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

Jui-Hung Tai[1,2#], Yu Ken Low[1,3#], Selina Cai-Ling Wang[4], Hsin-Fu Lin[1], Tzi-Yuan Wang[5], You-Yu Lin[1,2], Charles S.P. Foster[6], Sebastiaan J. van Hal[7,8], Ya-Yun Lai[9], Shiou-Hwei Yeh[10], Sui-Yuan Chang[11], Pei-Jer Chen[1,12,13,14], Shu-Miaw Chaw[2,5], Hurng-Yi Wang[1,2,3,15*]

[1]Graduate Institute of Clinical Medicine, College of Medicine, National Taiwan University, Taipei 10002, Taiwan

[2]Genome and Systems Biology Degree Program, National Taiwan University and Academia Sinica, Taipei 10617, Taiwan

[3]Institute of Ecology and Evolutionary Biology, National Taiwan University, Taipei 10617, Taiwan

[4]Taipei First Girls High School, Taipei 10006, Taiwan

[5]Biodiversity Research Center, Academia Sinica, Taipei 11529, Taiwan

[6]School of Medical Sciences, University of New South Wales, Randwick NSW 2031, Australia

[7]NSW Health Pathology, Department of Infectious Diseases and Microbiology, Royal Prince Alfred Hospital, Sydney, NSW, Australia

[8]Central Clinical School, University of Sydney, Sydney, NSW, Australia

[9]Center of Genomic and Precision Medicine, National Taiwan University, Taipei 10055, Taiwan

[10]Department of Clinical Laboratory Sciences and Medical Biotechnology, College of Medicine, National Taiwan University, Taipei 10002, Taiwan.

[11]Department of Microbiology, College of Medicine, National Taiwan University, Taipei 10617, Taiwan.

[12]Hepatitis Research Center, National Taiwan University College of Medicine and National Taiwan University Hospital, Taipei 10002, Taiwan

[13]Department of Internal Medicine, National Taiwan University College of Medicine and National Taiwan University Hospital, Taipei 10002, Taiwan

[14]Department of Medical Research, National Taiwan University College of Medicine and National Taiwan University Hospital, Taipei 10002, Taiwan

[15]Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, Taipei 10002, Taiwan

# These authors contributed equally

*Correspondence: hurngyi@ntu.edu.tw

ABSTRACT

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

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

Importance

In order to clarify the spatial and temporal origin of the third SARS-CoV-2 outbreak in Taiwan, we sequenced and reconstructed a phylogeny of SARS-CoV-2 genomes. Our comprehensive analyses demonstrated that the outbreak was caused by a unique lineage (T-III), descended from the alpha variant of concern (B.1.1.7), which most likely originated from Europe, perhaps Turkey, in February 2021. In addition, the T-III could have circulated in Taiwan in mid-March 2021. The virus was unnoticed while spreading within the community. Our results demonstrate that phylogenetic approaches can be used to trace the outbreak derived from local spread by infected travelers and even a small number of imported cases can undermine the strict control measures in Taiwan.

Introduction

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

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

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

109 While it has been suggested that the third SARS-CoV-2 outbreak in Taiwan was sparked by
110 two imported cases from USA on April 20, 2021 (10), which, in turn, was caused by the partial
111 relaxation of entry quarantine measures in early April 2021 (11, 12), several unresolved
112 observations remain to be answered. First, according to CECC, the dates of symptom onset in
113 two infection clusters are very close (April 16 for the first cluster and April 23 for the second)
114 (9). However, the relationships between the two clusters remain elusive. Second, several airline
115 pilots and their family members were PCR negative but with serum IgM-/IgG+ in late April
116 and early May 2021 (13, 14), suggesting they have probably been infected by SARS-CoV-2
117 earlier than April 20, 2021.
118
119 Understanding the origin of the outbreak is important not only for epidemiological study but
120 also for the design of future control measures. In order to clarify the origin(s) and
121 interrelationship between the clusters, we sequenced and reconstructed a phylogeny of SARS-
122 CoV-2 genomes. We find that the third outbreak is caused by a single virus lineage (T-III),
123 descended from the alpha variant of concern (B.1.1.7), which carries four distinctive mutations,
124 including spike M1237I (S-M1237I) and three silent changes, from its most closely related
125 sequences in Europe.
126

127 ## Methods

128 **Sample preparation**

129 Thirty-three specimens from SARS-CoV-2 positive patients were obtained (Table S1). One
130 sample from Royal Prince Alfred Hospital, Australia was sequenced directly from extract using
131 Nanopore (15). The rest 32 specimens from National Taiwan University Hospital (NTUH)
132 were maintained in viral transport medium. Virus in the specimens was propagated in VeroE6
133 cells in Dulbecco's modified Eagle's medium (DMEM) supplemented with 2 μg/mL
134 tosylsulfonyl phenylalanyl chloromethyl ketone (TPCK)-trypsin (Sigma-Aldrich, St. Louis, MI,
135 USA). Culture supernatant was harvested when cytopathic effects (CPE) were observed in
136 more than 70% of cells, and the culture supernatant was harvested for viral genomic sequencing.
137

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

139 SARS-CoV-2 cDNA was generated from 100 ng of RNA in a RT-PCR reaction buffer
140 containing 4 μL of 5X PrimeScript IV 1st strand cDNA Synthesis Mix (Takara Bio, Kusatsu,
141 Shiga, Japan), 2 μL of 50 μM random hexamer primer, and variable amount of DEPC water to

142 fill up to 20 µL of total reaction volume. Pre-heat at 30°C for 10 minutes, followed by 20
143 minutes of 42°C, and then 15 minutes of 70°C.
144
145 **Amplification of complete SARS-CoV-2 genomes with multiplex PCR**
146 The "Midnight" amplicon protocol was used to generate 1200 bp amplicons. Briefly, PCR was
147 used with 2·5 µL of cDNA product from RT-PCR in 22·5 µL buffer, containing 12·5 µL of Q5
148 Hot Start High-Fidelity 2X Master Mix (New England BioLabs, M0494S), 1·1 µL of 10 µM
149 SARS-Cov2-Midnight-1200 primer (either Pool 1 or Pool 2)(16) (Integrated DNA
150 Technologies, Coralville, IA USA), and 8·9 µL of nuclease-free water (Thermo Scientific,
151 Waltham, MA, USA). Amplifications were performed with 30 seconds of 98°C for initial
152 denaturation, followed by 25 cycles of 98°C for 15 seconds and 65°C for 5 minutes in a Veriti
153 96-Well Thermal Cycler machine (Applied Biosystems, Waltham, MA, USA). Each sample
154 was separately amplified using both Pool 1 and Pool 2 primers. 20 µL PCR products were then
155 purified with DNA Clean & Concentrator-5 (Zymo Research, Irvine, CA, USA). Amplicons
156 were eluted with 25 µL of nuclease-free water (Thermo Scientific). DNA quality checks were
157 done using a Nanodrop (Thermo Scientific) and 1·5% agarose gel electrophoresis.
158
159 **Library preparation for Nanopore MinION sequencing**
160 For each sample, 1 µL (0·5 µL from each pool of a same sample), approximately 50 ng, of
161 purified PCR amplicons were used for library preparation. The KAPA HyperPrep Kit (Roche,
162 Basel, Switzerland) was used in a 15 µL reaction for end repair and A-tailing. The reactions
163 contained 1 µL of amplicon, 1·75 µL of End-repair & A-tailing Buffer, 0·75 µL of End-repair
164 & A-tailing Enzyme, and 11·5 µL of nuclease-free water (Thermo Scientific). Reactions were
165 done in 30 minutes at 20 °C and 30 minutes at 65 °C. Each sample was then barcoded via
166 ligation in a 27·5 µL reaction at 20°C for 15 minutes, with 2·5 µL of DNA Ligase, 7·5 µL of
167 Ligation Buffer, 2·5 µL of Native Barcode (Oxford Nanopore Technologies, Oxford, UK), and
168 15 µL of A-tailed amplicon. After that, barcoded amplicons were purified with 33 µL (1.2X)
169 of KAPA Pure Beads (Roche) by following the official protocol and eluted with 11 µL of
170 nuclease-free water (Thermo Scientific, R0582). We then pooled 2·7 µL of each barcoded
171 amplicon. The 110 µL of reaction for adapter ligation contained 65 µL of pooled barcoded
172 amplicons, 5 µL of Adapter Mix II (Oxford Nanopore Technologies, EXP-NBD104), 30 µL of
173 Ligation Buffer and 10 µL of DNA Ligase. After incubating in 20 °C for 15 minutes, libraries
174 were purified with 110 µL (1X) of KAPA Pure Beads (Roche, 07983271001) and Short
175 Fragment Buffer (Oxford Nanopore Technologies, SQK-LSK109) according to the official

176 protocol and then eluted with 14 μL of Elution Buffer (Oxford Nanopore Technologies, SQK-
177 LSK109).
178
179 **Nanopore MinION sequencing**
180 We sequenced 220 ng of purified library with an Oxford Nanopore Technoogies MinION
181 R.9.4.1 flowcell (FLO-MIN106) with the software MinKNOW Core version 19.12.5. Super-
182 accurate basecalling was done using Guppy version 5.0.11 with default settings in GPU mode.
183 The Nanopore reads of each sample were mapped to the reference genome Wuhan-hu-1
184 (EPI_ISL_402125) by *BWA-mem* [preprint] (17) to generate an alignment file.
185 *BCFtools* (18) was applied to call variant (-i 'QUAL > 20') and generate consensus sequence.
186
187 **Data collection**
188 A total of 267 complete and high coverage SARS-CoV-2 genomes from Taiwan with complete
189 collection date were downloaded from the Global Initiative on Sharing Avian Influenza Data
190 (GISAID, https://www.gisaid.org/)(19) on December 8, 2021. To find the SARS-CoV-2
191 genomes most closely related to the third outbreak in Taiwan, we used the Audacity Instant
192 search tool in GISAID to search the database and used EPI_ISL_2455264 (case 1079) as the
193 query. After excluding 42 sequences from Taiwan, 5,812 foreign sequences with fewer than or
194 equal to 10 SNP differences were downloaded. All the sequences used in this study can be
195 found in https://github.com/ala98412/T-III.
196
197 **M1237I frequency in different genetic backgrounds**
198 We used tools within GISAID to provide a quick search of the database. We chose different
199 SARS-CoV-2 strains in the drop-down menu "Variant" to get the count of sequences in each
200 major variant strain (e.g., Alpha, Delta …), and further selected "Spike_M1237I" in the drop-
201 down menu "Substitutions" to receive the count of sequences with the M1237I mutation in
202 each genetic background. We calculated the frequency of M1237I in different genetic
203 backgrounds via the count of sequences with the M1237I mutation divided by the total number
204 of sequences representing each major variant.
205
206 **Sequence analysis and phylogenetic reconstruction**
207 All sequences were aligned against the reference genome Wuhan-hu-1 (EPI_ISL_402125) by
208 using MAFFT v7 (20). Nucleotide diversity, including number of segregating sites,

209 Watterson's estimator of θ (21), and nucleotide diversity (π)(22), was estimated using MEGA-
210 X (23). Phylogenetic trees were also constructed by using the neighbor-joining method (24)
211 based on Kimura's two-parameter model in MEGA-X. The Nexus file for the haplotype
212 network analysis was generated using DnaSP 6.0 (25) and input into PopART v1.7 (26) to
213 construct the haplotype network using TCS software (27).
214
215 The times to the most recent common ancestor (TMRCA) of T-III lineage were estimated using
216 an established Bayesian Markov chain Monte Carlo (MCMC) approach implemented in
217 BEAST version 2.5 (28). The sampling dates were incorporated into TMRCA estimation.
218 These analyses were performed using the Hasegawa-Kishino-Yano (HKY) model of nucleotide
219 substitution assuming an uncorrelated lognormal relaxed molecular clock (29). We linked
220 substitution rates for the first and second codon positions and allowed independent rates for the
221 third codon position. We performed two independent runs with 1×10^7 MCMC steps and
222 combined the results. Log files were checked using Tracer (http://beast.bgio.ed.ac.uk/Tracer).
223 Effective sample sizes were >300 for all parameters.
224

225 ## Results

226 A total of 33 SARS-CoV-2 genomes were sequenced, including 32 from NTUH, Taiwan, and
227 one from Royal Prince Alfred Hospital, Australia (Table S1). We also downloaded 267
228 genomes from GISAID as of 2021/12/08, to construct the phylogeny as shown in Fig. 2. Since
229 the cases from the first outbreak had a contact history or exposures to different SARS-CoV-2
230 infected patients, they do not form a single cluster in the phylogeny. The sequences derived
231 from the second local outbreak are presented in emerald green (Fig. 2). The third local outbreak
232 consisted of the alpha strain (B.1.1.7), divided into two main clusters, shown in Fig. 2b. All
233 sequences in the basal lineage of Fig. 2b were from imported cases, whereas all sequences in
234 the more advanced lineage were local to Taiwan. Hereafter, we name this particular lineage as
235 T-III, which stands for the third outbreak in Taiwan, and note that the T-III belongs to Pangolin
236 lineage B.1.1.7.
237

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

239 In order to search for the spatial origin of T-III, the sequence recovered from the earliest case
240 in the third outbreak (case 1079: EPI_ISL_2455264, Table S1) was used to search against the
241 GISAID database. There were 5,812 sequences with ≤ 10 nucleotide differences from

242 EPI_ISL_2455264 as of November 21, 2021. Phylogenetic reconstruction including all 5,812
243 sequences and T-III demonstrate that the latter is a distinctive lineage (Fig. S1). The vast
244 majority of sequences closely related to T-III were from Europe, including Turkey (Fig. S1b).
245
246 Haplotype network analyses of the T-III lineage reveals 48 haplotypes. The network shows that
247 T-III differs from the outgroups in four mutation steps (Fig. 3), including two synonymous
248 mutations, C5812T and C15895T, in Orf1ab, one nonsynonymous mutation G25273C
249 (M1237I) in Spike, and one T27869C mutation in a non-coding region (Table S2). The closest
250 outgroup haplotype consists of four sequences, including two sequences from Turkey that were
251 collected on February 8, 2021 (EPI_ISL_1097034, 1097035). The rest were collected after the
252 onset of the third outbreak in Taiwan (Table S3). Further database mining confirmed that of
253 1,140,328 alpha strain genomes examined as of 12/11/2021, only the lineage T-III possessed
254 the four above mentioned mutations, which form a distinctive genetic fingerprint. Within T-III,
255 the network forms a star-like shape centered on a core haplotype comprising 26 sequences.
256 Most of the remaining haplotypes are directly connected to this major haplotype. Of the three
257 cases identified on the first day of the third outbreak (April 20, 2020), the case from Australia
258 belongs to the major haplotype, with the rest (1078 and 1079) one mutation away.
259
260 The estimated substitution rate of T-III is 9.8×10^{-4} substitution per site per year (95% highest
261 posterior density [HPD] 5.6×10^{-4} – 1.4×10^{-3}) which is close to 8.4×10^{-4} /site/year of SARS-
262 CoV-2 (30), and 7.5×10^{-4} /site/year of strain B.1.1 (31). The date of the most recent common
263 ancestor (TMRCA) of T-III is in late March (3/23/2021; 95% HPD February 24 - April 13,
264 2021) (Fig. 4), almost one month before the first three confirmed cases on April 20, 2021. We
265 noticed that the mutation rate of T-III during this epidemic was almost constant (Fig. S2).
266 Effective population size of the T-III lineage increased approximately 20-fold after the onset
267 of the outbreak and reached a plateau in early June. The estimated demographic expansion of
268 T-III is consistent with epidemiological data (Fig. 1). We noticed that demography in Fig. 4
269 does not capture the population decline after July 2021 as shown in Fig. 1. This finding is
270 because most sequences used in this analysis were collected before June 2021 (Table S1), with
271 only four sequences obtained after July 2, 2021. As we are interested in the outbreak origin,
272 this sampling strategy should not affect our conclusions.
273
274 Rapid population expansion can also be revealed by contrasting patterns of genetic variation
275 estimated using different approaches. The Watterson's estimator of θ ($6{\cdot}95 \times 10^{-4}$) is

276 approximately seven times higher than the nucleotide diversity (π) ($1{\cdot}05 \times 10^{-4}$), leading to
277 significantly negative Tajima's D (-2·82, $p < 0{\cdot}001$) (Table S4). Because θ is strongly
278 influenced by rare mutations, which are common during recent population expansion or after
279 positive selection (32), it is a better estimator of genetic diversity for T-III.
280

281 **Discussion**

282 Our results provide solid evidence that the third local outbreak in Taiwan was caused by a
283 single lineage, T-III. This does not in itself mean that two clusters of infections (see
284 Introduction) have a single common origin. For example, among 293,742 sequences analyzed
285 during the first year of the SARS-CoV-2 pandemic, the most abundant haplotype was sampled
286 3,466 times from across 53 countries (Table S5). It is possible that the haplotype exhibited
287 some transmission advantage, making it widespread. Under this scenario, two clusters of
288 infection may be caused by the same lineage imported from different sources. However, as the
289 T-III lineage bearing a combination of four unique mutations is exclusive to Taiwan, it seems
290 highly unlikely that this lineage was imported from different sources.
291

292 Our estimation that T-III originated from Europe with TMRCA on March 23, 2021 (95% HPD
293 February 24, 2021 – April 13, 2021) reconciles several unresolved observations. First, the first
294 two cases (case 1078 and 1079) of the outbreak shared identical sequences, indicating that they
295 were from the same source. Nevertheless, the cycle threshold (Ct) values at the time of
296 diagnosis (April 20, 2021) were 29 and 17 for case 1078 and 1079, respectively, suggesting
297 that they were infected at different times (6). Second, several airline pilots and their family
298 members were PCR negative but with serum IgM-/IgG+ in late April and early May 2021 (13,
299 14). It has been shown that IgM levels increase during the first week after a SARS-CoV-2
300 infection, peak after two weeks, and then recede to near-background levels in most patients.
301 IgG is detectable one week after disease onset and is maintained at a high level for a long period
302 (33). Consequently, IgM negative but IgG positive individuals have probably been infected by
303 SARS-CoV-2 earlier than April 20, 2021. Third, according to CECC, the dates of symptom
304 onset in two seemingly unrelated infection clusters are very close (April 16 for the first cluster
305 and April 23 for the second) (9). As our phylogenetic analysis reveals that all sequences in the
306 third outbreak have a single origin, the occurrence of two infection clusters at similar time
307 without traceable connection demonstrates that the virus may have been cryptically circulating

308 in the community undetected. Consequently, the origin of the third outbreak was most likely
309 prior to April 20, 2021.
310
311 As most sequences closely related to T-III were from Europe, our results disagree with the
312 notion that the outbreak was imported by cases 1078 and 1079 from the USA (10). Among four
313 sequences closest to T-III, only two from Turkey were collected before April 20, 2021 (on
314 February 8, 2021). The rest appeared after April 20 and cannot be associated with the third
315 outbreak. Consequently, the lineage leading to the third outbreak was most likely introduced
316 from Europe, perhaps Turkey, by infected travelers in February 2021 (Fig. 3).
317
318 There are several possible routes by which SARS-CoV-2 can be introduced by incoming
319 travelers. First, some carriers may have the virus undetected during quarantine. Although the
320 mean incubation period ranges from five to seven days, it can be longer than 14 days (34, 35).
321 Indeed, there has been much debate about whether changing from the mandatory five day home
322 quarantine plus nine day autonomous health management (or the so-called "5+9" strategy) to
323 three day home quarantine plus 11 day autonomous health management ("3+11") for flight
324 crews on April 15, 2021 (36) was the cause of the third outbreak (11, 12). Since the outbreak
325 likely originated in mid-March, as our analyses demonstrate, the possibility that changing
326 strategies from "5+9" to "3+11" directly caused the outbreak can be ruled out.
327
328 Second, there may have transmission from people associated with quarantine hotels to the
329 community. For example, from December 2021 to March 2022 alone, 15 infection clusters
330 occurred in quarantine hotels (37). Since one of the two infection clusters in the third outbreak
331 directly link to a hotel where many flight crews stayed during their quarantine (see
332 Introduction), it is most likely that the outbreak was introduced from this hotel during late
333 February to early April 2021. The virus was then undetected while spreading within the
334 community until late April 2021.
335
336 Our results demonstrate that even a small number of imported cases can undermine the strict
337 control measures in Taiwan (38). We also show that phylogenetic approaches can be used to
338 trace the outbreak derived from local spread by infected travelers (39). Although Taiwan has
339 been very cautious of emerging disease and has strengthened its control measures, the SARS-
340 CoV-2 genomes have not been sampled in proportion to the actual size of infection and made
341 publicly available in a timely fashion. This situation severely hinders efforts to trace the

underlying transmission patterns of spread. Consequently, policies that include real-time public-sharing and transparency for infectious disease surveillance and control are critically and urgently needed. Blackouts in information sharing and transparency can cause information disparity and suspicion, which in turn hamper pandemic control efforts (40).

Declaration of Interests

The authors declare that they have no competing interests.

Availability of Data and Materials

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

Authors' Contributions

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

Funding

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

Reference

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

2. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC,

374 Zhang YZ. 2020. A new coronavirus associated with human respiratory disease in
375 China. Nature 579:265-269.

376 3. Chaw S-M, Tai J-H, Chen S-L, Hsieh C-H, Chang S-Y, Yeh S-H, Yang W-S, Chen P-
377 J, Wang H-Y. 2020. The origin and underlying driving forces of the SARS-CoV-2
378 outbreak. Journal of Biomedical Science 27:73.

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

382 5. Lin CY, Wang WH, Urbina AN, Tseng SP, Lu PL, Chen YH, Yu ML, Wang SF.
383 2020. Importation of SARS-CoV-2 infection leads to major COVID-19 epidemic in
384 Taiwan. Int J Infect Dis 97:240-244.

385 6. CECC. 2021. CECC confirms 2 more cases; sources of infection to be investigated,
386 *on* CDC.
387 https://www.cdc.gov.tw/En/Category/ListContent/tov1jahKUv8RGSbvmzLwFg?uaid
388 =NrqGhobGKPIYRPGw2w2z0Q. Accessed

389 7. CECC. 2021. CECC launches an investigation into notification concerning an airline
390 pilot testing positive for COVID-19 in Australia and contacts testing positive in
391 Taiwan, *on* CDC.
392 https://www.cdc.gov.tw/En/Bulletin/Detail/fDE2kCBYdnmbK0QiU_tSMg?typeid=1
393 58. Accessed

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

398 9. CECC. 2021. CECC confirms 185 more COVID-19 cases, including 180 indigenous
399 cases and 5 imported cases, *on* CDC.
400 https://www.cdc.gov.tw/En/Category/ListContent/tov1jahKUv8RGSbvmzLwFg?uaid
401 =70ryEYOTbvx8VaItygA1IA. Accessed

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

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

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

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

14. CECC. 2021. CECC confirms 3 more indigenous cases; contacts of confirmed cases found to have COVID-19, *on* CDC. https://www.cdc.gov.tw/En/Bulletin/Detail/nikDToV8jVW2V4m5ag9NcQ?typeid=158. Accessed

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

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

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

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

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

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

438 21. Watterson GA. 1975. On the number of segregating sites in genetical models without
439 recombination. Theor Popul Biol 7:256-276.

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

442 23. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular
443 evolutionary genetics analysis across computing platforms. Molecular biology and
444 evolution 35:1547.

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

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

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

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

454 28. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018.
455 Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus
456 Evolution 4.

457 29. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and
458 dating with confidence. PLoS Biol 4:e88.

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

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

466 32. Li W-H. 1997. Molecular evolution. Sinauer Associates, Sunderland, Mass.

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

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

474 35. Backer JA, Klinkenberg D, Wallinga J. 2020. Incubation period of 2019 novel
475 coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28
476 January 2020. Euro Surveill 25.

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

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

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

487 39. Chan W-M, Ip JD, Chu AW-H, Tse H, Tam AR, Li X, Kwan MY-W, Yau Y-S,
488 Leung W-S, Chik TS-H, To W-K, Ng AC-K, Yip CC-Y, Poon RW-S, Chan K-H,
489 Wong SC-Y, Choi GK-Y, Lung DC, Cheng VC-C, Hung IF-N, Yuen K-Y, To KK-
490 W. 2021. Phylogenomic analysis of COVID-19 summer and winter outbreaks in
491 Hong Kong: An observational study. The Lancet Regional Health - Western Pacific
492 10:100130.

493 40. Nabeshima T, Takazono T, Ashizawa N, Miyazaki T, Inoue S, Ngwe Tun MM,
494 Izumikawa K, Mukae H, Moi ML, Morita K. 2021. COVID-19 cryptic transmission
495 and genetic information blackouts: Need for effective surveillance policy to better
496 understand disease burden. Lancet Reg Health West Pac 7:100104.
497
498
499

Figure legends

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.

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.

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).

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.

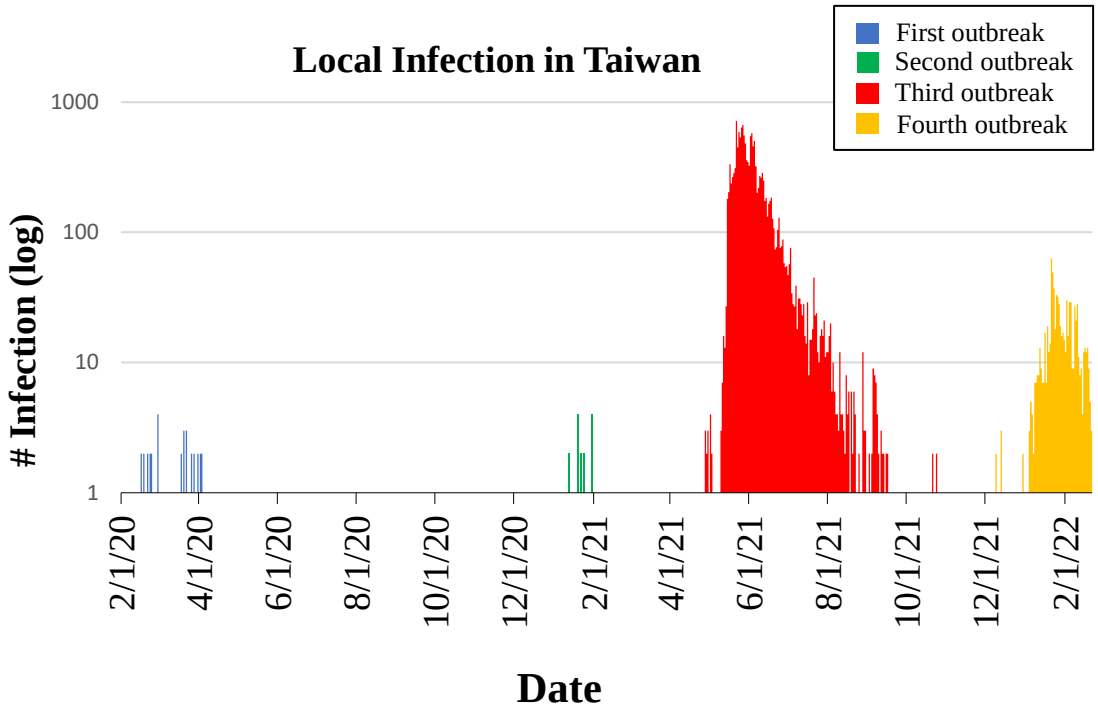

Figure 1 Number of local COVID-19 cases in Taiwan Data were downloaded from COVID-19 DASHBOARD (https://covid-19.nchc.org.tw/) on February 22, 2022.

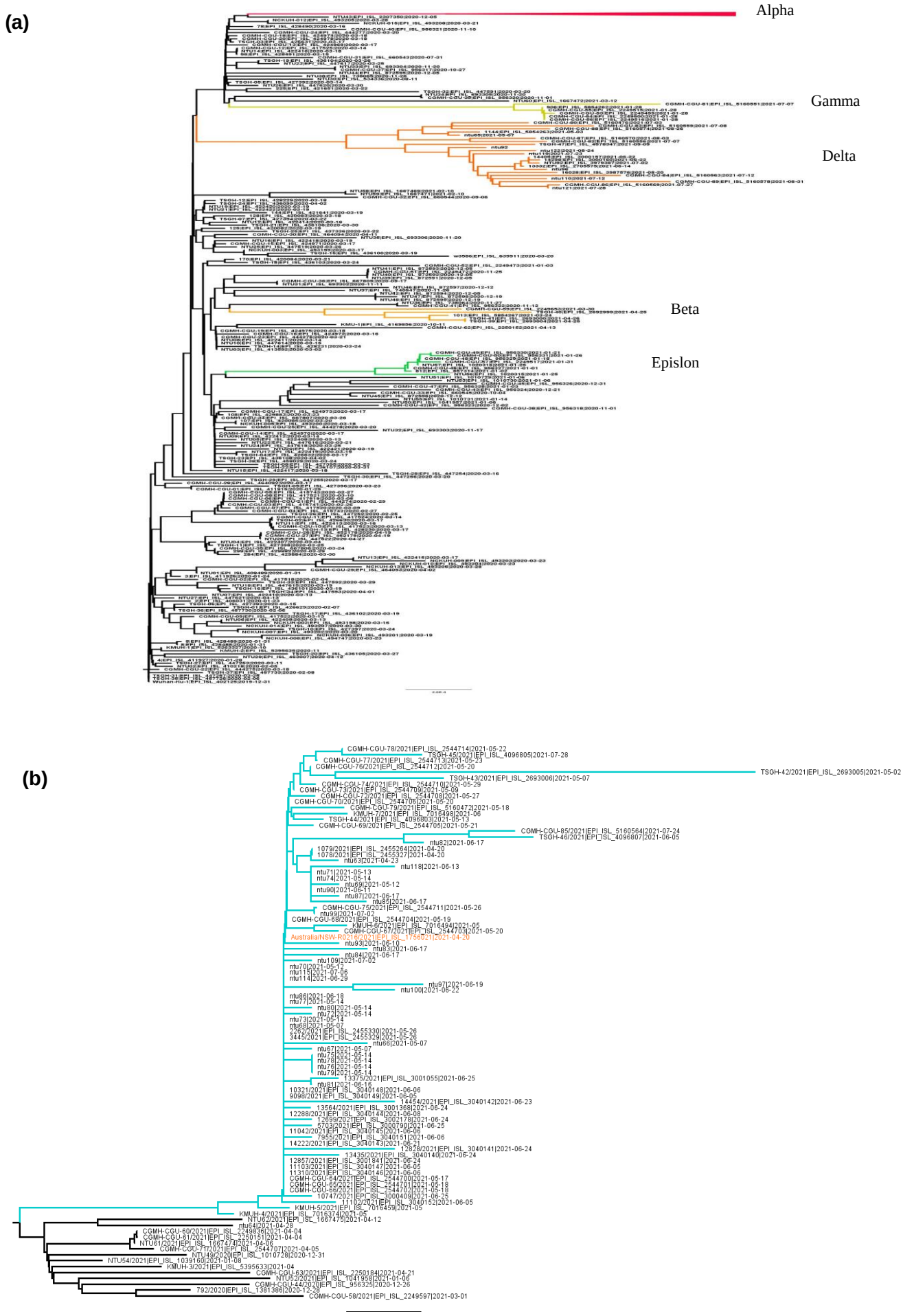

Figure 2 Phylogeny of (A) all SARS-CoV-2 and (B) Alpha strain only genomes from Taiwan. In (B) the black branches are imported, and colored branches are local cases.

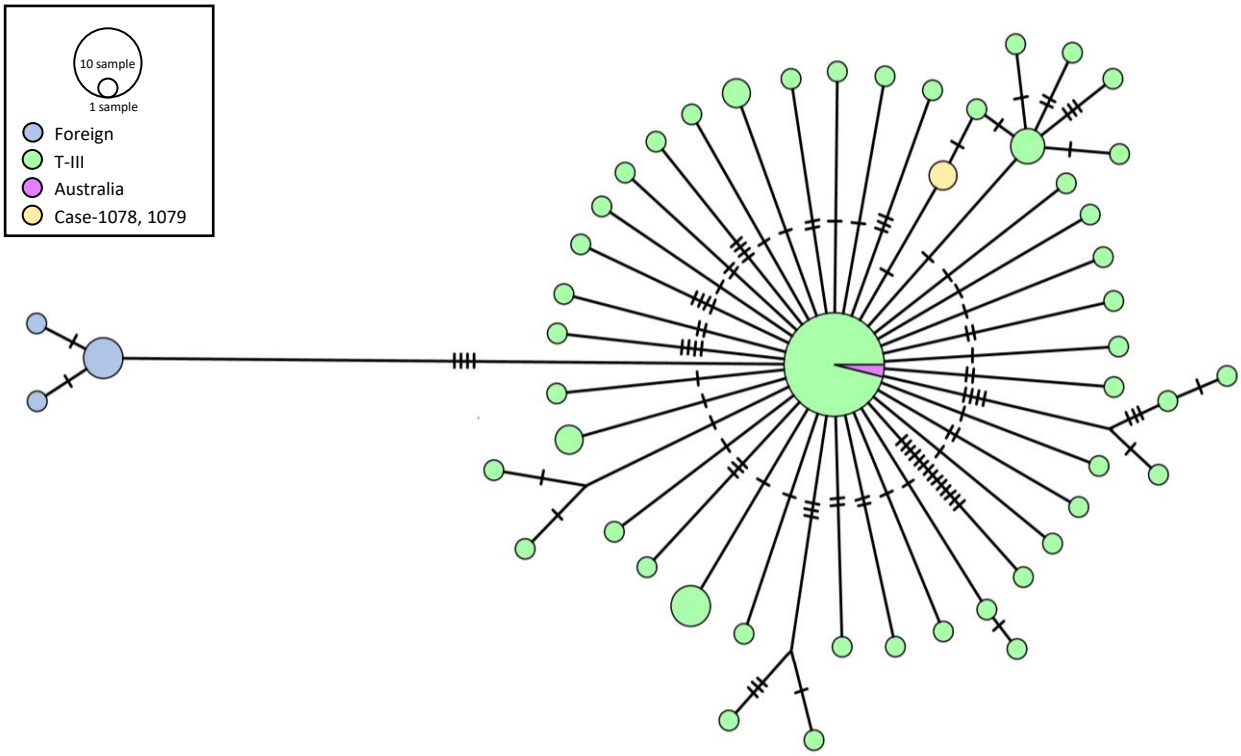

Figure 3 Haplotype network of SARS-CoV-2 genomes from the third local outbreak. Haplotypes in yellow and pink are from Airline flight crews, who were the first three cases during the third outbreak. Cases 1078 and 1079 (in yellow) were diagnosed as COVID-19 positive on April 20 (CECC 2021b). On the same day, one pilot of the same airline tested positive for COVID-19 in Australia (in pink) while on duty (CECC 2021f).

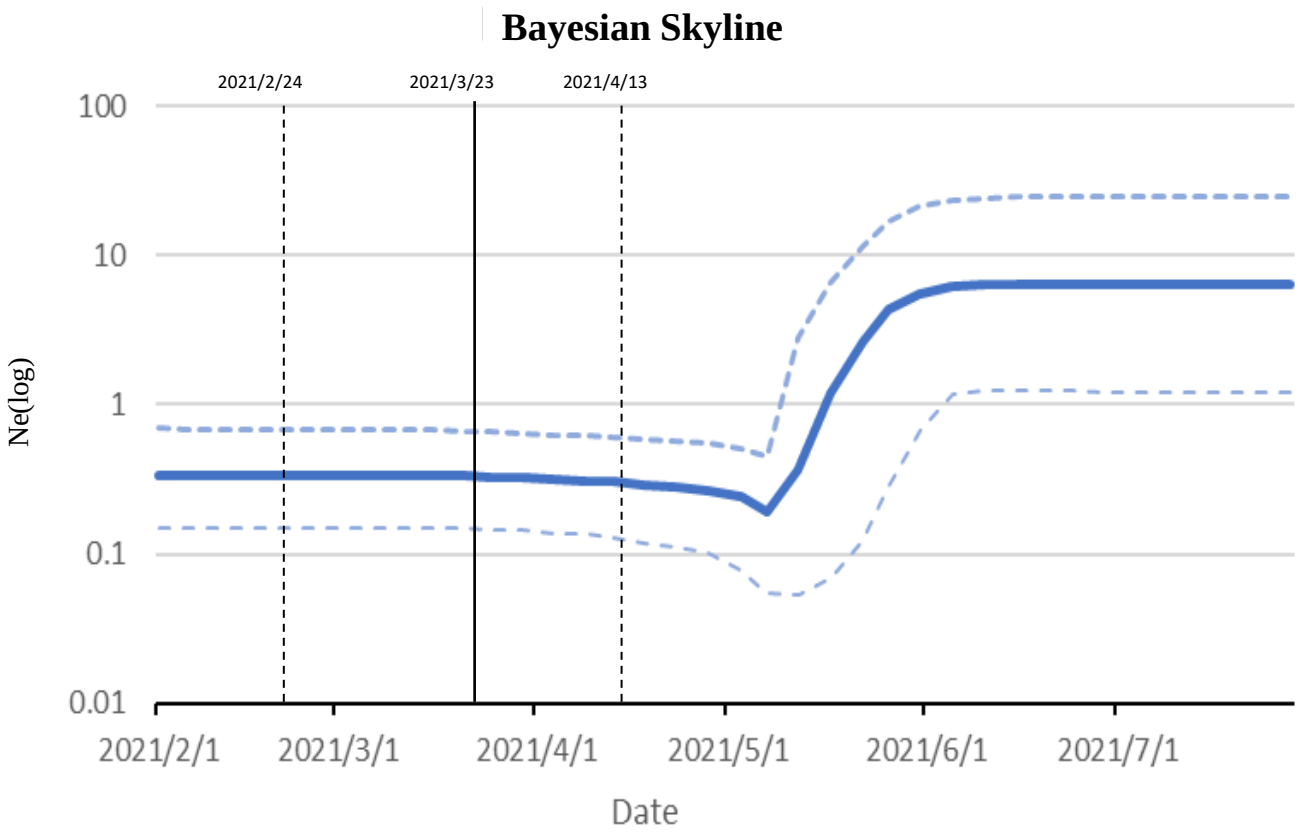

Figure 4 The epidemic growth curve of SARS-CoV-2 genomes from the third local outbreak. The three lines are the median (blue line) and 95% HPD intervals (dashed lines) of the Bayesian skyline plot (m = 5). Vertical solid line indicates the estimated time to the most recent common ancestor with 95% HPD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