

1 **Spatiotemporal Evolution of SARS-CoV-2 Alpha and Delta Variants during a** 2 **Large Nationwide Outbreak in Vietnam, 2021**

3 Nguyen Thi Tam[#], Nguyen To Anh[#], Trinh Son Tung[#], Pham Ngoc Thach, Nguyen Thanh Dung,
4 Van Dinh Trang, Le Manh Hung, Trinh Cong Dien, Nghiem My Ngoc, Le Van Duyet, Phan Manh
5 Cuong, Hoang Vu Mai Phuong, Pham Quang Thai, Nguyen Le Nhu Tung, Dinh Nguyen Huy Man,
6 Nguyen Thanh Phong, Vo Minh Quang, Pham Thi Ngoc Thoa, Nguyen Thanh Truong, Tran Nguyen
7 Phuong Thao, Dao Phuong Linh, Ngo Tan Tai, Ho The Bao, Vo Trong Vuong, Huynh Thi Kim
8 Nhung, Phan Nu Dieu Hong, Le Thi Phuoc Hanh, Le Thanh Chung, Nguyen Thi Thanh Nhan, Ton
9 That Thanh, Do Thai Hung, Huynh Kim Mai, Trinh Hoang Long, Nguyen Thu Trang, Nguyen Thi
10 Hong Thuong, Nguyen Thi Thu Hong, Le Nguyen Truc Nhu, Nguyen Thi Han Ny, Cao Thu Thuy,
11 Le Kim Thanh, Lam Anh Nguyet, Le Thi Quynh Mai, Tang Chi Thuong, Le Hong Nga, Tran Tan
12 Thanh, Guy Thwaites, H. Rogier van Doorn, Nguyen Van Vinh Chau, Thomas Kesteman, Le Van
13 Tan, for OUCRU COVID-19 research group*

14 **Author affiliations**

15 Oxford University Clinical Research Unit, Hanoi, Vietnam (NTT, TST, NTT, NTHT, TK, HRvD)

16 National Hospital for Tropical Diseases, Hanoi, Vietnam (PNT, VDT, LVD, PMC)

17 Vietnam Military Medical University, Hanoi, Vietnam (TCD)

18 National Institute of Hygiene and Epidemiology, Hanoi, Vietnam (HVMP, LTQM, PQT)

19 Hue National Hospital, Hue, Vietnam (PNDH, LTPH)

20 Da Nang Center for Disease Control, Da Nang, Vietnam (LTC, NTTN, TTT)

21 Pasteur Institute in Nha Trang, Nha Trang, Khanh Hoa, Vietnam (DTH, HKM, THL)

22 Oxford University Clinical Research Unit, Ho Chi Minh city, Vietnam (NTA, NTTH, LNTN, NTHN,

23 TTT, CTT, LKT, LAN, GT, LVT)

24 Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam (NTD, LMH, NLNT, DNHM, NMN,
25 NTP, VMQ, PTNT, NTT, TNPT, DPL, NTT, HTB, VTV, HTKN)
26 Department of Health, Ho Chi Minh City, Vietnam (NVVC, TCT)
27 HCMC Center for Disease Control, Ho Chi Minh city, Vietnam (LHN)
28 Centre for Tropical Diseases, Nuffield Department of medicine, University of Oxford, Oxford, UK
29 (HRvD, GT)

30 # These authors contributed equally to this article.

31 *members of the working group are presented in the Acknowledgements

32 **Correspondence:** Nguyen To Anh, Oxford University Clinical Research Unit, Ho Chi Minh City,
33 Vietnam; Tel: (+84 28) 8384009; Fax: (+84 8) 9238904; E-mail: anhnt@oucru.org

34 Le Van Tan, Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam; Tel: (+84 28)
35 8384009; Fax: (+84 8) 9238904; E-mail: tanlv@oucru.org

36 **Key words:** SAR-CoV-2, Alpha variant, Delta variant, evolution, Vietnam

37 **Figures:** 6, **Tables:** 1

38 **Running title:** Evolution of SARS-CoV-2 Variants in Vietnam, 2021

39 **Word count:** Abstract: 147 words, Max: 150 words; Main text: 3496 words, max 3500 words

40 **ABSTRACT**

41 In 2021, Vietnam experienced a large nationwide outbreak of COVID-19, with over 1.7 million
42 infections and 32,000 deaths. We generated 1,303 SARS-CoV-2 whole-genome sequences, and
43 mapped out the public health measures alongside the evolutionary trajectory of the pathogen. The
44 Alpha variant caused sporadic outbreaks in early 2021 prior to the upsurge in cases associated with
45 the Delta variant from June onward. The Delta variant outbreak was almost entirely confined to the
46 AY.57 lineage, accounting for 99.2% of 1,212 Delta variant sequences, and resulting from a single
47 introduction. Viral dispersal from the north, where it was first introduced into Vietnam, to the south,
48 marked the start of the nationwide outbreak, with the south subsequently seeding the virus back to the
49 other regions. Distinct AY.57 phylogenetic clusters in different regions of Vietnam were documented,
50 pointing to the impact of in-country lockdown. Genomic surveillance is critical to inform pandemic
51 response.

52 INTRODUCTION

53 SARS-CoV-2 Alpha and Delta variants of concern (VoC) were first detected in November 2020 in
54 England and India, respectively. Both variants carry key mutations (including N501Y, P681H and
55 H69/V70 deletion for the Alpha variant, and L452R, P681R and T478K for the Delta variant) in the
56 spike protein that make the viruses bind more efficiently to the angiotensin-converting-enzyme 2
57 receptor of human cells and/or escape antibody neutralisation induced by infection or vaccination
58 with ancestral strain [1-3]. Because of these features, the Alpha variant is estimated to be 50% more
59 transmissible than the ancestral strain, and the Delta variant is approximately 40-60% more
60 transmissible than the Alpha variant [4]. As a consequence, both variants rapidly spread around the
61 world and became the dominant viruses responsible for two consecutive COVID-19 waves in 2021.
62 Having reported only 1,465 PCR-confirmed cases of SARS-CoV-2 infection and 35 deaths in 2020
63 [5], Vietnam was one of the few countries to successfully control community transmission of SARS-
64 CoV-2 during the first year of the pandemic. The cornerstone of this success was the early
65 introduction of border and travel restrictions, meticulous contact tracing coupled with mass testing
66 and strict quarantine, and intensive mass communication to the public [5, 6]. However, the
67 importations of the highly transmissible Alpha variant followed by the Delta variant in early 2021
68 resulted in a large nationwide outbreak, with over 1.7 million infections and 32,000 deaths reported
69 by the end of 2021 [7].

70 Genomic surveillance has been one of the top priorities of the WHO. In addition to enabling early
71 detection of new SARS-CoV-2 variants, it has provided significant insights into the evolutionary
72 trajectories of SARS-CoV-2 in time and space [8], critical to inform pandemic response. Yet, most of
73 the studies published about genetic evolution of SARS-CoV-2 are based on datasets from high-
74 income countries with relatively open borders [9-11], and little is known about the transmission

75 dynamics of SARS-CoV-2 in settings with extremely limited number of viral importations such as
76 Vietnam. Here, we focused our analysis on spatiotemporal evolution of SARS-CoV-2 Alpha and
77 Delta variants in Vietnam in 2021. Our aim was also to map out the patterns of viral evolution and
78 diffusion against the public health measures implemented in Vietnam as the pandemic progressed
79 during the study period.

80 **MATERIALS AND METHODS**

81 **Vietnam COVID-19 containment approach in 2021**

82 After successfully controlling community transmission during the first eight months in 2020, Vietnam
83 cautiously reopened its borders, allowing some repatriation flights, while steadily lifting in-country
84 travel restrictions [5]. However, following the report of new SARS-CoV-2 variants in the UK and
85 India in late 2020, now known as Alpha and Delta variants, respectively, Vietnam suspended all
86 inbound flights from countries reporting the detection of community transmission associated with
87 these two variants. Additionally, all travelers entering Vietnam were subjected to 14-day isolation,
88 and testing on the day of arrival and on day 14 of quarantine. Yet, after 28 days of no community
89 transmission, on 18th January 2021, two clusters of SARS-CoV-2 infection with epidemiological links
90 were detected in two neighboring provinces in the north of Vietnam, Hai Duong and Quang Ninh
91 [12]. This was followed by the detection of several other community transmission clusters of unclear
92 origin in several provinces across the country in subsequent months before the start of the 2021 major
93 outbreak (Figure 1). At the national level, the responses were initially targeted e.g., lockdown at
94 community/province/city level where a community transmission cluster was detected. However, in
95 response to the escalation of the outbreak starting in July 2021 [7], Vietnam suspended all arriving
96 international flights in May 2021, and re-applied in-country travel restriction (Figure 1) until 10th
97 October 2021 when nearly 70% of the eligible population had received at least one dose of vaccine

98 [7]. In parallel, under the zero COVID-19 policy, Vietnam implemented the meticulous contact
99 tracing and mass testing of the cases and their contacts conducted by provincial Centers for Diseases
100 Control (CDC). This had enabled accurate identification of cases of community transmission origin
101 alongside the demographic data for analysis [13].

102 **Settings and sampling approach**

103 The present study was conducted at the National Hospital for Tropical Diseases (NHTD) in Hanoi,
104 and the Hospital for Tropical Diseases (HTD) in Ho Chi Minh City, Vietnam. Both NHTD and HTD
105 are tertiary referral hospitals for COVID-19 patients in the north and south of Vietnam, respectively.
106 The laboratories of these hospitals were responsible for SARS-CoV-2 diagnosis and sequencing in
107 Vietnam. Therefore, nasopharyngeal swab (NPS) samples submitted to NHTD and HTD laboratories
108 for testing and sequencing were either from provincial CDC or from inpatients being treated at these
109 two hospitals.

110 To increase the chance of successfully obtaining the virus genome, we first applied a preselection
111 criterion based on the cycle threshold (Ct) value of the tested samples generated by the Lightmix
112 Modular SARS-CoV-2 RdRp/E gene assay (Tib Molbiol, Berlin, Germany) [14]. Accordingly, at
113 NHTD, only NPS samples with a cycle threshold (Ct) value of ≤ 30 for the RdRp gene were eligible,
114 while at HTD, a sample with Ct value of ≤ 25 for the E gene was included. Additionally, because of
115 the availability of the resources, between January and June 2021 when community transmission
116 remained limited, our approach focused on new community clusters detected through contact tracing
117 under zero-COVID strategy. Between July and December 2021 during which community
118 transmission was escalating, the selection of samples for sequencing was carried out using WHO
119 recommendations [15]. Epidemiological data, and data on infections and deaths were retrieved from
120 e-hospital record or provided by the National Institute of Hygiene and Epidemiology and the

121 Vietnamese Ministry of Health. Here, we focused our analysis on sequences obtained cases of locally
122 acquired infection, and generated by NHTD or HTD laboratories, of which we had accurate sampling
123 date and demographic data.

124 **RNA extraction, whole genome sequencing and sequence assembly**

125 Total RNA was extracted from NPS using QIAamp viral RNA mini kit (Qiagen, Hilden, Germany)
126 following the manufacturer's instructions, and finally eluted in 50µl of elution buffer (provided with
127 the extraction kit). Whole genome amplification was performed on the extracted RNA using either
128 the long pooled amplicons protocol, developed by the University of Sydney [16, 17], or the ARTIC
129 V3 protocol [18] on an Illumina MiSeq platform as previously described. Library preparation was
130 carried out using the Nextera XT Library preparation kit (Illumina, USA), followed by library
131 quantification using KAPA Library Quant Kit (Kapa Biosystems, Wilmington, MA, USA), according
132 to the manufacturer's instructions. The prepared library was sequenced using MiSeq reagent kit V2
133 (300 cycles) on a Miseq platform (Illumina). For each run, tested samples were multiplexed and
134 differentiated by double indexes using IDT-ILMN Nextera DNA UD indexes (IDT).

135 Sequence assembly of the obtained sequencing data was carried out using a reference-based mapping
136 approach available in CLC genomics workbench (v.21.0.4) and Geneious 8.1.5 (Biomatters, San
137 Francisco, CA, USA). This method involved mapping of sequencing output of individual samples to a
138 reference genome (WuHan-Hu-1: NC_045512, Alpha: EPI_IS1_905782, Delta: EPI_ISL_1942165),
139 followed by manual editing of the obtained consensus, as described previously [16]. The consensus
140 sequences generated in this study were submitted to the National Center for Biotechnology
141 Information under the assigned accession numbers ON458864-ON459533, ON459545-ON459608,
142 ON755375-ON755859 and OP647358-OP647411.

143 **Lineage classification, sequences alignment and recombination detection**

144 SARS-CoV-2 lineage classification of the obtained consensus was determined using PANGO
145 lineage [19] with pangolin v4.1.2 and pangolin-data v1.13 [20]. The analysis was carried out using
146 the Ultrafast Sample Placement on Existing Trees option to assign the lineage based on the nearest
147 lineage on existing global tree. Sequence alignment was carried out using tool available on Nextclade
148 [21] and Minimap2 [22]. Recombination detection was inferred using a combination of Freyja v
149 1.3.10 [23] and sc2rf [24].

150 **Phylogenetic analysis**

151 To explore the phylogenetic relationship of Vietnamese Alpha and Delta variant sequences, we
152 focused our analysis on complete coding region (29,408 bp) of the obtained sequences, and applied
153 maximum likelihood (ML) method using TIM2 nucleotide substitution model with invariant for
154 Alpha variant, and UNREST+F0+R4 for Delta variant as suggested by IQ tree [25]. Support for
155 individual nodes was assessed using a bootstrap procedure with 1000 replicates. To assess the
156 placement of the Vietnamese variants in the context of global sequences, we employed the
157 phylogenetic framework incorporated in NextClade [21], taking into account representatives of global
158 sequences submitted to GISAID.

159 **Phylogeography and phylodynamics of Delta variant AY.57 lineage**

160 To study the phylogeographics and phylodynamics of the pathogen, we applied Bayesian
161 phylogenetic inference in BEAST v1.10.4 [26]. We first excluded identical sequences and
162 sequences of low quality (e.g., internal gaps). We then used TempEst 1.5 to assess the temporal signal
163 of the dataset [27]. Subsequently, we excluded the sequences not conforming to a linear evolutionary
164 pattern as suggested by TempEst. For phylogeographic analysis, we divided Vietnam into eight major

165 geographic regions (Figure 5). We employed a Bayesian Markov chain Monte Carlo framework
166 (available in BEAST) with 1 billion steps and sampling every 100,000 steps using the general time
167 reversible (GTR) nucleotide substitution model with invariant (as suggested by IQ-TREE to be the
168 best-fit model) under an uncorrelated relaxed clock model [28], and a Bayesian skygride coalescent
169 tree prior (10 groups) [29]. We assessed convergence using Tracer version 1.5 [30]. We selected a
170 burn-in threshold of 10% and accepted effective sample size values above 200. Maximum-clade
171 credibility (MCC) tree was then summarized with TreeAnnotator (available in the BEAST package)
172 and visualized in Figtree version 1.4.2 [31].

173 To assess the effective population size trajectory, we applied Bayesian Skyride model using the above
174 frame work. Finally, we used codon-based method (HyPhy) available in MEGA5 to measure the
175 selection pressure on coding sequences of the pathogen genome, by estimating the ratio of
176 nonsynonymous to synonymous substitution (dN/dS) [32].

177 **Ethics**

178 The study received approvals from the Institutional Review Board of the Hospital for Tropical
179 Diseases in Ho Chi Minh City, Vietnam and the Oxford Tropical Research Ethics Committee. Our
180 study formed part of the national COVID-19 response, and used deidentified throat swabs from
181 COVID-19 patients. Thus, the need for individual informed consent was waived.

182 **RESULTS**

183 **COVID-19 situation in Vietnam in 2021**

184 The COVID-19 situation in Vietnam in 2021 was characterized by the two distinct phases. The first
185 one was between January and April, during which a total of 1,632 infections and no deaths were

186 recorded. This was followed by a second phase occurring between May and December, with a large
187 upsurge in the number of infections (n=1,727,398) and deaths (n=32,359) (Figure 2A).

188 **Results of whole genome sequencing and the study subjects**

189 In total, 1,365 NPS were selected for WGS. Eventually, 1,303 genome sequences were successfully
190 obtained (Figure 3). No recombinants were detected. The majority belonged to Delta variant
191 (n=1,222, 93.8%), followed by Alpha (n= 74, 5.7%), A.23.1 (n=5, 0.4%), and B.1.637 (n=2, 0.2%)
192 variants. Of Delta sequences, 1,212 (99.2%) were assigned to AY.57 lineage, and the remaining was
193 assigned to AY.23 and AY.79 (n=3 each), AY.6, AY.38, AY.85, and B.1.617.2 (n=1 each).

194 Temporally, the obtained sequences distributed throughout the year, with sequences of the Alpha and
195 A.23.1 documented between January and May 2021. The first three Delta variant sequences,
196 including two AY.57 and one B.1.617.2, were detected late April 2021 (Figure 2B). From June
197 onward, Delta variant was the only variant detected, coinciding with the upsurge in the number of
198 infections and deaths of the 2021 outbreak (Figure 2A).

199 The demographic features of the study subjects are presented in Table 1. The characteristics of the
200 A.23.1 variant infected cases were previously reported [12].

201 **Phylogenetic relatedness between sequences of Alpha variant**

202 Phylogenetic analysis of Alpha variant sequences showed that they clustered into four major groups
203 (Figure 4 and Supplementary Figure 1), corresponding to the sporadic community transmission
204 clusters detected in the north, central, and south of Vietnam in early 2021 (Figure 1). This suggested
205 that multiple introductions of Alpha variant into Vietnam occurred between January and May 2021.
206 However, community wide transmission was not sustained, likely attributable to the rapid and
207 rigorous domestic responses (Figure 1 and Figure 2A).

208 **Phylogenetic relatedness between sequences of Delta variant**

209 The reconstructed phylogenetic trees derived from 1,222 Delta variant sequences confirmed the
210 dominance of AY.57 lineage (99.2%) detected in Vietnam throughout 2021, with a small number of
211 the sequences grouped into non-AY.57 groups (Supplementary Figure 2). This was consistent with
212 the results of pangolin lineage classification.

213 **Spatiotemporal evolution of Delta variant AY.57 lineage in Vietnam**

214 Because Delta variant AY.57 lineage was the dominant virus detected during the study period, we
215 focused our phylogeographic analysis on 1,212 sequences of this lineage detected among cases of
216 community transmission. After removing identical sequences, sequences of low quality, and outliers
217 as suggested by TempEst, a total of 748 non-identical sequences were available for analysis. TempEst
218 analysis indicated a positive correlation between genetic divergence and sampling time; the retrieved
219 R^2 value of 0.4538 ($F = 438.4117$, p value < 0.001), suggesting a moderate temporal signal in the
220 included sequences (Supplementary Figure 3). Results of phylogeographic analysis confirmed that
221 AY.57 lineage responsible for the first nationwide wave of COVID-19 in Vietnam in 2021 was first
222 introduced into the Northeast region in early 2021 (Figure 5) and the estimated time to the most
223 recent common ancestor (tMRCA) was March 14, 2021 (95%CI: February 22, 2021 to April 8, 2021).
224 This coincided with the detection of a community transmission cluster in Vinh Phuc and Yen Bai
225 provinces linked with travelers in the north in April 2021 (Figure 1). In the following months, the
226 northeast and Red river delta regions then acted as a source seeding the virus to other neighboring
227 regions, and especially to provinces in the southeastern region. However, viral dispersal between
228 provinces/cities was not frequently observed (Figure 5).

229 Our analysis also revealed that between July and December 2021, the southeastern region was the
230 main source seeding the virus back to the north and the rest of Vietnam. Additionally, we observed

231 the establishment of multiple localized clusters of AY.57 lineage elsewhere in the south central
232 coastal region and the Red river delta (Figure 5).

233 **Phylodynamics of Delta variant lineage AY.57**

234 Bayesian Skyride plot reflecting the dynamics of effective population size of AY.57 lineage
235 circulating within Vietnam is shown in Figure 6. Although the results should be interpreted with
236 caution, we observed a sharp increase in genetic diversity of the pathogen between April and August
237 2021 (Figure 6A), reflecting the expansion of the AY.57 lineage across the country, coinciding with
238 the start of the large nationwide outbreak from May onward (Figure 1 and Figure 2A). In the
239 following months, the viral population size remained relative stable, despite some remarkable
240 fluctuations in the number of viral sequences obtained (Figure 6B). This was followed by a slight
241 decrease in the genetic diversity of the pathogen in the end of November.

242 **Natural selection**

243 The estimated mean of dN/dS value of polyprotein coding sequences was 0.86. The dN/dS values for
244 each individual protein coding gene were as follows: Open Reading Frame (ORF) 1a: 0.69, ORF 1b:
245 1.45, spike protein coding sequence: 0.59, ORF 3a: 1.44, envelope protein coding sequence: 1.14,
246 Matrix protein coding sequence: 0.27, ORF: 0.86, ORF 7a: 6; 1.3, ORF 7b: 0.82, ORF 8: 0.37, and N
247 protein coding sequence: 0.53. Finally, our estimated evolutionary rate of AY.57 coding sequences
248 from Vietnamese community cases sampled in 2021 was estimated to be 5.29×10^{-4} (95% CI:
249 4.966×10^{-4} - 5.639×10^{-4}) substitutions per site per year.

250 **DISCUSSION**

251 We report the evolutionary history of SARS-CoV-2 variants and the potential impact of public health
252 measures on the spread of the virus during an explosive nationwide outbreak in Vietnam in 2021. The

253 Alpha variant was first introduced into Vietnam in early 2021, shortly after it was first discovered in
254 the UK in November 2020, via different importation events as suggested by the results of
255 phylogenetic analysis. However, we only documented sporadic clusters of community transmission
256 associated with the Alpha variant until May before the Delta variant took over from June onward.
257 This was likely attributable to the rigorous containment approach applied in Vietnam at the time, also
258 known as zero COVID-19 policy, and the transmissibility of the Alpha variant itself. Notably, during
259 the same period the Alpha variant was responsible for a new wave of COVID-19 elsewhere,
260 including in the UK and in Europe [33, 34].

261 Unlike the Alpha variant, the Delta variant was not detected in Vietnam until April 2021. Despite
262 rigorous public health measures applied during the next two months, the virus rapidly spread to other
263 provinces in the north, where it was first detected, and in the south before it caused an explosive
264 nationwide outbreak from May onward. Our genomic surveillance data revealed that AY.57 lineage
265 of the Delta variant was the responsible virus with a tMRCA dating back in March 2021.

266 Viruses of the Delta variant have been divided into over 240 sublineages [35]. Of these, AY.57
267 lineage is characterized by the presence of signature mutations in ORF1a (P380S, P1640L, A3209V,
268 V3718A, and T3750I), and in spike protein (A222V). The origin of AY.57 lineage causing the
269 outbreak in Vietnam remains to be determined. However, the result of tMRCA analysis suggested
270 that this lineage had emerged shortly after the discovery of the Delta variant in India in November
271 2020. We may however underestimated the tMRCA of the AY.57 lineage since we only sequenced
272 1212 AY.57 sequences, while over 1.7 million cases were reported during the study period.

273 Collectively, the data emphasize the importance of active genomic surveillance to inform pandemic
274 responses.

275 In spite of the rapid expansion of the AY.57 lineage and the escalation of the outbreak from July
276 onward, our phylogeographic analysis suggested that viral spread between localities in Vietnam only
277 sporadically occurred. Of note, between July and September, Vietnam implemented strict national
278 lockdown with very limited in-country travel, and further tightened border controls. This may have
279 influenced the dispersal of the virus and explained the establishment of distinct phylogenetic clusters
280 of AY.57 lineage in different geographic locations across Vietnam. Notably, in other countries, such
281 as the UK and the US, where travel restriction and border control were less stringent as compared to
282 those applied in Vietnam, viral dispersal across the localities was more apparent [36, 37].
283 Additionally, during the study period, 198 and 195 sublineages of the Delta variant were documented
284 in the UK and the US, respectively [35]. In contrast, the 2021 outbreak in Vietnam was almost
285 entirely confined to the AY.57 lineage, likely via a single introduction as suggested by the tight
286 cluster of the Vietnamese AY.57 lineage on the NextClade-based global phylogenetic tree.

287 We were able to reveal important insights in the evolutionary dynamics of the AY.57 lineage.
288 Accordingly, we showed that there was a sharp increase in relative genetic diversity of the pathogen
289 between April and July 2021, just around the start of the nationwide outbreak. This suggested that the
290 transmission into the community, followed by uncontrolled spread for the first time in Vietnam, was
291 likely attributable to the high transmissibility of the Delta variant. This had resulted in the natural
292 expansion in genetic diversity of the pathogen as it spread among a susceptible population. Notably,
293 between April and July 2021, only a very small proportion of Vietnamese people were vaccinated
294 against COVID-19 (51,000 (0.0005%) and 6.2 million (0.06%) by 1 April and 31 July of a population
295 of 97 million, respectively) [38]. Therefore, at the time Vietnamese people remained largely naïve to
296 SARS-CoV-2 infection [7]. Collectively, our data therefore reflect the natural evolutionary dynamics
297 of AY.57 lineage.

298 Our estimated dN/dS ratio for entire coding region suggested that the evolution of AY.57 lineage was
299 driven by a weak purifying selection, although this varied across the virus genome, with ORF1b,
300 ORF3a, E gene and ORF7a exhibiting a positive selection. Of note, a recent report showed that the
301 vaccines had played a role in selective adaptation of SARS-CoV-2 Delta variant [39]. As a
302 consequence, different mutation profiles were recorded between vaccine breakthrough infection and
303 non-breakthrough infection groups [39]. The role of vaccination in shaping the evolution of SARS-
304 CoV-2 merits further research. Finally, our estimated evolutionary rate of AY.57 lineage was within
305 the range of previously estimated values for the Delta variant [36, 40].

306 Our study has some limitations. We acknowledged a potential bias towards our referral-hospital
307 based sampling approach, especially from July 2021 onwards, when moderate and asymptomatic
308 cases started to be managed at home. As a consequence, we cannot exclude the possibility of the co-
309 circulation of other SARS-CoV-2 variants in Vietnam, which was not captured by our sampling
310 approach. However, this was very unlikely given the tight control border implemented in Vietnam
311 during the study period, and our results are consistent with sequences uploaded on GISAID [35].

312 In summary, we report how the Alpha and Delta variants were introduced, spread, and evolved during
313 the large nationwide outbreak of COVID-19 in Vietnam in 2021. Rigorous public health measures
314 successfully contained the Alpha variant during the first wave. The introduction of the highly
315 transmissible Delta variant AY.57 in an immunologically naïve population however resulted in an
316 explosive nationwide outbreak, coinciding with a rapid increase in genetic diversity of the pathogen.
317 Distinct clusters of AY.57 lineage were established in different geographic locations, pointing to the
318 impact of public health containment on the evolution and the diffusion of SARS-CoV-2. Our study
319 thus emphasizes how genomic surveillance is critical to inform pandemic response.

320 **ACKNOWLEDGEMENTS**

321 We would like to thank to Dr.Le Nguyen Minh Hoa and technicians at Department of Microbiology
322 and Molecular, National Hospital for Tropical Diseases for collecting swab samples and initial testing
323 of SARS-CoV-2 diagnostic. We are grateful to OUCRU team for their supporting for whole genome
324 sequencing and data entry. We thank the submitters from many laboratories for the sequences on the
325 GISAID database.

326 The genomic surveillance work was funded by Wellcome (222574/Z/21/Z). LVT and GT are
327 supported by the Wellcome Trust of Great Britain (204904/Z/16/Z and 106680/B/14/Z, respectively).

328 OUCRU COVID-19 research group: Chambers Mary, Choisy Marc, Dong Huu Khanh Trinh, Dong
329 Thi Hoai Tam, Du Hong Duc, Dung Vu Tien Viet, Fisher Jaom, Flower Barney, Geskus Ronald,
330 Hang Vu Thi Kim, Ho Quang Chanh, Ho Thi Bich Hai, Ho Van Hien, Hung Vu Bao, Huong Dang
331 Thao, Huynh le Anh Huy, Huynh Ngan Ha, Huynh Trung Trieu, Huynh Xuan Yen, Kestelyn
332 Evelyne, Kesteman Thomas, Lam Anh Nguyet, Lawson Katrina, Leigh Jones, Le Kim Thanh, Le
333 Dinh Van Khoa, Le Thanh Hoang Nhat, Le Van Tan, Lewycka Sonia Odette, Lam Minh Yen, Le
334 Nguyen Truc Nhu, Le Thi Hoang Lan, Nam Vinh Nguyen, Ngo Thi Hoa, Nguyen Bao Tran, Nguyen
335 Duc Manh, Nguyen Hoang Yen, Nguyen Le Thao My, Nguyen Minh Nguyet, Nguyen To Anh,
336 Nguyen Thanh Ha, Nguyen Than Ha Quyen, Nguyen Thanh Ngoc, Nguyen Thanh Thuy Nhien,
337 Nguyen Thi Han Ny, Nguyen Thi Hong Thuong, Nguyen Thi Hong Yen, Nguyen Thi Huyen Trang,
338 Nguyen Thi Kim Ngoc, Nguyen Thi Kim Tuyen, Nguyen Thi Ngoc Diep, Nguyen Thi Phuong Dung,
339 Nguyen Thi Tam, Nguyen Thi Thu Hong, Nguyen Thu Trang, Nguyen Thuy Thuong Thuong,
340 Nguyen Xuan Truong, Nhung Doan Phuong, Ninh Thi Thanh Van, Ong Phuc Thinh, Pham Ngoc
341 Thanh, Phan Nguyen Quoc Khanh, Phung Ho Thi Kim, Phung Khanh Lam, Phung Le Kim Yen,
342 Phung Tran Huy Nhat, Rahman Motiur, Thuong Nguyen Thi Huyen, Thwaites Guy, Thwaites Louise,

343 Tran Bang Huyen, Tran Dong Thai Han, Tran Kim Van Anh, Tran Minh Hien, Tran Phuong Thao,
344 Tran Tan Thanh, Tran Thi Bich Ngoc, Tran Thi Hang, Tran Tinh Hien, Trinh Son Tung, van Doorn
345 H. Rogier, Van Nuil Jennifer, Vidaillac Celine Pascale, Vu Thi Ngoc Bich, Vu Thi Ty Hang, Yacoub
346 Sophie. HTD COVID-19 research group: Nguyen Van Vinh Chau, Nguyen Thanh Dung, Le Manh
347 Hung, Huynh Thi Loan, Nguyen Thanh Truong, Nguyen Thanh Phong, Dinh Nguyen Huy Man,
348 Nguyen Van Hao, Duong Bich Thuy, Nghiem My Ngoc, Nguyen Phu Huong Lan, Pham Thi Ngoc
349 Thoa, Tran Nguyen Phuong Thao, Tran Thi Lan Phuong, Le Thi Tam Uyen, Tran Thi Thanh Tam,
350 Bui Thi Ton That, Huynh Kim Nhung, Ngo Tan Tai, Tran Nguyen Hoang Tu, Vo Trong Vuong, Dinh
351 Thi Bich Ty, Le Thi Dung, Thai Lam Uyen, Nguyen Thi My Tien, Ho Thi Thu Thao, Nguyen Ngoc
352 Thao, Huynh Ngoc Thien Vuong, Huynh Trung Trieu, Pham Ngoc Phuong Thao, Phan Minh Phuong.
353 EOCRU COVID-19 research group: Bachtiar, Andy, Baird, Kevin J., Dewi, Fitri, Dien, Ragil,
354 Djaafara, Bimandra A., Elyazar, Iqbal E., Hamers, Raph L., Handayani, Winahyu, Kurniawan, Livia
355 Nathania, Limato, Ralalicia, Natasha, Cindy, Nuraeni, Nunung, Puspatriani, Khairunisa, Rahadjani,
356 Mutia, Rimainar, Atika, Saraswati, Shankar, Anuraj H., Suhendra, Henry, Sutrisni, Ida Ayu, Suwarti,
357 Tarino, Nicolas, Timoria, Diana, Wulandari, Fitri. OUCRU-NP COVID-19 research group: Basnyat
358 Buddha, Duwal Manish, Gautum Amit, Karkey Abhilasha, Kharel Niharika, Pandey Aakriti, Rijal
359 Samita, Shrestha Suchita, Thapa Pratibha, Udas Summita.

360 **ABOUT THE AUTHOR**

361 Dr. Nguyen Thi Tam is a Postdoc Scientist at Oxford University Clinical Research Unit, Hanoi,
362 Vietnam. She is a key person of genomic surveillance of SARS-CoV-2 variants emerging in Vietnam
363 since 2020. She also interested in situation and molecular mechanisms of AMR of bacterial pathogens
364 in Vietnam.

REFERENCES

- 365 1. Tian, D., et al., *The Global Epidemic of the SARS-CoV-2 Delta Variant, Key Spike Mutations and*
366 *Immune Escape*. Front Immunol, 2021. **12**: p. 751778.
- 367 2. Starr, T.N., et al., *Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals*
368 *Constraints on Folding and ACE2 Binding*. Cell, 2020. **182**(5): p. 1295-1310 e20.
- 369 3. Hoffmann, M., H. Kleine-Weber, and S. Pohlmann, *A Multibasic Cleavage Site in the Spike Protein of*
370 *SARS-CoV-2 Is Essential for Infection of Human Lung Cells*. Mol Cell, 2020. **78**(4): p. 779-784 e5.
- 371 4. Washington, N.L., et al., *Genomic epidemiology identifies emergence and rapid transmission of SARS-*
372 *CoV-2 B.1.1.7 in the United States*. medRxiv, 2021.
- 373 5. Van Tan, L., *COVID-19 control in Vietnam*. Nat Immunol, 2021. **22**(3): p. 261.
- 374 6. Thai, P.Q., et al., *The First 100 Days of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)*
375 *Control in Vietnam*. Clin Infect Dis, 2021. **72**(9): p. e334-e342.
- 376 7. WHO, *CCOVID-19 in Viet Nam Situation Report 101, accessed on 24 October 2022*. available at
377 [https://www.who.int/vietnam/internal-publications-detail/covid-19-in-viet-nam-situation-report-](https://www.who.int/vietnam/internal-publications-detail/covid-19-in-viet-nam-situation-report-101)
378 [101](https://www.who.int/vietnam/internal-publications-detail/covid-19-in-viet-nam-situation-report-101), 2022.
- 379 8. Attwood, S.W., et al., *Phylogenetic and phylodynamic approaches to understanding and combating*
380 *the early SARS-CoV-2 pandemic*. Nat Rev Genet, 2022. **23**(9): p. 547-562.
- 381 9. du Plessis, L., et al., *Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK*.
382 Science, 2021. **371**(6530): p. 708-712.
- 383 10. da Silva Filipe, A., et al., *Genomic epidemiology reveals multiple introductions of SARS-CoV-2 from*
384 *mainland Europe into Scotland*. Nat Microbiol, 2021. **6**(1): p. 112-122.
- 385 11. Lemey, P., et al., *Untangling introductions and persistence in COVID-19 resurgence in Europe*. Nature,
386 2021. **595**(7869): p. 713-717.

- 387 12. Chau, N.V.V., et al., *Rapid whole-genome sequencing to inform COVID-19 outbreak response in*
388 *Vietnam*. J Infect, 2021. **82**(6): p. 276-316.
- 389 13. Van Vinh Chau, N., et al., *The Natural History and Transmission Potential of Asymptomatic Severe*
390 *Acute Respiratory Syndrome Coronavirus 2 Infection*. Clin Infect Dis, 2020. **71**(10): p. 2679-2687.
- 391 14. Corman, V.M., et al., *Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR*. Euro
392 *Surveill*, 2020. **25**(3).
- 393 15. Organization, W.H., *Operational considerations to expedite genomic sequencing component of GISRS*
394 *surveillance of SARS-CoV-2*. WHO. 1–10. 2021.
- 395 16. Nguyen, T.T., et al., *Genetic diversity of SARS-CoV-2 and clinical, epidemiological characteristics of*
396 *COVID-19 patients in Hanoi, Vietnam*. PLoS One, 2020. **15**(11): p. e0242537.
- 397 17. Eden, J.-S. and E. Sim, *SARS-CoV-2 Genome Sequencing Using Long Pooled Amplicons on Illumina*
398 *Platforms*. Available on [https://www.protocols.io/view/sars-cov-2-genome-sequencing-using-long-](https://www.protocols.io/view/sars-cov-2-genome-sequencing-using-long-pooled-amp-kxygxeob4v8j/v1)
399 [pooled-amp-kxygxeob4v8j/v1](https://www.protocols.io/view/sars-cov-2-genome-sequencing-using-long-pooled-amp-kxygxeob4v8j/v1). 2020.
- 400 18. Quick, J., *nCoV-2019 sequencing protocol V.3*. 2020.
- 401 19. Rambaut, A., et al., *A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic*
402 *epidemiology*. Nat Microbiol, 2020. **5**(11): p. 1403-1407.
- 403 20. <https://github.com/cov-lineages/pangolin>.
- 404 21. Aksamentov, I., et al., *Nextclade: clade assignment, mutation calling and quality control for viral*
405 *genomes*. Journal of Open Source Software, 2021. **6**(67).
- 406 22. Li, H., *Minimap2: pairwise alignment for nucleotide sequences*. Bioinformatics, 2018. **34**(18): p. 3094-
407 3100.
- 408 23. *Freyja v 1.3.10*, available at <https://github.com/andersen-lab/Freyja>.
- 409 24. *Sc2rf - SARS-Cov-2 Recombinant Finder*, available at <https://github.com/lenaschimmel/sc2rf>.

- 410 25. Nguyen, L.T., et al., *IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-*
411 *likelihood phylogenies*. Mol Biol Evol, 2015. **32**(1): p. 268-74.
- 412 26. Suchard, M.A., et al., *Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10*.
413 *Virus Evol*, 2018. **4**(1): p. vey016.
- 414 27. Rambaut, A., et al., *Exploring the temporal structure of heterochronous sequences using TempEst*
415 *(formerly Path-O-Gen)*. Virus Evol, 2016. **2**(1): p. vew007.
- 416 28. Drummond, A.J., et al., *Relaxed phylogenetics and dating with confidence*. PLoS Biol, 2006. **4**(5): p.
417 e88.
- 418 29. Minin, V.N., E.W. Bloomquist, and M.A. Suchard, *Smooth skyride through a rough skyline: Bayesian*
419 *coalescent-based inference of population dynamics*. Mol Biol Evol, 2008. **25**(7): p. 1459-71.
- 420 30. Available at <http://tree.bio.ed.ac.uk/software/tracer/>.
- 421 31. Figtree version 1.4.2, available at (<http://tree.bio.ed.ac.uk/software/figtree/>).
- 422 32. Tamura, K., et al., *MEGA5: molecular evolutionary genetics analysis using maximum likelihood,*
423 *evolutionary distance, and maximum parsimony methods*. Mol Biol Evol, 2011. **28**(10): p. 2731-9.
- 424 33. Walker, A.S., et al., *Tracking the Emergence of SARS-CoV-2 Alpha Variant in the United Kingdom*. N
425 Engl J Med, 2021. **385**(27): p. 2582-2585.
- 426 34. Stefanelli, P., et al., *Co-circulation of SARS-CoV-2 Alpha and Gamma variants in Italy, February and*
427 *March 2021*. Euro Surveill, 2022. **27**(5).
- 428 35. *Detect and analyze variants of SARS-CoV-2, available at* [https://cov-](https://cov-spectrum.org/explore/Vietnam/AllSamples/Past6M)
429 [spectrum.org/explore/Vietnam/AllSamples/Past6M](https://cov-spectrum.org/explore/Vietnam/AllSamples/Past6M).
- 430 36. Eales, O., et al., *SARS-CoV-2 lineage dynamics in England from September to November 2021: high*
431 *diversity of Delta sub-lineages and increased transmissibility of AY.4.2*. BMC Infect Dis, 2022. **22**(1): p.
432 647.

- 433 37. McCrone, J.T., et al., *Context-specific emergence and growth of the SARS-CoV-2 Delta variant*. Nature,
434 2022. **610**(7930): p. 154-160.
- 435 38. *Avaiable at* <https://ourworldindata.org/coronavirus/country/vietnam>
- 436 39. Duerr, R., et al., *Clinical and genomic signatures of SARS-CoV-2 Delta breakthrough infections in New*
437 *York*. EBioMedicine, 2022. **82**: p. 104141.
- 438 40. Benazi, N. and S. Bounab, *Comparison of the evolutionary phylodynamic of Delta and Omicron*
439 *variants of SARS-CoV-2*. Research Square, 2022.

Table 1. Demographics of the study participants

| | Number of sequences (N=1303) | Alpha (N = 74) | AY.57 (N = 1212) |
|--------------------|---|---------------------------|-----------------------------|
| Age (Years) | | | |
| Median (Q1-Q3) | 43 (29 - 61) | 35 (30-55) | 44 (29 - 61) |
| Missing | 38 (2.9%) | 31 (41.9%) | 2 (0.2%) |
| Gender | | | |
| Female | 689 (52.9%) | 22 (29.7%) | 661 (54.5%) |
| Male | 578 (44.4%) | 21 (28.4%) | 551 (45.5%) |
| Missing | 36 (2.8%) | 31 (41.9%) | |
| Region | | | |
| Centre | 120 (9.2%) | 24 (32.4%) | 94 (7.8%) |
| North | 504 (38.7%) | 44 (59.5%) | 450 (37.1%) |
| South | 679 (52.1%) | 6 (8.1%) | 668 (55.1%) |

LEGENDS TO FIGURES

Figure 1: Government directions and COVID-19 epidemiology events in Vietnam in 2021. **Note to**

Figure 1: #Directive-15: Suspension of non-essential services/businesses and mass gatherings, applying physical distance of 2 meters when contact with others, banning the gatherings of 20 people or more in one place and 10 people or more outside workplaces and limitation of movements.

[§]3 ring quarantine: apply Directive-16 whole city (bans gatherings of two more people in public and asks people to only leave homes for emergencies, food, medicine, work in factories, and businesses that involve essential goods and services), lockdown at residential areas with covid-19 case report, home quarantine for F1. HCMC: Ho Chi Minh city

Figure 2: A) Bar chart showing the number infections (brown bars) and deaths (grey bars) during the 2021 outbreak in Vietnam, B) monthly number of SARS-CoV-2 variants detected among cases of community transmission between January and December 2021. Delta variant non-AY.57 lineage includes B.1.617.2 (n=1), AY.23 (n=3), AY.79 (n=3), AY.85 (n=1), AY.6 (n=1), and AY.38 (n=1). Others include lineages B.1.637 (n=2) and A.23.1 (n=5).

Figure 3: Workflow of the study

Figure 4: Maximum likelihood tree of complete coding sequences of Alpha variant circulating in Vietnam in 2021. Phylogenetic clusters were named accordingly to community clusters recorded during the study period shown in Figure 1.

Figure 5: Maximum clade credibility trees illustrating the phylogeography of AY.57 lineage in Vietnam in 2021. Branches are color-coded according to locations of sampling. Posterior probabilities >0.8% and state probabilities >75% (black circles) are indicated at all nodes

Figure 6: A) Bayesian skyride plot illustrating the relative genetic diversity of AY.57 lineage in Vietnam during the year 2021, and B) the Number of AY.57 sequences used for analysis. Blue shading (Panel A) indicates 95% highest posterior density interval.

Government directions

Border control

Tighten immigration through unofficial/open access along border

Resuspension of flights from countries recorded Delta variant

Limitation of rescue flights to Vietnam

- HaiDuong: social distancing
- VanDon airport: lock downed

Vaccination for healthcare and frontline workers

HaNam: lock downed at infected areas

- **International border closed**
- **Domestic travel limitation and non-essential activities suspension**

BacGiang-BacNinh: Directive-15# social distancing
Lock downed industrial zones

HCMC: Directive-15# social distancing

Vaccination for whole community

HCMC: 3 ring quarantine[§]

HCMC: daily night curfew

HCMC: "stay at home" order

HCMC: relaxing social distancing

VietNam: living safely with COVID-19



Cluster I: HaiDuong - QuangNinh

Cluster II: HaNam
• index case: traveler from Japan

Cluster III: VinhPhuc-Yenbai
• index case: traveler from China

Cluster IV: HaNoi: Hospital clusters (unclear epi)
↳ **BacGiang:** index case related to HaNoi hospital clusters
↳ **BacNinh:** index cases linked to HaNoi hospital clusters

Cluster V: Danang
• index case: Hotel staff (unclear epi)

Cluster VI: HCMC
• A traveler from HaNam
• District 3 cluster (unclear epi)

Cluster VII: HCMC
• ThuDuc cluster (unclear epi)
• PhuHung religious mission cluster

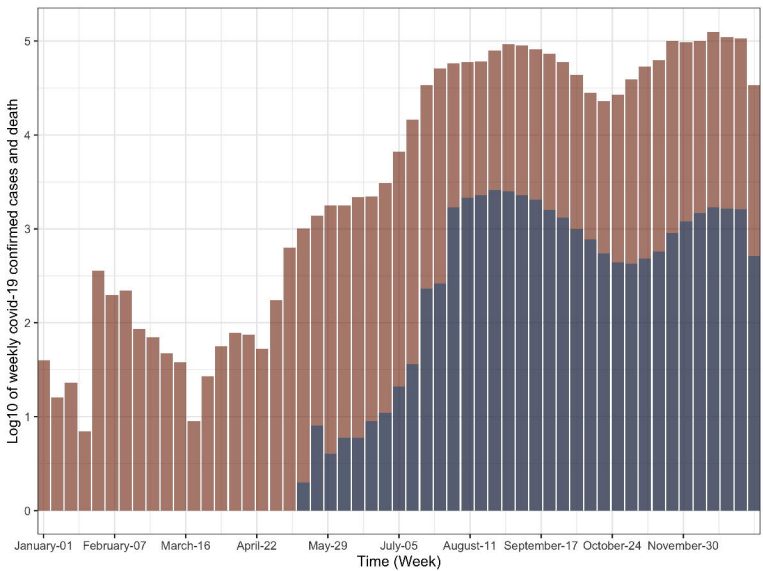
↳ Southeast and Mekong delta
↳ North and Central coast areas

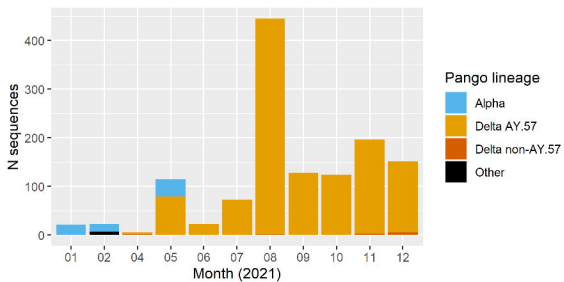
Epidemiology events

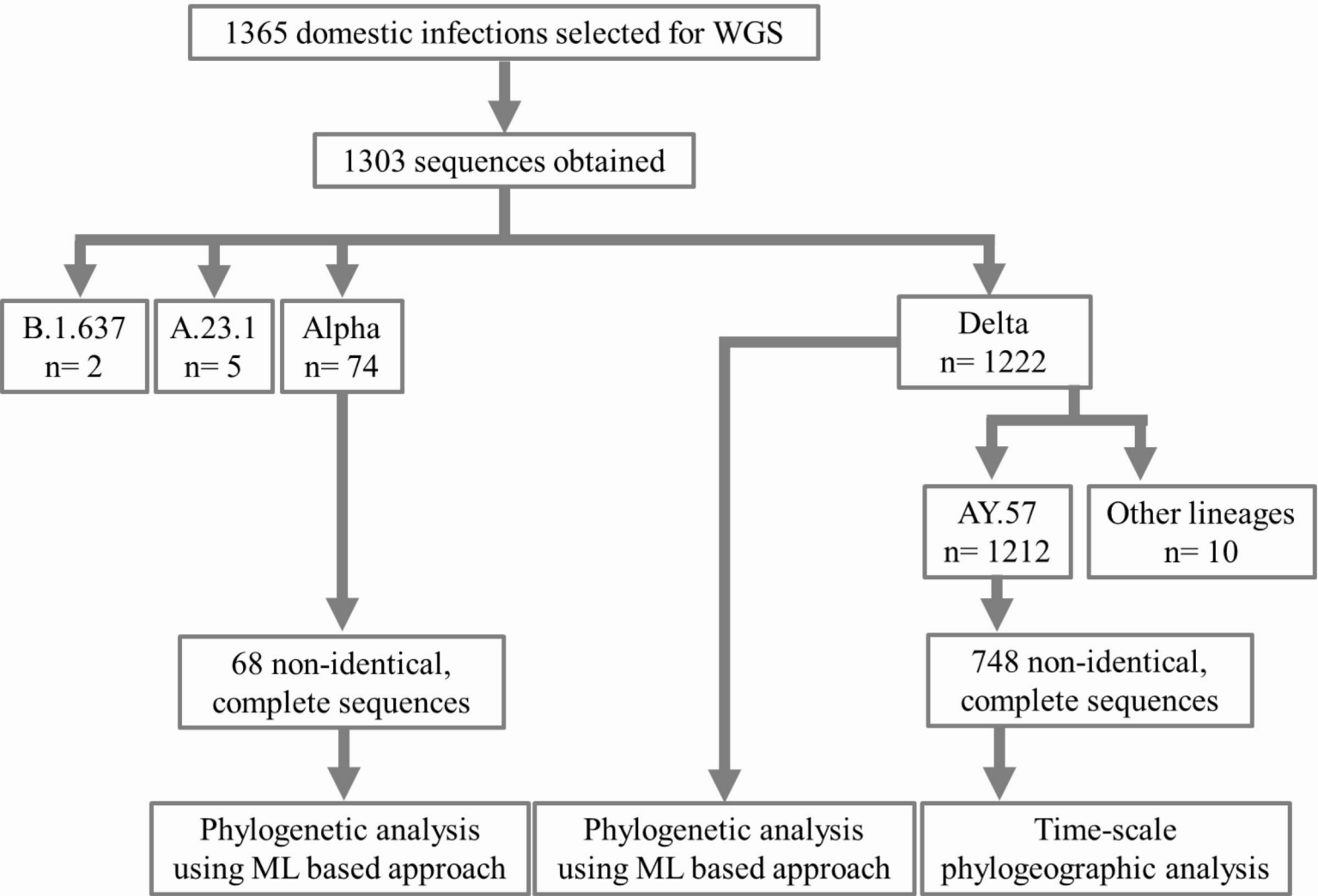
- Alpha
- Delta

• Alpha imported by flights

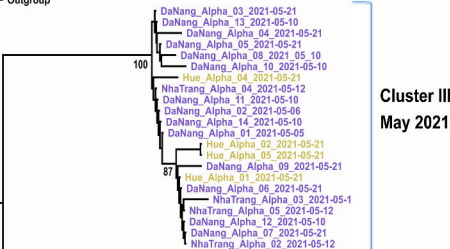
• A.23.1 cluster in TanSanNhat airport, HCMC



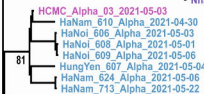




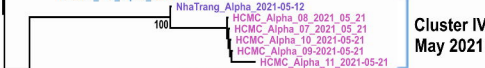
Outgroup



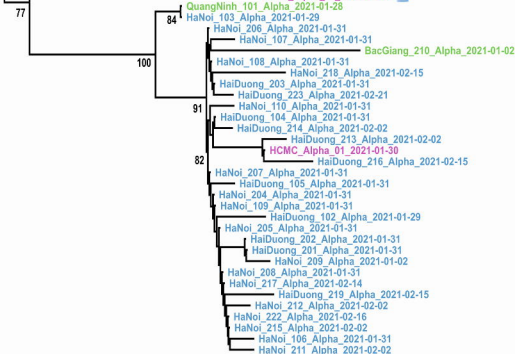
Cluster III
May 2021



Cluster II
May 2021



Cluster IV
May 2021



Cluster I
Jan-Apr 2021

1.0E-4

