SPATIAL AND TEMPORAL ORIGIN OF THE THIRD SARS-COV-2 OUTBREAK IN TAIWAN

Jui-Hung Tai[1,2], Yu Ken Low[1,3], Hsin-Fu Lin[1], Tzi-Yuan Wang[4], You-Yu Lin[1,2], Charles Foster[5], Ya-Yun Lai[6], Shiou-Hwei Yeh[7], Sui-Yuan Chang[8], Pei-Jer Chen[1,9,10,11], Hurng-Yi Wang[1,3,12]*

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

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

[6]Center of Precision Medicine, National Taiwan University, Taipei 10055, Taiwan

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

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

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

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

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

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

*Correspondence: hurngyi@ntu.edu.tw

ABSTRACT

Since the first report of SARS-CoV-2 in December 2019, Taiwan had gone through three local outbreaks. Unlike the first two, the spatial and temporal origin of the third outbreak (April 20 to November 5, 2021) is still unclear. We sequenced and reconstructed the phylogeny of SARS-CoV-2 genomes and find that the third outbreak was caused by a single virus lineage (T-III), which carries four genetic fingerprints, including spike M1237I (S-M1237I), and three silent changes. The T-III is closest to sequences derived from Turkey on February 8, 2021. The estimated date of divergence from the most recent common ancestor (TMRCA) of T-III is March 23, 2021 (95% 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. Consequently, the lineage leading to the third outbreak 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.

Keywords: Spike, M1237I, alpha/B.1.1.7

BACKGROUND

Since the first report of coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in December 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 the SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV) (Corman, et al. 2018; Cui, et al. 2019; Wu, et al. 2020).

After experiencing a series of SARS outbreaks in 2003 which caused 668 probable cases and 181 deaths (Chen, et al. 2005), Taiwan has been very cautious of emerging disease and has strengthened its pandemic control measures. For example, the Central Epidemic Command Center (CECC) was established after SARS 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 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 (Lin, et al. 2020). The second local outbreak started on January 12 and ended on February 9, 2021. It was sparked by an intrahospital 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 (CECC 2021b). On the same day, one pilot of the same Airline has tested positive for COVID-19 in

Australia while on duty (CECC 2021f). This cluster 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 (CECC 2021c). The second cluster involved several local incidences in New Taipei City and Yilan County that later spread to many counties (Fig. 1). 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 (CECC 2021e). Unlike the first two outbreaks, the spatial and temporal origin of this outbreak is still unclear. In addition, the relationships between the two clusters remain elusive, despite, according to the CECC, all local incidences being confirmed caused by the alpha strain (B.1.1.7) of SARS-CoV-2.

In order to clarify the origin(s) and interrelationship between the clusters, we sequenced and reconstructed the phylogeny of SARS-CoV-2 genomes. We find that the third outbreak was caused by a single virus lineage (T-III), which carries four distinctive mutations, including spike M1237I (S-M1237I) and three silent changes, from its closest-related sequences in Europe. S-M1237I is commonly found in different genetic backgrounds with similar frequencies, suggesting that this mutation has occurred frequently.

METHODS

Cell culture

Thirty-three respiratory specimens obtained from SARS-CoV-2-infected patients from National Taiwan University Hospital (NTUH) were maintained in viral transport medium. Virus in the specimens was propagated in VeroE6 cells in Dulbecco's modified Eagle's medium (DMEM) supplemented with 2 μg/mL tosylsulfonyl phenylalanyl chloromethyl ketone (TPCK)-trypsin (Sigma-Aldrich). Culture supernatant was harvested when cytopathic

effects (CPE) were observed in more than 70% of cells, and the culture supernatant was harvested for viral genomic sequencing.

Reverse transcription polymerase chain reaction (RT-PCR)

SARS-CoV-2 cDNA was generated from 100 ng of RNA in a RT-PCR reaction buffer containing 4 µℓ of 5X PrimeScript IV 1st strand cDNA Synthesis Mix (Takara Bio, 6215A), 2 µℓ of 50 µM random hexamer primer, and variable amount of DEPC water to fill up to 20 µℓ of total reaction volume. Pre-heat at 30℃ for 10 minutes, followed by 20 minutes of 42℃, and then 15 minutes of 70℃.

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

1,200 bp amplicon (Freed, et al. 2020) was generated by PCR with 2.5 µℓ of cDNA product from RT-PCR in 22.5 µℓ buffer, containing 12.5 µℓ of Q5 Hot Start High-Fidelity 2X Master Mix (New England BioLabs, M0494S), 1.1 µℓ of 10 µM SARS-Cov2-Midnight-1200 primer (either Pool 1 or Pool 2)(Freed, et al. 2020)(Integrated DNA Technologies, 10007184), and 8.9 µℓ of nuclease-free water (Thermo Scientific, R0582). Amplifications were performed with 30 seconds of 98℃ for initial denaturation, followed by 25 cycles of 98℃ for 15 seconds and 65℃ for 5 minutes in a Veriti 96-Well Thermal Cycler machine (Applied Biosystems, 4375786). Each sample was separately amplified using both Pool 1 and Pool 2 primers. PCR products were then cleaned up with DNA Clean & Concentrator-5 (Zymo Research, D4014). 20 µℓ of PCR products were used for clean-up. Amplicons were eluted with 25 µℓ of nuclease-free water (Thermo Scientific, R0582). DNA quality checks were done using the Nanodrop (Thermo Scientific, ND1000) and 1.5% agarose gel electrophoresis.

Library preparation for Nanopore MinION sequencing

For each sample, 1 µℓ (0.5 µℓ from each Pool of a same sample), approximately 50 ng, of purified PCR amplicons were used for library preparation. The KAPA HyperPrep Kit (Roche, 07962347001) was used in a 15 µℓ reaction for end repair and A-tailing. The reactions contained 1 µℓ of amplicon, 1.75 µℓ of End-repair & A-tailing Buffer, 0.75 µℓ of End-repair & A-tailing Enzyme, and 11.5 µℓ of nuclease-free water (Thermo Scientific, R0582). Reactions were done in 30 minutes at 20℃ and 30 minutes at 65℃. Each sample was then barcoded via ligation in a 27.5 µℓ reaction at 20℃ for 15 minutes, with 2.5 µℓ of DNA Ligase (KAPA HyperPrep Kit), 7.5 µℓ of Ligation Buffer (KAPA HyperPrep Kit), 2.5 µℓ of Native Barcode (Oxford Nanopore Technologies, EXP-NBD104 & EXP-NBD114), and 15 µℓ of A-tailed amplicon. After that, barcoded amplicons were purified with 33 µℓ (1.2X) of KAPA Pure Beads (Roche, 07983271001) by following the official protocol and eluted with 11 µℓ of nuclease-free water (Thermo Scientific, R0582). We then pooled 2.7 µℓ of each barcoded amplicon. The 110 µℓ of reaction for adapter ligation contained 65 µℓ of pooled barcoded amplicons, 5 µℓ of Adapter Mix II (Oxford Nanopore Technologies, EXP-NBD104), 30 µℓ of Ligation Buffer (KAPA HyperPrep Kit), and 10 µℓ of DNA Ligase (KAPA HyperPrep Kit). After incubating in 20℃ for 15 minutes, libraries were purified with 110 µℓ (1X) of KAPA Pure Beads (Roche, 07983271001) and Short Fragment Buffer (Oxford Nanopore Technologies, SQK-LSK109) according to the official protocol and then eluted with 14 µℓ of Elution Buffer (Oxford Nanopore Technologies, SQK-LSK109).

Nanopore MinION sequencing

220 ng of purified library was sequenced by using MinION R.9.4.1 flowcell (FLO-MIN106) with the software MinKNOW Core version 19.12.5. Basecalling was done using Guppy version 5.0.11 (Super High Accuracy) with default settings in GPU mode.

151 Data collection

152 A total of 267 complete and high coverage SARS-CoV-2 genomes from Taiwan with complete
153 collection date were downloaded from the Global Initiative on Sharing Avian Influenza Data
154 (GISAID, https://www.gisaid.org/)(Shu and McCauley 2017) on December 8, 2021. In
155 addition, 1,193 Alpha strains with the M1237I mutation were downloaded on November 14,
156 2021. To find the SARS-CoV-2 genomes most closely related to the third outbreak in Taiwan,
157 we used Audacity Instant search tool in GISAID to search the database and used
158 EPI_ISL_2455264 (case 1079) as the query. After excluding 42 sequences from Taiwan, 5,812
159 foreign sequences with fewer than or equal to 10 SNP differences were downloaded.

160

161 M1237I frequency in different genetic backgrounds

162 GISAID provided a quick search of the database. We chose different SARS-CoV-2 strains in
163 the drop-down menu "Variant" to get the count of sequences in each variant strain (e.g., Alpha,
164 Delta …), and further chose "Spike_M1237I" in the drop-down menu "Substitutions" to
165 receive the count of sequences with the M1237I mutation in each genetic background. We
166 calculated the frequency of M1237I in different genetic backgrounds via the count of sequences
167 with the M1237I mutation divided by the total number of sequences representing each major
168 variant.

169

170 Sequence analysis and phylogenetic reconstruction

171 All sequences were aligned against the reference genome Wuhan-hu-1 (EPI_ISL_402125) by
172 using MAFFT v7 (Katoh, et al. 2019). Nucleotide diversity, including number of segregating
173 sites, Watterson's estimator of θ (Watterson 1975), and nucleotide diversity (π)(Nei and Li
174 1979), was estimated using MEGA-X (Kumar, et al. 2018). Phylogenetic trees were also
175 constructed by using MEGA-X using the neighbor-joining method (Saitou and Nei 1987) based

176 on Kimura's two-parameter model. The Nexus file for the haplotype network analyses was
177 generated using DnaSP 6.0 (Rozas, et al. 2017) and input into PopART v1.7 (Leigh and Bryant
178 2015) to construct the haplotype network using TCS network (Clement, et al. 2002).
179
180 Times to the most recent common ancestor (TMRCA) of virus isolates were estimated using
181 an established Bayesian MCMC approach implemented in BEAST version 2.5 (Suchard, et al.
182 2018). The sampling dates were incorporated into TMRCA estimation. The analyses were
183 performed using the HKY model of nucleotide substitution assuming an uncorrelated
184 lognormal molecular clock (Drummond, et al. 2006). We linked substitution rates for the first
185 and second codon positions and allowed independent rates for the third codon position. We
186 performed two independent runs with 1×10^7 MCMC steps and combined the results. Log files
187 were checked using Tracer (http://beast.bgio.ed.ac.uk/Tracer). Effective sample sizes were
188 >300 for all parameters.
189

RESULTS

We used 299 SARS-CoV-2 genomes, including 32 from NTUH (Table S1) and 267 downloaded from GISAID as of 2021/12/08, to construct the phylogeny as shown in Fig. 2. Since the cases from the first outbreak had a contact history or exposure to different SARS-CoV-2 infected patients, they do not form a single cluster in the phylogeny. The sequences derived from the second local outbreak are presented in emerald green (Fig. 2a). The third local outbreak consisted of the alpha strain (B.1.1.7) shown in Fig. 2B. All sequences in the basal lineage of Fig. 2B were from imported cases, whereas all sequences in the more advanced lineage (T-III) are local.

Spatial and temporal origin of the third local outbreak in Taiwan

In order to search for the spatial origin of T-III, the sequence (EPI_ISL_2455264) recovered from the earliest case in the third outbreak (case 1079, Table S1) was used to search against the database. There were 5812 sequences with ≤ 10 nucleotide differences from EPI_ISL_2455264 as of 11/21/2021. Phylogenetic reconstruction including all 5812 sequences and T-III demonstrate that the latter is a distinctive lineage (Fig. S1). The vast majority of sequences closely related to T-III were from Europe, including Turkey (Fig. S1B).

Haplotype network analyses of the T-III lineage reveals 48 haplotypes. The network shows that T-III differs from the outgroups in four mutation steps (Fig. 3), including two synonymous mutations, C5812T and C15895T, in Orf1ab, one nonsynonymous mutation G25273C (M1237I) in Spike, and one T27869C mutation in a non-coding region (Table S2). The closest outgroup haplotype consists of four sequences, with the two sequences from Turkey were collected on February 8, 2021. The rest were collected after the onset of the third outbreak in Taiwan (Table S3). Further database mining confirmed that of 1,140,328 alpha strain genomes

examined as of 12/11/2021, only the lineage T-III possessed the four above mentioned mutations, which form a distinctive genetic fingerprint. Within T-III, the network forms a star-like shape centered on a core haplotype comprising 26 sequences. Most of the remaining haplotypes are directly connected to this major haplotype. Of the three cases identified on the first day of the third outbreak (April 20 2020), the case from Australia belongs to the major haplotype, with the rest (1078 and 1079) one mutation away.

The estimated date of the most recent common ancestor (TMRCA) of T-III is March 23, 2021 (95% HPD February 24 - April 13, 2021) (Fig. 4), almost one month before the first three confirmed cases on April 20, 2021. Effective population size of the T-III lineage increased approximately 20-fold after the onset of the outbreak and reached a plateau in early June. The estimated demographic expansion of T-III is consistent with epidemiological data (Fig. 1). We noticed that demography in Fig. 4 does not capture the population decline after July 2021 as shown in Fig. 1. That is because most sequences used in this analysis were collected before June 2021 (Table S1), with only four sequences obtained after July 2, 2021. As we are interested in the outbreak origin, this sampling strategy should not affect our conclusions.

Rapid population expansion can also be revealed by contrasting patterns of genetic variation estimated using different approaches. The Watterson's estimator of θ (6.95×10^{-4}) is approximately seven times higher than nucleotide diversity (π) (1.05×10^{-4}), leading to significantly negative Tajima's D (-2.82, $p < 0.001$) (Table S4). Because θ is strongly influenced by rare mutations which are common during recent population expansion or after positive selection (Li 1997), it is a better estimator of genetic diversity for T-III.

DISCUSSION

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

Our estimation that T-III originated from Europe with TMRCA on Mar 23 2021 (95% HPD February 24 2021 - April 13 2021) reconciles several unresolved observations. First, the first two cases (case 1078 and 1079) of the outbreak shared identical sequences, indicating they were from the same source. Nevertheless, the Ct values at the time of diagnosis (April 20, 2021) were 29 and 17 for case 1078 and 1079, respectively, suggesting they were infected at different times (CECC 2021b). Second, several Airline pilots and their family members were PCR negative but serum IgM-/IgG+ in late April and early May 2021 (CECC 2021a, d). It has been shown that IgM levels increase during the first week after SARS-CoV-2 infection, peak after two weeks, and then recede to near-background levels in most patients. IgG is detectable after one week after disease onset and is maintained at a high level for a long period (Hou, et al. 2020). Consequently, IgM negative but IgG positive individuals have probably been infected by SARS-CoV-2 earlier than April 20, 2021. Third, according to CECC, the dates of symptom onset in two seemingly unrelated infection clusters are very close (April 16 for the first cluster and April 23 for the second). As our phylogenetic analysis reveals that all sequences in the third outbreak have a single origin, the occurrence of two infection clusters at similar time

265 without traceable connection suggests that the virus may have been cryptically circulating in
266 the community undetected. Consequently, the origin of the third outbreak should be prior to
267 April 20, 2021.
268
269 In addition, as most sequences closely related to T-III are from Europe, our results also argue
270 against the notion that the outbreak was imported by case 1078 and 1079 from the USA. Among
271 four sequences closest to T-III only two from Turkey were collected on February 8, 2021. The
272 rest appeared after April 20 and cannot be associated with the third outbreak. Consequently,
273 the lineage leading to the third outbreak is most likely originated from Europe, perhaps Turkey,
274 in February 2021 (Fig. 3). Without further information, it is difficult to decide where the four
275 additional mutations were accumulated. However, T-III likely circulated within Taiwan in mid-
276 March 2021. The virus was undetected while spreading within the community.
277
278 **ACKNOWLEDGEMENT**
279 This study was supported by grants from the Ministry of Science and Technology (MOST),
280 Taiwan (111-2321-B-002-017, 111-2634-F-002-017, 109-2311-B-002 -023 -MY3) and the
281 "Center of Precision Medicine" from The Featured Areas Research Center Program within the
282 framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in
283 Taiwan.
284

285 **REFERENCE**

286

287 6 pilots test positive for antibodies in expanded testing program for airline pilots [Internet].
288 Taiwan: CDC; 2021a. Available from:
289 https://www.cdc.gov.tw/En/Bulletin/Detail/D3vIMViyvOwtrL91OSOlfw?typeid=158

290 CECC confirms 2 more cases; sources of infection to be investigated [Internet]. Taiwan:
291 CDC; 2021b. Available from:
292 https://www.cdc.gov.tw/En/Category/ListContent/tov1jahKUv8RGSbvmzLwFg?uaid=NrqG
293 hobGKPIYRPGw2w2z0Q

294 CECC confirms 2 more indigenous cases; contractor of airport quarantine hotel and family
295 member of airline pilot found to have COVID-19 [Internet]. Taiwan: CDC; 2021c. Available
296 from: https://www.cdc.gov.tw/En/Bulletin/Detail/Rk3a6olNXLGuSKn6-YSy4Q?typeid=158

297 CECC confirms 3 more indigenous cases; contacts of confirmed cases found to have COVID-
298 19 [Internet]. Taiwan: CDC; 2021d. Available from:
299 https://www.cdc.gov.tw/En/Bulletin/Detail/nikDToV8jVW2V4m5ag9NcQ?typeid=158

300 CECC confirms 185 more COVID-19 cases, including 180 indigenous cases and 5 imported
301 cases [Internet]. Taiwan: CDC; 2021e. Available from:
302 https://www.cdc.gov.tw/En/Category/ListContent/tov1jahKUv8RGSbvmzLwFg?uaid=70ryE
303 YOTbvx8VaItygA1IA

304 CECC launches an investigation into notification concerning an airline pilot testing positive
305 for COVID-19 in Australia and contacts testing positive in Taiwan [Internet]. Taiwan: CDC;
306 2021f. Available from:
307 https://www.cdc.gov.tw/En/Bulletin/Detail/fDE2kCBYdnmbK0QiU_tSMg?typeid=158

308 Chen K-T, Twu S-J, Chang H-L, Wu Y-C, Chen C-T, Lin T-H, Olsen SJ, Dowell SF, Su I-J.
309 2005. SARS in Taiwan: an overview and lessons learned. International Journal of Infectious
310 Diseases 9:77-85.

311 Clement M, Snell Q, Walker P, Posada D, Crandall K editors. Parallel and Distributed
312 Processing Symposium, International. 2002.

313 Corman VM, Muth D, Niemeyer D, Drosten C. 2018. Hosts and Sources of Endemic Human
314 Coronaviruses. Adv Virus Res 100:163-188.

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

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

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

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

325 Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: multiple sequence
326 alignment, interactive sequence choice and visualization. Briefings in Bioinformatics
327 20:1160-1166.

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

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

332 Li W-H. 1997. Molecular evolution. Sunderland, Mass.: Sinauer Associates.

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

336 Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of
337 restriction endonucleases. Proc Natl Acad Sci U S A 76:5269-5273.

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

341 Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing
342 phylogenetic trees. Molecular biology and evolution 4:406-425.

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

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

348 Watterson GA. 1975. On the number of segregating sites in genetical models without
349 recombination. Theoretical Population Biology 7:256-276.

350 Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, et al.
351 2020. A new coronavirus associated with human respiratory disease in China. Nature
352 579:265-269.
353
354

Figure legend

Figure 1 Number of local COVID-19 cases in Taiwan

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.

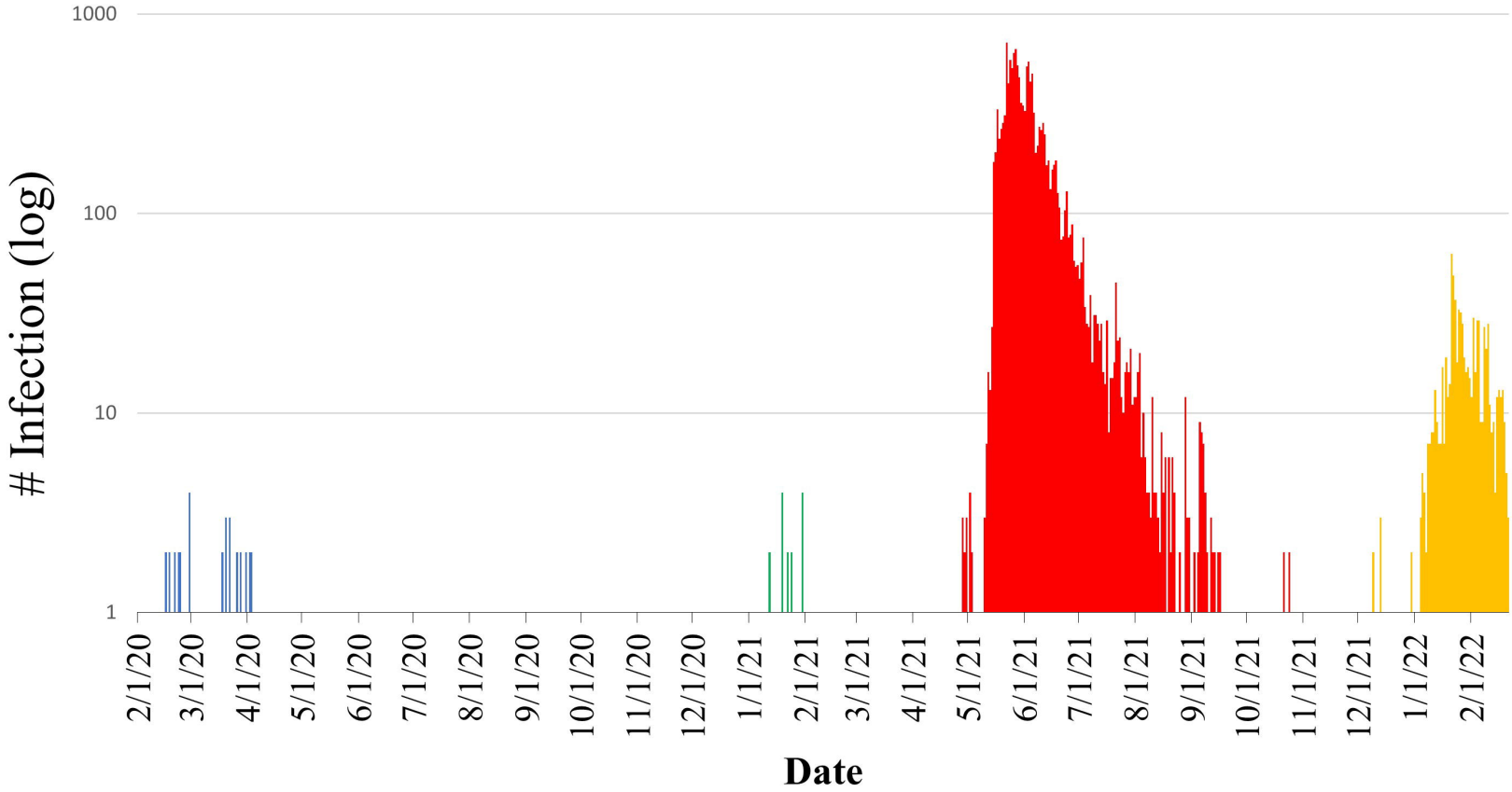
Figure 3 Haplotype network of SARS-CoV-2 genomes from the third local outbreak

Haplotypes in yellow and pink are 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).

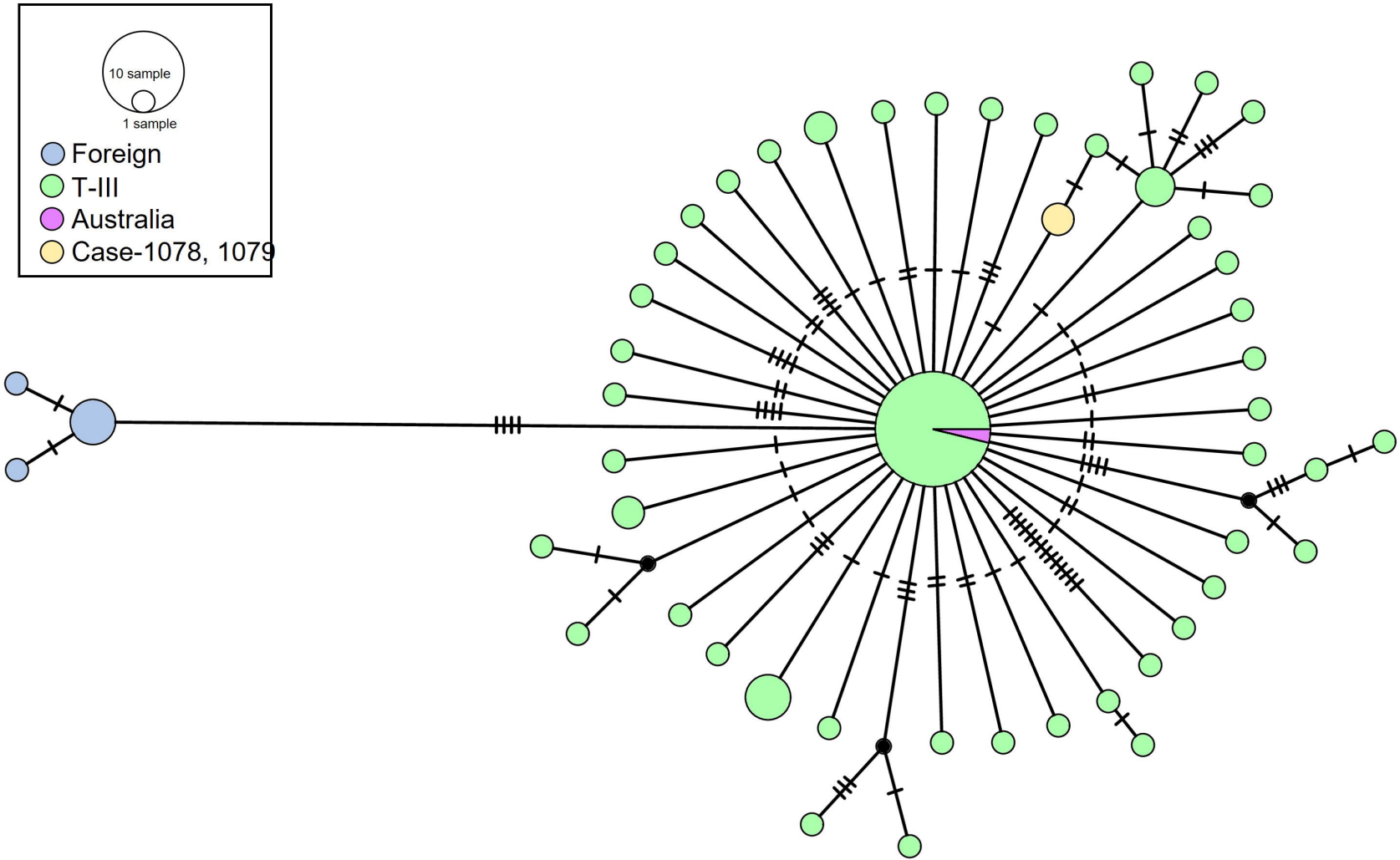
Figure 4 The epidemic growth curve of SARS-CoV-2 genomes from the third local outbreak.

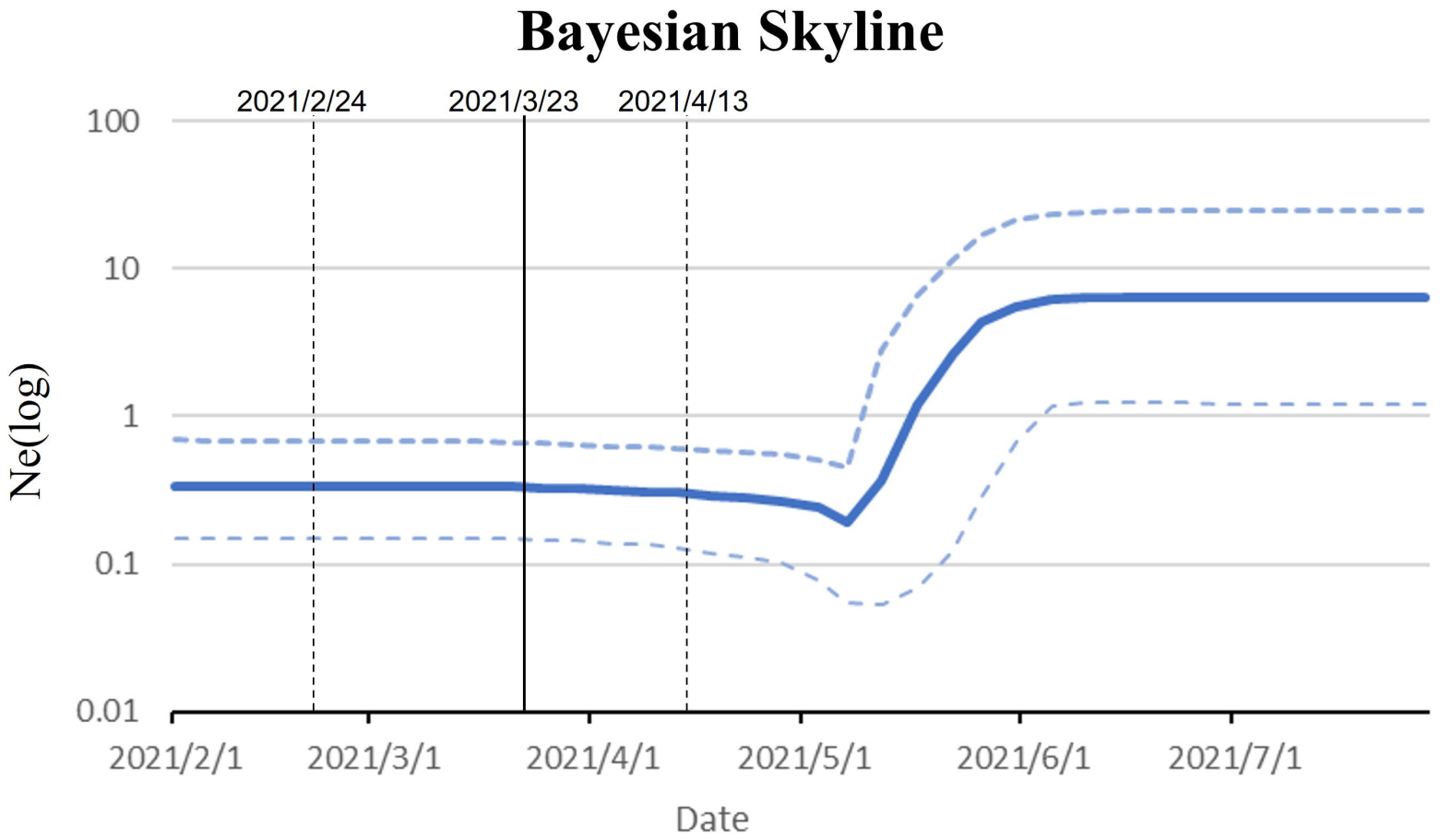
The three lines are the median (blue line) and 95% HPD intervals (dashed lines) of the Bayesian skyline plot (m = 5). Vertical solid line indicates the estimated time to the most recent common ancestor with 95% HPD intervals in dashed lines.

Local Infection in Taiwan




(A)
(B)
CGMH-CGU-78/2021|EPI_ISL_2544714|2021-05-22
TSGH-45/2021|EPI_ISL_4096805|2021-07-28
CGMH-CGU-77/2021|EPI_ISL_2544713|2021-05-23
CGMH-CGU-76/2021|EPI_ISL_2544712|2021-05-20
TSGH-42/2021|EPI_ISL_2693005|2021-05-02
TSGH-43/2021|EPI_ISL_2693006|2021-05-07
CGMH-CGU-74/2021|EPI_ISL_2544710|2021-05-29
CGMH-CGU-73/2021|EPI_ISL_2544709|2021-05-09
CGMH-CGU-72/2021|EPI_ISL_2544708|2021-05-27
CGMH-CGU-70/2021|EPI_ISL_2544706|2021-05-20
CGMH-CGU-79/2021|EPI_ISL_5160472|2021-05-18
KMUH-7/2021|EPI_ISL_7016498|2021-06-06
TSGH-44/2021|EPI_ISL_4096680|2021-05-13
CGMH-CGU-69/2021|EPI_ISL_2544705|2021-05-21
CGMH-CGU-85/2021|EPI_ISL_5160564|2021-07-24
TSGH-46/2021|EPI_ISL_4096807|2021-06-05
ntu82|2021-06-17
1079/2021|EPI_ISL_2455264|2021-04-20
1078/2021|EPI_ISL_2455327|2021-04-20
ntu63|2021-04-23
ntu118|2021-06-13
ntu71|2021-05-13
ntu74|2021-05-14
ntu69|2021-05-12
ntu90|2021-06-11
ntu87|2021-06-17
ntu85|2021-06-17
CGMH-CGU-75/2021|EPI_ISL_2544711|2021-05-26
ntu99|2021-07-02
CGMH-CGU-68/2021|EPI_ISL_2544704|2021-05-19
KMUH-6/2021|EPI_ISL_7016494|2021-05
CGMH-CGU-67/2021|EPI_ISL_2544703|2021-05-20
Australia/NSW-R0216/2021|EPI_ISL_1756021|2021-04-20
ntu93|2021-06-10
ntu83|2021-06-17
ntu84|2021-06-17
ntu109|2021-07-02
ntu70|2021-05-12
ntu115|2021-07-06
ntu114|2021-06-29
ntu97|2021-06-19
ntu100|2021-06-22
ntu86|2021-06-18
ntu77|2021-05-14
ntu80|2021-05-14
ntu72|2021-05-14
ntu73|2021-05-14
ntu68|2021-05-07
2262/2021|EPI_ISL_2455330|2021-05-26
3445/2021|EPI_ISL_2455329|2021-05-26
ntu66|2021-05-07
ntu67|2021-05-07
ntu75|2021-05-14
ntu78|2021-05-14
ntu76|2021-05-14
ntu79|2021-05-14
13375/2021|EPI_ISL_3001055|2021-06-25
ntu81|2021-06-16
10321/2021|EPI_ISL_3040148|2021-06-06
9098/2021|EPI_ISL_3040149|2021-06-05
14454/2021|EPI_ISL_3040142|2021-06-23
13564/2021|EPI_ISL_3001368|2021-06-24
12288/2021|EPI_ISL_3040144|2021-06-08
12699/2021|EPI_ISL_3002178|2021-06-24
5703/2021|EPI_ISL_3000790|2021-06-25
11042/2021|EPI_ISL_3040145|2021-06-06
7955/2021|EPI_ISL_3040151|2021-06-06
14222/2021|EPI_ISL_3040143|2021-06-21
12828/2021|EPI_ISL_3040141|2021-06-24
13435/2021|EPI_ISL_3040140|2021-06-24
12857/2021|EPI_ISL_3001841|2021-06-24
11103/2021|EPI_ISL_3040147|2021-06-05
11310/2021|EPI_ISL_3040146|2021-06-06
CGMH-CGU-64/2021|EPI_ISL_2544700|2021-05-17
CGMH-CGU-65/2021|EPI_ISL_2544701|2021-05-18
CGMH-CGU-66/2021|EPI_ISL_2544702|2021-05-18
10747/2021|EPI_ISL_3000409|2021-06-25
11102/2021|EPI_ISL_3040152|2021-06-05
KMUH-5/2021|EPI_ISL_7016459|2021-05
KMUH-4/2021|EPI_ISL_7016374|2021-05
NTU62/2021|EPI_ISL_1667475|2021-04-12
ntu64|2021-04-28
CGMH-CGU-60/2021|EPI_ISL_2249936|2021-04-04
CGMH-CGU-61/2021|EPI_ISL_2250151|2021-04-04
NTU61/2021|EPI_ISL_1667474|2021-04-06
CGMH-CGU-71/2021|EPI_ISL_2544707|2021-04-05
NTU49/2020|EPI_ISL_1010728|2020-12-31
NTU54/2021|EPI_ISL_1039160|2021-01-08
KMUH-3/2021|EPI_ISL_5395633|2021-04
CGMH-CGU-63/2021|EPI_ISL_2250184|2021-04-21
NTU52/2021|EPI_ISL_1041958|2021-01-06
CGMH-CGU-44/2020|EPI_ISL_956325|2020-12-26
792/2020|EPI_ISL_1381386|2020-12-28
CGMH-CGU-58/2021|EPI_ISL_2249597|2021-03-01
9.0E-5


10 sample
1 sample
Foreign
T-III
Australia
Case-1078, 1079


Bayesian Skyline
2021/2/24
2021/3/23
2021/4/13
100
10
1
0.1
0.01
Ne(log)
2021/2/1
2021/3/1
2021/4/1
2021/5/1
2021/6/1
2021/7/1
Date