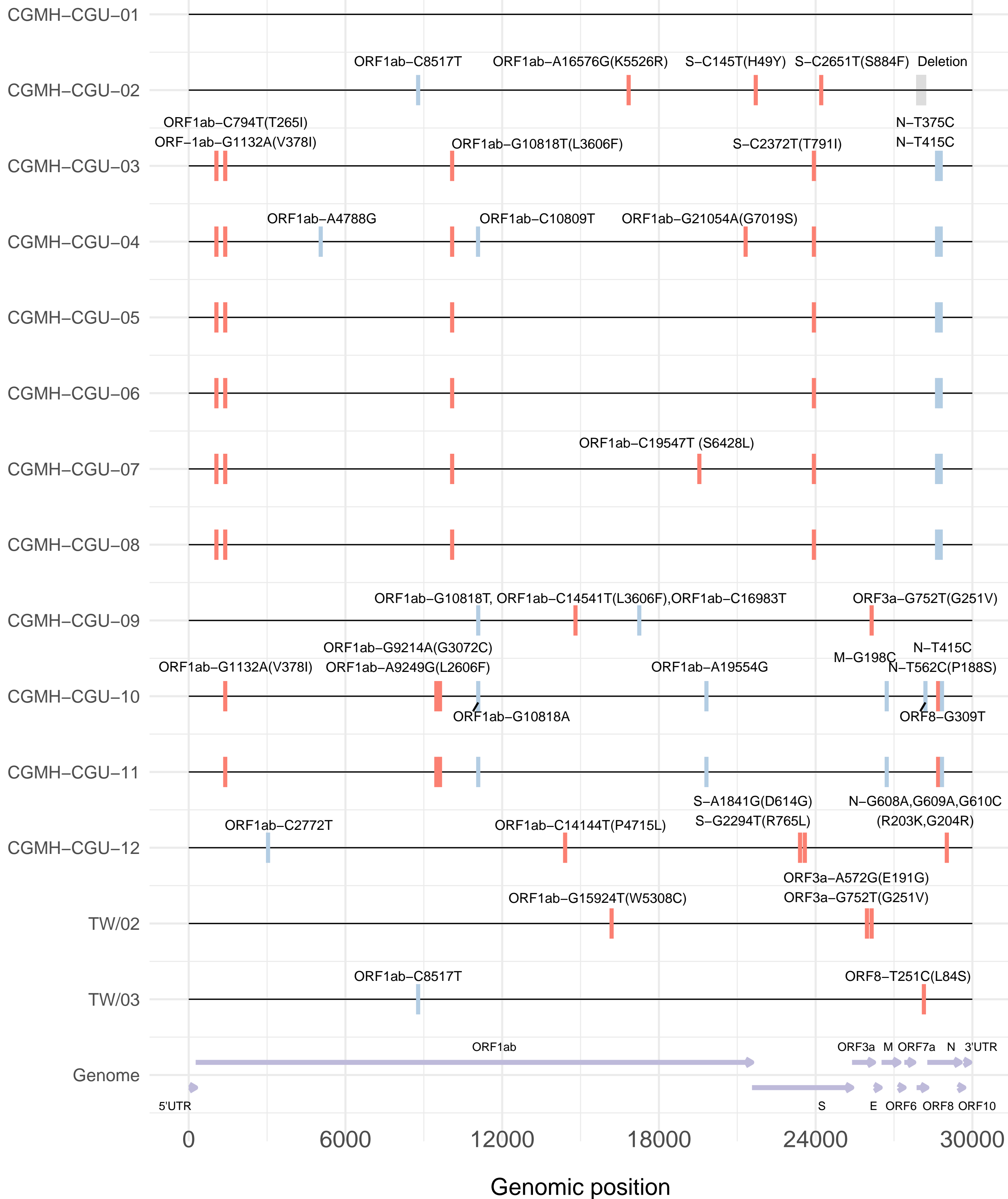
- ORF1ab−V378I
- ORF8−L84S
- S−D614G
- ORF3a−G251V




A
CGMH–CGU–01
NGS depth
Genomic position
5'UTR
ORF1ab
S
E
ORF6
ORF8
ORF10
ORF3a
M
ORF7a
N
3'UTR


B
CGMH–CGU–02
NGS depth
Genomic position
5'UTR
ORF1ab
S
E
ORF6
ORF10
ORF3a
M
ORF7a
N
3'UTR


C
CGMH-CGU-01
Pos. 27894
Pos. 28259
ORF8 366-nt
CGMH-CGU-02
Pos. 27848
Pos. 28229
Deletion in 382-nt



CGMH–CGU–01
CGMH–CGU–02
ORF1ab–C8517T
ORF1ab–A16576G(K5526R)
S–C145T(H49Y)
S–C2651T(S884F)
Deletion
CGMH–CGU–03
ORF1ab–C794T(T265I)
ORF–1ab–G1132A(V378I)
ORF1ab–G10818T(L3606F)
S–C2372T(T791I)
N–T375C
N–T415C
CGMH–CGU–04
ORF1ab–A4788G
ORF1ab–C10809T
ORF1ab–G21054A(G7019S)
CGMH–CGU–05
CGMH–CGU–06
CGMH–CGU–07
ORF1ab–C19547T (S6428L)
CGMH–CGU–08
CGMH–CGU–09
ORF1ab–G10818T, ORF1ab–C14541T(L3606F),ORF1ab–C16983T
ORF3a–G752T(G251V)
CGMH–CGU–10
ORF1ab–G9214A(G3072C)
ORF1ab–G1132A(V378I)
ORF1ab–A9249G(L2606F)
ORF1ab–A19554G
M–G198C
N–T415C
N–T562C(P188S)
ORF1ab–G10818A
ORF8–G309T
CGMH–CGU–11
CGMH–CGU–12
ORF1ab–C2772T
ORF1ab–C14144T(P4715L)
S–A1841G(D614G)
S–G2294T(R765L)
N–G608A,G609A,G610C
(R203K,G204R)
TW/02
ORF1ab–G15924T(W5308C)
ORF3a–A572G(E191G)
ORF3a–G752T(G251V)
TW/03
ORF1ab–C8517T
ORF8–T251C(L84S)
Genome
ORF1ab
ORF3a
M
ORF7a
N
3'UTR
5'UTR
S
E
ORF6
ORF8
ORF10
0
6000
12000
18000
24000
30000
Genomic position