

1 **Phylodynamics of SARS-CoV-2 transmissions in France,** 2 **Europe and the world during 2020**

3 Romain Coppée^{1§}, François Blanquart^{1,2}, Aude Jary³, Valentin Leducq³, Valentine Marie
4 Ferré^{1,4}, Anna Maria Franco Yusti^{1,4}, Lena Daniel^{1,4}, Charlotte Charpentier^{1,4}, Samuel
5 Lebourgeois^{1,4}, Karen Zafilaza^{3,5}, Vincent Calvez^{3,5}, Diane Descamps^{1,4}, Anne-Geneviève
6 Marcelin^{3,5}, Benoit Visseaux^{1,4}, Antoine Bridier-Nahmias¹

7 ¹ Université Paris Cité and Sorbonne Paris Nord, Inserm, IAME, F-75018 Paris, France

8 ² Centre for Interdisciplinary Research in Biology (CIRB), Collège de France, CNRS Inserm,
9 PSL Research University, Paris, France

10 ³ Sorbonne Université, Inserm, iPLESP, F-75013 Paris, France

11 ⁴ Service de Virologie, Hôpital Bichat Claude Bernard, AP-HP, F-75018 Paris, France

12 ⁵ Service de Virologie, Hôpital de la Pitié-Salpêtrière, AP-HP, F-75013 Paris, France

13 [§] Corresponding author: Romain Coppée (romain.coppee@gmail.com)

14 Abstract

15 **Background:** Although France was one of the most affected European countries by the
16 COVID-19 pandemic in 2020, the dynamics of SARS-CoV-2 transmissions within France,
17 Europe and worldwide remain only partially characterized during the first year of the pandemic.

18 **Methods:** Here, we analyzed GISAID deposited sequences from January to December 2020
19 (n = 638,706 sequences). To tackle the huge number of sequences without the bias of
20 analyzing a single sequence subset, we produced 100 independent and randomly selected
21 sequence datasets and related phylogenetic trees for different geographic scales (worldwide,
22 European countries and French administrative regions) and time periods (first and second half
23 of 2020). We applied a maximum likelihood discrete trait phylogeographic method to date
24 transmission events and to estimate the geographic spread of SARS-CoV-2 to, from and within
25 France, Europe and worldwide.

26 **Results:** The results unraveled two different patterns of inter- and intra-territory transmission
27 events between the first and second half of 2020. Throughout the year, Europe was
28 systematically associated with most of the intercontinental transmissions, for which France has
29 played a pivotal role. SARS-CoV-2 transmissions with France were concentrated with North
30 America and Europe (mainly Italy, Spain, United Kingdom, Belgium and Germany) during the
31 first wave, and were limited to neighboring countries without strong intercontinental
32 transmission during the second one. Regarding French administrative regions, the Paris area
33 was the main source of transmissions during the first wave. But, for the second epidemic wave,
34 it equally contributed to virus spread with Lyon and Marseille area, the two other most densely
35 populated cities in France.

36 **Conclusion:** By enabling the inclusion of tens of thousands of viral sequences, this original
37 phylogenetic strategy enabled us to robustly depict SARS-CoV-2 transmissions through
38 France, Europe and worldwide in 2020.

39 **Keywords**

- 40 SARS-CoV-2, COVID-19, phylodynamics, genomic epidemiology, France, transmission,
41 maximum likelihood

42 Background

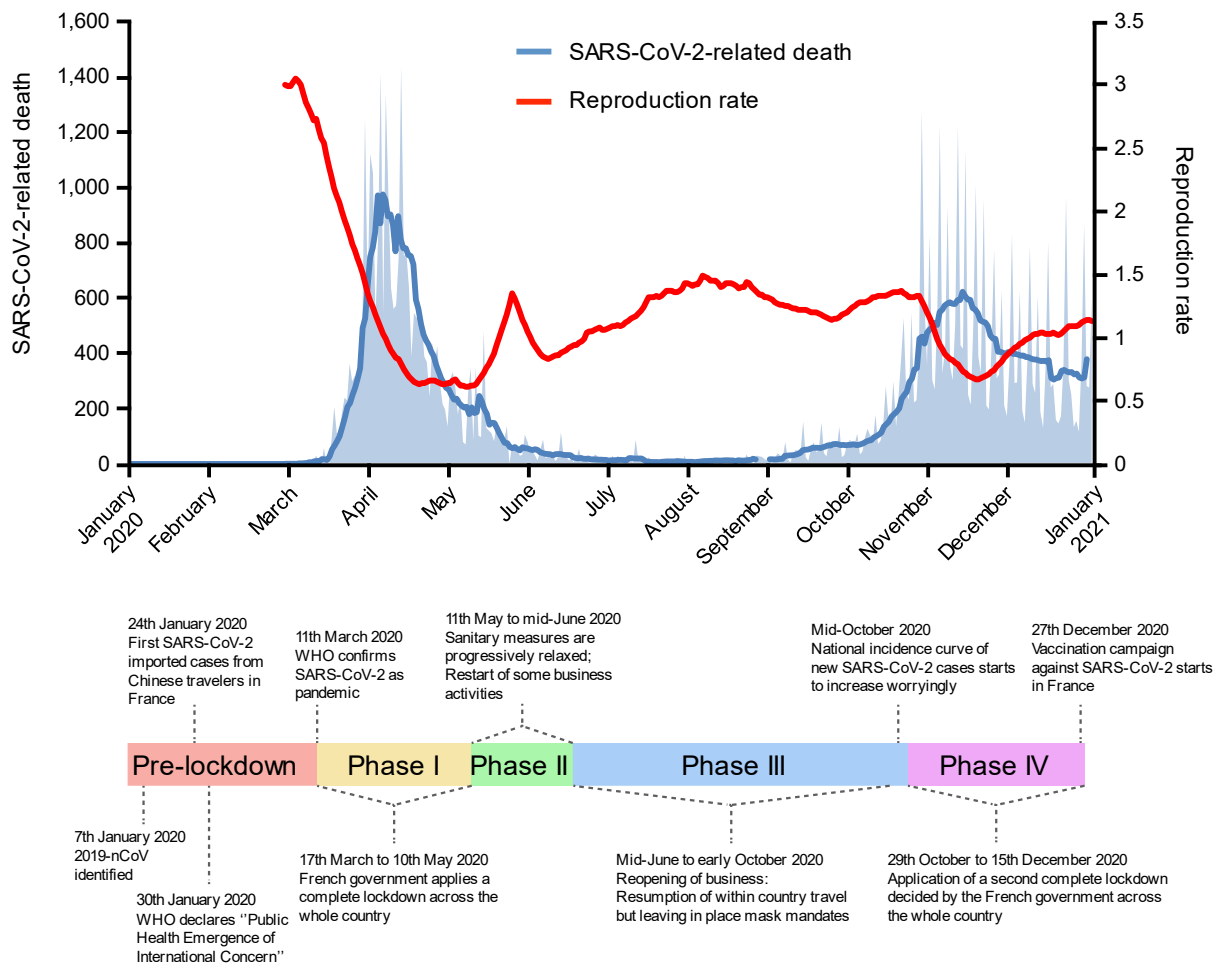
43 On 1st December 2019, an outbreak of severe respiratory disease was identified in the city of
44 Wuhan, China (Huang et al., 2020). Rapidly, the agent of the disease was identified as the
45 severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Zhu et al., 2020),
46 responsible for the ongoing global pandemic of coronavirus disease 2019 (COVID-19). By the
47 end of 2020, the virus had caused over 1.8 million deaths worldwide including ~65k deaths in
48 France, concomitantly with social and economic devastations in many regions of the world
49 (Mofijur et al., 2021; Santomauro et al., 2021). Since the beginning of COVID-19 pandemic,
50 the scientific community has stepped up to thoroughly characterize the virus, including its
51 pathogenesis, the monitoring of its circulation in human populations, and the development of
52 several treatments or vaccines (Cevik et al., 2020; Krammer, 2020). The development of
53 epidemiological models have been particularly helpful to evaluate the viral spread both in short-
54 and long-terms and inform the public health decisions (Hoertel et al., 2020; Kissler et al., 2020).

55 In addition to clinical and epidemiological insights, the viral whole-genome sequencing has
56 become a powerful and invaluable tool to better understand the infection dynamics (Volz et al.,
57 2013). The number of SARS-CoV-2 whole-genome sequences available have rapidly grown
58 thanks to an altruistic and international effort of scientist departments gathered via the Global
59 Initiative on Sharing All Influenza Data, GISAID (<https://www.gisaid.org/>) (Khare et al., 2021).
60 These genomic sequences are essential to effectively reconstruct the global viral spread and
61 the origins of variants. Today, genomic data have become a strong asset in addition to
62 epidemiological data to inform governments, helping to define and assess the most appropriate
63 public health measures (Attwood et al., 2022; Rife et al., 2017). Until now, the existing
64 phylogenetic tools cannot include large number of data such as those generated by
65 widespread viral sequencing and the development of specific frameworks is necessary.

66 In France, the first COVID-19 suspected case has been identified in late December 2019
67 (Deslandes et al., 2020), and the first confirmed cases of SARS-CoV-2 infection were detected

68 on the 24th January 2020, in individuals who were recently arrived from China (Bernard
69 Stoecklin et al., 2020). COVID-19 cases remained scarce until the end of February, when the
70 national incidence curve of new SARS-CoV-2 infections started to rise (Figure 1). By the end
71 of February, reinforced measures were announced, including social distancing, cessation of
72 passenger flights to France, school closure, and finally, a complete lockdown across the entire
73 country from 17th March to 10th May 2020. The reported daily incidence and numbers of
74 severe cases peaked at the beginning of April 2020 before decreasing steadily until August
75 2020. However, after relaxation of social distancing in June, a second wave of infections
76 occurred in early September peaking at more than 100k positive cases in a single day on the
77 2nd November 2020 (Figure 1). As for the first wave, daily incidence and severe COVID-19
78 cases gradually diminished down to a number of positive daily cases varying between 2k and
79 25k at the end of the year after a second national lockdown between the 29th October and
80 15th December 2020. These epidemiological tendencies were somehow similar in most
81 European countries except for Russia or Romania, where high rates of SARS-CoV-2-related
82 deaths were reported even during the summer of 2020. Of note, the other continents showed
83 different patterns of virus circulation: the number of deaths increased about two weeks later in
84 North America; and from early May, Asia, North America and South America were highly
85 impacted by the pandemic (Figure 1—figure supplement 1).

86 Elucidating the SARS-CoV-2 dynamic throughout the various phases of the pandemic is
87 paramount to get lessons for future viral epidemics (Rife et al., 2017). Here, we analyzed
88 GISAID deposited sequences to elucidate the origins and spread of the virus in France, Europe
89 and worldwide from January to December 2020. Through a maximum likelihood discrete trait
90 phylogeographic method, we estimated the main geographical areas that contributed to viral
91 introductions in France and Europe, the countries/continents to which France transmitted
92 SARS-CoV-2 the most and the contribution of the different French regions to the national
93 circulation of the virus. We explored the differences in viral circulation during each of the two
94 European epidemic waves of 2020 independently.



95

96 **Figure 1. Timeline of SARS-CoV-2-related deaths and reproduction rate in France, 2020.** Key
 97 events are indicated on the timeline. Official lockdowns included stay home orders and closure of
 98 schools and daycares. The two first French epidemic waves are respectively dated of March to end-
 99 June 2020, and September to December 2020. SARS-CoV-2-related deaths are displayed as the daily
 100 number of deaths (light blue area) and as the mean of deaths in a week (curve).

101 Results

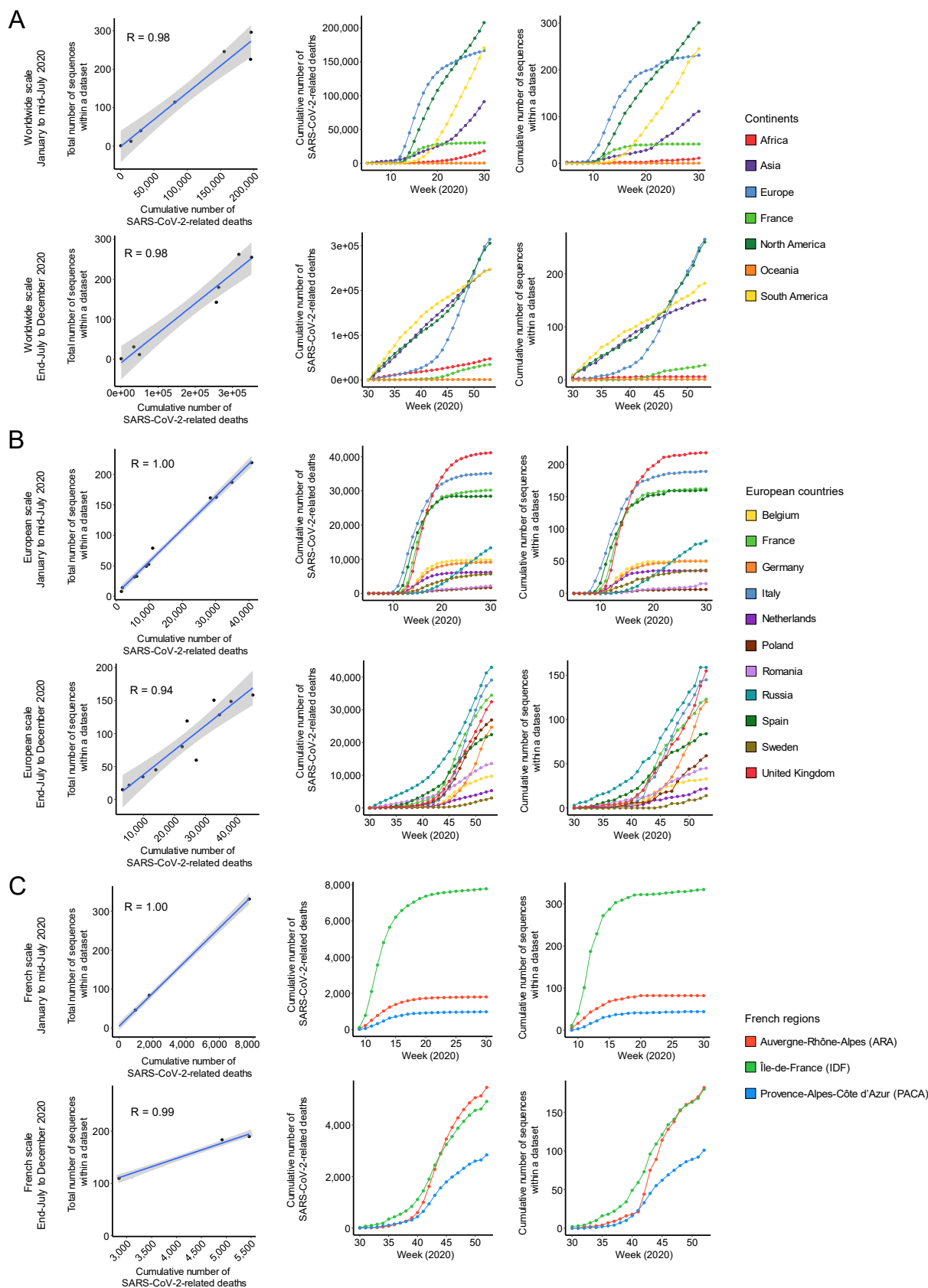
102 Description of the datasets and global diversity of SARS-CoV-2 sequences

103 Prior to any phylodynamics investigation, we checked whether the sets of sequences produced
 104 for each geographic scale (i.e., worldwide, Europe and French regions) and time period (i.e.,
 105 the first and second European epidemic waves) matched with the number of SARS-CoV-2
 106 infections in 2020. This step was necessary to ensure that our sampling method did not
 107 introduce too strong biases from the start. For each geographic scale and time period, we
 108 noticed a positive correlation with the estimated SARS-CoV-2-related deaths (Spearman's

109 rank correlation, $p < 0.001$; $r = 0.94$ for the lowest correlation). We also confirmed that the
110 number of sequences per territory was, on average, properly temporally distributed within each
111 time period (**Figure 2**).

112 Some countries and French administrative regions were however discarded in the analyses
113 because they were not sufficiently represented in the GISAID database. Overall, a total of
114 41,566 and 43,368 distinct SARS-CoV-2 sequences were included across the 100 phylogenies
115 for the worldwide dataset, respectively for the first and the second time periods investigated
116 (**Table 1**). At the European scale, 28,235 and 29,369 different SARS-CoV-2 sequences
117 covering eleven countries were analyzed across the 100 replicates (**Table 1**). Focusing on
118 French administrative regions, sequences available on the GISAID database were very
119 sparse. The Provence-Alpes-Côte d'Azur (PACA, Southeast of France, Marseille area) was
120 the only region that highly sequenced SARS-CoV-2 in 2020. Île-de-France (center of France,
121 Paris area) and Auvergne-Rhône-Alpes (ARA, East of France, Lyon area) have sequenced
122 much less than PACA, but provided sufficient data to investigate SARS-CoV-2 transmission
123 events within France. The remaining French administrative regions were discarded since no
124 sufficient sequences were available to properly match the number of weekly deaths (**Figure**
125 **2—figure supplement 1**). We thus considered 2,810 unique sequences across the 100
126 replicates for the first half of 2020, and 3,639 unique sequences for the second time period
127 studied (**Table 1**).

128 The genomic diversity of circulating SARS-CoV-2 in the different continents, countries and
129 French regions was found to be globally uniform (**Figure 2—figure supplement 2**). Overall,
130 genomes showed high sequence conservation compared to the Wuhan-Hu-1 reference in
131 2020 (mean and median of ~13 SNPs with 95% of the distribution comprised between 4 and
132 25 SNPs).



133

134 **Figure 2. Positive correlation between the number of SARS-CoV-2-related deaths and the number**
 135 **of sequences included within a dataset at (A) worldwide, (B) European and (C) French scales.**
 136 For each item, the first row of plots corresponds to the data from January to mid-July 2020, and the
 137 second one covers the period of end-July to December 2020. The *left* plots show the positive correlation

138 between the total number of sequences per territory and the number of SARS-CoV-2-related deaths
139 during the period investigated. The coefficient of correlation R is indicated inside the plots. The *central*
140 and *right* plots show the cumulative number of SARS-CoV-2-related deaths and sequences within a
141 dataset over time. Each color corresponds to a territory.

142 **Table 1. Number of SARS-CoV-2 sequences investigated within each dataset**

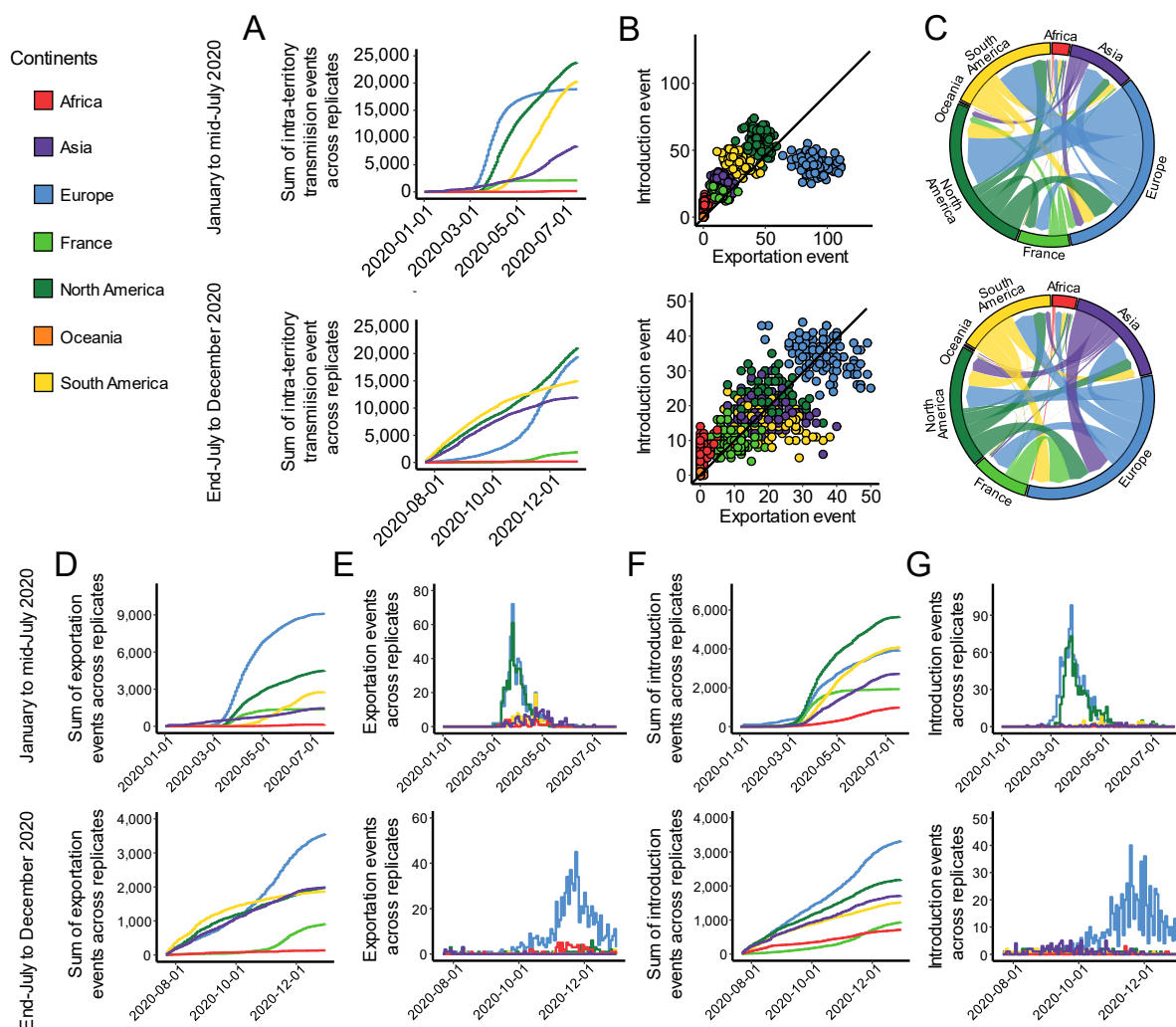
Dataset	Geographical categories	Period investigated	Average number of sequences within a set	Number of sequences investigated
World	Africa, Asia, Europe, France, North America, Oceania, South America	January to mid-July 2020	935	41 566
		End-July to December 2020	882	43 368
Europe	Belgium, France, Germany, Italy, The Netherlands, Poland, Romania, Russia, Spain, Sweden, United Kingdom	January to mid-July 2020	999	28 235
		End-July to December 2020	961	29 369
France	Auvergne-Rhône-Alpes (ARA), Île-de-France (IDF), Provence-Alpes-Côte d'Azur (PACA)	January to mid-July 2020	462	2 810
		End-July to December 2020	483	3 639

143

144 Which continents exchanged SARS-CoV-2 with Europe and France?

145 Through several dozen distinct, dated and ancestrally reconstructed phylogenetic trees, we
146 first studied the worldwide origins of SARS-CoV-2 transmissions into France and Europe for
147 each of the time periods studied (**Figure 3**). Intra-continent and intra-France transmissions
148 accounted for most of the total transmission events, representing 79.3% and 87.4% during the
149 first and second half of 2020, respectively (**Figure 3—figure supplement 1**). From March
150 2020, the number of intra-territory transmissions lowly increased first in Asia before an
151 acceleration of virus circulation in May. The first exponential increase of SARS-CoV-2
152 transmissions was observed in some European countries in early-March, then concomitantly
153 in France and North America two weeks later. Intra-Europe and intra-France transmission
154 levels were then very low between July and end-September until the beginning of the second
155 European wave from October to December 2020. The rate of intra-North America transmission
156 events remained very high from March to December. Intra-territory transmissions in South
157 America started to increase from mid-April while those in Africa remained very limited. South
158 America and Asia had a similar profile of intra-territory transmissions, as occurring at high

159 levels between July and October followed by a slackening until December (**Figure 3A**). The
 160 profiles of intra-territory transmission events across the replicates and for each period
 161 investigated were positively correlated to the estimated number of deaths related to SARS-
 162 CoV-2 in each geographic region (Spearman's rank correlation, $p < 0.001$; $r = 0.96$ for the
 163 lowest correlation; **Figure 3—figure supplement 1**), thus confirming that our methodology
 164 adequately reflects spatial diffusion of COVID-19.



165

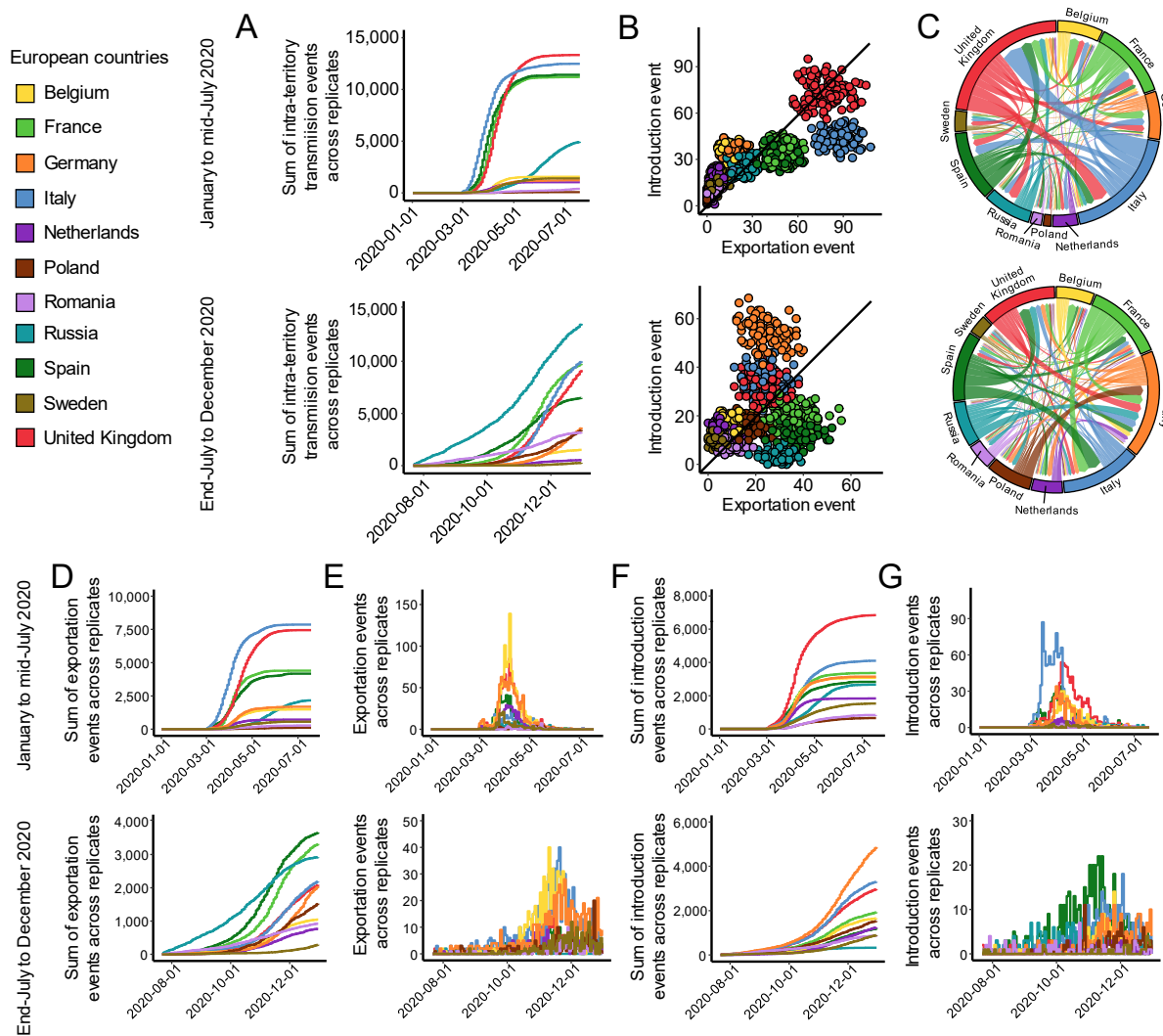
166 **Figure 3. SARS-CoV-2 intra- and inter-territory transmissions worldwide.** Each territory is
 167 associated with a specific color. Transmission events were calculated through 100 phylogenies as
 168 replicates between January and mid-July 2020, and between end-July to December 2020. When
 169 unspecified, transmission values are based on the sum of the replicates. **(A)** Evolution of intra-territory
 170 transmissions over time. **(B)** Number of introduction and exportation events for each replicate and for
 171 each continent and France. **(C)** SARS-CoV-2 exchange flows between continents and France during
 172 the two time periods investigated. In these plots, migration flow out of a particular location starts close
 173 to the outer ring and ends with an arrowhead more distant from the destination location. Migration flow
 174 out correspond to the mean of the worldwide sets of sequences. **(D)** Cumulative exportation and **(F)**
 175 introduction events per territory over time. **(E)** Exportation events over time originated from France. **(G)**
 176 Introduction events into France over time.

177 When focusing on introduction and exportation events during the first wave, we found that
178 Europe was associated with the highest number of exportation events in all locations (**Figure**
179 **3B** and **3C**, and **Figure 3—figure supplement 2**). Both North America and South America
180 also highly participated in virus exportation, while France and Asia had lower but similar
181 numbers of exportation events across the replicates by the end of July 2020 (**Figure 3D** and
182 **Figure 3—figure supplement 2**). The exportations from France were almost all headed
183 towards Europe and North America (**Figure 3E** and **Figure 3—figure supplement 2**). North
184 America accounted for a large proportion of SARS-CoV-2 introductions, followed by South
185 America and Europe at similar rates by the end of July 2020 (**Figure 3F**). The first introduction
186 cases into France were sparse and originated from Asia before March 2020, then introductions
187 came mostly from Europe followed by North America and increased in number to reach a peak
188 in end-March (**Figure 3G**). We noticed that the number of new introductions was low after mid-
189 May 2020 in France until the second European epidemic wave.

190 The second half of 2020 (from end-July to December 2020) showed a very different pattern of
191 inter-territory transmission events worldwide. The rate of exportation events was quite similar
192 between North America, South America, Asia and Europe until mid-October, then this rate
193 drastically increased only in Europe until the end of the time period investigated (**Figure 3B**,
194 **3C** and **3D**, and **Figure 3—figure supplement 2**). Before October 2020, rare exportation
195 events originated from France, consistent with the very low incidence of the virus during this
196 period in this country. Then, the number of virus exportations drastically increased and formed
197 a peak in the mid of November 2020. Those exportations were nearly all headed towards
198 Europe, accompanied with rare events into Africa and North America (**Figure 3E**). In the same
199 way, only rare introduction events into France were originated from Asia, North America and
200 South America until end-October, while high rates of introductions came from Europe from
201 early-October to December 2020 (**Figure 3G**). SARS-CoV-2 introductions in Europe originated
202 from Asia, North America, South America and France at similar levels (**Figure 3—figure**
203 **supplement 2**).

204 **How did the virus spread in Europe?**

205 We then aimed to get a finer view of SARS-CoV-2 transmission events between France and
 206 other European countries with the same strategy as the one used at the continental scale
 207 (**Figure 4**). Here, we only focused on European countries associated with a high incidence and
 208 without under-sampling due to a lack of data on GISAID.



209

210 **Figure 4. SARS-CoV-2 intra- and inter-territory transmissions at the European scale.** Each territory
 211 is associated with a specific color. Transmission events were calculated through 100 phylogenies as
 212 replicates between January and mid-July 2020, and between end-July to December 2020. When
 213 unspecified, transmission values are based on the sum of the replicates. (A) Evolution of intra-territory
 214 transmissions over time. (B) Number of introduction and exportation events for each replicate and for
 215 each European country. (C) SARS-CoV-2 exchange flows between European countries during the two
 216 time periods investigated. In these plots, migration flow out of a particular location starts close to the
 217 outer ring and ends with an arrowhead more distant from the destination location. Migration flow out
 218 correspond to the mean of the sets of sequences. (D) Cumulative exportation and (F) introduction events
 219 per territory over time. (E) Exportation events over time originated from France. (G) Introduction events
 220 into France over time.

221 Intra-country transmissions accounted for most of the total transmission events, representing
222 65.7% and 75.6% of the transmissions during the first and second half of 2020, respectively
223 **(Figure 4—figure supplement 1)**. Between March and July 2020, the number of intra-country
224 transmissions sequentially – and drastically – increased first in Italy, then Spain, France and
225 the United Kingdom. Germany was associated with a number of intra-country transmissions
226 as low as Belgium and the Netherlands **(Figure 4A and Figure 4—figure supplement 1)**.
227 While the number of intra-country transmission events reached a plateau in early-June 2020
228 across nearly all Europe, Russia and Romania were associated with an increased number of
229 cases from April which coincided with a similar increase from Asian countries bordering Russia.
230 The number of intra-country transmission events in Russia and Romania remained high until
231 the end of December 2020. In early September 2020, Spain was the first country associated
232 with a drastic increase of intra-territory transmissions, then followed by France, the United
233 Kingdom, Italy and Poland. The increase was observed a bit later for Germany and Belgium,
234 around end-October 2020. By the end of 2020, Russia accounted for most of the intra-country
235 transmissions between July and December, followed by Italy and France, the United Kingdom
236 and Spain. The profiles of intra-country transmissions across the replicates and for each wave
237 investigated were, as for the continents, positively correlated to the estimated number of
238 deaths related to SARS-CoV-2 for each country (Spearman's rank correlation, $p < 0.001$; $r =$
239 0.98 for the lowest correlation; **Figure 4—figure supplement 1**).

240 By calculating the count of introduction and exportation events during the first epidemic wave,
241 we observed that Italy and the United Kingdom were the major contributor to virus
242 dissemination towards other European countries, followed by France and Spain **(Figure 4B,**
243 **4C and 4D, and Figure 4—figure supplement 2)**. Italy was the first country to be associated
244 with a drastic increase of exportation events in early-March **(Figure 4D)**, followed by Spain,
245 the United Kingdom and France about one week later. France and Spain were however
246 surpassed by the United Kingdom by the end of April 2020. Most of exportation events from
247 France were directed towards Belgium, the United Kingdom, and Germany, and a little towards

248 Spain, the Netherlands and other European countries (**Figure 4C** and **Figure 4—figure**
249 **supplement 2**), mostly before the official lockdown in France (i.e., before end-March) (**Figure**
250 **4E**). The rate of exportations from France then diminished until the second epidemic wave, as
251 well as for other European countries except Russia. For introduction events, the United
252 Kingdom accounted for a quarter of the total number of events, followed by Italy, while France,
253 Germany, Spain and Belgium shared a strikingly similar pattern of transmissions (**Figure 4F**).
254 In France, most of the introductions first came from Italy in early-March, while a few early
255 introductions into France also originated from Spain, Germany and United Kingdom (**Figure**
256 **4G**). The United Kingdom, Spain, Belgium and Germany were associated with a high number
257 of introductions into France, but mostly a few days before and at the beginning of the French
258 national lockdown.

259 The second European epidemic wave showed a different pattern of transmission events.
260 France, Spain and Russia were the countries associated with most of the exportation events,
261 followed by Germany, Italy and The United Kingdom (**Figure 4B, 4C** and **4D**, and **Figure 4—**
262 **figure supplement 2**). The profile of exportation events was globally uniform in Russia with a
263 continuous increase through the whole year 2020 (**Figure 4D**). This was also the case in
264 Romania, although the incidence remained lower than Russia. Spain, shortly followed by
265 France, were the two first countries associated with a steep increase of exportations from
266 September to December 2020. Later in September and October, the United Kingdom, Italy,
267 followed by Poland, Germany, Belgium, and the Netherlands have each seen a sharp increase
268 in their rate of exportations, sustained until the end of 2020. France moderately transmitted the
269 virus until the end of October in some European countries, then a high rate of transmissions
270 was mainly directed towards Belgium, Italy and Germany during November 2020 (**Figure 4E**).
271 Concerning introduction events, Germany accounted for a large proportion of SARS-CoV-2
272 transmissions, followed by Italy and the United Kingdom (**Figure 4F** and **Figure 4—figure**
273 **supplement 2**). For France, a few introductions came from Russia and Romania between

274 August and September, while most of these events originated from Spain, Italy, Germany
275 Belgium and the United Kingdom during November and December 2020 (**Figure 4G**).

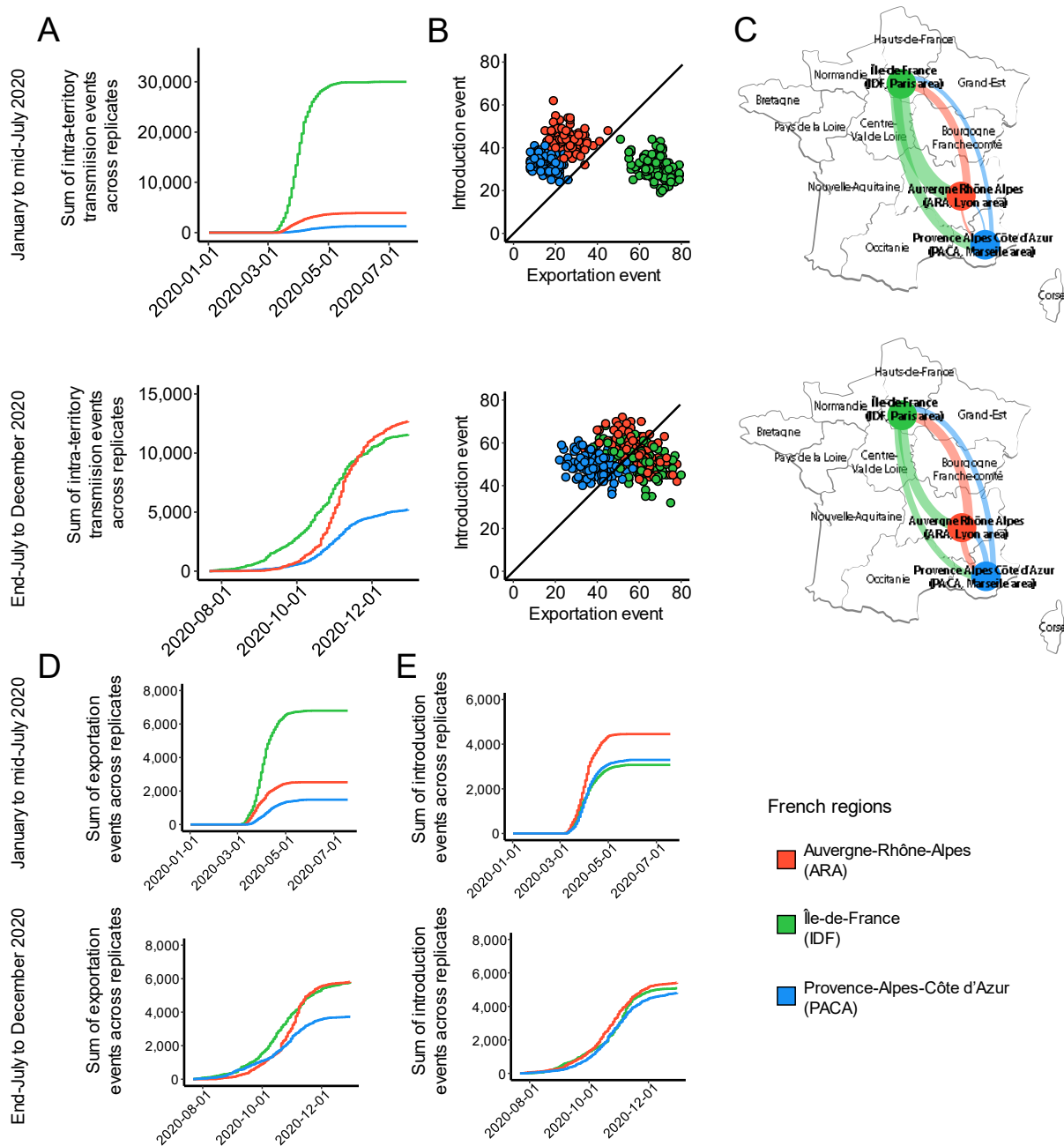
276 Altogether, these results highlight that most governments adopted variable measures to
277 contain COVID-19 pandemic between the first and second European epidemic waves with
278 varying success rates. French people have, in all likelihood, traveled less abroad in Europe
279 during the second wave.

280 **The first and second French epidemic waves: two patterns of SARS-CoV-2** 281 **transmissions within France**

282 We finally conducted an in-depth analysis of virus spread inside France by studying
283 transmission events between French administrative regions (**Figure 5**). As previously
284 explained, only three regions (Auvergne-Rhône-Alpes, ARA, Île-de-France, IDF, and
285 Provence-Alpes-Côte d'Azur, PACA, which cover the most densely populated cities in France,
286 i.e., Paris, Lyon and Marseille) were investigated since only a few or in some cases no
287 sequence at all were available on GISAID for the other regions (**Figure 2—figure supplement**
288 **1**).

289 Similarly to previous scales investigated, intra-region transmissions represented 75.5% of the
290 transmissions during the first wave, but decreased to 66.3% during the second half of 2020
291 likely because of a much less strict national lockdown implemented by the end of the year
292 compared to the first wave (**Figure 5—figure supplement 1**). IDF (Paris area) was associated
293 with much more intra-region transmissions than ARA (Lyon area) and PACA (Marseille area)
294 (**Figure 5A**). Both ARA and PACA had lower intra- than inter-region transmission events during
295 the first wave. During this period, we counted seven and twenty-three times less transmission
296 events in ARA and PACA than in IDF respectively, which was globally consistent with the
297 reported count of SARS-CoV-2-related deaths reported in French hospitals within these
298 regions (Spearman's rank correlation, $p < 0.001$; $r = 0.97$ for the lowest correlation; **Figure 5—**
299 **figure supplement 1**). Furthermore, we observed that the first intra-region transmission

300 events were found in IDF in early-March, then in ARA and PACA about one week later. The
 301 number of intra-region transmissions started to decrease from the end of March, fitting with the
 302 complete lockdown across the entire country that was enforced the 17th March 2020.
 303 Transmissions were very low during the summer 2020. Then, our results indicated a rise of
 304 intra-region transmissions from early September 2020 first in IDF, and three weeks later in
 305 ARA and PACA. During October, intra-region transmissions in ARA were higher than in IDF
 306 and PACA, an observation that contrasts with the first half of 2020. Intra-region transmissions
 307 in PACA were a bit lower than in IDF (**Figure 5—figure supplement 1**).



308

309 **Figure 5. SARS-CoV-2 intra- and inter-territory transmissions in France.** Each French
310 administrative region is associated with a specific color. Transmission events were calculated through
311 100 phylogenies as replicates between January and mid-July 2020, and between end-July to December
312 2020. When unspecified, transmission values are based on the sum of the replicates. **(A)** Evolution of
313 intra-region transmissions over time. **(B)** Number of introduction and exportation events for each
314 replicate and for each French administrative region. **(C)** SARS-CoV-2 exchange flows between French
315 administrative regions during the two time periods investigated. In these plots, migration flow out of a
316 particular location starts close to the outer ring and ends with an arrowhead more distant from the
317 destination location. Migration flow out correspond to the mean of the France sets of sequences. **(D)**
318 Cumulative exportation and **(E)** introduction events per French administrative region over time.

319 We then focused on introduction and exportation events during the first half of 2020. We
320 observed three and four times more exportation events from IDF (Paris area) compared to
321 ARA (Lyon area) and PACA (Marseille area), respectively (**Figure 5B** and **Figure 5—figure**
322 **supplement 2**). Exportation events from IDF were a bit higher in ARA than in PACA (**Figure**
323 **5C**). ARA and PACA had more introduction events compared to IDF, while exportation events
324 from ARA and PACA were mostly oriented towards IDF (**Figure 5C**). The first exportation
325 events originated from IDF in early March, quickly followed, but at much lower rates, by ARA
326 and a few days later by PACA (**Figure 5D**), while the first introduction events were observed
327 in all locations (**Figure 5E**). Viral circulation started to decrease after the implementation of the
328 national lockdown (17th March 2020). During the second half of 2020, IDF was no longer the
329 single epicenter of SARS-CoV-2 inter-region transmissions. Both ARA and PACA participated
330 concomitantly with IDF in virus spread at similar levels (**Figure 5B** to **5E**). Rare exportation
331 and introduction events were detected in early August in all regions. The rate of exportation
332 events rapidly increased first in IDF and PACA from mid-September, then in ARA from end-
333 September and even exceeded other regions by the end of the year (**Figure 5D** and **5E**).

334 Altogether, SARS-CoV-2 transmission events within France showed two distinct patterns
335 between the first and second half of 2020. While most of transmissions and subsequent deaths
336 related to SARS-CoV-2 were localized in the vicinity of Paris during the first half of 2020, those
337 indicators were more balanced after the summer 2020 with an important part of transmissions
338 and deaths around Lyon (ARA) and Marseille (PACA) regions. This may be attributable to
339 French people limiting journeys across Europe during the summer vacations to the benefit of
340 enjoying the French landscapes.

341 Discussion

342 Until now, the dynamics of SARS-CoV-2 transmissions within France or between France and
343 other countries (in Europe or worldwide) remain only partially characterized. To our knowledge,
344 three studies have explored the virus spread in France, but only focused on the first French
345 epidemic wave and/or disregarded transmission events SARS-CoV-2 introductions from other
346 continents or other European countries (Danesh et al., 2020; Elie and Alizon, 2020; Gámbaro
347 et al., 2020). Other studies have focused on larger scale such as the whole Europe, but did
348 not deeply explored the dynamics especially in France (Lemey et al., 2021; Nadeau et al.,
349 2021). Here, we studied introduction and exportation events in France at different scales during
350 the two first European epidemic waves. We think our approach provides a way to tackle large
351 genomic data for future epidemics and allows a deeper comprehension of viral circulation in
352 France and Europe.

353 Methodologically, most phylodynamics studies on SARS-CoV-2, as well as on other viruses or
354 pathogens, are based on Bayesian approaches (notably with the popular and robust BEAST
355 tool; Drummond and Rambaut, 2007) using a unique phylogenetic tree composed of several
356 thousand sequences. To reconstruct the evolution parameters and ancestral states at nodes
357 (here, the geographical area), Bayesian methods require high computational power,
358 representing their main limitation (Drummond and Rambaut, 2007). Overall, the generation of
359 a phylogenetic tree and the ancestral state reconstruction for even a few thousands of
360 sequences is no small feat because requiring very long timeframe, or the parameter estimation
361 can fail to converge. Furthermore, conclusions obtained will be based on a single sampling
362 with a relatively limited number of sequences, but different samplings may lead to some
363 contradictory results (Hall et al., 2016). Since the GISAID database stores several hundred
364 thousand SARS-CoV-2 sequences (n = 638,704 on the 8th May 2020 for the year 2020 alone),
365 we believe that one phylogeny of a few thousand sequences only may not always accurately
366 reflect transmission events. In order to obtain more comprehensive trends drawing conclusions
367 from a maximum of SARS-CoV-2 sequences, we constructed 100 independent maximum

368 likelihood phylogenies of relatively limited size (from 462 to 999 sequences depending on the
369 geographic scale and time period investigated) and inferred ancestral state reconstruction by
370 a maximum likelihood discrete trait phylogeographic method. Bayesian approach seems not
371 to be more accurate than maximum likelihood for the reconstruction of ancestral states
372 (Hanson-Smith et al., 2010), and we gain a lot in terms of sampling, computational time, and
373 trends through distinct tens of replicates. As a direct validation of our approach, we observed
374 that the number of intra-transmission events (i.e., a SARS-CoV-2 transmission within a same
375 territory) was directly correlated to the number of SARS-CoV-2-related deaths reported by
376 national health agencies. Also, the method was recently used to depict with accuracy the
377 genomic epidemiology of the virus in Canada (McLaughlin et al., 2022).

378 The pipeline conducted here allowed to get an in-depth view of SARS-CoV-2 transmissions
379 implying France at both local and international contexts during 2020. During the first European
380 epidemic wave, France mostly exchanged SARS-CoV-2 with Europe and North America. The
381 observation contrasts with the second European epidemic where France exclusively shared
382 the virus with Europe. This result is quite logical insofar as the borders of France have gradually
383 reopened with European countries, but not with the rest of the world. A very low rate of SARS-
384 CoV-2 transmissions between Europe (including France) and Asia or South America was
385 found in 2020, which is not surprising since the virus started to worryingly spread in these
386 continents only from May 2020 where all borders of Europe were closed. At the European
387 scale, France was unambiguously a major actor of SARS-CoV-2 transmission. It was ranked
388 third regarding exportation events (mainly directed towards neighboring countries) only behind
389 the United Kingdom and Italy during the first wave, and second behind Spain during the second
390 one. Temporally, the first introduced European cases into France came from Italy and Spain,
391 consistent with the beginning of SARS-CoV-2 spread in Europe. These results are globally in
392 line with the observations made across Europe through a Bayesian inference (Lemey et al.,
393 2021). For the second European wave, a few introduced cases into France came from Russia
394 (and at lower rates, from Romania) which was associated with an impressive increase in virus

395 incidence since the summer 2020. Then, most virus introductions into France originated from
396 Spain which initiated the second wave in European Union, followed by Italy and Germany.
397 Interestingly, a much higher number of intra-country transmission events was systematically
398 noticed during the second wave compared to the first one, indicating the effectiveness of
399 border control measures and that French people globally not travelled outside France after the
400 first national lockdown.

401 Our results at the French scale should be interpreted in light of several limitations. In addition
402 to a limited overall size, genomic data only cover three out of thirteen administrative regions
403 (**Figure 2—figure supplement 1**). The North and East of France (i.e., regions Hauts-de-
404 France and Grand-Est, respectively) that border Belgium, Luxembourg and Germany, were
405 not investigated. More importantly, the region Grand-Est was associated with early large
406 clusters that participated in virus dissemination across the country during the first European
407 epidemic wave, such as the evangelical gathering of the Christian Open-Door Church
408 (Mulhouse, Haut-Rhin) that took place from February 17th to 21th 2020 and consisted of
409 approximately a thousand contaminated believers. Keeping these limitations in mind, we
410 observed two distinct patterns of SARS-CoV-2 transmissions within France during 2020.
411 During the first European wave, most of the transmissions originated from Île-de-France (IDF;
412 Paris area). Many people living in IDF decided to leave the Paris region at the beginning of the
413 epidemic to live in their second homes in the provinces, which seems to have resulted in a
414 wide spread of the virus seeded from Paris. For the second epidemic wave, we observed a
415 balanced transmission contribution between IDF and the other regions, that can be explained
416 as most people from IDF still resided in the provinces without returning to IDF during the
417 second part of 2020.

418 Overall, using an original approach allowing deeply investigate SARS-CoV-2 phylodynamics
419 taking into account a very large number of sequences, our findings allow a stronger
420 understanding of SARS-CoV-2 transmission events related to France and Europe in 2020 and
421 the changes observed between the two epidemic waves.

422 **Methods**

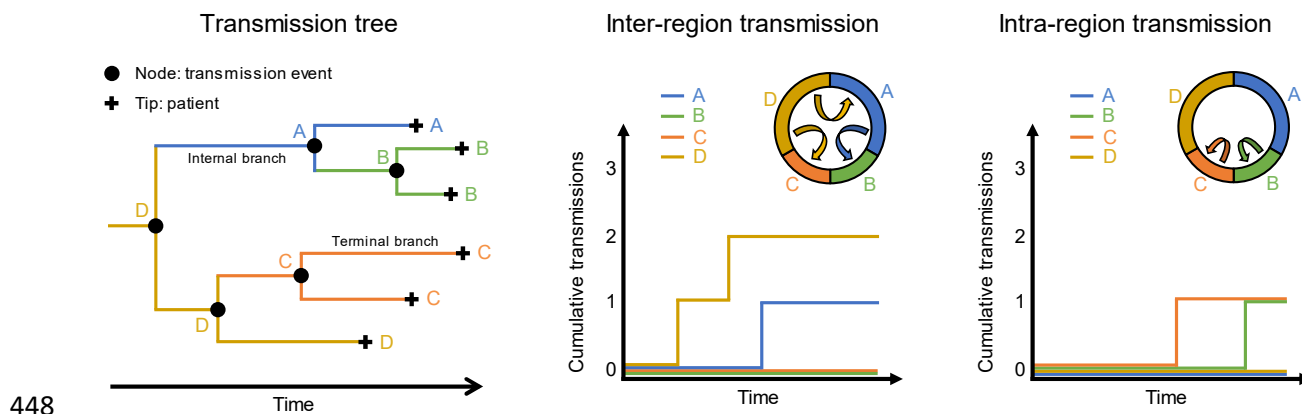
423 **Ethics statement**

424 This study was carried out in accordance with the Declaration of Helsinki. It is a retrospective
425 non-interventional study based on data from the public database repository GISAID
426 (<https://www.gisaid.org/>). The Institutional Review Board – IRB 00006477 – of HUPNVS,
427 Université Paris Cité (AP-HP), has reviewed and approved the research project. Laboratories
428 and organisms that participated in depositing SARS-CoV-2 sequences used in this study can
429 be found on the GISAID database from the identifiers listed in **Supplementary Dataset 1**.

430 **Sequence acquisition and curation, and dataset production**

431 A phylodynamics study (which aims to depict with accuracy transmission events between
432 locations; **Figure 6**), requires a large number of sequences (Attwood et al., 2022). SARS-CoV-
433 2 genome sequences were downloaded from the GISAID database on the 8th May 2022. The
434 following inclusion and exclusion criteria were applied to select samples for analysis: (1) only
435 complete viral genomes from infected individuals (more than 29 kb in length) were included;
436 (2) the maximal proportion of undetermined nucleotide bases was fixed to 5%; (3) sequences
437 with unknown or incomplete sampling dates, or unknown geographical location, were excluded;
438 (4) sampling dates were comprised between January 1st 2020 and December 31st 2020; and
439 (5) sequences with host listed as anything but were excluded. The final dataset was constituted
440 of 638,704 sequences.

441 We studied the epidemic at three different geographic scales: (1) worldwide, to identify SARS-
442 CoV-2 transmissions from continents to France and conversely; (2) at the European scale, to
443 determine more precisely which European countries were mostly involved in transmission
444 flows with France; and finally (3) at the French administrative region level, to get a better
445 understanding of virus spread inside the country. For each geographic scale, the analyses
446 were independently run for the two European epidemic waves (January to mid-July 2020
447 (weeks 1 to 30), and end-July to December 2020 (weeks 31 to 52); **Figure 1**).



448
 449 **Figure 6. Conceptual overview of phylodynamic analysis.** Each tip of a phylogenetic tree
 450 corresponds to a patient infected by SARS-CoV-2 in a given geographical location and at a given date.
 451 Through maximum likelihood, each node may be assigned with a geographical location and each branch
 452 from this node correspond to a transmission event within this location, if the same location is also
 453 observed at the next node, or, in the contrary, from the ancestral node location to the following node
 454 location. The sum of branch lengths to root may be used to calculate the date of transmission. The
 455 number of transmissions inside or outside a same geographical location could then be established.

456 Considering the large number of sequences in GISAID that passed our criteria ($n = 638,704$),
 457 we produced for each investigated geographic scale analysis (i.e., worldwide, Europe, and
 458 French administrative regions) and time period (January to mid-July 2020 and end-July to
 459 December 2020) 100 distinct sets of sequences. To determine the number of SARS-CoV-2
 460 genome sequences to be included so as to be representative of each location epidemic, we
 461 used the number of SARS-CoV-2-related deaths per territory and per week as a proxy for the
 462 number of infections taking place two weeks earlier. For each set, sequences were drawn
 463 randomly from our GISAID extraction without replacement. Some countries and French
 464 regions had to be discarded because of their under-representation in the GISAID database.
 465 When a set was constituted, all the sequences were replaced in the whole dataset for the
 466 random selection of the next one. For the first time period investigated (January to mid-July
 467 2020), an average of 935, 999 and 462 SARS-CoV-2 sequences were included in each set at
 468 the worldwide, European and French scales, respectively (**Table 1**). When considering the
 469 second time period (end-July to December 2020), we used about 300 sequences from the first
 470 wave to anchor phylogenies with the previously acquired viral diversity. In addition to those
 471 sequences, we added worldwide, European and French scale sequence dataset containing
 472 882, 961 and 483 genomic sequences covering July to December 2020 (**Table 1**).

473 **Multiple sequence alignment and maximum likelihood phylogenetic tree**

474 For each set, genome sequences were each independently aligned against the Wuhan-Hu-1
475 reference genome (GenBank accession: NC_045512.2) using MAFFT v7.450 then merged
476 altogether (Kato and Standley, 2013). From each alignment, we inferred a maximum
477 likelihood phylogenetic tree using FastTree v.2.1.11 under a general time-reversible (GTR)
478 model with a discrete gamma distribution to model inter-site rate variation and 1,000 bootstraps
479 to compute support values (Price et al., 2010). The choice of FastTree was mainly due to its
480 short runtime while giving results very close to IQTree 2.1.2 (Minh et al., 2020). Each
481 phylogenetic tree was dated using TreeTime v.0.8.5 (Sagulenko et al., 2018) with a molecular
482 clock rate fixed at 6.00×10^{-3} substitutions per site per year. Note that this evolutionary rate is
483 not in agreement with the one largely reported in other studies (clock rate of $\sim 6.00 \times 10^{-4}$
484 substitutions per site per year; Xia, 2021), but it permits to date the tree without overestimation
485 of branch lengths (**Figure 6—figure supplement 1**). Phylogenetic trees were manipulated and
486 viewed using treeio and ggtree R packages respectively (Wang et al., 2020; Yu et al., 2017).

487 **Worldwide, Europe and French region SARS-CoV-2 phylodynamics analyses**

488 For each phylogenetic tree we produced, ancestral state reconstructions of discrete,
489 geographical data were performed with the ace function of ape package in R (Paradis et al.,
490 2004). Three distinct Markov models of discrete character evolution through maximum
491 likelihood were tested and compared: the Equal Rates (ER) which assumes a single rate of
492 transition among all possible states, the All Rates Different (ARD) which allows a distinct rate
493 for each position transition between two states, and the Symmetrical Rates (SYM) where
494 forward and reverse transitions share the same parameter. In this study, we only present the
495 results obtained with the SYM model since the ARD model led to high variations across
496 replicates probably because the model is too sensible to outliers (**Figure 6—figure
497 supplement 2**), while the ER model led to nearly identical observations compared to the SYM
498 model (data not showed).

499 With the reconstruction, we assigned nodes to a location when supported by $\geq 50\%$ of the state
500 assignment. To detect and count each transition event, we checked for each node of the
501 phylogenetic tree whether children nodes corresponded to the same geographic region as the
502 current node, i.e., a SARS-CoV-2 intra-territory transmission (transmission event that does not
503 cross any administrative border). If they correspond to different geographic locations, we then
504 assumed the transmission from the parent node-state to the given child node-state (SARS-
505 CoV-2 inter-territory transmission). The midpoint of the parent-child branch was chosen as the
506 date of transmission in all cases (**Figure 6**).

507 **Declaration**

508 **Consent for publication**

509 There are no case presentations that require disclosure of respondent's confidential
510 data/information in this study.

511 **Competing interests**

512 The authors declare that they have no competing interests with the current work.

513 **Funding**

514 This study was supported in part by the ANRS MIE (Agence Nationale de la Recherche sur le
515 SIDA et les hépatites virales – Maladies Infectieuses Emergentes), the FRM (Fondation pour
516 la Recherche Médicale) and the Inserm UMR1137 unit.

517 **Availability of data and materials**

518 All sequences used in the study are publicly available on the GISAID database. The list of the
519 GISAID identifiers for each dataset is provided in the **Supplementary Data 1**. All the scripts
520 developed for this study were deposited in the following GitHub repository:
521 <https://github.com/Rcoppee/PhyloCoV>. Large metadata can be shared on demand.

522 **Author's contributions**

523 **Romain Coppée:** Conceptualization, Methodology, Software, Validation, Formal analysis,
524 Investigation, Resources, Data curation, Writing – Original Draft, Writing – Review & Editing,
525 Visualization. **François Blanquart:** Software, Data curation, Writing – Review & Editing. **Aude**
526 **Jary:** Writing – Review & Editing. **Valentin Leducq:** Writing – Review & Editing. **Valentine**
527 **Marie Ferré:** Writing – Review & Editing. **Anna Maria Franco Yusti:** Writing – Review &
528 Editing. **Lena Daniel:** Writing – Review & Editing. **Charlotte Charpentier:** Writing – Review &
529 Editing. **Samuel Lebourgeois:** Writing – Review & Editing. **Karen Zafilaza:** Writing – Review
530 & Editing. **Vincent Calvez:** Writing – Review & Editing. **Diane Descamps:** Writing – Review

531 & Editing. **Anne-Geneviève Marcellin**: Writing – Review & Editing. **Benoit Visseaux**:
532 Conceptualization, Methodology, Validation, Writing – Review & Editing, Supervision, Project
533 administration, Funding acquisition. **Antoine Bridier-Nahmias**: Conceptualization,
534 Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation,
535 Writing – Review & Editing, Supervision, Project administration

536 **Acknowledgements**

537 We gratefully thank all the laboratories and organisms that deposited SARS-CoV-2 sequences
538 on the GISAID dataset, without which this study would not have been possible. All providers
539 of sequences used in this study are available on GISAID from the accession numbers listed in
540 the **Supplementary Data 1**.

541 References

- 542 Attwood SW, Hill SC, Aanensen DM, Connor TR, Pybus OG. 2022. Phylogenetic and
543 phylodynamic approaches to understanding and combating the early SARS-CoV-2
544 pandemic. *Nat Rev Genet* 1–16. doi:10.1038/s41576-022-00483-8
- 545 Bernard Stoecklin S, Rolland P, Silue Y, Mailles A, Campese C, Simondon A, Mechain M,
546 Meurice L, Nguyen M, Bassi C, Yamani E, Behillil S, Ismael S, Nguyen D, Malvy D,
547 Lescure FX, Georges S, Lazarus C, Tabai A, Stempfelet M, Enouf V, Coignard B,
548 Levy-Bruhl D, Investigation Team. 2020. First cases of coronavirus disease 2019
549 (COVID-19) in France: surveillance, investigations and control measures, January
550 2020. *Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull* **25**.
551 doi:10.2807/1560-7917.ES.2020.25.6.2000094
- 552 Cevik M, Kuppalli K, Kindrachuk J, Peiris M. 2020. Virology, transmission, and pathogenesis
553 of SARS-CoV-2. *BMJ* **371**:m3862. doi:10.1136/bmj.m3862
- 554 Danesh G, Elie B, Michalakakis Y, Sofonea MT, Bal A, Behillil S, Destras G, Boutolleau D,
555 Burrel S, Marcelin A-G, Plantier J-C, Thibault V, Simon-Lorier E, Werf S van der,
556 Lina B, Josset L, Enouf V, Alizon S, Group the CSP. 2020. Early phylodynamics
557 analysis of the COVID-19 epidemic in France. doi:10.1101/2020.06.03.20119925
- 558 Deslandes A, Berti V, Tandjaoui-Lambotte Y, Alloui C, Carbonnelle E, Zahar JR, Brichtler S,
559 Cohen Y. 2020. SARS-CoV-2 was already spreading in France in late December
560 2019. *Int J Antimicrob Agents* **55**:106006. doi:10.1016/j.ijantimicag.2020.106006
- 561 Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees.
562 *BMC Evol Biol* **7**:214. doi:10.1186/1471-2148-7-214
- 563 Elie B, Alizon S. 2020. [Sars-CoV-2 genomic and phylodynamic analyses]. *Rev Francoph*
564 *Lab RFL* **2020**:57–62. doi:10.1016/S1773-035X(20)30314-2
- 565 Gámbaro F, Behillil S, Baidaliuk A, Donati F, Albert M, Alexandru A, Vanpeene M, Bizard M,
566 Brisebarre A, Barbet M, Derrar F, van der Werf S, Enouf V, Simon-Lorier E. 2020.
567 Introductions and early spread of SARS-CoV-2 in France, 24 January to 23 March
568 2020. *Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull* **25**.
569 doi:10.2807/1560-7917.ES.2020.25.26.2001200
- 570 Hall MD, Woolhouse MEJ, Rambaut A. 2016. The effects of sampling strategy on the quality
571 of reconstruction of viral population dynamics using Bayesian skyline family
572 coalescent methods: A simulation study. *Virus Evol* **2**:vew003.
573 doi:10.1093/ve/vew003
- 574 Hanson-Smith V, Kolaczkowski B, Thornton JW. 2010. Robustness of ancestral sequence
575 reconstruction to phylogenetic uncertainty. *Mol Biol Evol* **27**:1988–1999.
576 doi:10.1093/molbev/msq081
- 577 Hoertel N, Blachier M, Blanco C, Olfson M, Massetti M, Rico MS, Limosin F, Leleu H. 2020.
578 A stochastic agent-based model of the SARS-CoV-2 epidemic in France. *Nat Med*
579 **26**:1417–1421. doi:10.1038/s41591-020-1001-6
- 580 Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T,
581 Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G,
582 Jiang R, Gao Z, Jin Q, Wang J, Cao B. 2020. Clinical features of patients infected
583 with 2019 novel coronavirus in Wuhan, China. *Lancet Lond Engl* **395**:497–506.
584 doi:10.1016/S0140-6736(20)30183-5
- 585 Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7:
586 Improvements in Performance and Usability. *Mol Biol Evol* **30**:772–780.
587 doi:10.1093/molbev/mst010
- 588 Khare S, Gurry C, Freitas L, Schultz MB, Bach G, Diallo A, Akite N, Ho J, Lee RT, Yeo W,
589 Curation Team GC, Maurer-Stroh S. 2021. GISAID's Role in Pandemic Response.
590 *China CDC Wkly* **3**:1049–1051. doi:10.46234/ccdcw2021.255
- 591 Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. 2020. Projecting the transmission
592 dynamics of SARS-CoV-2 through the postpandemic period. *Science* **368**:860–868.
593 doi:10.1126/science.abb5793

- 594 Krammer F. 2020. SARS-CoV-2 vaccines in development. *Nature* **586**:516–527.
595 doi:10.1038/s41586-020-2798-3
- 596 Lemey P, Ruktanonchai N, Hong SL, Colizza V, Poletto C, Van den Broeck F, Gill MS, Ji X,
597 Levasseur A, Oude Munnink BB, Koopmans M, Sadilek A, Lai S, Tatem AJ, Baele G,
598 Suchard MA, Dellicour S. 2021. Untangling introductions and persistence in COVID-
599 19 resurgence in Europe. *Nature* **595**:713–717. doi:10.1038/s41586-021-03754-2
- 600 McLaughlin A, Montoya V, Miller RL, Mordecai GJ, Canadian COVID-19 Genomics Network
601 (CanCOGen) Consortium, Worobey M, Poon AF, Joy JB. 2022. Genomic
602 epidemiology of the first two waves of SARS-CoV-2 in Canada. *eLife* **11**:e73896.
603 doi:10.7554/eLife.73896
- 604 Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear
605 R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in
606 the Genomic Era. *Mol Biol Evol* **37**:1530–1534. doi:10.1093/molbev/msaa015
- 607 Mofijur M, Fattah IMR, Alam MA, Islam ABMS, Ong HC, Rahman SMA, Najafi G, Ahmed SF,
608 Uddin MA, Mahlia TMI. 2021. Impact of COVID-19 on the social, economic,
609 environmental and energy domains: Lessons learnt from a global pandemic. *Sustain
610 Prod Consum* **26**:343–359. doi:10.1016/j.spc.2020.10.016
- 611 Nadeau SA, Vaughan TG, Scire J, Huisman JS, Stadler T. 2021. The origin and early spread
612 of SARS-CoV-2 in Europe. *Proc Natl Acad Sci* **118**. doi:10.1073/pnas.2012008118
- 613 Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R
614 language. *Bioinformatics* **20**:289–290. doi:10.1093/bioinformatics/btg412
- 615 Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – Approximately Maximum-Likelihood
616 Trees for Large Alignments. *PLOS ONE* **5**:e9490. doi:10.1371/journal.pone.0009490
- 617 Rife BD, Mavian C, Chen X, Ciccozzi M, Salemi M, Min J, Prosperi MC. 2017. Phylodynamic
618 applications in 21st century global infectious disease research. *Glob Health Res
619 Policy* **2**:13. doi:10.1186/s41256-017-0034-y
- 620 Sagulenko P, Puller V, Neher RA. 2018. TreeTime: Maximum-likelihood phylodynamic
621 analysis. *Virus Evol* **4**:vex042. doi:10.1093/ve/vex042
- 622 Santomauro DF, Herrera AMM, Shadid J, Zheng P, Ashbaugh C, Pigott DM, Abbafati C,
623 Adolph C, Amlag JO, Aravkin AY, Bang-Jensen BL, Bertolacci GJ, Bloom SS,
624 Castellano R, Castro E, Chakrabarti S, Chattopadhyay J, Cogen RM, Collins JK, Dai
625 X, Dangel WJ, Dapper C, Deen A, Erickson M, Ewald SB, Flaxman AD, Frostad JJ,
626 Fullman N, Giles JR, Giref AZ, Guo G, He J, Helak M, Hulland EN, Idrisov B,
627 Lindstrom A, Linebarger E, Lotufo PA, Lozano R, Magistro B, Malta DC, Månsson JC,
628 Marinho F, Mokdad AH, Monasta L, Naik P, Nomura S, O'Halloran JK, Ostroff SM,
629 Pasovic M, Penberthy L, Jr RCR, Reinke G, Ribeiro ALP, Sholokhov A, Sorensen
630 RJD, Varavikova E, Vo AT, Walcott R, Watson S, Wiysonge CS, Zigler B, Hay SI, Vos
631 T, Murray CJL, Whiteford HA, Ferrari AJ. 2021. Global prevalence and burden of
632 depressive and anxiety disorders in 204 countries and territories in 2020 due to the
633 COVID-19 pandemic. *The Lancet* **398**:1700–1712. doi:10.1016/S0140-
634 6736(21)02143-7
- 635 Volz EM, Koelle K, Bedford T. 2013. Viral Phylodynamics. *PLOS Comput Biol* **9**:e1002947.
636 doi:10.1371/journal.pcbi.1002947
- 637 Wang L-G, Lam TT-Y, Xu S, Dai Z, Zhou L, Feng T, Guo P, Dunn CW, Jones BR, Bradley T,
638 Zhu H, Guan Y, Jiang Y, Yu G. 2020. Treeio: An R Package for Phylogenetic Tree
639 Input and Output with Richly Annotated and Associated Data. *Mol Biol Evol* **37**:599–
640 603. doi:10.1093/molbev/msz240
- 641 Xia X. 2021. Dating the Common Ancestor from an NCBI Tree of 83688 High-Quality and
642 Full-Length SARS-CoV-2 Genomes. *Viruses* **13**:1790. doi:10.3390/v13091790
- 643 Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. ggtree: an r package for visualization and
644 annotation of phylogenetic trees with their covariates and other associated data.
645 *Methods Ecol Evol* **8**:28–36. doi:10.1111/2041-210X.12628
- 646 Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P,
647 Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W, China Novel Coronavirus
648 Investigating and Research Team. 2020. A Novel Coronavirus from Patients with

649 Pneumonia in China, 2019. *N Engl J Med* **382**:727–733.
650 doi:10.1056/NEJMoa2001017
651

652 **Figure supplement legends**

653 **Figure 1—figure supplement 1.** SARS-CoV-2-related deaths per week of 2020 for each (A)
654 continent and (B) European country. Each color corresponds to a territory.

655 **Figure 2—figure supplement 1.** Differential profiles of SARS-CoV-2 sequences wanted and
656 available for each French region. In this analysis, we requested 500 SARS-CoV-2 for each
657 time period (January to mid-July 2020 and end-July to December 2020) to generate a dataset
658 covering the different French administrative regions. We calculated the differential between
659 the number of sequences available and the number of sequences that are required to properly
660 represent the number of SARS-CoV-2 infections per week. Green and red colors correspond
661 to an excess and a lack of sequences for a week, respectively. For each region, we also
662 calculated the percentage of weeks where a lack of sequences was observed. For each time
663 period, we also indicated the number of sequences wanted.

664 **Figure 2—figure supplement 2.** Within-territory pairwise genetic distance across the
665 replicates for each period at (A) worldwide, (B) European and (C) French administrative region
666 scales. The first row of plots corresponds to the data from January to mid-July 2020, and the
667 second one covers the period of end-July to December 2020. Each color corresponds to a
668 territory. Distances were calculated from the dated trees (with TreeTime) using the *distTips*
669 function of adephylo R package. Abbreviations: ARA, Auvergne-Rhône-Alpes; IDF, Île-de-
670 France; PACA, Provence-Alpes-Côte d'Azur.

671 **Figure 3—figure supplement 1.** Intra-territory SARS-CoV-2 transmissions and correlation
672 with epidemiological data at the worldwide scale. (A) SARS-CoV-2 inter- and intra-territory
673 flows during the first and second half of 2020. In these plots, migration flow out of a particular
674 location starts close to the outer ring and ends with an arrowhead more distant from the
675 destination location. Migration flow out correspond to the mean of the worldwide sets of
676 sequences. (B) Dynamics of intra-territory transmission events across the 100 independent
677 replicates during the first (*left plot*) and second (*right plot*) half of 2020. Each territory is

678 associated with a specific color. Profiles are very similar to the estimated numbers of SARS-
679 CoV-2-related deaths. The curves were smoothed based on a window of seven days. **(C)**
680 Correlation between intra-territory events and the number of SARS-CoV-2-related deaths over
681 time. In a graph, a point corresponds to the cumulative numbers of intra-territory transmissions
682 and the number of deaths at a given date. Each plot and corresponding color correspond to a
683 territory. The coefficient of correlation R is indicated inside the plots.

684 **Figure 3—figure supplement 2.** Inter-territory SARS-CoV-2 transmissions at the worldwide
685 scale. **(A)** Variations of exportation and introduction events across the 100 independent
686 replicates. The two first plots show the variation of the events on the data covering January to
687 mid-July 2020, while the two last plots display the variation between End-July and December
688 2020. Each color corresponds to a territory. The standard deviation corresponds to the sum of
689 the previous replicates. The first replicates are associated with high variations, but tendencies
690 are rapidly reached. **(B)** Mean number of introduction and exportation events for each continent
691 and France. During the first half of 2020, Europe has much more exportation than introduction
692 events. Between end-July and December 2020, all territories harbored nearly equal
693 exportation and introduction events. **(C)** Number of exportation events from France and **(D)**
694 introduction events into France according to the different continents. **(E)** Worldwide maps of
695 continental and France transmissions. The width of the links is proportional to the number of
696 transmission events. The color, specific to each territory, indicates the flow direction. The
697 points that serve to anchor the territories were randomly placed.

698 **Figure 4—figure supplement 1.** Intra-territory SARS-CoV-2 transmissions and correlation
699 with epidemiological data at the European scale. **(A)** SARS-CoV-2 inter- and intra-territory
700 flows during the first and second half of 2020. In these plots, migration flow out of a particular
701 location starts close to the outer ring and ends with an arrowhead more distant from the
702 destination location. Migration flow out correspond to the mean of the sets of sequences. **(B)**
703 Dynamics of intra-territory transmission events across the 100 independent replicates during
704 the first (*left plot*) and second (*right plot*) half of 2020. Each European country is associated

705 with a specific color. Profiles are very similar to the estimated numbers of SARS-CoV-2-related
706 deaths. The curves were smoothed based on a window of seven days. **(C)** Correlation between
707 intra-territory events and the number of SARS-CoV-2-related deaths over time. In a graph, a
708 point corresponds to the cumulative numbers of intra-territory transmissions and the number
709 of deaths at a given date. Each plot and corresponding color correspond to an European
710 country. The coefficient of correlation R is indicated inside the plots.

711 **Figure 4—figure supplement 2.** Inter-territory SARS-CoV-2 transmissions at the European
712 scale. **(A)** Variations of exportation and introduction events across the 100 independent
713 replicates. The two first plots show the variation of the events on the data covering January to
714 mid-July 2020, while the two last plots display the variation between End-July and December
715 2020. Each color corresponds to an European country. The standard deviation corresponds to
716 the sum of the previous replicates. The first replicates are associated with high variations, but
717 tendencies are rapidly reached. **(B)** Mean number of introduction and exportation events for
718 each European country. **(C)** Number of exportation events from France and **(D)** introduction
719 events into France according to the different European countries. **(E)** Maps of European SARS-
720 CoV-2 transmissions. The width of the links is proportional to the number of transmission
721 events. The color, specific to each European country, indicates the flow direction. The points
722 that serve to anchor the territories were randomly placed.

723 **Figure 5—figure supplement 1.** Intra-territory SARS-CoV-2 transmissions and correlation
724 with epidemiological data at the France scale. **(A)** SARS-CoV-2 inter- and intra-territory flows
725 during the first and second half of 2020. In these plots, migration flow out of a particular location
726 starts close to the outer ring and ends with an arrowhead more distant from the destination
727 location. Migration flow out correspond to the mean of the worldwide sets of sequences. **(B)**
728 Dynamics of intra-territory transmission events across the 100 independent replicates during
729 the first (*left plot*) and second (*right plot*) half of 2020. Each territory is associated with a specific
730 color. Profiles are very similar to the estimated numbers of SARS-CoV-2-related deaths. The
731 curves were smoothed based on a window of seven days. **(C)** Correlation between intra-

732 territory events and the number of SARS-CoV-2-related deaths over time. In a graph, a point
733 corresponds to the cumulative numbers of intra-territory transmissions and the number of
734 deaths at a given date. Each plot and corresponding color correspond to a territory. The
735 coefficient of correlation R is indicated inside the plots.

736 **Figure 5—figure supplement 2.** Inter-territory SARS-CoV-2 transmissions at the France
737 scale. **(A)** Variations of exportation and introduction events across the 100 independent
738 replicates. The two first plots show the variation of the events on the data covering January to
739 mid-July 2020, while the two last plots display the variation between End-July and December
740 2020. Each color corresponds to a territory. The standard deviation corresponds to the sum of
741 the previous replicates. The first replicates are associated with high variations, but tendencies
742 are rapidly reached. **(B)** Mean number of introduction and exportation events for each French
743 administrative region. During the first half of 2020, IDF has much more exportation than
744 introduction events. Between end-July and December 2020, all territories harbored nearly
745 equal exportation and introduction events.

746 **Figure 6—figure supplement 1.** Fixing branch lengths using a fixed clock-rate. We used the
747 worldwide dataset covering the whole year 2020, and we focused on the first replicate as an
748 example. **(A)** Maximum likelihood phylogenetic trees obtained after dating the trees using
749 TreeTime, with estimated (*left*) and fixed (*right*) clock-rates. Colors at nodes correspond to
750 dates (gradient color). **(B)** Date of tips and nodes retrieved from the phylogenies after applying
751 an estimated (*left*) or a fixed (*right*) clock-rate. With the estimated clock-rate, consistent with
752 the known evolutionary rate of SARS-CoV-2 (here, 5.80×10^{-4} substitutions per site per year),
753 most nodes are dated between February and April 2020. Very rare nodes were dated after
754 October 2020, despite that some samples were collected after this date. Most samples had
755 very long branches. With the fixed clock-rate (here, 6.00×10^{-3} substitutions per site per year),
756 branch lengths were much short and node dating was consistent with the date of sample
757 collection.

758 **Figure 6—figure supplement 2.** High variations in the number transmission events using the
759 ARD model. We used the worldwide dataset covering the whole year 2020 as an example. **(A)**
760 Variations of exportation and introduction events across the 100 independent replicates. Each
761 color corresponds to a territory. The standard deviation corresponds to the sum of the previous
762 replicates. **(B)** Number of introduction and exportation events for each replicate and for each
763 continent and France. The number of exportation events is very dispersal in the ARD model
764 compared to both SYM and ER models, especially for Africa and Oceania. **(C)** Cumulative
765 exportation events per territory over time. Across the replicates, Africa had similar exportation
766 events than South America. This result is very unlikely since the virus much more circulated in
767 South America compared to Africa. Another unlikely observation is that France had only two
768 times less exportation events than Europe, suggesting that the single France contributed to
769 the half of the exportation events.