

1 **SARS-CoV-2 phylogeny during the early outbreak in the Basel area, Switzerland: import and spread dominated by a**
2 **single B.1 lineage variant (C15324T)**

3
4 Madlen Stange^{1,2,3*}, Alfredo Mari^{1,2,3*}, Tim Roloff^{1,2,3*}, Helena MB Seth-Smith^{1,2,3*}, Michael Schweitzer^{1,2},
5 Myrta Brunner⁴, Karoline Leuzinger^{5,6}, Kirstine K. Søgaaard^{1,2}, Alexander Gensch¹, Sarah Tschudin-Sutter⁷,
6 Simon Fuchs⁸, Julia Bielicki⁹, Hans Pargger¹⁰, Martin Siegemund¹⁰, Christian H Nickel¹¹, Roland Bingisser¹¹,
7 Michael Osthoff¹², Stefano Bassetti¹², Rita Schneider-Sliwa⁴, Manuel Battegay⁷, Hans H Hirsch^{5,6,7}, Adrian
8 Egli^{1,2,+}

9
10 ¹ Applied Microbiology Research, Department of Biomedicine, University of Basel, Basel, Switzerland

11 ² Clinical Bacteriology and Mycology, University Hospital Basel, Basel, Switzerland

12 ³ Swiss Institute for Bioinformatics, Basel, Switzerland

13 ⁴ Human Geography, University of Basel, Basel, Switzerland

14 ⁵ Clinical Virology, University Hospital Basel, Basel, Switzerland

15 ⁶ Transplantation & Clinical Virology, Department of Biomedicine, University of Basel, Basel, Switzerland

16 ⁷ Infectious Diseases and Hospital Epidemiology, University Hospital Basel and University of Basel, Basel,
17 Switzerland

18 ⁸ Health Services for the City of Basel, Basel, Switzerland

19 ⁹ Pediatric Infectious Diseases, University Children's Hospital Basel, Basel, Switzerland

20 ¹⁰ Intensive Care Unit, University Hospital Basel, Basel, Switzerland

21 ¹¹ Emergency Medicine, University Hospital Basel, Basel, Switzerland

22 ¹² Internal Medicine, University Hospital Basel, Basel, Switzerland

23 *these four authors contributed equally to this work

24

25 + correspondence

26 Adrian Egli, MD PhD

27 University Hospital Basel

28 Petersgraben 4

29 4031 Basel, Switzerland

30 Email: adrian.egli@usb.ch

31 Phone: +41 61 556 5749

32

33 **Abstract**

34 **Background.** The first case of SARS-CoV-2 in Basel, Switzerland, was detected on February 26th 2020. We
35 present a phylogenetic longitudinal study and explore viral introduction and evolution during the exponential
36 early phase of the local COVID-19 outbreak from February 26th until March 23rd.

37 **Methods.** We sequenced SARS-CoV-2 from naso-oropharyngeal swabs, generated 468 high quality genomes,
38 and called variants with our COVID-19 Pipeline (COVGAP). We analysed viral genetic diversity using
39 PANGOLIN taxonomic lineages. To identify introduction and dissemination events we incorporated global
40 SARS-CoV-2 genomes and inferred a time-calibrated phylogeny.

41 **Findings.** The early outbreak in Basel was dominated by lineage B.1 (83.6%), detected from March 2nd,
42 although the first lineage identified was B.1.1. Within B.1, a clade containing 68.2% of our samples, defined by
43 the SNP C15324T, suggests local spreading events. We infer the geographic origin of this mutation to our tri-
44 national region. The remaining genomes map broadly over the global phylogenetic tree, evidencing several
45 events of introduction from and/or dissemination to other regions of the world. We also observe family
46 transmission events.

47 **Interpretation.** A single lineage dominated the outbreak in the City of Basel while other lineages such as the
48 first (B1.1) did not propagate. Thus spreading events seem to have contributed most to viral spread, while travel
49 returners and family transmissions were better controlled by the recommended measures. This phylogenetic
50 analysis enriches epidemiological and contact tracing data, allowing connection of seemingly unconnected
51 events, and can inform public health interventions.

52 **Funding.** No dedicated funding was used for this work.

53

54 Introduction

55 The COVID-19 pandemic has rapidly spread around the globe during the first six months of 2020. The
56 causative coronavirus, SARS-CoV-2, is the subject of many studies using genomic analysis providing key
57 insights into viral diversity across cities¹, provinces²⁻⁵, countries⁶⁻¹¹, and globally¹². SARS-CoV-2 has an
58 estimated mutation rate of $0.71-1.40 \times 10^{-3}$ ¹³, which translates to 21-42 mutations per year. Due to the
59 accumulation of mutations, phylogenetic analysis of SARS-CoV-2 is becoming more granular over time¹⁴,
60 providing increasing resolution of transmission dynamics and events. Comparisons of single nucleotide
61 polymorphisms (SNPs) allows us to explore transmission events with highest resolution across communities.
62 The identification of transmission routes is important, especially with various public health measures being
63 introduced, such as lockdown policies, which have been implemented on country or regional levels to limit viral
64 transmission. The impact of public health measures can be monitored through phylogenies³. Genomic data can
65 also deliver insights into mutations and whether they alter virulence, or aid adaptation to novel hosts. The spike
66 protein D614G mutation, for example, has been implicated in more effective transmission¹⁵, although the actual
67 impact may be through fixation in an expanding lineage rather than conferring increased transmissibility per
68 se¹⁶.

69 The scope of this study is to provide a more granular picture of the phylogenetic diversification and propagation
70 of the early-stage SARS-CoV-2 pandemic on a local scale. The City of Basel has a population of 175,350
71 inhabitants (median over the past five years) with half a million people in the Basel area. Situated in North-
72 Western Switzerland, directly bordering both Germany and France, Basel has almost 34,000 workers
73 commuting daily across the international borders¹⁷. Given this large exchange of people in this tri-national
74 region, the fact that the neighbouring region Alsace, France, was already experiencing an intense epidemic¹⁸,
75 and a low threshold testing strategy implemented weeks before the first case, we aim to explore the early stage
76 of SARS-CoV-2 transmission dynamics in Basel and the surrounding area from the first case to one week post
77 border closure.

78 Materials and Methods

79 Characteristics of the longitudinal study

80 This cohort study includes patient samples from Basel-City and the surrounding area during the initial 26 days
81 of the local outbreak, between February 26th and March 23rd. Only single, non-repeated tests per patient were
82 considered eligible for phylogenetic analysis. This timeframe covers the first two positively tested cases in Basel
83 on February 26th, which we were able to capture via early implementation of PCR-based detection by routine
84 diagnostics, until the date of border closure plus seven days (March 23rd).

85 From the first case on February 26th 2020 until March 23rd we had performed 6,943 PCR tests. Of these, 746
86 samples (10.7%) were SARS-CoV-2 positive (**Figure 1**). March 23rd had the maximum number of positive tests,
87 with 66 cases. Of all PCR tested patients and positively tested patients during the study period, a majority were
88 female with a median age of 42 and 49 years, respectively (**Table 1**). Of the PCR-confirmed cases, only 17
89 (2.3%) were in patients younger than 18 years. (**Table 1, Figure S1**). 418 (56%) were living in the canton of
90 Basel-City (City of Basel, Riehen, Bettingen) and 328 (44%) were from the surrounding area.

91 Whole genome sequencing, assembly, and variant calling

92 We included potentially all SARS-CoV-2 positive isolates in the subsequent molecular epidemiological study.
93 SARS-CoV-2 genomes were amplified following the amplicon sequencing strategy of the ARTIC protocol
94 (<https://artic.network/ncov-2019>) with V.1 or V.3 primers¹⁹ and 150 nucleotides paired-end sequenced, on an
95 Illumina platform. We developed a new bioinformatic pipeline, COVID-19 Genome Analysis Pipeline
96 (COVGAP), to generate high-quality genomes and call genomic variants from our Illumina raw sequencing
97 data. This pipeline maps quality filtered reads to the reference genome Wuhan-Hu-1 (MN908947.3), calls
98 variants and retains high quality genomes (< 10% ambiguous nucleotides, > 50x read coverage) (**Figure S2**),
99 generally reflecting those from high viral load samples (low cycle threshold (Ct) values; **Figure S3**).

100 In order to validate COVGAP, 15 mock mutated genomes were generated, featuring in total 23 single nucleotide
101 polymorphisms (SNPs), five deletions and two insertions compared to the reference Wuhan-Hu-1. Database
102 sample SARS-CoV-2/human/USA/AZ-ASU2923/2020 (MT339040.1) with seven SNPs and an 81-nucleotide
103 deletion was also included. COVGAP analysed artificially generated reads, correctly identifying all 30 SNPs,
104 four of the six deletions, including the 81-nucleotide deletion, and one of the two insertions.

105 **SARS-CoV-2 phylogenetic analyses**

106 COVGAP yielded 468 unique-patient, high-quality SARS-CoV-2 genomes from the Basel area. Low-quality
107 genomes were excluded. High-quality and full-length consensus sequences from global viruses were
108 downloaded from GISAID^{20,21} on June 22nd, 2020. 30 genomes per country and month, totalling 2,017 genomes,
109 were retained for joint analysis with our cohort sequences, which were performed using custom R-scripts,
110 PANGOLIN¹⁴ and the nextstrain command line interface analysis pipeline v.2.0.0 (nextstrain.org) and augur
111 v.8.0.0²².
112 More details on sample processing, genome generation, variant calling, and phylogenetic analyses can be found
113 in **Supplementary material**.

114 **Data accessibility**

115 Sequencing data (viral reads only) was submitted to European Nucleotide Archive (ENA) under accession
116 number PRJEB39887, consensus sequences were submitted to GISAID, bioinformatic pipelines are accessible
117 on Github.

118 **Ethics**

119 The study was conducted according to good laboratory practice and in accordance with the Declaration of
120 Helsinki and national and institutional standards and was approved by the ethical committee (EKNZ 2020-
121 00769). The clinical trial accession number is NCT04351503 (clinicaltrials.gov).
122

123 **Results**

124 **COVGAP pipeline validation**

125 No false positive variants were called (**Figure S4, Table 2, Table S1**). Ambiguous mapping was responsible for
126 the failure to call the three indels, resulting in insufficient (<70%) coverage to be reliably called as a variant.
127 This validation allowed us to determine the specificity (100%), sensitivity (94.2%), and accuracy (100%) of
128 COVGAP, thus confirming its accuracy in SNP detection, as well as its ability to call the majority of indels
129 from short read data.

130

131 The COVGAP pipeline produced 468 (63%) high quality genomes for subsequent analysis, of the 746 samples
132 taken. The remaining samples were either not available (N = 57), did not pass sequencing quality control (N =
133 156), or were duplicates from the same patient (N = 65). These 468 samples are subsequently referred to as the
134 Basel area cohort. Of these, 240 (51.9%) were from female patients (**Table 1, Figure S1**), and 12 (2.6%) were
135 from patients younger than 18 years.

136 **Lineages observed over time in the Basel area cohort**

137 Over the 26-day study period, 13 out of 91 globally circulating phylogenetic lineages were recorded in the Basel
138 area; only one additional lineage was recorded in all Swiss sequences.

139 Lineage B.1 dominated the cases during the initial phase of the outbreak (**Figure 2**), with 83.6% (N = 391) of
140 sequenced samples (**Table 3**), being recorded for the first time in Basel on March 2nd. The first patient
141 diagnosed at our hospital, on February 26th, had a virus belonging to lineage B.1.1. This lineage is seen
142 sporadically through the outbreak with a maximum of six sequenced cases from March 16th. Lineage B.1 is
143 associated with the Italian outbreak¹⁴, yet both B.1.1 (35.7%) and B.1 (51.0%) were the most prevalent lineages
144 in Italy during this time span (**Figure 3**). From March 13th, rarer lineages are seen in the Basel area, such as
145 B.1.1.6, a lineage that is associated with an Austrian origin¹⁴. Only two cases from the A lineage and sub-
146 lineages were sequenced in the Basel area cohort (**Table 3**).

147 **Lineage diversity in Basel-City, Switzerland, and neighbouring countries**

148 Comparison with lineages found in neighbouring countries over this period shows that lineage B.1 dominates in
149 all, but the numbers of lineages identified, and the proportions vary (**Figure 3**). Switzerland (77.1%) has a
150 similarly large proportion of B.1 lineage to France (90.0%). We describe the viral diversity based on abundance
151 of lineages (retrieved from GISAID, for details see **Supplementary material**) using a range of diversity indices
152 (**Table S5**). Simpson diversity, which accounts for differences in sample abundance between countries, was
153 highest in Germany (3.87) and Austria (3.73), followed by Italy (2.52); it was lowest in Switzerland (1.62) and
154 France (1.23). The share of our samples that originates from Basel-City residents (376 samples) excluding
155 sequences that were obtained from commuters mirrors the lineage proportions of Switzerland (**Figure 3**), while
156 contributing 56% of all available sequencing data for Switzerland within this timeframe (GISAID database as of
157 June 22nd,^{20,21}).

158 **Basel samples in global phylogenetic context**

159 In order to better contextualize our findings, we analysed our virus genomes phylogenetically with a subset of
160 global publicly available sequences (see **Supplementary material; Figure 4**). While phylogenetic lineages may
161 show some geographical signal, lineages do not exclusively correspond to continents (**Figure 4A**), illustrating
162 the degree of global interconnectivity and speed of spreading. The phylogenetic lineages recorded in the Basel
163 area are distributed across the global phylogenetic tree (**Figure 4B**). Mismatch of 'taxonomic' assignment and
164 phylogenetic origin of genomes assigned to B.1 can be seen, with several as yet unnamed sub-lineages apparent
165 in the phylogeny (**Figure 4B**).

166 We can identify a major clade, within lineage B.1, comprising 68.2% of our samples (319/468 samples with 264
167 (82.8%) from patients from cantons Basel-City and Basel-Landschaft; **Figure 5A**). The remaining Basel area
168 sequences (31.8%) (**Figure 5 B-C, Figure S6 B-C**) are spread throughout the phylogeny and cluster with global
169 genomes. Introductions and features of some of the clades are analysed in the following sections and in
170 **Supplementary material**.

171 **The first identified introduction of SARS-CoV-2 to Basel**

172 The first two positively diagnosed patients with SARS-CoV-2 in Basel and first identified cases of COVID-19,
173 *Patient 1* and *Patient 2*, travelled together to Italy. In our analysis, both *Patient 1* and *Patient 2* carried viruses
174 from the B.1.1 lineage, the second most prevalent lineage in Italy at that time (**Figure 3**). Interestingly, the two
175 virus genomes from *Patient 1* and *Patient 2* are separated by two SNPs, suggesting two independent infections
176 (**Figure 5B**). Moreover, we did not identify minor alleles within these samples that would hint at double
177 infection of the patients. In this case, the epidemiological cluster of *Patient 1* and *Patient 2* is not congruent with
178 the phylogenetic inference.

179
180 The virus genome of *Patient 1* carried a synonymous mutation at C313T in *ORF1ab*, which is found in samples
181 from Israel, Hungary, Japan, USA, Argentina, Greece, India, Brazil, Morocco, and Netherlands among others
182 (nextstrain.org), all sharing an unsampled common ancestor that emerged around February 25th (CI February 23-
183 26th). This SNP is also found in ten other Basel area cohort genomes: eight of these were from a family and
184 social friends cluster unrelated to *Patient 1* that were diagnosed between March 13th and March 22nd.

185
186 The virus genome of *Patient 2* carried a synonymous mutation at T19839C in *ORF1ab* and not C313T, and
187 clusters together with eight identical virus samples sampled between February 26th and March 23rd. Two of
188 these are from family members of *Patient 2*, who tested positive two and six days later, suggesting a family
189 route of infection. The epidemiological data of the other six patients suggests no direct transmission via *Patient*
190 *2*. One of the six returned from a Swiss ski resort three days prior to onset of symptoms. Two additional samples
191 forming a family cluster (*Family 1*, **Figure 5B**) diagnosed on March 3rd, carried the T19839C plus non-
192 synonymous mutation G28179A leading to amino acid change ORF8-G96S. One member of *Family 1* travelled
193 with *Patients 1* and *2* to Italy and possibly got infected there with a yet different virus variant.

194 **Introduction of the Basel cluster**

195 The clade within lineage B.1, into which 68.2% (N = 319) of our Basel area cohort sequences fall, is
196 characterized by a synonymous SNP C15324T in *ORF1ab*, henceforth referred to as the “Basel cluster”. *Patient*
197 *1* and *Patient 2* are not linked to this Basel cluster. The first sample within the Basel cluster, from March 2nd,
198 was from a patient residing in Central Switzerland, who was transferred to a care facility in Basel where six
199 further people tested positive between March 5th and March 17th with identical viral genomes. The source of the
200 first infection is unknown. The second sample within the Basel cluster was from a patient diagnosed on March
201 3rd, who had attended a religious event in Alsace, France, that took place between February 17th and 21st. One
202 other patient, who tested positive on March 9th and had symptoms 14 days prior to testing, also attended this
203 event.

204 **Description of the Basel cluster**

205 The 319 genomes in the Basel cluster (**Figure 5A**) show a divergence of up to five SNPs up until March 23rd,
206 although 157 genomes are identical and located at the root of the clade. This infers that the common ancestor
207 originated between February 9th and February 17th. That the clade defining SNP C15324T (GISAID emerging
208 clades label 20A/15324T) was registered globally for the first time on March 2nd simultaneously in the Basel
209 area sample 42173111 and GISAID sample Germany/FrankfurtFFM7/2020 suggests unsampled circulation of
210 this variant from mid-February. By searching all genomes available on GISAID (N = 80,189; as of August 12th,
211 2020), filtering for genomes belonging emerging clade 20A/15324 and sampled until March 23rd (N = 2,856),
212 we find that subsequently the C15324T mutation has also been observed in other countries but remains most
213 prevalent in Switzerland (N_{GISAID} = 57/213, 26.8%; N_{GISAID+this study} = 386/675, 57.2%; first genome 42173111
214 from March 2nd) (**Table S3**). Of the 57 Swiss genomes with this mutation that were already deposited on
215 GISAID, 26 also originate from the cantons Basel-City and Basel-Landschaft (**Table S4**). The mutation is also
216 found in higher proportions in genomes from France (N = 69/369, 18.7%, first from March 3rd sample
217 France/HF1870/2020), Luxembourg (N = 24/116, 20.7%, from March 8th sample
218 Luxembourg/LNS2614631/2020), and Belgium (N = 40/268, 14.9%, from March 6th sample Belgium/NKR-
219 030645/2020), but date later than the first recorded occurrence from Switzerland and Germany (**Table S3**).
220 Subsequently, as of March 9th, descendants of this variant were recorded outside Europe (**Table S4**) in smaller
221 proportions than in Basel (**Table S3**), which suggests dissemination from the Basel area.

222 **Potential ski-holiday related cluster (C1059T)**

223 Previous reports have identified viruses in lineage B.1 carrying SNP C1059T (amino acid change ORF1a-
224 T265I) in travel returners from ski holidays in Ischgl, Austria^{9,23}. We detected 21 (4.5%) viral samples within
225 our cohort with these features (**Figure 5C**), divergent by up to two SNPs from the common ancestor. Samples
226 date from March 11th to March 23rd with an inferred internal node age of February 21st (CI: February 20th-
227 February 21st, 2020), fitting with possible infections in Ischgl from end of February to March 13th²⁴.
228 Epidemiological data confirms that five patients from that cluster had returned from Austria, and three
229 specifically from Ischgl, the fourth from Tyrol, before testing positive. Two additional patients returned from
230 skiing in Swiss ski resorts.

231 **Spike protein mutation prevalent in Basel patients**

232 The spike protein S-D614G mutation is associated with the B.1 lineage and all those derived from this (**Figure**
233 **4**). As such, it occurs in 448 of the 468 (95.7%) samples from the Basel area. The SNP responsible has not been
234 lost once in our sub-sampled dataset, but it is not present in B.2 or B.3 or other sister lineages to B.1 (**Table 3**).
235 Among our samples, we found no significant difference in viral loads between patients with and without the S-
236 D614G mutation ($z = -0.881$, $p = 0.38$). However, our cohort is biased to samples with higher viral load, as
237 these were those that were successfully sequenced.

238 **Discussion**

239 We reconstruct the early events focusing on introduction and spread of SARS-CoV-2 in Basel and the
240 surrounding area, Switzerland, from a phylogenetic perspective. We present COVGAP, a new combination of
241 existing tools to effectively and efficiently mine SARS-CoV-2 genomes from Illumina paired end reads. Unlike
242 other currently available tools²⁵, COVGAP shows higher sensitivity levels in SNP calling from raw reads
243 (100%), failing only in ambiguously mapped deletions and insertions. In such cases, it adopts a coverage-
244 conservative approach, needed to reliably call variants in real world scenarios.
245

246 The majority of genome variants in Basel are similar to those from France, Italy, and Germany. We found the
247 presence of 13 SARS-CoV-2 lineages in our samples, with the beginning of the Basel outbreak being powered
248 by the European B.1 lineage. In particular, a B.1 lineage variant with the C15324T mutation dominated the early
249 phase of the local spread with 70% of samples forming a large Basel cluster. Compared to Victoria, Australia⁵,
250 the UK²⁶, or Austria (this study) the diversity seen arriving in Switzerland and Basel as determined using
251 Simpson diversity is more limited, reflecting European rather than intercontinental connections. This diversity
252 measure can be used to monitor viral introductions as an effect of travel restrictions in the future.
253

254 The Basel cluster virus variant 20A/C15324T was first detected in Europe on March 2nd in Germany and
255 Switzerland simultaneously. We locate its geographic origin to our tri-national region between February 9th and
256 February 17th. Our epidemiologically-informed phylogenetic analysis indicates that the Basel cluster represents
257 a larger transmission chain that was unchecked and spread effectively among unrelated people throughout Basel
258 and eventually outside of Europe. The first recognized case in the large Basel cluster goes back to a patient in a
259 care facility, in which several more infections occurred.
260

261 The beginning of the COVID-19 outbreak described here coincided with the winter school holidays in Basel,
262 from February 22nd to March 8th. During this time, many residents take the opportunity to travel, in particular to
263 skiing resorts. Viral introductions from ski resorts are known from contract tracing data to have affected
264 Germany²⁷, Denmark²³, Iceland⁹, France, Spain, and UK²⁸, with Ischgl, Austria being a described source of
265 many cases. Our data supports the finding that skiing resorts in Austria and Switzerland served as dissemination
266 hotspots. This school holiday also provided the opportunity for the first identified SARS-CoV-2 introductions to
267 Basel through two jointly returned travellers, notably each with different viral variants. Overall, however travel
268 returners did not drive the outbreak in Basel evidenced by the low diversity of variants and proportion of such
269 variants in our sample. A second likely source represents the many workers travelling daily across borders from
270 France and Germany, particularly from heavily affected areas such as Alsace¹⁸. As the B.1 and B.1.1 lineages
271 were dominant in France and Germany, these may be some sources of cases and transmissions.
272

273 The timing of the epidemic in Basel also coincided with three major events. Firstly, a religious event from
274 February 17th to 21st in Alsace that was described as a super-spreading event in France²⁹. We confirmed that the
275 virus genomes of two patients known to have attended are indeed situated at the root of the clade that constitutes

276 the Basel outbreak. Secondly, carnival in the Basel area is celebrated over several weeks from early January,
277 with numerous events, and thousands of active participants. The culmination is the UNESCO world-heritage
278 ‘Basler Fasnacht’, scheduled this year (2020) for March 2nd-4th, but cancelled due to COVID-19. Notably, the
279 active participants practice the piccolo and drums over weeks in closed rooms as a preparation for their
280 performance during the carnival, and unofficial events are likely to have taken place. Thirdly, Basel hosted three
281 international soccer events at the St. Jakob Stadium on February 15th (20,675 spectators), 23rd (20,265
282 spectators), and 27th (14,428 spectators). All three major events fell around the inferred date of origin of the
283 Basel mutation C15324T and subsequent dissemination phase.

284
285 A limitation of the current study is that we are likely to have missed some cases as not all symptomatic people
286 were advised to be tested, especially children younger than 18 years old. Nevertheless, our cohort represents a
287 very high sequencing density per detected case for a city (468 genomes from 746 PCR-confirmed cases in Basel
288 area (62.7%) and from 10,680 PCR-confirmed cases nationwide (4.4%)) for this early phase of the pandemic³⁰.

289
290 The availability and integration of epidemiological data in the interpretation of phylogenetic clades underlines
291 the validity of instrumentalising those tools for improving the understanding of SARS-CoV-2 outbreak
292 dynamics. Utilizing the clades as the backbone for targeted epidemiological analysis of specific cases helped in
293 grasping how mass gatherings, travel returners, and care facilities may influence an outbreak within a city. The
294 epidemiological data that was collected for the Federal Office of Public Health (FOPH), as requested by law,
295 helped tremendously to verify travel related links; however it was not designed to obtain data on local super-
296 spreading events such as attendance to soccer games, visiting clubs, restaurants, bars, and concerts and future
297 versions could be improved. The classical epidemiological context is very important to further explain
298 molecular epidemiological links especially in a still not very diversified virus.

299
300 In conclusion, the start of the outbreak of SARS-CoV-2 in the Basel area was characterized by a dominant
301 variant, C15324T, within the B.1 lineage, which we infer to have arisen in mid-February in our tri-national
302 region. Large gatherings (potential super spreading events) could have had profound effects on outbreak
303 dynamics. Improved surveillance measures are needed in the management of an outbreak, including large-scale,
304 active screening in the broader public, including more children to assess their role in transmission. Our analysis
305 shows the potential of molecular epidemiology to support classical contact tracing, even retrospectively, in order
306 to evaluate and improve measures to contain epidemics like COVID-19.

307 **Acknowledgements**

308 We thank Daniel Gander, Christine Kiessling, Magdalena Schneider, Elisabeth Schultheiss, Clarisse Straub, and
309 Rosa-Maria Vesco (University Hospital Basel) for excellent technical assistance with sequencing. Calculations
310 were performed at sciCORE (<http://scicore.unibas.ch/>) scientific computing center at University of Basel, the
311 support from the sciCORE team for the analysis is greatly appreciated. Support for the creation of schematic
312 figures (S2) was provided by BioRender.com. We thank all authors who have shared their genomic data on
313 GISAID. A full table outlining the originating and submitting labs is included as a supplementary file. No
314 dedicated funding was used for this work.

315 **Authors contributions**

316 AE and HH devised the project. KL and AG collected and prepared samples and associated data. MS performed
317 the phylogenetic analysis and interpretation, and led the writing and revising of the report. AM constructed the
318 COVGAP bioinformatic pipeline and released it on Github. TR and HSS prepared viral RNA for sequencing,
319 directed the phylogenetic analysis and deposited genomic data to GISAID and ENA. MSch and KKS collected
320 clinical and epidemiological data. MyB and RSS provided geographical expertise. JB, STS and SF provided
321 public health and epidemiological expertise. HP, MSi, CHN, RB, MO, SB, and MB provided clinical expertise
322 and valuable discussion on the results.

323 All authors commented on the draft report and contributed to the final version.

324

325 References

326

- 327 1. Tayoun A, Loney T, Khansaheb H, et al. Genomic surveillance and phylogenetic analysis reveal
328 multiple introductions of SARS-CoV-2 into a global travel hub in the Middle East. *BioRxiv* 2020;
329 **2020.05.06.080606**.
- 330 2. Banu S, Jolly B, Mukherjee P, et al. A distinct phylogenetic cluster of Indian SARS-CoV-2 isolates.
331 *bioRxiv* 2020: 2020.05.31.126136.
- 332 3. Lu J, du Plessis L, Liu Z, et al. Genomic Epidemiology of SARS-CoV-2 in Guangdong Province,
333 China. *Cell* 2020; **181**(5): 997-1003.e9.
- 334 4. Meredith LW, Hamilton WL, Warne B, et al. Rapid implementation of SARS-CoV-2 sequencing to
335 investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *Lancet Infect*
336 *Dis* 2020.
- 337 5. Seemann T, Lane C, Sherry N, et al. Tracking the COVID-19 pandemic in Australia using genomics.
338 *medRxiv* 2020: 2020.05.12.20099929.
- 339 6. Candido DS, Claro IM, de Jesus JG, et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil.
340 *Science* 2020: eabd2161.
- 341 7. Díez-Fuertes F, Iglesias-Caballero M, Monzón S, et al. Phylodynamics of SARS-CoV-2 transmission
342 in Spain. *bioRxiv* 2020: 2020.04.20.050039.
- 343 8. Gámbaro F, Behillil S, Baidaliuk A, et al. Introductions and early spread of SARS-CoV-2 in France.
344 *bioRxiv* 2020: 2020.04.24.059576.
- 345 9. Gudbjartsson DF, Helgason A, Jonsson H, et al. Spread of SARS-CoV-2 in the Icelandic Population. *N*
346 *Engl J Med* 2020; **382**(24): 2302-15.
- 347 10. Kumar P, Pandey R, Sharma P, et al. Integrated genomic view of SARS-CoV-2 in India. *bioRxiv* 2020:
348 2020.06.04.128751.
- 349 11. Zehender G, Lai A, Bergna A, et al. Genomic characterization and phylogenetic analysis of SARS-
350 COV-2 in Italy. *J Med Virol* 2020.
- 351 12. team TN. Genomic epidemiology of novel coronavirus - Global subsampling. 2020.
352 <https://nextstrain.org/ncov/global>.
- 353 13. Hill V, Rambaut A. Phylodynamic analysis of SARS-CoV-2 | Update 2020-03-06. 2020.
354 <https://virological.org/t/phylodynamic-analysis-of-sars-cov-2-update-2020-03-06/420>.
- 355 14. Rambaut A, Holmes EC, O'Toole Á, et al. A dynamic nomenclature proposal for SARS-CoV-2
356 lineages to assist genomic epidemiology. *Nat Microbiol* 2020.
- 357 15. Zhang L, Jackson CB, Mou H, et al. The D614G mutation in the SARS-CoV-2 spike protein reduces
358 S1 shedding and increases infectivity. *bioRxiv* 2020.
- 359 16. van Dorp L, Richard D, Tan CCS, Shaw LP, Acman M, Balloux F. No evidence for increased
360 transmissibility from recurrent mutations in SARS-CoV-2. *bioRxiv* 2020: 2020.05.21.108506.
- 361 17. Grenzgänger. <https://www.statistik.bs.ch/haeufig-gefragt/arbeiten/grenzgaenger.html> (accessed Jul 30,
362 2020).
- 363 18. Swissinfo.ch. Swiss hospitals to take French coronavirus patients. [https://www.swissinfo.ch/eng/cross-](https://www.swissinfo.ch/eng/cross-border-care_swiss-hospitals-take-french-coronavirus-patients/45634674)
364 [border-care_swiss-hospitals-take-french-coronavirus-patients/45634674](https://www.swissinfo.ch/eng/cross-border-care_swiss-hospitals-take-french-coronavirus-patients/45634674).
- 365 19. Quick J. nCoV-2019 sequencing protocol. protocols.io 2020.
366 [dx.doi.org/10.17504/protocols.io.bdp7i5rn](https://doi.org/10.17504/protocols.io.bdp7i5rn).
- 367 20. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality.
368 *Euro Surveill* 2017; **22**(13): 30494.
- 369 21. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global
370 health. *Glob Chall* 2017; **1**(1): 33-46.
- 371 22. Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution.
372 *Bioinformatics* 2018; **34**(23): 4121-3.
- 373 23. Bluhm A, Christandl M, Gesmundo F, et al. SARS-CoV-2 Transmission Chains from Genetic Data: A
374 Danish Case Study. *bioRxiv* 2020: 2020.05.29.123612.
- 375 24. Versteeg B, Bruisten SM, Pannekoek Y, et al. Genomic analyses of the *Chlamydia trachomatis* core
376 genome show an association between chromosomal genome, plasmid type and disease. *BMC genomics* 2018;
377 **19**(1): 130.
- 378 25. Xing Y, Li X, Gao X, Dong Q. MicroGMT: A Mutation Tracker for SARS-CoV-2 and Other Microbial
379 Genome Sequences. *Front Microbiol* 2020; **11**: 1502.
- 380 26. Pybus OG, Rambaut A, du Plessis L, et al. Preliminary analysis of SARS-CoV-2 importation &
381 establishment of UK transmission lineages. 2020. [https://virological.org/t/preliminary-analysis-of-sars-cov-2-](https://virological.org/t/preliminary-analysis-of-sars-cov-2-importation-establishment-of-uk-transmission-lineages/507)
382 [importation-establishment-of-uk-transmission-lineages/507](https://virological.org/t/preliminary-analysis-of-sars-cov-2-importation-establishment-of-uk-transmission-lineages/507).

- 383 27. Felbermayr G, Hinz J, S C. Après-ski: The Spread of Coronavirus from Ischgl through Germany. 2020.
384 [https://www.ifw-kiel.de/fileadmin/Dateiverwaltung/IfW-Publications/Gabriel_Felbermayr/Apres-](https://www.ifw-kiel.de/fileadmin/Dateiverwaltung/IfW-Publications/Gabriel_Felbermayr/Apres-ski_The_Spread_of_Coronavirus_from_Ischgl_through_Germany/coronavirus_from_ischgl.pdf)
385 [ski_The_Spread_of_Coronavirus_from_Ischgl_through_Germany/coronavirus_from_ischgl.pdf](https://www.ifw-kiel.de/fileadmin/Dateiverwaltung/IfW-Publications/Gabriel_Felbermayr/Apres-ski_The_Spread_of_Coronavirus_from_Ischgl_through_Germany/coronavirus_from_ischgl.pdf).
386 28. Hodcroft EB. Preliminary case report on the SARS-CoV-2 cluster in the UK, France, and Spain. *Swiss*
387 *Med Wkly* 2020; **150**(9-10).
388 29. Zeitung B. Wir haben in Mulhouse die Epidemie-Phase erreicht.
389 [https://www.bazonline.ch/basel/region/gottesdienst-in-mulhouse-entwickelt-sich-zu-moeglichem-](https://www.bazonline.ch/basel/region/gottesdienst-in-mulhouse-entwickelt-sich-zu-moeglichem-coronaherd/story/29069253)
390 [coronaherd/story/29069253](https://www.bazonline.ch/basel/region/gottesdienst-in-mulhouse-entwickelt-sich-zu-moeglichem-coronaherd/story/29069253).
391 30. Joseph SJ, Didelot X, Gandhi K, Dean D, Read TD. Interplay of recombination and selection in the
392 genomes of *Chlamydia trachomatis*. *Biol Direct* 2011; **6**: 28.
393
394

395 **Figure legends**

396 **Figure 1. Epidemiological curve of the first COVID-19 wave in the city of Basel and hinterland,**
 397 **Switzerland.** Positive (red line, dark grey area) and negative (light grey area) SARS-CoV-2 PCR tests are
 398 depicted from the beginning of the outbreak in February to March 23, 2020. Major events and imposed
 399 restrictions are marked by horizontal lines. First confirmed cases in Switzerland and Basel were on February
 400 25th and February 26th, respectively.

401
 402 **Figure 2. Detection of SARS-CoV-2 lineages found in the Basel area cohort from the first detected case on**
 403 **February 26th to March 23rd 2020.** Major events and imposed restrictions are marked by horizontal lines.
 404 Detection of low abundant lineages increases over time.

405
 406 **Figure 3. SARS-CoV-2 lineage diversity in neighbouring countries to Switzerland from first detected case**
 407 **until March 23rd, 2020.** Number of lineages (L) and total number of genomes (N) per country in brackets,
 408 values within charts represent percentages. France was the first country in Europe that had confirmed COVID-
 409 19 cases on January 24th, followed by Germany on January 27th, Italy on January 31st, and some weeks later
 410 Austria and Switzerland followed on February 25th. Simpson diversity based on the available genomes and
 411 PANGOLIN lineage assignments is largest in Germany (3.87) and Austria (3.73), followed by Italy (2.52), it is
 412 smallest in Switzerland (1.62) and France (1.23). Basel-City mirrors the lineage proportions of Switzerland,
 413 while contributing half of Switzerland's sequence data.

414
 415 **Figure 4. SARS-CoV-2 phylogeny of Basel area samples and genetic lineages (PANGOLIN) in a global**
 416 **context.** **A.** Time tree of SARS-CoV-2 genomes from the Basel area cohort as well as subsampled global
 417 genomes (30 genomes per country and month), coloured by continent of origin. Amino acid mutations at
 418 internal nodes representing clade defining mutations are shown. **B.** Mirrored time tree coloured by genetic
 419 lineages sensu PANGOLIN v.May19 (<https://github.com/cov-lineages/>). Each tip with a circle represents a
 420 genome from the Basel area cohort, branches without circled tips represent global genomes, included to confer
 421 the global context of the Basel genomes.

422
 423 **Figure 5. Divergence trees plotting nucleotide divergence between 468 genomes and expanded clusters of**
 424 **genomes in selected phylogenetic lineages.** **A.** Genomes from Basel area cohort in global context. Tree
 425 composition is identical to the time tree from Figure 4. Branches with circles at the tip represent genomes from
 426 the present study; branches without circles represent global genomes from GISAID. The major Basel cluster
 427 contains samples with up to five mutations. **B.** Zoom into a mixed cluster derived from B.1.1 with seven to nine
 428 mutations difference to the root. A single genomes assigned to lineage B.1.1.6 has an assumed origin in Austria;
 429 two genomes (B.1.1.10) most likely originate from the UK. **C.** Potential ski-holiday related cluster (C1059T)
 430 with seven samples having a confirmed association with skiing destinations. Note: Divergence can be translated
 431 to number of mutations difference to the root (Wuhan-Hu-1) by multiplication by with the SARS-CoV-2
 432 genome size (29903 bases).

433
 434
 435
 436 **Table 1. Number and age summary of all tested patients, positively tested patients, and patients with**
 437 **successfully sequenced SARS-CoV-2 genomes, by sex.**

		Number	%	Median Age [years]	IQR [years]	< 18 years old
All tests*	Males	3067	44.2	44	31-60	396 (5.7%)
	Females	3867	55.8	42	29-56	
Positive tested*	Males	363	48.7	49	33-61.5	17 (2.3%)
	Females	383	51.3	47	32-60	
In study cohort*	Males	222	48.1	49	34-60	12 (2.6%)
	Females	240	51.9	47	33-60	

438 * six patients with no information regarding sex

439
 440
 441
 442
 443
 444

445 **Table 2. Sensitivity, specificity, and accuracy of COVGAP.**

	MOCK POSITIVE	MOCK NEGATIVE
COVGAP POSITIVE	TP=180	FN=11
COVGAP NEGATIVE	FP=0	TN=2541564

446 *TP: true positive, FP: false positive, FN: false negative. Numbers represent cumulative counts of bases that*
 447 *were or were not mutated over the 16 test genomes.*

448

449

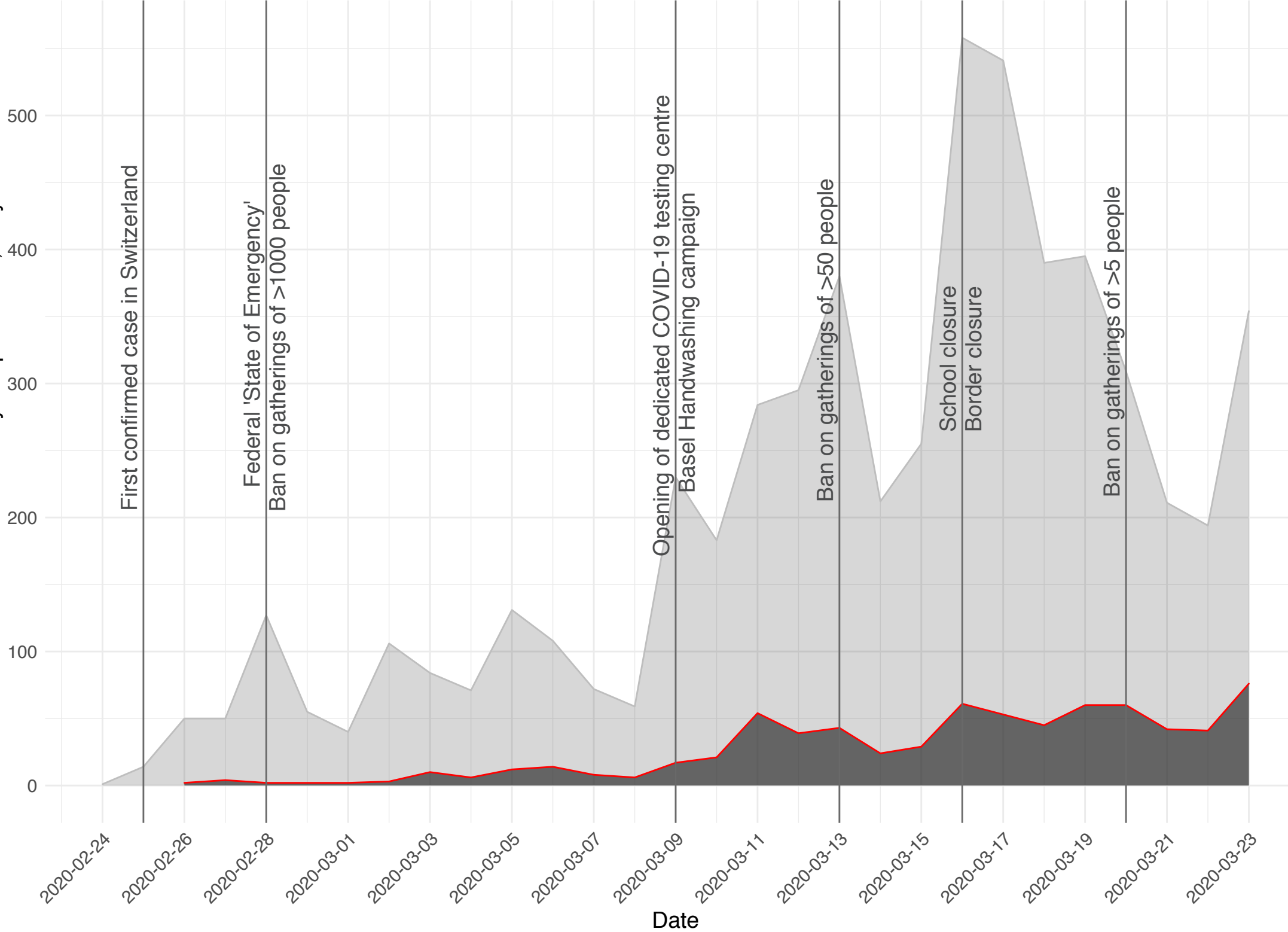
450 **Table 3. Number of cases harbouring the S-D614G mutation in spike protein encoding gene in each**
 451 **phylogenetic lineage (PANGOLIN definition ver. May 19) and total count, in Basel area cohort by March**
 452 **23rd 2020.**

Phylogenetic lineage	Number of samples S ^{G614} (derived)	Number of samples S ^{D614} (ancestral)	Total counts
A.2	0	1	1
A.5	0	1	1
B	2	6	8
B.1	391	0	391
B.1.1	36	0	36
B.1.1.1	1	0	1
B.1.1.10	2	0	2
B.1.1.6	1	0	1
B.1.5	12	0	12
B.1.8	3	0	3
B.10	0	1	1
B.2	0	8	8
B.2.1	0	2	2
B.3	0	1	1
Sum	448	20	468

453

454

Number of tests at University Hospital Basel, daily counts



First confirmed case in Switzerland

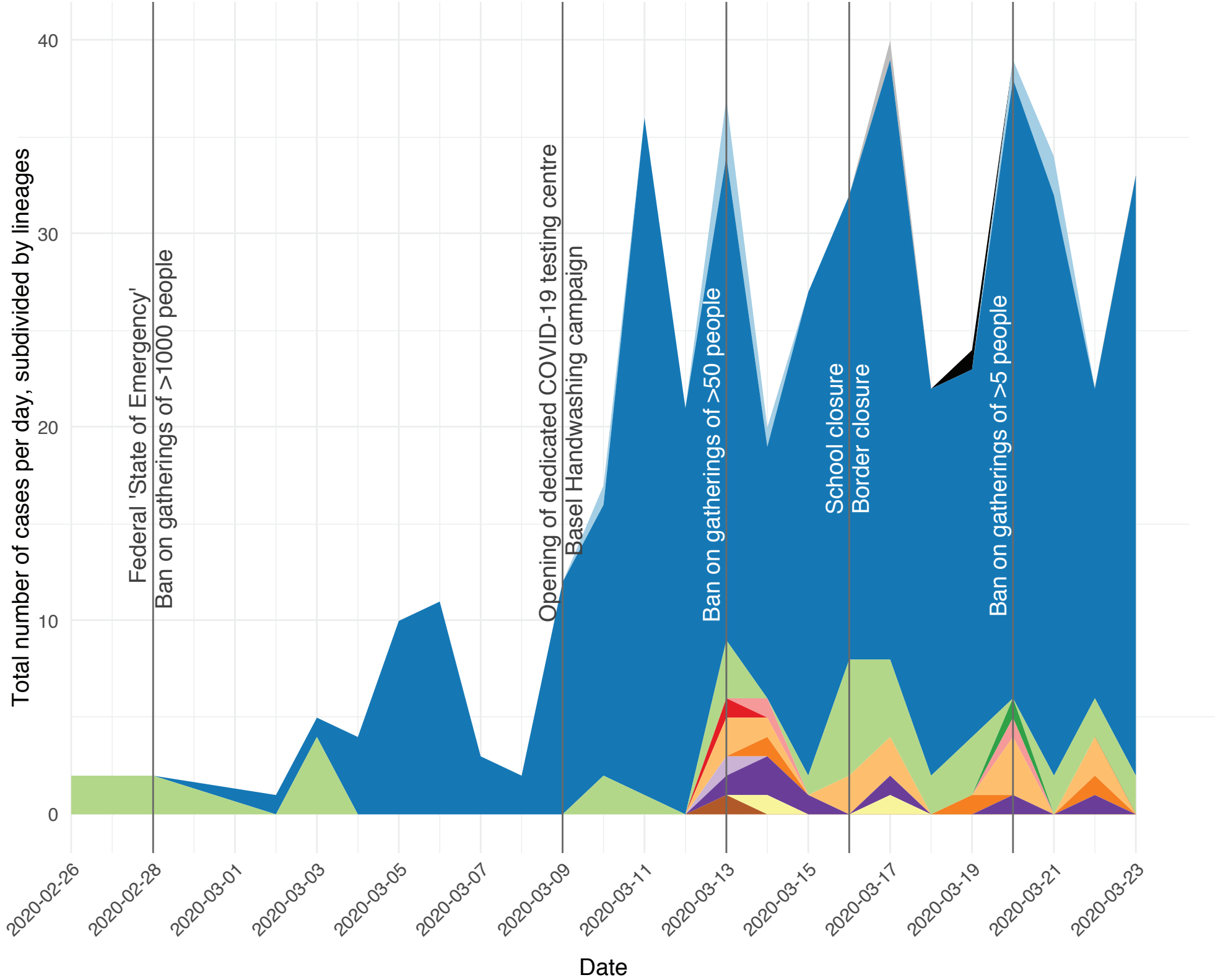
Federal 'State of Emergency'
Ban on gatherings of >1000 people

Opening of dedicated COVID-19 testing centre
Basel Handwashing campaign

Ban on gatherings of >50 people

School closure
Border closure

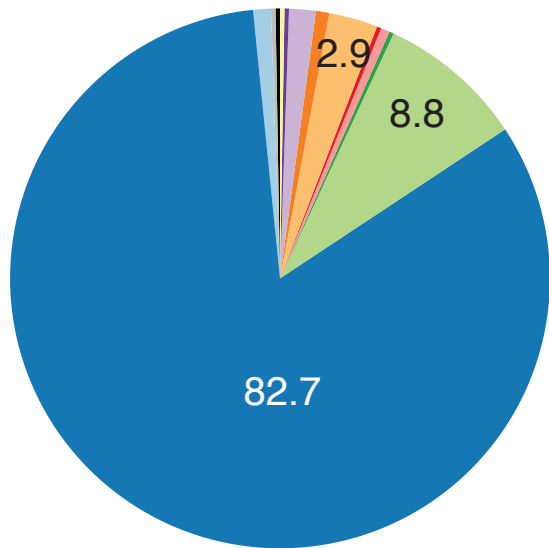
Ban on gatherings of >5 people



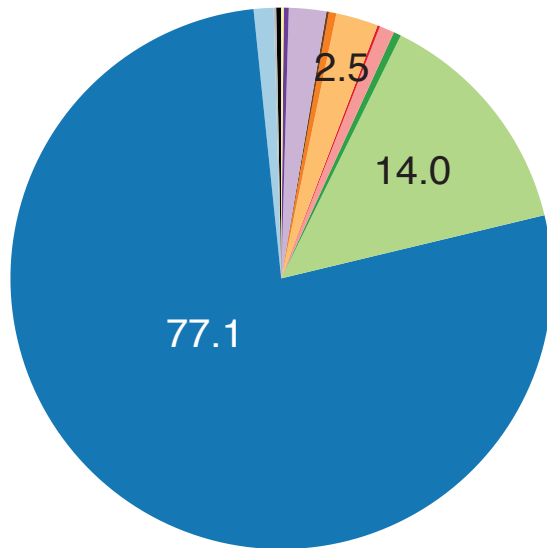
Sars-CoV-2 genetic lineages (and putative origin)

- | | | |
|---|---|----------------------------|
| A.2 (Spain/ UK/ Australia) | B.1.1.1 (UK/ Belgium/ Netherlands) | B.2 (UK/ USA/ Netherlands) |
| A.5 (Spain/ UK/ Uruguay) | B.1.1.10 (UK/Iceland) | B.2.1 (Global) |
| B(UK, China, USA SNPs T8782C, C28144T) | B.1.1.6 (Austria) | B.3 (UK/ Belgium/ Denmark) |
| B.1 (European, Italian outbreak) | B.1.5 (UK/ Spain/Turkey/Australia/USA/Brasil) | |
| B.1.1 (Europe/ UK SNPs G28881A,G28882A,G28883C) | B.1.8 (Netherlands/ Iceland/ Denmark) | |

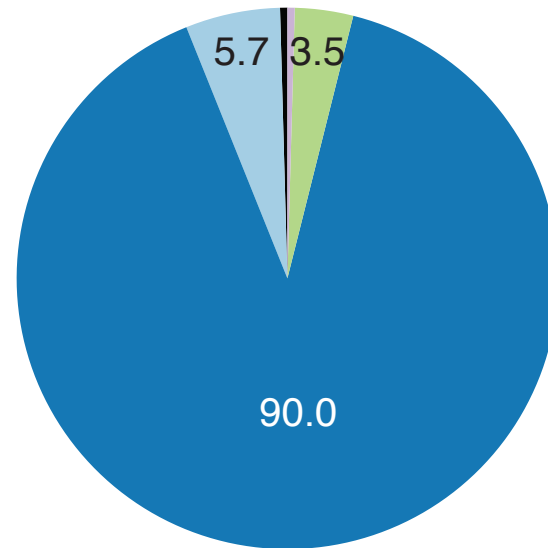
Basel-City (L = 13, N = 376)



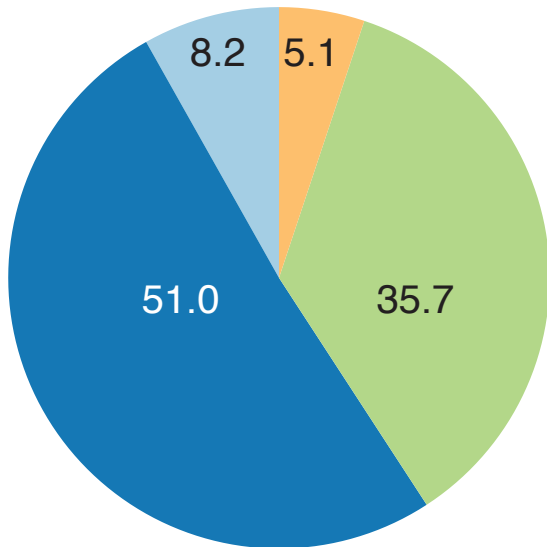
Switzerland (L = 14, N = 673)



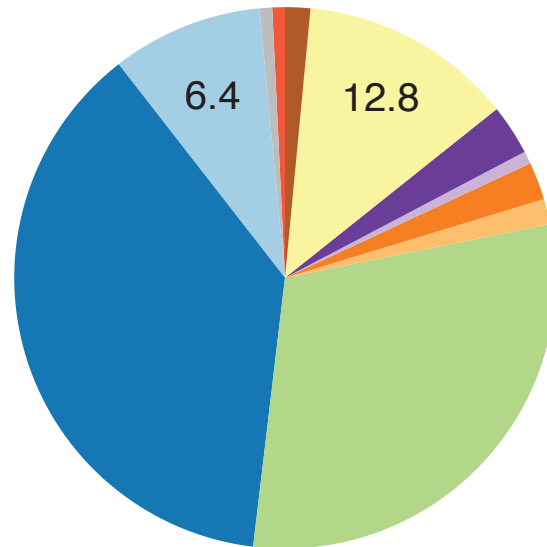
France (L = 5, N = 230)



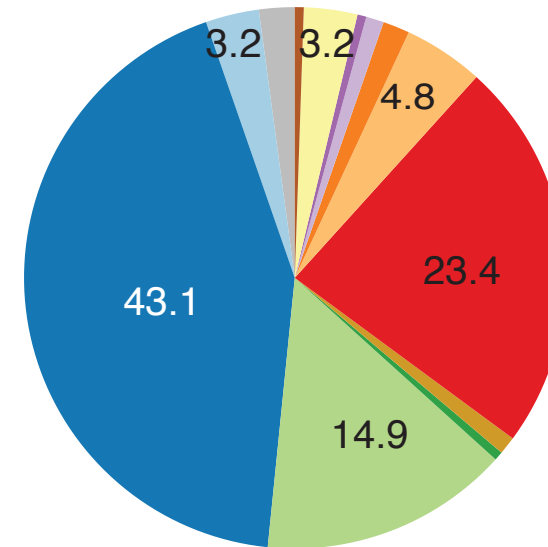
Italy (L = 4, N = 98)

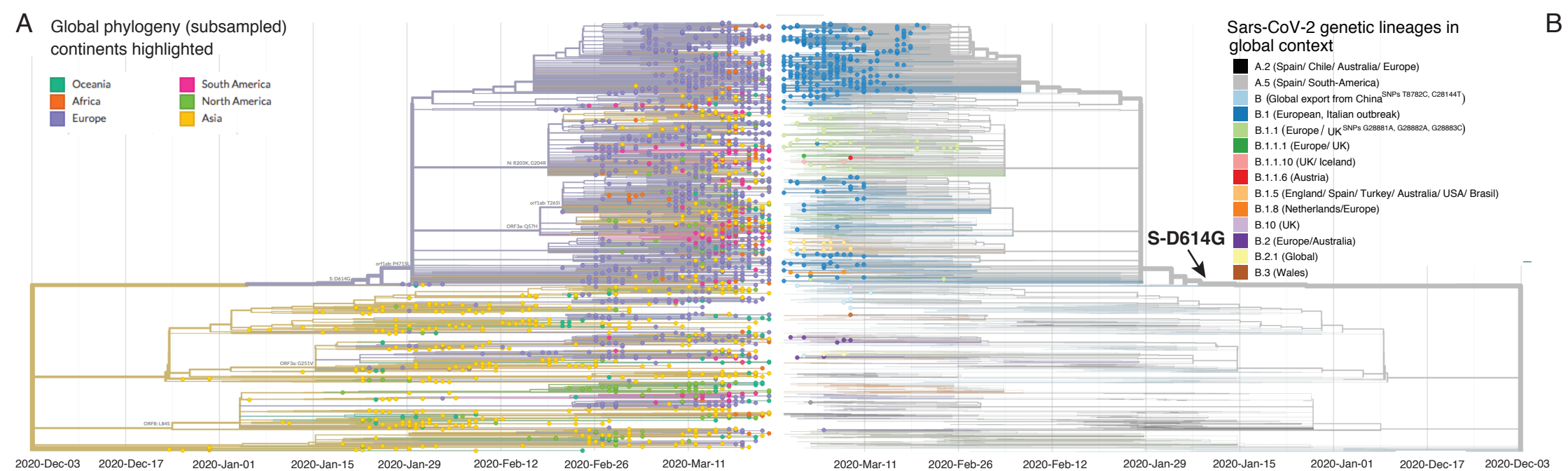


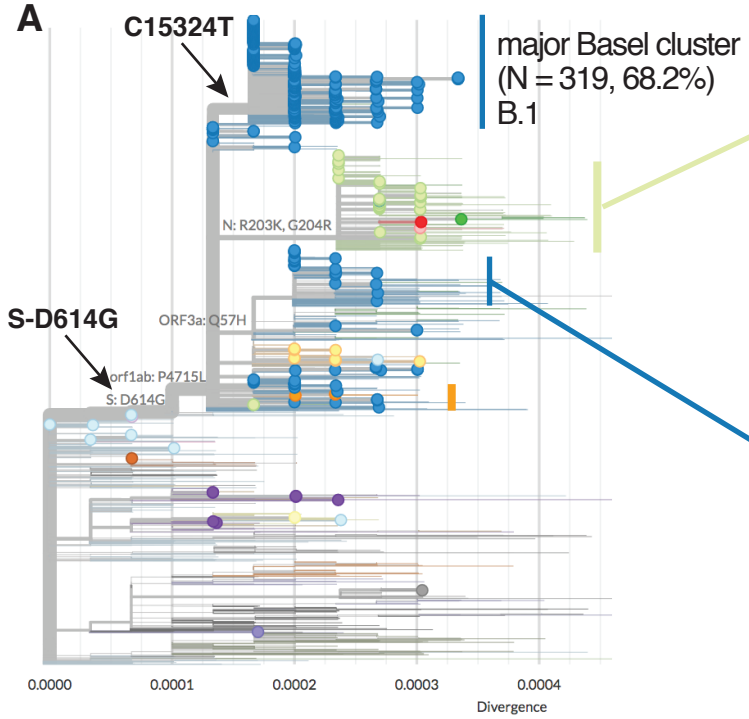
Germany (L = 11, N = 133)



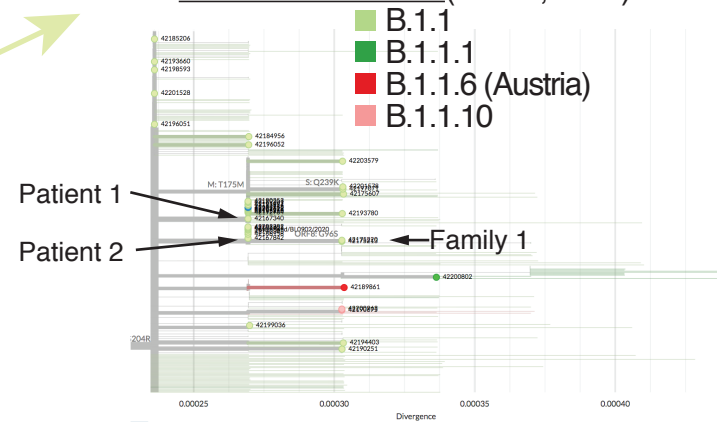
Austria (L = 13, N = 188)





A**B**

B.1.1 derived cluster (N = 39, 8.3%):

**C**

C1059T cluster, B.1 (N = 21, 4.5%)

