bioRxiv preprint doi: https://doi.org/10.1101/2022.07.04.498645; this version posted April 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 **Spatial and Temporal Origin of The Third SARS-Cov-2 Outbreak in**
2 **Taiwan**
3

4 Jui-Hung Tai[1,2#], Yu Ken Low[1,3#], Selina Cai-Ling Wang[4], Hsin-Fu Lin[1], Tzi-Yuan Wang[5],
5 Jann-Tay Wang[6], Yu-Shu Liu[7], You-Yu Lin[1,2], Charles S.P. Foster[8], Sebastiaan J. van Hal[9, 10],
6 Ya-Yun Lai[11], Shiou-Hwei Yeh[12], Sui-Yuan Chang[7], Pei-Jer Chen[1,13,14,15*], Shu-Miaw
7 Chaw[2,5]
8
9 [1]Graduate Institute of Clinical Medicine, College of Medicine, National Taiwan University,
10 Taipei 10002, Taiwan
11 [2]Genome and Systems Biology Degree Program, National Taiwan University and Academia
12 Sinica, Taipei 10617, Taiwan
13 [3]Institute of Ecology and Evolutionary Biology, National Taiwan University, Taipei 10617,
14 Taiwan
15 [4]Taipei First Girls High School, Taipei 10006, Taiwan
16 [5]Biodiversity Research Center, Academia Sinica, Taipei 11529, Taiwan
17 [6]Department of Internal Medicine, National Taiwan University Hospital and National Taiwan
18 University College of Medicine, Taipei, Taiwan
19 [7]Department of Clinical Laboratory Sciences and Medical Biotechnology, College of
20 Medicine, National Taiwan University, Taipei 10002, Taiwan.
21 [8]School of Medical Sciences, University of New South Wales, Randwick NSW 2031,
22 Australia
23 [9]NSW Health Pathology, Department of Infectious Diseases and Microbiology, Royal Prince
24 Alfred Hospital, Sydney, NSW, Australia
25 [10]Central Clinical School, University of Sydney, Sydney, NSW, Australia
26 [11]Center of Genomic and Precision Medicine, National Taiwan University, Taipei 10055,
27 Taiwan
28 [12]Department of Microbiology, College of Medicine, National Taiwan University, Taipei
29 10617, Taiwan.
30 [13]Hepatitis Research Center, National Taiwan University College of Medicine and National
31 Taiwan University Hospital, Taipei 10002, Taiwan
32 [14]Department of Internal Medicine, National Taiwan University College of Medicine and
33 National Taiwan University Hospital, Taipei 10002, Taiwan

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.04.498645; this version posted April 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

34 [15]Department of Medical Research, National Taiwan University College of Medicine and
35 National Taiwan University Hospital, Taipei 10002, Taiwan
36
37 # These authors contributed equally
38
39 *Correspondence: peijerchen@ntu.edu.tw
40

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.04.498645; this version posted April 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

41 # ABSTRACT

42 Since the first report of SARS-CoV-2 in December 2019, Taiwan has gone through three
43 local outbreaks. Unlike the first two outbreaks, the spatial and temporal origin of the third
44 outbreak (April 20 to November 5, 2021) is still unclear. We assembled and analyzed a data
45 set of more than 6,000 SARS-CoV-2 genomes, including 300 from Taiwan and 5812 related
46 sequences downloaded from GISAID as of 2021/12/08. We found that the third outbreak in
47 Taiwan was caused by a single virus lineage belonging to Alpha (B.1.1.7) strain. This lineage,
48 T-III (the third outbreak in Taiwan), carries a distinct genetic fingerprint, consisting of spike
49 M1237I (S-M1237I) and three silent mutations, C5812T, C15895T, and T27869C. The T-III
50 is closest to the sequences derived from Turkey on February 8, 2021. The estimated age of
51 the most recent common ancestor (TMRCA) of T-III is March 23, 2021 (95% highest
52 posterior density [HPD] February 24 - April 13, 2021), almost one month before the first
53 three confirmed cases on April 20, 2021. The effective population size of the T-III showed
54 approximately 20-fold increase after the onset of the outbreak and reached a plateau in early
55 June 2021. Our results reconcile several unresolved observations, including the occurrence of
56 two infection clusters at the same time without traceable connection and several airline pilots
57 who were PCR negative but serum IgM-/IgG+ for SARS-CoV-2 in late April. Therefore, in
58 contrast to the general notion that the third SARS-CoV-2 outbreak in Taiwan was sparked by
59 two imported cases from USA on April 20, 2021, which, in turn, was caused by the partial
60 relaxation of entry quarantine measures in early April 2021, our comprehensive analyses
61 demonstrated that the outbreak was most likely originated from Europe in February 2021.

62

63

64

65 ***Keywords:*** Spike protein, M1237I, Alpha/B.1.1.7, time to the most recent common ancestor
66 (TMRCA), haplotype network

67

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.04.498645; this version posted April 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

68 ## Introduction

69 Since the first report of coronavirus disease 2019 (COVID-19) caused by *Severe acute*
70 *respiratory syndrome coronavirus* 2 (SARS-CoV-2) in the December of 2019 in Wuhan, the
71 virus has rapidly sparked an ongoing pandemic. SARS-CoV-2 is the third coronavirus
72 causing severe respiratory illness in humans after SARS-CoV and Middle East respiratory
73 syndrome coronavirus (MERS-CoV) (1-3).

74

75 After experiencing a series of SARS outbreaks in 2003 which caused 668 probable cases and
76 181 deaths (4), Taiwan has been exceedingly cautious of emerging disease and has
77 strengthened its pandemic control measures. For example, the Central Epidemic Command
78 Center (CECC) was established after the SARS epidemic in 2003, and was activated on 20
79 January 2020, before the first case of COVID-19 was identified in Taiwan. The control
80 strategy implemented by the CECC was based on three essential components: border control,
81 case identification and contact tracing, and containment.

82

83 As of February 2022, Taiwan has ended three local COVID-19 outbreaks, while the fourth is
84 ongoing (Fig. 1). The first local outbreak was between January 28 and April 11, 2020 and
85 involved 55 confirmed cases. Most of these local cases had a contact history or exposure to
86 SARS-CoV-2 infected patients (5). The second local outbreak started on January 12 and
87 ended on February 9, 2021. It was sparked by an intra-hospital infection and involved 21
88 cases. The third outbreak consists of two infection clusters and lasted for at least five months
89 with more than 14,000 cases. Among them, all of the 614 cases tested by CECC are Alpha
90 strain (B.1.1.7) of SARS-CoV-2 (6). The first cluster began with two airline flight crews
91 (case 1078 and 1079) showing symptoms on April 17 and 18, respectively, after returning
92 from the USA on April 16, 2021 (Table S1). They were diagnosed as COVID-19 positive on
93 April 20, 2021 (7). On the same day, one pilot of the same airline tested positive for COVID-
94 19 in Australia while on duty (8). This cluster was subsequently linked to staff working in a
95 hotel in Northern Taiwan close to Taoyuan Airport where airline pilots and flight crews
96 stayed during their quarantine (9) (Cluster I). The second cluster involved several local
97 incidences in New Taipei City and Yilan County that later spread to many other counties
98 (Cluster II). This cluster was first recognized on May 11, 2021, but a later survey found that
99 the first case (case 1424) in this cluster showed symptoms as early as April 23, 2021 (10).

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.04.498645; this version posted April 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

100 Unlike the first two outbreaks, the spatial and temporal origin of this outbreak is still under
101 debate.
102
103 While it has been suggested that the third SARS-CoV-2 outbreak in Taiwan was sparked by
104 two imported cases from USA on April 20, 2021 (11), which, in turn, was caused by the
105 partial relaxation of entry quarantine measures in early April 2021 (12, 13), several
106 unresolved observations remain to be answered. First, according to CECC, the dates of
107 symptom onset in two infection clusters are very close (April 16 for the first cluster and April
108 23 for the second) (10). However, the relationships between the two clusters remain elusive.
109 Second, several airline pilots and their family members were PCR negative but with serum
110 IgM-/IgG+ in late April and early May 2021 (14, 15), suggesting they have probably been
111 infected by SARS-CoV-2 earlier than April 20, 2021. While it is not entirely impossible that
112 these cases were not linked to the third outbreak in Taiwan, the fact that there were only 1077
113 confirmed COVID-19 cases in Taiwan prior to April 20, 2021 (Fig. 1) makes it more
114 probable that these cases are associated to the third outbreak.
115
116 Understanding the origin of the outbreak is important not only for epidemiological study but
117 also for the design of future control measures. In order to clarify the origin(s) and
118 interrelationship between the clusters, we sequenced and reconstructed a phylogeny of
119 SARS-CoV-2 genomes. We find that the third outbreak is caused by a single virus lineage (T-
120 III), descended from the Alpha variant of concern (B.1.1.7), which carries four distinctive
121 mutations, including spike M1237I (S-M1237I) and three silent changes, from its most
122 closely related sequences in Europe.
123

124 Methods

125 Sample preparation

126 Thirty-three specimens from SARS-CoV-2 positive patients were obtained (Table S1). One
127 sample from Royal Prince Alfred Hospital, Australia was sequenced directly from extract
128 using Nanopore (16). The remaining 32 specimens from individuals who were admitted to
129 National Taiwan University Hospital (NTUH) were maintained in viral transport medium.
130 Virus in the specimens was propagated in VeroE6 cells in Dulbecco's modified Eagle's
131 medium (DMEM) supplemented with 2 μg/mL tosylsulfonyl phenylalanyl chloromethyl
132 ketone (TPCK)-trypsin (Sigma-Aldrich, St. Louis, MI, USA). After one passage, culture

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.04.498645; this version posted April 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

133 supernatant was harvested when cytopathic effects (CPE) were observed in more than 70% of
134 cells, and the culture supernatant was harvested for viral genomic sequencing. Viral RNA
135 was extracted by QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany) according to the
136 instruction manual. In brief, 140 microliter of culture supernatant was used for RNA
137 extraction and the extracted RNA was eluted using 40 μL RNase-free H2O. The RNA was
138 stored at -20°C until use.
139

140 **Reverse transcription polymerase chain reaction (RT-PCR)**

141 SARS-CoV-2 cDNA was generated from 100 ng of RNA in a RT-PCR reaction buffer
142 containing 4 μL of 5X PrimeScript IV 1st strand cDNA Synthesis Mix (Takara Bio, Kusatsu,
143 Shiga, Japan), 2 μL of 50 μM random hexamer primer, and variable amount of DEPC water
144 to fill up to 20 μL of total reaction volume. Pre-heat at 30□ for 10 minutes, followed by 20
145 minutes of 42□, and then 15 minutes of 70□.
146

147 **Amplification of complete SARS-CoV-2 genomes with multiplex PCR**

148 The “Midnight” amplicon protocol was used to generate 1200 bp amplicons. Briefly, PCR
149 was used with 2·5 μL of cDNA product from RT-PCR in 22·5 μL buffer, containing 12·5 μL
150 of Q5 Hot Start High-Fidelity 2X Master Mix (New England BioLabs, M0494S), 1·1 μL of
151 10 μM SARS-Cov2-Midnight-1200 primer (either Pool 1 or Pool 2)(17) (Integrated DNA
152 Technologies, Coralville, IA USA), and 8·9 μL of nuclease-free water (Thermo Scientific,
153 Waltham, MA, USA). Amplifications were performed with 30 seconds of 98□ for initial
154 denaturation, followed by 25 cycles of 98□ for 15 seconds and 65□ for 5 minutes in a Veriti
155 96-Well Thermal Cycler machine (Applied Biosystems, Waltham, MA, USA). Each sample
156 was separately amplified using both Pool 1 and Pool 2 primers. 20 μL PCR products were
157 then purified with DNA Clean & Concentrator-5 (Zymo Research, Irvine, CA, USA).
158 Amplicons were eluted with 25 μL of nuclease-free water (Thermo Scientific). DNA quality
159 checks were done using a Nanodrop (Thermo Scientific) and 1·5% agarose gel
160 electrophoresis (Fig. S1).
161

162 **Library preparation for Nanopore MinION sequencing**

163 For each sample, 1 μL (0·5 μL from each pool of a same sample), approximately 50 ng, of
164 purified PCR amplicons were used for library preparation. The KAPA HyperPrep Kit (Roche,
165 Basel, Switzerland) was used in a 15 μL reaction for end repair and A-tailing. The reactions
166 contained 1 μL of amplicon, 1·75 μL of End-repair & A-tailing Buffer, 0·75 μL of End-repair

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.04.498645; this version posted April 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

167 & A-tailing Enzyme, and 11·5 µL of nuclease-free water (Thermo Scientific). Reactions were
168 done in 30 minutes at 20 □ and 30 minutes at 65 □. Each sample was then barcoded via
169 ligation in a 27·5 µL reaction at 20□ for 15 minutes, with 2·5 µL of DNA Ligase, 7·5 µL of
170 Ligation Buffer, 2·5 µL of Native Barcode (Oxford Nanopore Technologies, Oxford, UK),
171 and 15 µL of A-tailed amplicon. After that, barcoded amplicons were purified with 33 µL
172 (1.2X) of KAPA Pure Beads (Roche) by following the official protocol and eluted with 11 µL
173 of nuclease-free water (Thermo Scientific, R0582). We then pooled 2·7 µL of each barcoded
174 amplicon. The 110 µL of reaction for adapter ligation contained 65 µL of pooled barcoded
175 amplicons, 5 µL of Adapter Mix II (Oxford Nanopore Technologies, EXP-NBD104), 30 µL
176 of Ligation Buffer and 10 µL of DNA Ligase. After incubating in 20 □ for 15 minutes,
177 libraries were purified with 110 µL (1X) of KAPA Pure Beads (Roche, 07983271001) and
178 Short Fragment Buffer (Oxford Nanopore Technologies, SQK-LSK109) according to the
179 official protocol and then eluted with 14 µL of Elution Buffer (Oxford Nanopore
180 Technologies, SQK-LSK109).
181

182 **Nanopore MinION sequencing**

183 We sequenced 220 ng of purified library with an Oxford Nanopore Technoogies MinION
184 R.9.4.1 flowcell (FLO-MIN106) with the software MinKNOW Core version 19.12.5. Super-
185 accurate basecalling was done using Guppy version 5.0.11 with default settings in GPU mode.
186 The Nanopore reads of each sample were mapped to the reference genome Wuhan-hu-1
187 (EPI_ISL_402125) by *BWA-mem* [preprint] (18) to generate an alignment file. The average
188 read depth is 992 per sample (Table S1). *BCFtools* (19) was applied to call variant (-i 'QUAL
189 > 20') and generate consensus sequence.
190

191 **Data collection**

192 A total of 267 complete and high coverage SARS-CoV-2 genomes from Taiwan with
193 complete collection date were downloaded from the Global Initiative on Sharing Avian
194 Influenza Data (GISAID, https://www.gisaid.org/)(20) on December 8, 2021. To find the
195 SARS-CoV-2 genomes most closely related to the third outbreak in Taiwan, we used the
196 Audacity Instant search tool in GISAID to search the database and used EPI_ISL_2455264
197 (case 1079) as the query. After excluding 42 sequences from Taiwan, 5,812 foreign
198 sequences with fewer than or equal to 10 SNP differences were downloaded. All the
199 sequences used in this study can be found in https://github.com/ala98412/T-III.

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.04.498645; this version posted April 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

200

201 **M1237I frequency in different genetic backgrounds**

202 We used tools within GISAID to provide a quick search of the database. We chose different
203 SARS-CoV-2 strains in the drop-down menu “Variant” to get the count of sequences in each
204 major variant strain (e.g., Alpha, Delta …), and further selected “Spike_M1237I” in the drop-
205 down menu “Substitutions” to receive the count of sequences with the M1237I mutation in
206 each genetic background. We calculated the frequency of M1237I in different genetic
207 backgrounds via the count of sequences with the M1237I mutation divided by the total
208 number of sequences representing each major variant.

209

210 **Sequence analysis and phylogenetic reconstruction**

211 All sequences were aligned against the reference genome Wuhan-hu-1 (EPI_ISL_402125) by
212 using MAFFT v7 (21). Nucleotide diversity, including number of segregating sites,
213 Watterson’s estimator of θ (22), and nucleotide diversity (π)(23), was estimated using
214 MEGA-X (24). Phylogenetic trees were also constructed by using the neighbor-joining
215 method (25) based on Kimura’s two-parameter model in MEGA-X. The Nexus file for the
216 haplotype network analysis was generated using DnaSP 6.0 (26) and input into PopART v1.7
217 (27) to construct the haplotype network using TCS software (28).

218

219 The times to the most recent common ancestor (TMRCA) of T-III lineage were estimated
220 using an established Bayesian Markov chain Monte Carlo (MCMC) approach implemented in
221 BEAST version 2.5 (29). The sampling dates were incorporated into TMRCA estimation.
222 These analyses were performed using the Hasegawa-Kishino-Yano (HKY) model of
223 nucleotide substitution assuming an uncorrelated lognormal relaxed molecular clock (30). We
224 linked substitution rates for the first and second codon positions and allowed independent
225 rates for the third codon position. We performed two independent runs with 1×10^7 MCMC
226 steps and combined the results. Log files were checked using Tracer
227 (http://beast.bgio.ed.ac.uk/Tracer). Effective sample sizes were >300 for all parameters.

228

229 ## Results

230 A total of 33 SARS-CoV-2 genomes were sequenced, including 32 from NTUH, Taiwan, and
231 one from Royal Prince Alfred Hospital, Australia (Table S1). We also downloaded all
232 available 267 genomes from Taiwan from GISAID as of 2021/12/08, to construct the

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.04.498645; this version posted April 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

233 phylogeny as shown in Fig. 2. Since the cases from the first outbreak had a contact history or
234 exposures to different SARS-CoV-2 infected patients, they do not form a single cluster in the
235 phylogeny. The sequences derived from the second local outbreak are presented in emerald
236 green (Fig. 2). The third local outbreak consisted of the Alpha strain (B.1.1.7), divided into
237 two main clusters, shown in Fig. 2b. All sequences in the basal lineage of Fig. 2b were from
238 imported cases, whereas all 80 sequences in the more advanced lineage were local to Taiwan.
239 Cluster I and II cases prior to May 10, 2021, are marked on the tree, and there is no clear
240 differentiation between them. Subsequently, a steep rise in case numbers made it impossible
241 to distinguish between the two clusters. Hereafter, we name this particular lineage as T-III,
242 which stands for the third outbreak in Taiwan, and note that the T-III belongs to Pangolin
243 lineage B.1.1.7. The nucleotide changes among T-III are listed in Table S2.
244

245 **Spatial and temporal origin of the third local outbreak in Taiwan**

246 In order to search for the spatial origin of T-III, the sequence recovered from the earliest case
247 in the third outbreak (case 1079: EPI_ISL_2455264, Table S1) was used to search against the
248 GISAID database. There were 5,812 sequences with ≤ 10 nucleotide differences from
249 EPI_ISL_2455264 as of November 21, 2021. Phylogenetic reconstruction including all 5,812
250 sequences and T-III demonstrate that the latter is a distinctive lineage (Fig. S2). The vast
251 majority of sequences closely related to T-III were from Europe, including Turkey (Fig. S2b).
252 We also tested whether the sequence from Australia is unique to T-III. All available Alpha
253 strain genomes from Australia collected during April 2021 were downloaded from GISAID
254 as of February 17, 2023 and compared with T-III (Fig S3). The sole sequence
255 (EPI_ISL_1756021) belonging to T-III from Australia is distinct from the other Alpha strains
256 collected at the same time. As EPI_ISL_1756021 was from a flight crew from Taiwan, the
257 result supports the notion that the T-III is unique to the third outbreak in Taiwan.
258

259 Haplotype network analyses of the T-III lineage reveals 48 haplotypes. The network shows
260 that T-III differs from the outgroups in four mutation steps (Fig. 3), including two
261 synonymous mutations, C5812T and C15895T, in Orf1ab, one nonsynonymous mutation
262 G25273C (M1237I) in Spike, and one T27869C mutation in a non-coding region (Table S3).
263 The closest outgroup haplotype consists of four sequences, including two sequences from
264 Turkey that were collected on February 8, 2021 (EPI_ISL_1097034, 1097035). The rest were
265 collected after the onset of the third outbreak in Taiwan (Table S4). Further database mining
266 confirmed that of 1,140,328 Alpha strain genomes examined as of December 11, 2021, only

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.04.498645; this version posted April 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

267 the lineage T-III possessed the four above mentioned mutations, which form a distinctive
268 genetic fingerprint. Within T-III, the network forms a star-like shape centered on a core
269 haplotype comprising 26 sequences. Most of the remaining haplotypes are directly connected
270 to this major haplotype. Of the three cases identified on the first day of the third outbreak
271 (April 20, 2020), the case from Australia belongs to the major haplotype, with the rest (1078
272 and 1079) one mutation away.

273

274 The estimated substitution rate of T-III is $9.8x10^{-4}$ substitution per site per year (95% highest
275 posterior density [HPD] $5.6x10^{-4}$ – $1.4x10^{-3}$) which is close to 8.4 x 10^{-4} /site/year of SARS-
276 CoV-2 (31), and 7.5□×□10^{-4} /site/year of strain B.1.1 (32). The results also suggest that a
277 single passage in Vero E6 cells would not significantly influence the number of mutations on
278 SARS-CoV-2 sequences. The date of the most recent common ancestor (TMRCA) of T-III is
279 in late March (3/23/2021; 95% HPD February 24 - April 13, 2021) (Fig. 4), almost one
280 month before the first three confirmed cases on April 20, 2021. We noticed that the mutation
281 rate of T-III during this epidemic was almost constant (Fig. S4). Effective population size of
282 the T-III lineage increased approximately 20-fold after the onset of the outbreak and reached
283 a plateau in early June. The estimated demographic expansion of T-III is consistent with
284 epidemiological data (Fig. 1). We noticed that demography in Fig. 4 does not capture the
285 population decline after July 2021 as shown in Fig. 1. This finding is because most sequences
286 used in this analysis were collected before June 2021 (Table S1), with only four sequences
287 obtained after July 2, 2021. As we are interested in the outbreak origin, this sampling strategy
288 should not affect our conclusions.

289

290 Rapid population expansion can also be revealed by contrasting patterns of genetic variation
291 estimated using different approaches. The Watterson's estimator of θ (6·95 x 10^{-4}) is
292 approximately seven times higher than the nucleotide diversity (π) (1·05 x 10^{-4}), leading to
293 significantly negative Tajima's D (-2·82, $p < 0·001$) (Table S5). Because θ is strongly
294 influenced by rare mutations, which are common during recent population expansion or after
295 positive selection (33), it is a better estimator of genetic diversity for T-III.

296

297 **Discussion**

298 Our results provide solid evidence that the third local outbreak in Taiwan was caused by a
299 single lineage, T-III. This does not in itself mean that two clusters of infections (see

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.04.498645; this version posted April 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

300 Introduction) have a single common origin. For example, among 293,742 sequences analyzed
301 during the first year of the SARS-CoV-2 pandemic, the most abundant haplotype was
302 sampled 3,466 times from across 53 countries (Table S6). It is possible that the haplotype
303 exhibited some transmission advantage, making it widespread. Under this scenario, two
304 clusters of infection may be caused by the same lineage imported from different sources.
305 However, as the T-III lineage bearing a combination of four unique mutations is exclusive to
306 Taiwan, it seems highly unlikely that this lineage was imported from different sources.
307
308 Our estimation that T-III originated from Europe with TMRCA on March 23, 2021 (95%
309 HPD February 24, 2021 □ April 13, 2021) reconciles several unresolved observations. First,
310 the first two cases (case 1078 and 1079) of the outbreak shared identical sequences,
311 indicating that they were from the same source. Nevertheless, the cycle threshold (Ct) values
312 at the time of diagnosis (April 20, 2021) were 29 and 17 for case 1078 and 1079, respectively,
313 suggesting that they were infected at different times (7). Second, several airline pilots and
314 their family members were PCR negative but with serum IgM-/IgG+ in late April and early
315 May 2021 (14, 15). It has been shown that IgM levels increase during the first week after a
316 SARS-CoV-2 infection, peak after two weeks, and then recede to near-background levels in
317 most patients. IgG is detectable one week after disease onset and is maintained at a high level
318 for a long period (34). Consequently, IgM negative but IgG positive individuals have
319 probably been infected by SARS-CoV-2 earlier than April 20, 2021. Third, according to
320 CECC, the dates of symptom onset in two seemingly unrelated infection clusters are very
321 close (April 16 for the first cluster and April 23 for the second) (10). As our phylogenetic
322 analysis reveals that all sequences in the third outbreak have a single origin, the occurrence of
323 two infection clusters at similar time without traceable connection demonstrates that the virus
324 may have been cryptically circulating in the community undetected. Consequently, the origin
325 of the third outbreak was most likely prior to April 20, 2021.
326
327 As most sequences closely related to T-III were from Europe, our results disagree with the
328 notion that the outbreak was imported by cases 1078 and 1079 from the USA (11). Among
329 four sequences closest to T-III, only two from Turkey were collected before April 20, 2021
330 (on February 8, 2021). The rest appeared after April 20 and cannot be associated with the
331 third outbreak. Consequently, the lineage leading to the third outbreak was most likely
332 introduced from Europe, perhaps Turkey, by infected travelers in February 2021 (Fig. 3).
333

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.04.498645; this version posted April 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

334 There are several possible routes by which SARS-CoV-2 can be introduced by incoming
335 travelers. First, some carriers may have the virus undetected during quarantine. Although the
336 mean incubation period ranges from five to seven days, it can be longer than 14 days (35, 36).
337 Indeed, there has been much debate about whether changing from the mandatory five day
338 home quarantine plus nine day autonomous health management (or the so-called "5+9"
339 strategy) to three day home quarantine plus 11 day autonomous health management ("3+11")
340 for flight crews on April 15, 2021 (37) was the cause of the third outbreak (12, 13). Since the
341 outbreak likely originated in mid-March, as our analyses demonstrate, the possibility that
342 changing strategies from "5+9" to "3+11" directly caused the outbreak can be ruled out.
343
344 Second, there may have transmission from people associated with quarantine hotels to the
345 community. For example, from December 2021 to March 2022 alone, 15 infection clusters
346 occurred in quarantine hotels (38). Since one of the two infection clusters in the third
347 outbreak directly link to a hotel where many flight crews stayed during their quarantine (see
348 Introduction), it is most likely that the outbreak was introduced from this hotel during late
349 February to early April 2021. The virus was then undetected while spreading within the
350 community until late April 2021.
351
352 Our results demonstrate that even a small number of imported cases can undermine the strict
353 control measures in Taiwan (39). We also show that phylogenetic approaches can be used to
354 trace the outbreak derived from local spread by infected travelers (40). Although Taiwan has
355 been very cautious of emerging disease and has strengthened its control measures, the SARS-
356 CoV-2 genomes have not been sampled in proportion to the actual size of infection and made
357 publicly available in a timely fashion. This situation severely hinders efforts to trace the
358 underlying transmission patterns of spread. Consequently, policies that include real-time
359 public-sharing and transparency for infectious disease surveillance and control are critically
360 and urgently needed. Blackouts in information sharing and transparency can cause
361 information disparity and suspicion, which in turn hamper pandemic control efforts (41).
362

363 ## Declaration of Interests

364 The authors declare that they have no competing interests.
365

366 ## Availability of Data and Materials

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.04.498645; this version posted April 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

367 The National Taiwan University Hospital, Taiwan and the Royal Prince Alfred Hospital,
368 Australia sequences have uploaded to Global Initiative on Sharing Avian Influenza Data
369 (GISAID, https://www.gisaid.org/). The remaining genome sequences were downloaded from
370 GISAID as well. All the sequences used in this study can be downloaded from
371 https://github.com/ala98412/T-III.
372

373 ## Authors' Contributions

374 JHT, SCLW, and HFL analyzed the data. SYC and SHY guided the analyses. JHT, YKL,
375 SCLW, PJC, HYW, CSPF drafted and revised the manuscript. YKL, Ya-Yun Lin, You-Yu
376 Lin, JTW, YSL, and SYC prepared samples. TYW, CSPF, SJH, and You-Yu Lin sequenced
377 samples. PJC, SMC, and HYW designed the study, obtained funding, and wrote the
378 manuscript.
379

380 ## Funding

381 This study was supported by grants from the Ministry of Science and Technology (MOST),
382 Taiwan (111-2321-B-002-017, 111-2634-F-002-017, 109-2311-B-002-023-MY3), Ministry
383 of Education, and Academia Sinica. The funding sources had no role in the study design, data
384 collection, analysis, interpretation, or writing of the report.
385

386 ## Acknowledgement

387 The authors thank four anonymous reviewers for constructive comments. We also thank Prof.
388 Chwan-Chuen King for her comments on the early version of this manuscript.
389

390 ## Reference

391 1. Cui J, Li F, Shi ZL. 2019. Origin and evolution of pathogenic coronaviruses. Nat Rev
392 Microbiol 17:181-192.

393 2. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei
394 YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC,
395 Zhang YZ. 2020. A new coronavirus associated with human respiratory disease in
396 China. Nature 579:265-269.

397 3. Chaw S-M, Tai J-H, Chen S-L, Hsieh C-H, Chang S-Y, Yeh S-H, Yang W-S, Chen P-
398 J, Wang H-Y. 2020. The origin and underlying driving forces of the SARS-CoV-2
399 outbreak. J Biomed Sci 27:73.

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.04.498645; this version posted April 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

400 4. Chen K-T, Twu S-J, Chang H-L, Wu Y-C, Chen C-T, Lin T-H, Olsen SJ, Dowell SF,
401 Su I-J. 2005. SARS in Taiwan: an overview and lessons learned. International Journal
402 of Infectious Diseases 9:77-85.

403 5. Lin CY, Wang WH, Urbina AN, Tseng SP, Lu PL, Chen YH, Yu ML, Wang SF.
404 2020. Importation of SARS-CoV-2 infection leads to major COVID-19 epidemic in
405 Taiwan. Int J Infect Dis 97:240-244.

406 6. Liu MT. 2021. Analysis of pathogens caused severe pneumonia, encephalitis and
407 unknown infectious diseases. Control CfD, CECC, Taiwan.

408 7. CECC. 2021. CECC confirms 2 more cases; sources of infection to be investigated,
409 *on* CDC.
410 https://www.cdc.gov.tw/En/Category/ListContent/tov1jahKUv8RGSbvmzLwFg?uaid
411 =NrqGhobGKPIYRPGw2w2z0Q. Accessed

412 8. CECC. 2021. CECC launches an investigation into notification concerning an airline
413 pilot testing positive for COVID-19 in Australia and contacts testing positive in
414 Taiwan, *on* CDC.
415 https://www.cdc.gov.tw/En/Bulletin/Detail/fDE2kCBYdnmbK0QiU_tSMg?typeid=1
416 58. Accessed

417 9. CECC. 2021. CECC confirms 2 more indigenous cases; contractor of airport
418 quarantine hotel and family member of airline pilot found to have COVID-19, *on*
419 CDC. https://www.cdc.gov.tw/En/Bulletin/Detail/Rk3a6olNXLGuSKn6-
420 YSy4Q?typeid=158. Accessed

421 10. CECC. 2021. CECC confirms 185 more COVID-19 cases, including 180 indigenous
422 cases and 5 imported cases, *on* CDC.
423 https://www.cdc.gov.tw/En/Category/ListContent/tov1jahKUv8RGSbvmzLwFg?uaid
424 =70ryEYOTbvx8VaItygA1IA. Accessed

425 11. Liu L-T, Tsai J-J, Chang K, Chen C-H, Lin P-C, Tsai C-Y, Tsai Y-Y, Hsu M-C,
426 Chuang W-L, Chang J-M, Hwang S-J, Chong I-W. 2022. Identification and Analysis
427 of SARS-CoV-2 Alpha Variants in the Largest Taiwan COVID-19 Outbreak in 2021.
428 Frontiers in Medicine 9.

429 12. Tong C, Shi W, Zhang A, Shi Z. 2021. Tracking and Controlling the Spatiotemporal
430 Spread of SARS-CoV-2 Lineage B.1.1.7 in COVID-19 Reopenings. Geohealth
431 5:e2021GH000517.

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.04.498645; this version posted April 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

432 13. Wu C, Wu E, Wu C, Wu KC. 2022. How Taiwan Succeeded in Containing Its 2021
433 COVID-19 Outbreak. Asian Social Science 18.

434 14. CECC. 2021. 6 pilots test positive for antibodies in expanded testing program for
435 airline pilots, *on* CDC.
436 https://www.cdc.gov.tw/En/Bulletin/Detail/D3vIMViyvOwtrL91OSOlfw?typeid=158.
437 Accessed

438 15. CECC. 2021. CECC confirms 3 more indigenous cases; contacts of confirmed cases
439 found to have COVID-19, *on* CDC.
440 https://www.cdc.gov.tw/En/Bulletin/Detail/nikDToV8jVW2V4m5ag9NcQ?typeid=15
441 8. Accessed

442 16. Bull RA, Adikari TN, Ferguson JM, Hammond JM, Stevanovski I, Beukers AG,
443 Naing Z, Yeang M, Verich A, Gamaarachchi H, Kim KW, Luciani F, Stelzer-Braid S,
444 Eden J-S, Rawlinson WD, van Hal SJ, Deveson IW. 2020. Analytical validity of
445 nanopore sequencing for rapid SARS-CoV-2 genome analysis. Nature
446 Communications 11:6272.

447 17. Freed NE, Vlková M, Faisal MB, Silander OK. 2020. Rapid and inexpensive whole-
448 genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford
449 Nanopore Rapid Barcoding. Biology Methods and Protocols 5:bpaa014.

450 18. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with
451 BWA-MEM. arXiv: Genomics.

452 19. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A,
453 Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and
454 BCFtools. GigaScience 10:giab008.

455 20. Shu Y, McCauley J. 2017. GISAID: Global initiative on sharing all influenza data -
456 from vision to reality. Euro surveillance : bulletin Europeen sur les maladies
457 transmissibles = European communicable disease bulletin 22:30494.

458 21. Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: multiple sequence
459 alignment, interactive sequence choice and visualization. Briefings in Bioinformatics
460 20:1160-1166.

461 22. Watterson GA. 1975. On the number of segregating sites in genetical models without
462 recombination. Theor Popul Biol 7:256-276.

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.04.498645; this version posted April 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

463 23. Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of
464 restriction endonucleases. Proc Natl Acad Sci U S A 76:5269-73.

465 24. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular
466 evolutionary genetics analysis across computing platforms. Molecular biology and
467 evolution 35:1547.

468 25. Saitou N, Nei M. 1987. The neighbor-joining method: a new method for
469 reconstructing phylogenetic trees. Molecular Biology and Evolution 4:406-425.

470 26. Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-
471 Onsins SE, Sanchez-Gracia A. 2017. DnaSP 6: DNA Sequence Polymorphism
472 Analysis of Large Data Sets. Mol Biol Evol 34:3299-3302.

473 27. Leigh JW, Bryant D. 2015. popart: full-feature software for haplotype network
474 construction. Methods in Ecology and Evolution 6:1110-1116.

475 28. Clement M, Snell Q, Walker P, Posada D, Crandall K. TCS: estimating gene
476 genealogies, p 0184-0184. *In* (ed), IEEE Computer Society,

477 29. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018.
478 Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus
479 Evolution 4.

480 30. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and
481 dating with confidence. PLoS Biol 4:e88.

482 31. Day T, Gandon S, Lion S, Otto SP. 2020. On the evolutionary epidemiology of
483 SARS-CoV-2. Current Biology 30:R849-R857.

484 32. Sonnleitner ST, Prelog M, Sonnleitner S, Hinterbichler E, Halbfurter H, Kopecky
485 DBC, Almanzar G, Koblmüller S, Sturmbauer C, Feist L, Horres R, Posch W, Walder
486 G. 2022. Cumulative SARS-CoV-2 mutations and corresponding changes in
487 immunity in an immunocompromised patient indicate viral evolution within the host.
488 Nature Communications 13:2560.

489 33. Li W-H. 1997. Molecular evolution. Sinauer Associates, Sunderland, Mass.

490 34. Hou H, Wang T, Zhang B, Luo Y, Mao L, Wang F, Wu S, Sun Z. 2020. Detection of
491 IgM and IgG antibodies in patients with coronavirus disease 2019. Clinical &
492 Translational Immunology 9:e1136.

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.04.498645; this version posted April 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

493 35. McAloon C, Collins Á, Hunt K, Barber A, Byrne AW, Butler F, Casey M, Griffin J,
494 Lane E, McEvoy D, Wall P, Green M, O'Grady L, More SJ. 2020. Incubation period
495 of COVID-19: a rapid systematic review and meta-analysis of observational research.
496 BMJ Open 10:e039652.

497 36. Backer JA, Klinkenberg D, Wallinga J. 2020. Incubation period of 2019 novel
498 coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28
499 January 2020. Euro Surveill 25.

500 37. CECC. 2021. Starting April 15, CECC to adjust quarantine and disease prevention
501 and control measures for flight crews of Taiwanese airlines aboard flights arriving in
502 Taiwan. https://www.cdc.gov.tw/En/Bulletin/Detail/a-
503 SVOMqxKLi4HbPU9tPpgw?typeid=158. Accessed

504 38. CECC. 2022. 2022/3/4 14:00 The press conference of SARS-CoV-2, Central
505 Epidemic Command Center, *on* Taiwan center for disease control.
506 https://youtu.be/F7bJCGgUyr4?t=946. Accessed

507 39. Ruan Y, Luo Z, Tang X, Li G, Wen H, He X, Lu X, Lu J, Wu C-I. 2020. On the
508 founder effect in COVID-19 outbreaks: how many infected travelers may have started
509 them all? National Science Review 8.

510 40. Chan W-M, Ip JD, Chu AW-H, Tse H, Tam AR, Li X, Kwan MY-W, Yau Y-S,
511 Leung W-S, Chik TS-H, To W-K, Ng AC-K, Yip CC-Y, Poon RW-S, Chan K-H,
512 Wong SC-Y, Choi GK-Y, Lung DC, Cheng VC-C, Hung IF-N, Yuen K-Y, To KK-W.
513 2021. Phylogenomic analysis of COVID-19 summer and winter outbreaks in Hong
514 Kong: An observational study. The Lancet Regional Health - Western Pacific
515 10:100130.

516 41. Nabeshima T, Takazono T, Ashizawa N, Miyazaki T, Inoue S, Ngwe Tun MM,
517 Izumikawa K, Mukae H, Moi ML, Morita K. 2021. COVID-19 cryptic transmission
518 and genetic information blackouts: Need for effective surveillance policy to better
519 understand disease burden. Lancet Reg Health West Pac 7:100104.

520

521

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.04.498645; this version posted April 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

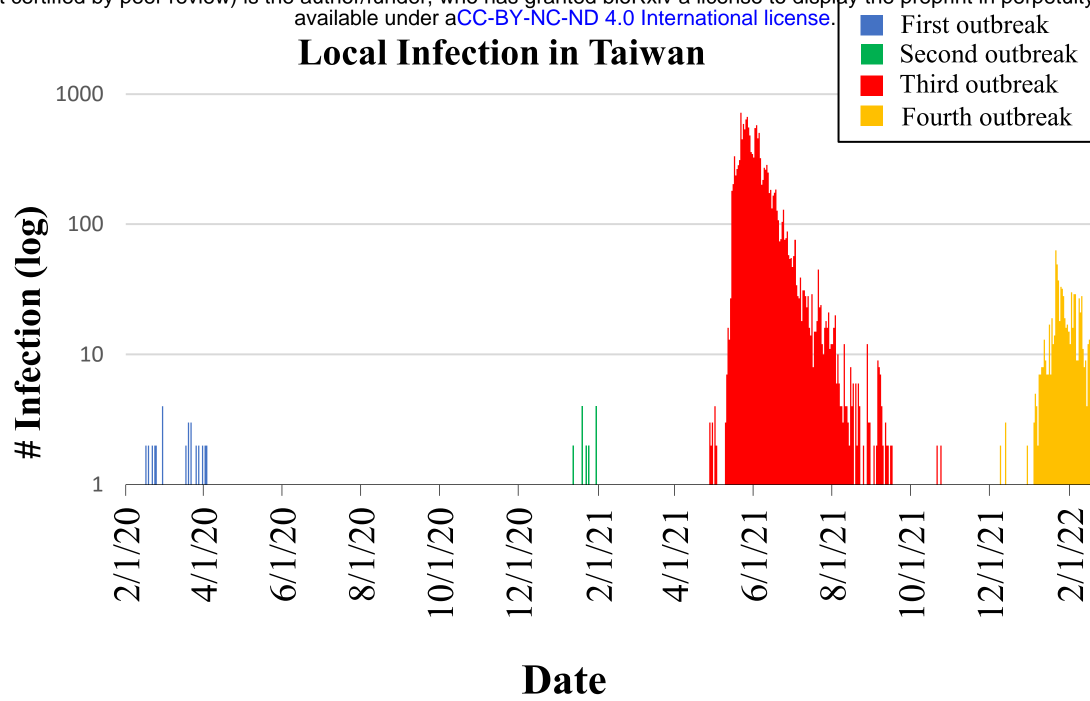

Figure 1 Number of local COVID-19 cases in Taiwan. Data were downloaded from COVID-19 DASHBOARD (https://covid-19.nchc.org.tw/) on February 22, 2022.

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.04.498645; this version posted April 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

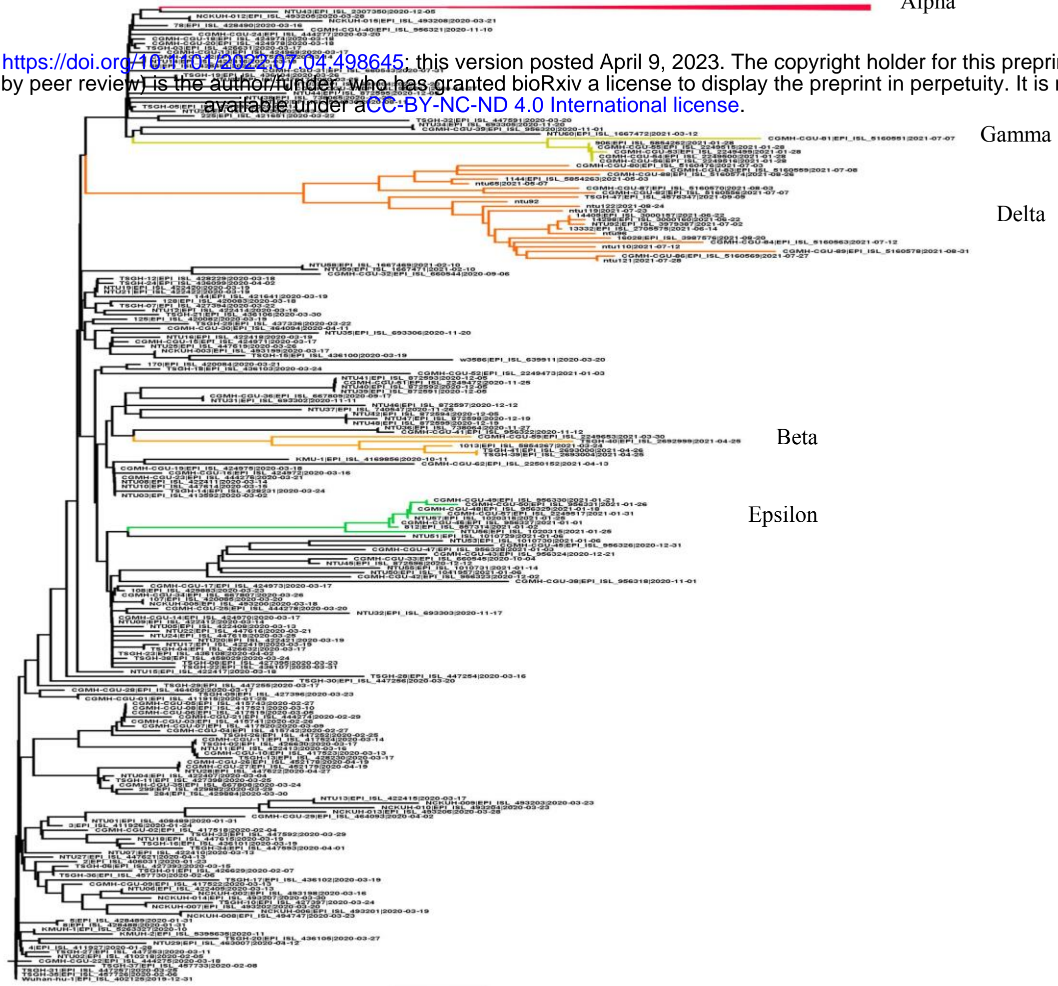

Figure 2 Phylogeny of (A) all SARS-CoV-2 and (B) Alpha strain only genomes from Taiwan. In (B) the black branches are imported, and colored branches are local cases. Australia sequence was colored in green. Cluster I and cluster II, both of which were collected prior to May 11, 2021, are colored in purple and orange, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.04.498645; this version posted April 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

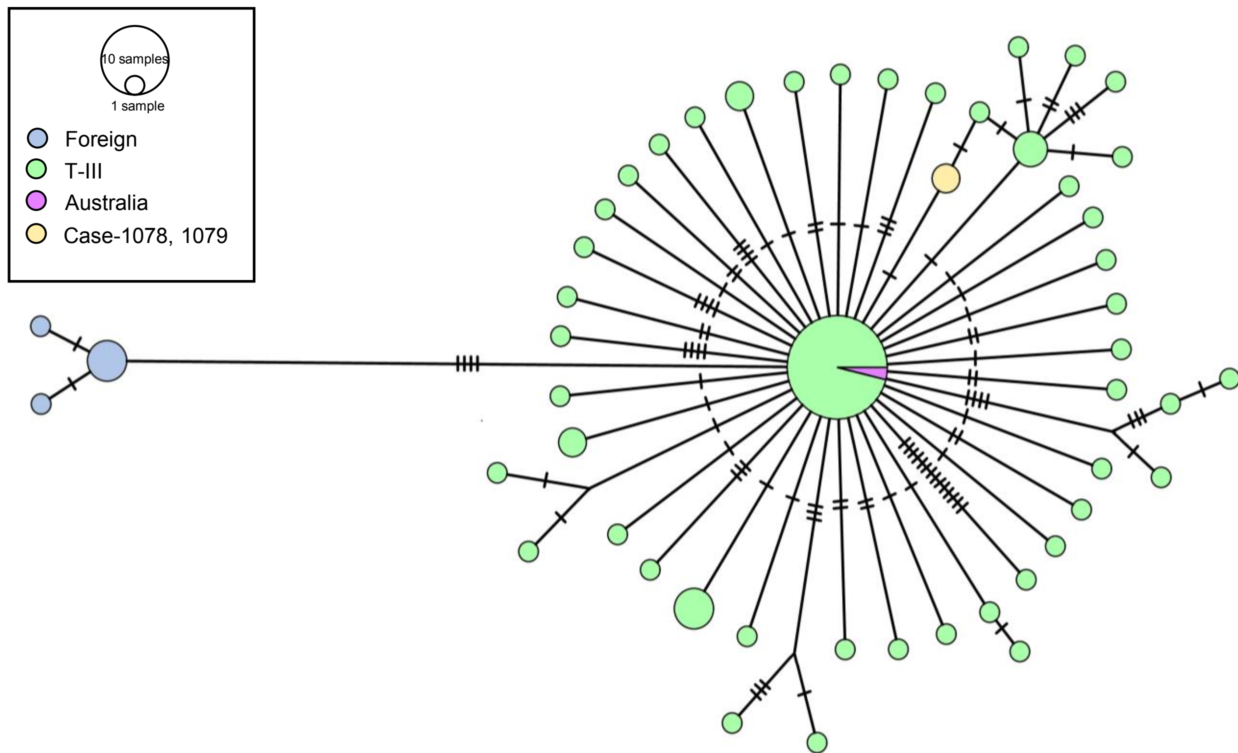

Figure 3 Haplotype network of SARS-CoV-2 genomes from the third local outbreak. The haplotype network was constructed by the median joining algorithm. Circle areas are proportional to the number of sequences. Mutational steps between haplotypes are symbolized by dashes. Haplotypes in yellow and pink are from Airline flight crews, who were the first three cases during the third outbreak. Cases 1078 and 1079 (in yellow) were diagnosed as COVID-19 positive on April 20 (CECC 2021b). On the same day, one pilot of the same airline tested positive for COVID-19 in Australia (in pink) while on duty (CECC 2021f).

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.04.498645; this version posted April 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

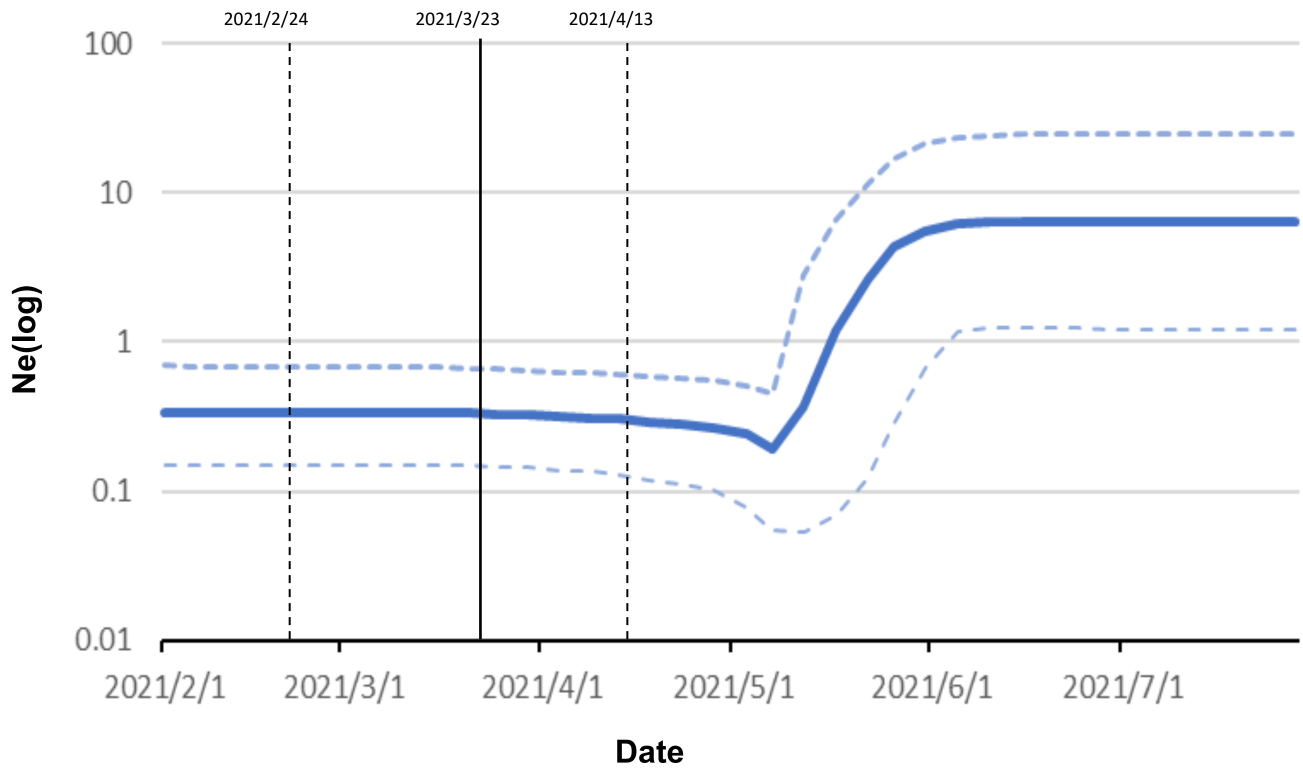

Figure 4 The epidemic growth curve of SARS-CoV-2 genomes from the third local outbreak. The three lines are the median (blue line) and 95% HPD intervals (dashed lines) of the Bayesian skyline plot (m = 5). Vertical solid line indicates the estimated time to the most recent common ancestor with 95% HPD intervals in dashed lines.