

Article

1 Emergence and spread of SARS-CoV-2 lineages B.1.1.7 and P.1 in Italy

2
3 Francesca Di Giallonardo^{1*}, Ilaria Puglia², Valentina Curini², Cesare Cammà², Iolanda Mangone², Paolo
4 Calistri², Joanna C.A. Cobbin³, Edward C. Holmes³, Alessio Lorusso²

5
6 ¹ The Kirby Institute, UNSW Sydney, Sydney, NSW 2052, Australia. (F.D.G)

7 fdigiallonardo@kirby.unsw.edu.au

8 ² Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise G. Caporale, Teramo, TE 641000, Italy.

9 (I.P.) i.puglia@izs.it, (V.C.) v.curini@izs.it, (C.C.) c.camma@izs.it, (I.M.) i.mangone@izs.it, (P.C.)

10 p.calistri@izs.it, (A.L.) a.lorusso@izs.it

11 ³ Marie Bashir Institute for Infectious Diseases and Biosecurity, School of Life & Environmental

12 Sciences and School of Medical Sciences, The University of Sydney, Sydney, NSW 2006, Australia.

13 (J.C.A.C.) Joanna.cobbin@sydney.edu.au, (E.C.H.) edward.holmes@sydney.edu.au

15 Abstract

16 Italy's second wave of SARS-CoV-2 has hit hard, with more than 3 million cases and over 100,000 deaths,
17 representing an almost ten-fold increase on the numbers reported by August 2020. Herein, we present the
18 analysis of 6,515 SARS-CoV-2 sequences sampled in Italy between 29th January 2020 and 1st March 2021
19 and show how different lineages emerged multiple times independently despite lockdown restrictions.
20 Virus lineage B.1.177 became the dominant variant in November 2020, when cases peaked at 40,000 a day,
21 but since January 2021 this is being replaced by the B.1.1.7 'variant of concern'. In addition, we report a
22 sudden increase in another documented variant of concern – lineage P.1 – from December 2020 onwards,
23 most likely caused by a single introduction into Italy. We again highlight how international importations
24 drive the emergence of new lineages and that genome sequencing should remain a top priority for
25 ongoing surveillance in Italy.

27 Introduction

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.¹

Article

28 COVID-19, caused by infection with severe respiratory coronavirus 2 (SARS-CoV-2), has had a devastating
29 global impact, with an estimated 116 million cases and 2.5 million deaths¹. Of these reported infections,
30 more than 78% have occurred during the so-called 'second wave' of the epidemic between September
31 2020 – March 2021, greatly outnumbering the 25 million cases and ~800k deaths reported globally prior to
32 31 August 2020². This second wave of infections is also characterised by the emergence of numerous new
33 lineages of SARS-CoV-2, three of which are commonly referred to global 'variants of concern' – B.1.1.7,
34 B.1.351, and P.1³. The B.1.1.7 lineage was first detected in the United Kingdom in late September 2020⁴
35 and is characterised by 17 polymorphisms across the genome including the N501Y amino acid mutation
36 and a two amino acid deletion at positions 69 and 70 of the spike protein⁵. The B.1.351 lineage was first
37 reported in South Africa and shares similar substitutions to B.1.1.7 including N501Y but not the 69/70
38 deletion⁶, while the P.1 lineage was first identified in Brazil and shares two substitutions with B.1.351 –
39 E484K and N501Y in the spike protein⁴. All of these lineages have seemingly replaced previous circulating
40 variants in their geographic regions and have spread to other countries in Europe, the Americas, and Asia⁷
41 ⁹. There is also considerable concern about the possibility of reinfection with these new lineages due to
42 reduced cross-protective immunity¹⁰⁻¹², which may also have implications for vaccine efficacy^{13,14}. Thus,
43 increased surveillance to document and understand the spread of different SARS-CoV-2 lineages is of
44 utmost importance.

45 The first cases of SARS-CoV-2 in Italy were reported in January 2020 and the country was hit hard
46 with a peak of >6,500 cases a day in late March 2020. Italy was also the first European country to impose a
47 nation-wide lockdown from 9th March to the 18th May 2020. In retrospect, this first wave was of relatively
48 low magnitude compared to the huge number of infections reported between September 2020 and March
49 2021. Indeed, during this second wave, Italy experienced a maximum of ~40,000 cases a day in mid-
50 November 2020 (Fig. 1a), and an estimated 3.1 million cases and ~100,000 deaths have been reported to
51 date (8th March 2021)¹⁵. Beginning in October 2020, a colour-coded system was established to restrict the
52 mobility of residents in each administrative region in Italy, depicting increasing the levels of restriction
53 from yellow to red. In addition, in attempt to reduce viral transmission during the holiday season, further
54 restrictions to social mobility and on possible sources of infection (schools, restaurants, etc.) were put in

Article

55 place for the whole of Italy between 21st December 2020 and the 6th January 2021. On 20th December
56 2020, the Italian Minister of Health signed an order prohibiting all air flights from/to United Kingdom. On
57 16th January 2021 the same restriction was applied to the air flight from/to Brazil.

58 We previously showed that virus importation associated with travel, followed by local
59 transmission, were key drivers of viral spread during the first wave of SARS-CoV-2 in Italy¹⁶. After a year
60 into the pandemic, lockdown restrictions have proven effective in reducing virus transmission and limiting
61 geographic spread¹⁷, although international travel is still a major source for the introduction of new
62 lineages into Italy¹⁸. Here we show how the B.1.1.7 and P.1 variants entered Italy before the
63 implementation of border closures and that these two variants have subsequently spread rapidly across
64 the country.

65

66 Results

67 Emerging lineages in Italy

68 Our analysis included 6,515 Italian SARS-CoV-2 sequences covering a time span between 29th January
69 2020 to 1st March 2021. These comprised 103 different lineages, with B.1 (11%), B.1.1.74 (5%), B.1.1.7
70 (34%), and B.1.177 (24%) the most commonly identified. Sequence sampling was not equal between
71 North, Central, and South Italy, as a large proportion of SARS-CoV-2 sequence data from North Italy was
72 only available for the beginning of the pandemic (March – May 2020) and again the most recent months
73 (January – March 2021) (Fig.1a). Despite this, the time range of data availability overlaps between these
74 three macro areas, minimising the effect of sampling bias.

75 Lineages B.1 and B.1.1.74 both appeared in Italy during early 2020 and were replaced by lineage
76 B.1.177 during the second half of 2020 (Fig. 1b). Lineage B.1 was the first reported in all three macro areas
77 on 20th February, 24th February, and 3rd March in North, Central, and South Italy, respectively. A similarly
78 narrow time period of introduction was observed for lineage B.1.1.74 which first appeared on 24th
79 February, 27th February, 4th March 2020 in North, Central, and South Italy, respectively. This latter lineage
80 persisted in Central and South Italy until February 2021, although at low levels, while no sequences from
81 this lineage were sampled in North Italy after 27th May 2020. Subsequently, lineage B.1.177 emerged in

Article

82 August in all three macro areas becoming the dominant variant in the second wave, with the earliest
83 documentation in North Italy (13th August), followed by South Italy on 15th August and Central Italy on
84 August 18th.

85 Although these data show that different lineages of SARS-CoV-2 were first reported within a short
86 time period throughout the country, we also identified region-specific sampling biases with unusual
87 abundances of sequences obtained on individual days (Supplementary Data S1). These sampling 'peaks'
88 are most likely the result of targeted sampling after an outbreak. For example, 193 sequences were
89 sampled on 26th October 2020 alone, of which 104 were B.1.177 all sampled in the region Campania in
90 South Italy. Also, in this region, 42 B.1.1.187 sequences were sampled on 29th June 2020 alone, and which
91 represents 51% of all sequence data available for this lineage in Italy. Notably, according to GISAID, this
92 lineage has only been sampled in the UK (n=14) and Japan (n=1), and South Italy (n=81).

93

94 **Lineages of global concern - B.1.1.7, P.1, and B.1.351**

95 At the time of writing (8th March 2021), a total of 2,184 B.1.1.7, 109 P.1, and 8 B.1.351 sequences sampled in
96 Italy were available for analysis and had been identified in all three Italian macro areas. The B.1.1.7 lineage
97 appeared in all three areas within a five-day period, first in Central Italy (14th December) and five days later
98 in both the North and South (19th December for both lineages) (Fig. 1b). The first P.1 sequences were
99 identified on 7th January 2021 in Central Italy in travellers returning from Brazil²⁹. This lineage was
100 subsequently found on 21st January and 12th February in the North and South Italy, respectively.

101 Notably, the Italian B.1.1.7 sequences were distributed across the global phylogeny indicating
102 multiple independent introductions into the country (Fig. 2). In addition, we observed numerous Italian-
103 specific transmission clades (i.e. extended transmission chains), although most (n=73, 78%) were area
104 specific, the largest containing 104 infections all sampled in South Italy (Fig. 2). Of the remaining clades
105 (n=22), 82% contained infections from two and 18% from all three macro areas. For 73% (n=16) of these,
106 the clade in question comprised ≥80% of infections from a single area only and only a small number of
107 infections from a second or third macro area. For example, one clade with 64 sequence comprised 63 from
108 Central Italy, one from South Italy and none from North Italy. Similarly, another clade contained 76

Article

109 infections from North Italy and one from Central Italy. In addition, we identified six clades with more inter-
110 area mixing, including one example consisting of 35 infections (10 North, 12 Central, 13 South). Notably,
111 this is the only example of prolonged transmission between macro areas for this time period (Fig. 2).

112 Three infections of the P.1 lineage were identified in international travellers on 18th January in
113 Central Italy. Fortunately, these did not lead to ongoing local transmission (Fig. 3). However, infections of
114 P.1 were also identified on 7th January, and which formed a distinct clade (clade I) within the P.1 global
115 phylogeny distinguished by a single T->C synonymous mutation in ORF1ab (nt 13,577 / aa 4,526 Ref
116 EPI_ISL_833137). In contrast to B.1.1.7, almost all of Italian P.1 sequences (n=104) fell within this single
117 node which also contained 10 sequences from Germany (Fig. 3). Within this clade I, 61 Italian sequences
118 (clade II) contained an additional distinct synonymous mutation in the spike gene (spike nt 3,357 / aa 1,118
119 Ref EPI_ISL_833137), and 43 of these Italian sequences (clade III) also contained a non-synonymous
120 mutation in the spike leading to an S813N mutation (spike nt 2,437 / aa 813 Ref EPI_ISL_833137). This
121 S813N mutation is located between two fusion peptides and not in the receptor binding domain where the
122 E484K and N501Y mutations are located (Supplementary Data S2). P.1. viruses with the mutation were
123 first sampled on 13th January 2021 in Central Italy and the most recent occurrence was found in an infection
124 sampled on 1st March. To date, the mutation has only been found in infections from Central and South
125 Italy. Of note, one additional sequence (EPI_ISL_1169907, Central Italy, 19-Feb) outside of both clade II
126 and III also contained this amino acid substitution (Fig. 3).

127 Finally, five of the eight B.1.351 sequences were scattered across the global phylogeny, confirming
128 that these are independent infections associated with returned travellers (30th January and 11th February in
129 North Italy n=2, 23rd February in Central Italy n=1, 19th and 26th February in South Italy n=2, Supplementary
130 Data S3). However, we also found B.1.351 sequences from North Italy sampled on 11th and 13th February
131 2021 who formed a phylogenetic cluster, suggestive of some local transmission.

132

133 Discussion

134 We document the complex patterns of virus transmission and lineage turnover during the second
135 wave of SARS-CoV-2 in Italy, with an analysis based on >6,500 sequences sampled from January 2020 –

Article

136 March 2021. First, we observed the replacement of B.1 and B.1.1.74 by B.1.177, the latter being the most
137 sampled lineage during the second wave. Second, and more notably, we observed the more recent
138 appearance of the emerging 'variants of concern' B.1.1.7 and P.1, with the former being the most sampled
139 lineage during January and February 2021.

140 Our phylogenetic analysis confirms extensive virus transmission within North, Central, and South
141 Italy, but only limited transmission between these regions, consistent with lockdown restrictions imposed
142 in December 2020 that prohibited travel between regions. However, the limited sequence data available
143 makes it challenging to accurately identify the directionality of transmission and lineage spread across the
144 country. Nevertheless, our data shows how different lineages appeared almost simultaneously in North,
145 Central, and South Italy (i.e., within a five-day time frame), suggesting a complex pattern of multiple
146 introductions into the different macro areas.

147 We focused our analysis on lineages B.1.1.7, B.1.351, and P.1 as these have been the subject of
148 considerable discussion and concern. In the case of B.1.351, we only found one instance of local
149 transmission in mid-February 2021, with no evidence for additional ongoing transmission. In contrast, we
150 found strong evidence for local ongoing transmission of both the B.1.1.7 and P.1 lineages, with the former
151 potentially replacing the previous dominant strain B.1.177. Of note, from 8th March 2021 onwards, B.1.1.7
152 represented on average 50% of all sequences sampled. This is compatible with observations that B.1.1.7
153 has enhanced transmissibility⁵ and has become the dominant variant in numerous other European
154 countries, e.g. Denmark²⁰.

155 After Brazil, Italy contains the second highest number of P.1 infections reported to date³. This
156 variant was first reported in Manaus (Brazil), which was hit hard by the first wave of SARS-CoV-2 with an
157 estimated 76% of the population infected²¹. Thus, the surge of P.1 in this area strongly implies the
158 possibility of re-infection⁹. Our molecular epidemiological analysis provided strong evidence for a single
159 introduction of P.1 into Italy, followed by extensive local transmission. As expected, our phylogenetic
160 analysis depicts importation of the P.1 variant into Italy, although from a yet unknown source, followed by
161 possible onward spread from Italy to Germany. Importantly, within the Italian network we also found a
162 unique, clade defining amino acid substitution - S813N. Although the function of this mutation is currently

Article

163 unknown, it is likely unrelated to virus binding and hence does not represent an immune escape variant
164 although this should be assessed further^{33,22}. No clade defining amino acid substitution was found for the
165 B.1.1.7 lineage in spike gene of Italian sequences.

166 In conclusion, we depict the rapid emergence and replacement of new lineages in Italy despite
167 imposed lockdown restriction, suggesting that the disease management policies employed were
168 insufficient to halt the spread of emerging variants. Notably, both B.1.1.7 and P.1 spread across Italy
169 before international border closers were imposed for the UK and Brazil. This highlights the importance of a
170 rapid and inclusive vaccine roll-out on a global scale.

171

172 Methods

173 **Sequence data.** All Italian sequences of SARS-CoV-2 available on the GISAID EpiCov™ database were
174 downloaded (8th March 2021, acknowledgment Table S1). Sequences without a complete date of
175 collection were removed. As per our previous study, cases were grouped according to geographical non-
176 administrative macro areas in North (n=1,195), Central (n= 1,403), and South Italy (n= 3,917)¹⁶. SARS-CoV-
177 2 lineages were assigned to each sequence using the Pangolin COVID-19 Lineage Assigner tool v2.0.7
178 (github.com/cov-lineages/pangolin). All available global sequence data for B.1.351 and P.1 was also
179 downloaded from the GISAID EpiCov™ database. Due to the large amount of data available for B.1.1.7
180 (>170,000), a random subset of 3,000 global sequence data was used in combination with all available
181 Italian sequence data (this was extracted from the global SARS-CoV-2 tree²³).

182

183 **Phylogenetic analysis.** Separate nucleotide sequence alignments were constructed for the B.1.1.7, P.1,
184 and B.1.351 lineages. A representative sequence for each lineage was used as a reference and as an
185 outgroup for tree rooting (B.1.1.7=EPI_ISL_862239, B.1.351=EPI_ISL_660190, P.1=EPI_ISL_833137). Three
186 separate alignments were performed using MAFFT implementing the L-INS-I algorithm and manually
187 inspected for accuracy using Geneious Prime® 2021.1.1 (<https://www.geneious.com>)²⁴. Full genome
188 sequences with >5% ambiguity and no exact sampling date were removed. The final data sets comprised:
189 P.1 n=614 (Italy n=111), B.1.351 n=2,317 (Italy n=8), and B.1.1.7 n=4,361 (Italy n=1,461). A maximum

Article

190 likelihood tree was estimated for each lineage using IQ-TREE implementing the following: Hasegawa-
191 Kishino-Yano nucleotide substitution model with a gamma distributed rate variation among sites (HKY+ Γ)
192 and an ultrafast bootstrap method (1000 repetitions)^{25,26}.

193

194 **References**

195

196 1 WHO. Weekly operational update on COVID-19 - 8 March 2021. (2021).

197 2 WHO. Weekly epidemiological update - 31 August 2020. (2020).

198 3 Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic
199 epidemiology. *Nat. Microbiol.* **5**, 1403-1407 (2020). doi:10.1038/s41564-020-0770-5

200 4 CDC. Science brief: emerging SARS-CoV-2 variants, [https://www.cdc.gov/coronavirus/2019-](https://www.cdc.gov/coronavirus/2019-ncov/more/science-and-research/scientific-brief-emerging-variants.html)
201 [ncov/more/science-and-research/scientific-brief-emerging-variants.html](https://www.cdc.gov/coronavirus/2019-ncov/more/science-and-research/scientific-brief-emerging-variants.html). Accessed 1 Mar 2021

202 5 Davies, N. G. *et al.* Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in
203 England. *Science* (2021). doi:10.1126/science.abg3055

204 6 Tegally, H. *et al.* Emergence and rapid spread of a new severe acute respiratory syndrome-related
205 coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv*
206 (2020). doi:10.1101/2020.12.21.20248640

207 7 O'Toole, A. *et al.* Tracking the international spread of SARS-CoV-2 lineages B.1.1.7 and
208 B.1.351/501Y-V2, [https://virological.org/t/tracking-the-international-spread-of-sars-cov-2-](https://virological.org/t/tracking-the-international-spread-of-sars-cov-2-lineages-b-1-1-7-and-b-1-351-501y-v2/592)
209 [lineages-b-1-1-7-and-b-1-351-501y-v2/592](https://virological.org/t/tracking-the-international-spread-of-sars-cov-2-lineages-b-1-1-7-and-b-1-351-501y-v2/592). Accessed 1 Mar 2021

210 8 Walensky, R. P., Walke, H. T. & Fauci, A. S. SARS-CoV-2 variants of concern in the United States-
211 challenges and opportunities. *JAMA* (2021). doi:10.1001/jama.2021.2294

212 9 Sabino, E. C. *et al.* Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *Lancet*
213 **397**, 452-455 (2021). doi:10.1016/S0140-6736(21)00183-5

214 10 Resende, P. C. *et al.* Spike E484K mutation in the first SARS-CoV-2 reinfection case confirmed in
215 Brazil, 2020, [https://virological.org/t/spike-e484k-mutation-in-the-first-sars-cov-2-reinfection-](https://virological.org/t/spike-e484k-mutation-in-the-first-sars-cov-2-reinfection-case-confirmed-in-brazil-2020/584)
216 [case-confirmed-in-brazil-2020/584](https://virological.org/t/spike-e484k-mutation-in-the-first-sars-cov-2-reinfection-case-confirmed-in-brazil-2020/584). Accessed 1 Mar 2021

Article

- 217 11 Naveca, F. *et al.* SARS-CoV-2 reinfection by the new Variant of Concern (VOC) P.1 in Amazonas,
218 Brazil, [https://virological.org/t/sars-cov-2-reinfection-by-the-new-variant-of-concern-voc-p-1-in-](https://virological.org/t/sars-cov-2-reinfection-by-the-new-variant-of-concern-voc-p-1-in-amazonas-brazil/596)
219 [amazonas-brazil/596](https://virological.org/t/sars-cov-2-reinfection-by-the-new-variant-of-concern-voc-p-1-in-amazonas-brazil/596). Accessed 1 Mar 2021
- 220 12 Wibmer, C. K. *et al.* SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor
221 plasma. (2021). doi:10.1101/2021.01.18.427166
- 222 13 Williams, T. C. & Burgers, W. A. SARS-CoV-2 evolution and vaccines: cause for concern? *Lancet*
223 *Respir. Med.* (2021). doi:10.1016/S2213-2600(21)00075-8
- 224 14 Xie, X. *et al.* Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by
225 BNT162b2 vaccine-elicited sera. *Nat. Med.* (2021). doi:10.1038/s41591-021-01270-4
- 226 15 Statistiche coronavirus. Statistiche coronavirus in Italia,
227 <https://statistichecoronavirus.it/coronavirus-italia/>. Accessed 8 Mar 2021
- 228 16 Di Giallonardo, F. *et al.* Genomic epidemiology of the first wave of SARS-CoV-2 in Italy. *Viruses* **12**
229 (2020). doi:10.3390/v12121438
- 230 17 Schlosser, F. *et al.* COVID-19 lockdown induces disease-mitigating structural changes in mobility
231 networks. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 32883-32890 (2020). doi:10.1073/pnas.2012326117
- 232 18 Swadi, T. *et al.* Genomic evidence of in-flight transmission of SARS-CoV-2 despite predeparture
233 testing. *Emerg. Infect. Dis.* **27**, 687-693 (2021). doi:10.3201/eid2703.204714
- 234 19 Delli Compagni, E. *et al.* Genome sequence of three SARS-CoV-2 P.1 strains identified from
235 patients returning from Brazil to Italy. *Microbiol. Resour. Announc.* (2021). doi:accepted for
236 publication
- 237 20 Danish Covid-19 Genome Consortium. Genomic overview of SARS-CoV-2 in Denmark,
238 <https://www.covid19genomics.dk/statistics>. Accessed 15 Mar 2021
- 239 21 Buss, L. F. *et al.* Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely
240 unmitigated epidemic. *Science* **371**, 288-292 (2021). doi:10.1126/science.abeg728
- 241 22 Wise, J. Covid-19: The E484K mutation and the risks it poses. *BMJ* **372**, n359 (2021).
242 doi:10.1136/bmj.n359

Article

- 243 23 Lanfear, R. A global phylogeny of hCoV-19 sequences from GISAID, <https://www.gisaid.org/>.
- 244 Accessed 10 Mar 2020
- 245 24 Kuraku, S., Zmasek, C. M., Nishimura, O. & Katoh, K. aLeaves facilitates on-demand exploration of
- 246 metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity.
- 247 *Nucleic Acids Res.* **41**, W22-28 (2013). doi:10.1093/nar/gkt389
- 248 25 Trifinopoulos, J., Nguyen, L. T., von Haeseler, A. & Minh, B. Q. W-IQ-TREE: a fast online
- 249 phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **44**, W232-235 (2016).
- 250 doi:10.1093/nar/gkw256
- 251 26 Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective
- 252 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268-274
- 253 (2015). doi:10.1093/molbev/msu300
- 254
- 255
- 256 **Author Contributions:** The project was conceptualized by F.D.G, E.C.H. and A.L. The manuscript was
- 257 written by F.D.G, E.C.H. and A.L. Sequence data was generated by I.P., V.C., M.M., P.C., A.L. Analyses
- 258 were performed by F.D.G, J.C.A.C. Results were visualized by F.D.G.
- 259
- 260 **Funding:** AL was funded by the Italian Ministry of Health IZS AM 08/19 Ricerca Corrente 2019, NGS e
- 261 diagnostica molecolare in Sanità Animale: Fast D2" and IZSAM 05/20 Ricerca Corrente 2020 "PanCO:
- 262 epidemiologia e patogenesi dei coronavirus umani ed animali". ECH was funded by an Australian Research
- 263 Council Australian Laureate Fellowship (FL17010002).
- 264
- 265 **Acknowledgments:** We are thankful to all individuals who submit data to GISAID. We thank the Pango
- 266 team for their efforts in maintaining a freely available tool for lineage classification. The authors would like
- 267 to acknowledge all doctors, nurses, technicians, medical staff, administrators, food and cleaning service
- 268 workers, pharmacists, and all other members of the COVID-19 diagnostic group at IZSAM.
- 269

Article

270 **Corresponding authors.** Correspondence to F Di Giallonardo (fdigiallonardo@kirby.unsw.edu.au)

271

272 **Conflicts of Interest:** The authors declare no conflict of interest.

273

274 **Supplementary Information:** Fig. S1: Changing frequency of SARS-CoV-2 lineages sampled over time in
275 Italy. Fig. S2: Amino acid alignment of the SARS-CoV-2 spike protein. Fig. S3: Molecular epidemiology of
276 global SARS-CoV-2 lineage B.1.351. Table S1: GISAID EpiCov™ acknowledgment.

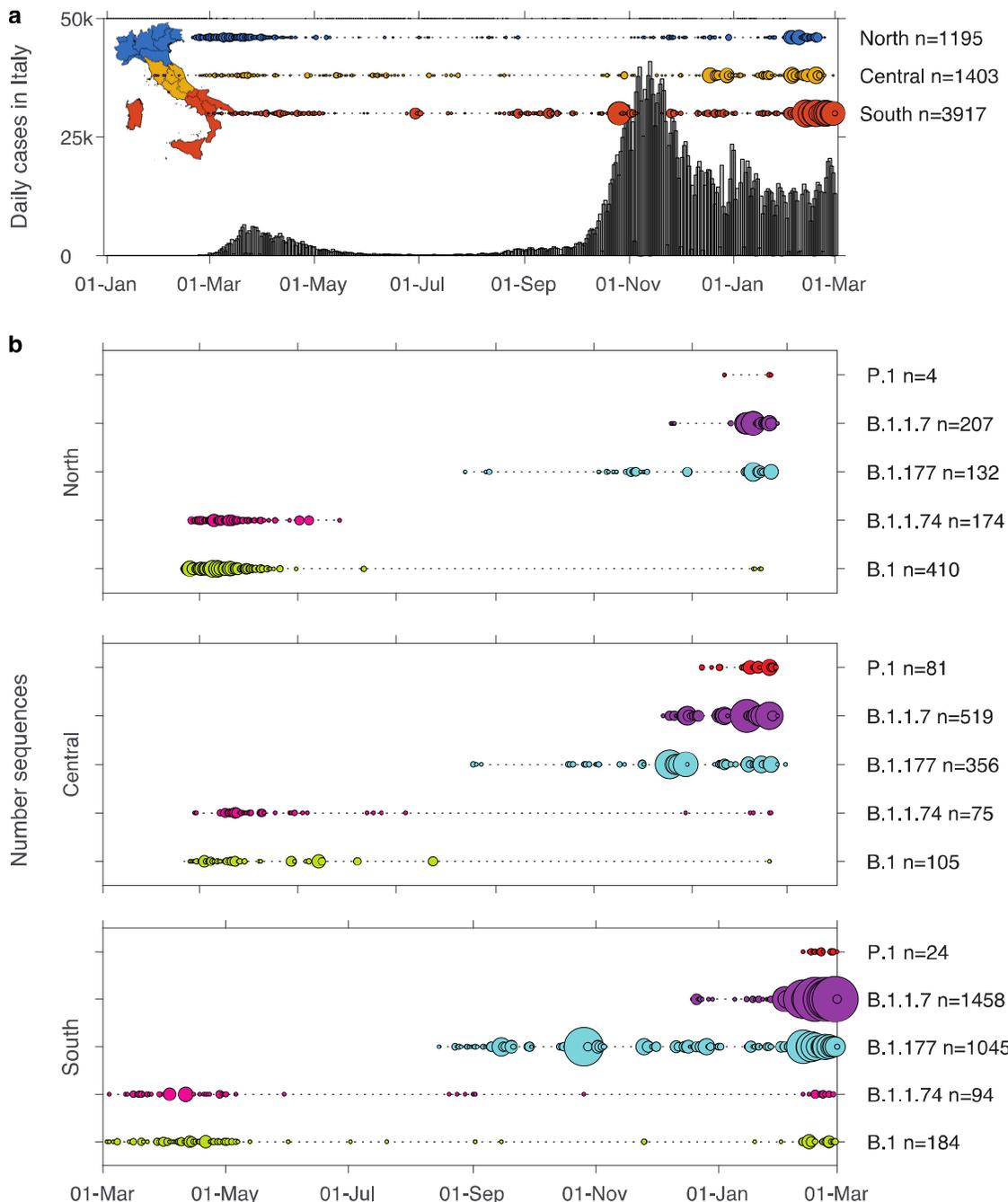
277

278 **Figures**

279

Article

280 **Fig. 1. Number of SARS-CoV-2 genomes sampled over time in Italy.**



281

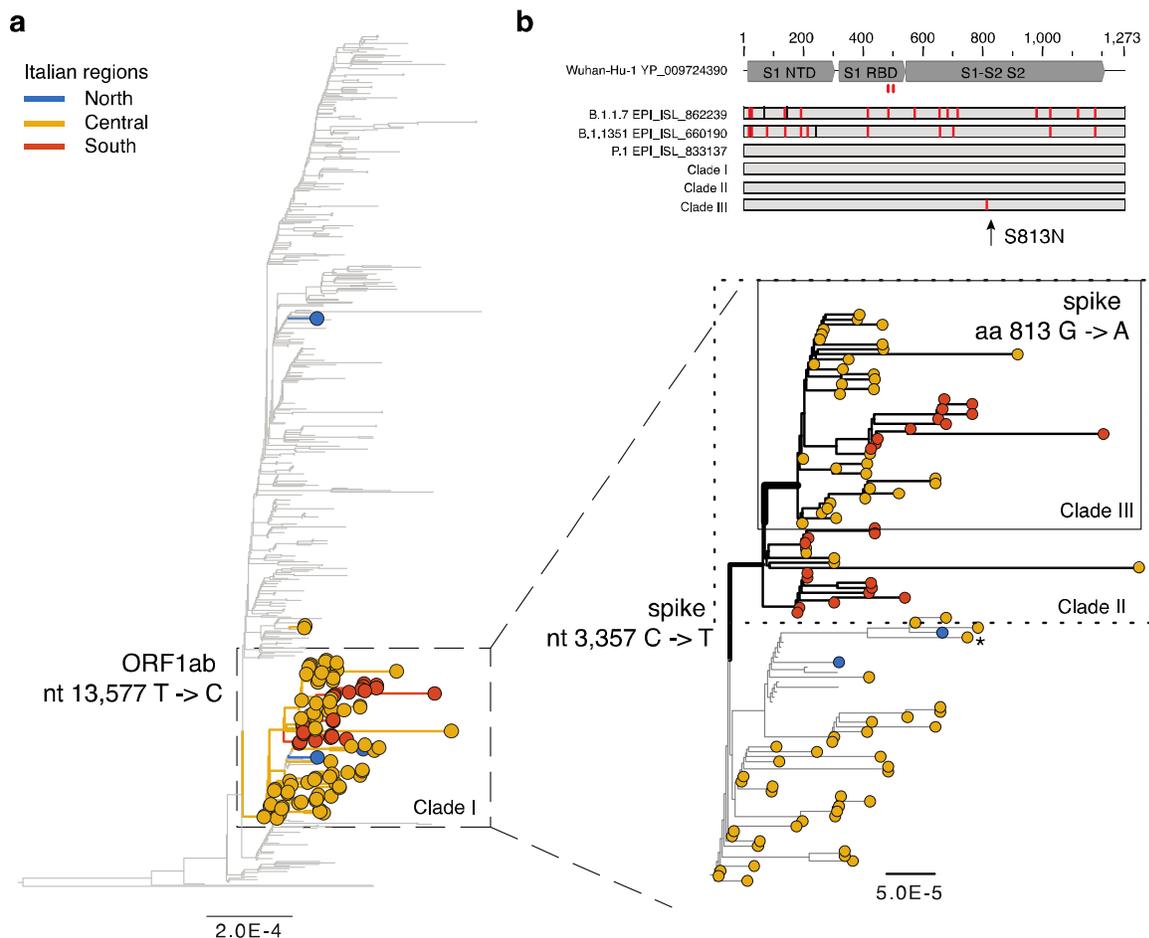
282 **a** Bars show the number of cases reported in Italy between 1st January 2020 – 1st March 2021. Horizontally
 283 aligned circles show the number of sequences sampled for each day in North (blue), Central (yellow),
 284 South (red) Italy. The size of the circle is equivalent to the total number of sequences (max = 256). Italian
 285 regions are coloured according to their macro areas: blue = North (Valle d’Aosta, Piemonte, Liguria,
 286 Lombardia, Emilia-Romagna, Veneto, Friuli-Venezia-Giulia, and Trentino-Alto Adige), yellow = Central
 287 (Lazio, Marche, Toscana, Umbria, and Abruzzo), red = South (Puglia, Basilicata, Calabria, Campania,
 288 Molise, Sicilia, and Sardegna). **b** Number of sequences sampled in Italy for the lineages B.1. B.1.1.74,
 289 B.1.177, B.1.1.7, and P.1 for North (top), Central (middle), and South (bottom) Italy. The total number of
 290 sequences per day (circles) and the time range between first and most recent sequence (dotted line) is
 291 shown (max = 146).

292

Article

305 **Fig. 3. Genomic epidemiology of SARS-CoV-2 lineage P.1.**

306



307

308 **a** Maximum likelihood phylogeny of SARS-CoV-2 full genome sequences (n=503 global, n=111 Italian).
 309 Lineage B.1.351 was used as an outgroup (EPI_ISL_660190). Branch lengths are scaled according to the
 310 number of nucleotide substitutions per site and are coloured according to geographic macro area: grey =
 311 global, blue = North Italy, yellow = Central Italy, red = South Italy. The node containing an T to C mutation
 312 in ORF1ab is marked with a box. **b (bottom)** Enlargement of this node and additional clades and their
 313 nucleotide or amino acid substitutions are indicated. Note, a single sequence containing also the S813N
 314 mutation but falling outside the main clade is marked with an asterisk. Branches without a tip circle
 315 represent sequences from Germany. **(top)** Schematic view of the spike protein alignment for different
 316 reference sequences as well as the clade containing the S813N substitution. The protein structural regions
 317 are marked: S1 NTD = N-terminal domain of the S1 subunit, S1 RBD = receptor binding domain of S1, S1-
 318 S2 S2 = S1/S2 cleavage region and S2 fusion subunit. A detailed view of the protein alignment can be
 319 found in Supplementary Data S2.