



HAL
open science

Analyses génomiques et phylodynamiques du SARS-Cov-2

Baptiste Elie, Samuel Alizon

► **To cite this version:**

Baptiste Elie, Samuel Alizon. Analyses génomiques et phylodynamiques du SARS-Cov-2. Revue Francophone des Laboratoires, 2020, 2020 (526), pp.57-62. 10.1016/S1773-035X(20)30314-2 . hal-02893568v2

HAL Id: hal-02893568

<https://hal.science/hal-02893568v2>

Submitted on 12 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analyses génomiques et phylodynamiques du SARS-Cov-2

Baptiste Elie^{1,2}, Samuel Alizon^{1,*}

¹ Laboratoire MIVEGEC (UMR CNRS 5290, IRD, UM), Montpellier, France

² Département de Biologie, École Normale Supérieure Paris-Saclay, Université Paris-Saclay, Cachan, France

* samuel.alizon@cirs.fr

Résumé

Dans la lignée des épidémies récentes des virus Ebola et Zika, la pandémie de COVID-19 a donné lieu à une avalanche de données génomiques. Celles-ci ont permis de mieux comprendre l'origine de ce virus, de dater son apparition en Chine mais aussi en France et d'analyser la propagation de l'épidémie grâce aux techniques issues du champ émergent qu'est la phylodynamique.

1 Une origine animale

En Décembre 2019, une épidémie de pneumonies d'allure virale d'étiologie inconnue a émergé dans la ville de Wuhan (province de Hubei, Chine) [1]. Le 9 janvier 2020, les autorités sanitaires chinoises et l'Organisation Mondiale de la Santé (OMS) annoncent la découverte d'un nouveau beta-coronavirus, appelé SARS-Cov2, et présenté comme l'agent responsable de ces pneumonies (Covid-19).

Les premières séquences du SARS-Cov2 sont publiées dès le 12 janvier. Les séquences virales les plus proches connues ont 96 % d'identité et proviennent du virus RaTG13, identifié chez une

14 espèce de chauve-souris très répandue dans le sud-est asiatique, du genre *Rhinolophe* (*Rhinolophus*
15 *affinis*)[2]. Les virus humains les plus proches de SARS-Cov-2 sont le virus du SARS et du MERS,
16 qui ont été à l'origine d'épidémies très mortelles mais heureusement limitées, en 2003 et 2012 res-
17 pectivement. On pense que ces deux virus ont comme espèce réservoir les chauve-souris, mais leur
18 propagation à l'homme s'est faite grâce à un hôte intermédiaire, la civette pour le SARS, et le chameau
19 pour le MERS.

20 En plus de montrer que nous avons affaire à un nouveau virus, l'analyse génome a aussi apporté
21 des informations au niveau de son histoire évolutive. Tout d'abord, il faut souligner que le virus de la
22 chauve-souris RaTG13 n'est probablement pas directement la source du SARS-CoV2, avec 4 % de
23 séquences divergentes : le virus animal le plus proche du SARS de 2002, identifié chez des civettes,
24 n'avait que 0,2 % de divergence [3]. Cela signifie que le virus SARS-CoV2 a divergé des virus de
25 chauve-souris voisins que l'on connaît il y a entre 40 et 70 ans [4]. En particulier, la région de la
26 protéine virale « Spike » (ou S), impliquée dans l'entrée des Coronavirus dans les cellules humaines,
27 présente 5 mutations clés et une insertion qui ne sont pas présentes dans le génome du virus RaTG13
28 [5]. En revanche, des pangolins illégalement importés dans la province voisine du Guangdong étaient
29 porteurs de Coronavirus également très similaires au SARS-Cov2 (80 à 88 % d'identité) et contenant
30 en particulier ces mutations. Des hypothèses ont donc été avancées quant à la possibilité d'hôtes
31 intermédiaires ou de recombinaisons entre virus. Cependant, la découverte récente d'un coronavirus
32 de chauve-souris présentant, comme le SARS-Cov-2, un site de clivage au sein la protéine S renforce
33 l'hypothèse que les chiroptères sont les réservoirs du virus [6].

34 Signalons aussi la présence d'un site de clivage polybasique dans S, qui n'a jamais été décrit dans
35 cette lignée de betacoronavirus [5]. La manière dont le clivage de la protéine d'ancrage a lieu est
36 cruciale chez de nombreux virus pour déterminer la spécificité de l'hôte et leur infectivité. Chez les
37 virus de la grippe, ce site de clivage polybasique dans la protéine d'ancrage hémagglutinine (HA)
38 est connu pour évoluer spontanément, notamment dans les élevages industriels de volailles. Il permet

39 une forte augmentation de la virulence du virus, qui peut infecter les voies respiratoires basses. Ces
40 mutations suggèrent une adaptation élevée du virus à l'exploitation de son hôte humain, ce qui est
41 assez cohérent avec la symptomatologie. En effet, l'infection est bien moins virulente que celles
42 causées par le SARS ou le MERS, tout en étant extrêmement contagieuse.

43 Ces analyses de génomique comparative se situent à un niveau multi-spécifique et permettent d'en
44 savoir plus sur les origines lointaines du virus et son adaptation à l'homme. Un important travail reste
45 à effectuer afin d'identifier et de séquencer des virus circulant dans la faune sauvage et domestique
46 potentiellement plus proches du SARS-CoV2. En parallèle, l'analyse de génomes issus d'infections
47 humaines permet elle d'en savoir plus sur l'épidémie en cours.

48 **2 Épidémie à Wuhan**

49 L'intense effort de recherche a permis rapidement de disposer de nombreuses séquences du virus
50 dans la population humaine : fin janvier, plus d'une cinquantaine de séquences étaient publiées. Tout
51 d'abord, ceci a permis de révéler que le virus présentait chez les humains une très faible diversité. Le
52 scénario le plus probable est donc une origine très récente du virus dans la population humaine, avec
53 une introduction unique.

54 Les coronavirus sont des virus à ARN qui mutent relativement rapidement. L'accumulation ré-
55 gulière de mutations au cours des infections peut être utilisée comme une « horloge moléculaire »,
56 permettant de dater des événements épidémiologiques. Une des hypothèses clés est que ces muta-
57 tions sont « neutres », autrement dit elle ne modifient pas la capacité du virus à se propager dans la
58 population. Grâce aux modèles phylogénétiques, on peut ainsi estimer à quoi ressemblait le plus vrai-
59 semblablement le génome des ancêtres communs aux virus circulant dans les populations humaines.
60 Il ne reste plus qu'à compter le nombre de mutations entre les génomes ainsi inférés et les génomes
61 des virus dépistés. Avec environ 2 substitutions par mois qui s'accumulent dans leurs génome, les pre-
62 mières estimations ont cerné l'origine de l'épidémie entre août et décembre 2019 [7], en cohérence

63 avec l'observation des premiers cas de pneumonie de Covid-19 à Wuhan début décembre.

64 Ces approches de datation sont maintenant classiques mais de nouvelles techniques développées
65 depuis une vingtaine d'année permettent d'extraire plus d'information à partir des séquences géné-
66 tiques de virus, en particulier en lien avec la dynamique des populations. Ce champ s'appelle d'ailleurs
67 la « phylodynamique » [8–11]. Ainsi, des modèles probabilistes comme celui du coalescent ou celui
68 des processus de naissance et de mort permettent d'utiliser la structure d'un arbre phylogénétique
69 pour estimer si une épidémie croît ou décroît. Pour l'épidémie de SARS-CoV2, qui au début a cru de
70 manière exponentielle dans la province du Hubei, on a pu calculer que l'épidémie doublait en taille
71 environ toutes les semaines (intervalle de confiance à 95 % entre 4,7 et 16,3 jours) [7, 12]. Il est
72 possible aussi d'estimer le fameux R_0 , c'est-à-dire le nombre de reproduction dans une population
73 entièrement sensible, avec des valeurs dans Wuhan proche de 4 avant le confinement [13]

74 **3 La pandémie**

75 En grande partie du fait de la transmission du virus avant l'apparition des symptômes, l'épidémie
76 dans Wuhan s'est propagée à la Chine avant de devenir pandémique, selon les termes de l'OMS le 11
77 mars 2020. L'analyse des génomes viraux permet de retracer cette expansion grâce aux méthodes de
78 phylogéographie [14]. L'idée est qu'en plus d'inférer la séquence génétique la plus vraisemblable des
79 virus ancestraux, on peut aussi inférer leur localisation géographique.

80 Cette méthode a été popularisée par la plateforme Nextstrain [17], qui permet à tout un chacun de
81 visualiser des phylogénies telles que celle représentée sur la figure 1. D'un point de vue technique,
82 leur approche se décompose en 5 étapes : i) « nettoyer » les données, afin de par exemple retirer
83 les génomes contenant beaucoup d'erreurs de séquençage ; ii) aligner ces séquences, étape clé pour
84 l'inférence phylogénétique ; iii) inférer l'arbre phylogénétique, que l'on peut interpréter comme une
85 sorte d'arbre généalogique reliant les différentes infections dont les virus ont été séquencés ; iv) caler
86 l'arbre dans le temps car à la base sa distance se mesure en nombre de mutations et v) inférer un état

87 géographique pour les nœuds internes de l'arbre, que l'on peut interpréter comme l'endroit où étaient
88 les ancêtres communs aux infections dépistées.

89 Les trois premières étapes effectuées par Nextstrain sont relativement robustes au vu du nombre
90 de séquences analysées (plusieurs milliers). Toutefois, les deux dernières étapes sont à nuancer et ce
91 d'autant plus que ce sont sur elles que s'appuient beaucoup de conclusions. En effet, la calibration
92 dans le temps pour un arbre aussi grand est délicate et nécessite des hypothèses simplificatrices.
93 Cela signifie que les dates des embranchements internes de l'arbre sont à prendre avec précautions.
94 Mais c'est surtout l'estimation de la position géographique des états ancestraux qui peut se révéler
95 particulièrement trompeuse. Imaginons un scénario improbable où aucune séquence de Chine n'aurait
96 été disponible et où le premier pays touché ait été la France, avant que l'épidémie ne touche les
97 USA (comme elle l'a fait). La conclusion de l'analyse phylogéographique serait alors que l'épidémie
98 provient de France. Ce biais lié à l'échantillonnage se retrouve dans toutes ces analyses. Ainsi, il est
99 plus probable de retrouver dans les branches internes de l'arbre des pays qui ont partagé beaucoup
100 plus de séquences que les autres.

101 Malgré cela, les analyses de nextstrain ont révélé une structure mondiale de pandémie très cohé-
102 rente avec les relevés épidémiologiques. Une nomenclature a aussi été proposée afin de décrire au
103 mieux cette structure [18]. D'autres approches phylogéographiques ont aussi pu montrer comment se
104 sont installées les épidémies, avec par exemple des introductions multiples en Caroline du Nord aux
105 États-Unis [19].

106 **4 L'épidémie en France**

107 La France a été le premier pays d'Europe à partager des séquences génomiques de SARS-Cov-2.
108 Il faut dire que le pays a connu les premiers cas de COVID-19 dépistés en janvier. Au 4 avril, la base
109 de donnée en ligne GISAID comptait 196 génomes issus de patients COVID-19 diagnostiqués en
110 France. La plupart ont été obtenus par les deux centres nationaux de référence à partir d'échantillons

111 collectés dans différents centre hospitaliers français.

112 Nos analyses ont permis de montrer plusieurs points. D'une part, la grande majorité des séquences
113 se regroupent dans un clade central (en jaune sur la figure 1). Comme les séquences les plus anciennes,
114 notamment les cas importés détectés en janvier, appartiennent à un clade majoritairement asiatique
115 (en bleu sur la figure 1), l'interprétation la plus parcimonieuse est que ce dernier correspond à la vague
116 épidémique. Les séquences isolées seraient elles associées à des événements d'importation isolés qui
117 n'ont pas conduit à des chaînes de transmission pérennes. À noter que les séquences les plus récentes
118 associées à la vague épidémique ne datent que de fin février, ce qui limite la portée des analyses.
119 À ce jour d'ailleurs, aucun génome issu d'une infection dépistée en France avant fin janvier n'est
120 disponible (et encore, pour fin janvier il s'agit de voyageurs revenant de Chine).

121 Comme pour l'origine de l'épidémie en Chine, nous avons pu dater cette vague épidémique [20].
122 L'incertitude est grande du fait du nombre limité de séquences et du peu de diversité génétique du
123 virus. On peut néanmoins la dater entre la mi-janvier et le début février, ce qui est cohérent avec les
124 estimations obtenues à partir des données d'incidence [21]. Il faut cependant tempérer ces résultats.
125 D'une part, plusieurs régions telles que l'Occitanie n'étaient pas représentées. De plus, il se pourrait
126 qu'il y ait eu des introductions multiples du virus en France, auquel cas l'ancêtre commun de la vague
127 épidémique que nous détectons se situerait alors vraisemblablement hors de France.

128 En plus de dater cette vague épidémique, nous avons aussi pu estimer sa vitesse de propagation
129 [20]. À l'aide de modèles de naissance et de mort, nous avons estimé le nombre de reproduction de
130 l'épidémie (R_t) sur trois périodes temporelles : avant le 19 février, du 19 février au 7 mars et du 7 au
131 24 mars. Ce nombre de reproduction correspond au nombre moyen d'infections engendrées par une
132 personne infectée au cours de son infection [22]. Sur la première période, il est impossible d'estimer
133 la valeur de R_t du fait du manque de données. Sur la seconde période, celui-ci est élevé avec une valeur
134 médiane de 2,6, ce qui est cohérent avec les données d'incidence [21, 23]. Enfin, sur la période plus
135 récente, ce nombre tombe à 1,4. Ces résultats sont cohérents avec la mise en place d'un confinement

136 national au 17 mars.

137 Enfin, un des apports particulièrement intéressant des modèles phylodynamiques est qu'ils per-
138 mettent d'estimer la durée de la période infectieuse [24]. Ces informations sont impossibles à extraire
139 des seules séries temporelles et nécessitent des suivis de chaînes de transmission sur le terrain afin de
140 mesurer le nombre de jours entre l'apparition des symptômes chez une personne infectante et chez une
141 personne infectée (aussi appelé intervalle sériel). Le résultat de l'analyse des génomes associés à la
142 vague épidémique est une durée de contagiosité médiane de 5,1 jours avec un intervalle de confiance
143 à 95 % entre 2,9 et 7,1 jours. Ceci est particulièrement cohérent avec la valeur estimée dans [25] à
144 l'aide de suivi de chaîne de transmission qui est de 5 jours.

145 **5 Perspectives**

146 L'épidémie d'Ebola en Afrique de l'Ouest avait déjà constitué un tournant au niveau de la géo-
147 mique avec la publication d'une étude phylogénétique très rapidement [26] et la démonstration qu'il
148 était possible de séquencer plusieurs génomes entiers par jour dans un pays ravagé par une crise sani-
149 taire [27]. Avec la pandémie de COVID-19, le partage des données génomiques a été particulièrement
150 impressionnant. Non seulement la séquence de ce nouveau virus était publiée le 12 janvier, moins de
151 deux semaines après la détection de l'épidémie, mais en plus beaucoup d'instituts de recherche ont
152 accepté de partager leurs données grâce à la base de données génomiques de grippe GISAID. Les
153 résultats tels qu'ils sont présentés dans cet article sont directement issus de ce partage.

154 Le cas de la France illustre que si des améliorations ont été faites, tout reste loin d'être parfait.
155 Ainsi, la couverture du territoire national dans la base de donnée GISAID est imparfaite. Mais surtout,
156 depuis la fin mars, la contribution de la France a stagné passant de plus de 13,5 % des séquences
157 européennes (67 sur 499) le 23 mars à 1,1 % le 23 juin (394 sur 35.162) derrière la Belgique ou
158 l'Espagne et très loin derrière le Royaume-Uni et l'Islande.

159 Il ne s'agit évidemment pas de séquencer des génomes pour le plaisir. Améliorer la couverture

160 géographique mais surtout temporelle de l'échantillonnage permettrait tout d'abord de mieux ré-
161 pondre à une éventuelle seconde vague. En effet, sans une représentation exhaustive de la diversité
162 génétique en France, il sera difficile de déterminer si les nouveaux cas sont liés à des importations
163 d'autres pays ou au contraire au redémarrage de chaînes de transmission locales. Pour le moment,
164 notre vision phylogénétique de l'épidémie en France est fixée à la fin mars avec en plus des inconnues
165 pour certaines régions. De plus, des données supplémentaires permettraient aussi de mieux analyser
166 la propagation de l'épidémie, à l'heure où les modèles basés sur la détection du nombre de nouveaux
167 cas perdent en puissance avec le reflux de l'épidémie. De plus, ils permettraient d'affiner l'intervalle
168 sériel, c'est-à-dire le nombre de jours séparant deux infections, données toujours inconnues pour
169 l'épidémie en France.

170 L'évolution et la diversité génomique du SARS-Cov-2 au sein des personnes infectées peut aussi
171 avoir des conséquences cliniques, que ce soit pour des questions de résistance aux traitements ou
172 d'échappement aux vaccins [28]. Les premières analyses semblent détecter une diversité génétique
173 non négligeable dans les populations virales intra-patient [29, 30]. Ceci peut sembler surprenant pour
174 des infections aiguës, mais une des explications pourrait être la co-transmission de particules virales
175 lors de l'infection [30].

176 Enfin, une dernière raison de surveiller la diversité génomique est que toutes les mutations fixées
177 dans les génomes ne sont pas neutres, comme l'ont illustré les épidémies de virus Chikungunya à la
178 Réunion [31] ou Ebola en Afrique de l'Ouest [32]. Dans les deux cas, une mutation dans le génome
179 du virus a conféré à ce dernier une capacité accrue à exploiter pour le premier le moustique tigre et
180 pour le second l'homme. Le souci est que le virus fixe plusieurs mutations par mois en moyenne par
181 hasard et distinguer des effets adaptatifs est ardu. Dans le cas du SARS-Cov-2, à ce jour une mutation
182 est candidate à un tel scénario mais les démonstrations sont encore limitées.

183 En conclusion, les virus ne sont pas des être figés. Ils évoluent continuellement et rapidement.
184 Cette évolution les rend particulièrement difficiles à contrôler du fait des mutations adaptatives. Mais

185 elle permet aussi de suivre leur propagation à la trace dans leur génome par le biais de la phylodyna-
186 mique. D'où l'intérêt à combiner microbiologie, évolution et épidémiologie [33].

187 **Encadré : 5 points à retenir**

- 188 — La manière dont les virus se propage laisse des traces dans leurs génomes,
- 189 — la phylodynamie permet de dater des événements d'une épidémie virale et d'estimer la valeur
190 de paramètres clés (nombre de reproduction),
- 191 — les progrès technologiques ont conduit à une production et un partage extrêmement rapide des
192 données de séquence virale,
- 193 — l'origine de l'épidémie a été datée en Chine entre fin août et début décembre 2019, en France
194 la vague épidémique aurait elle commencé vers la mi-janvier,
- 195 — ces résultats sont limités par les biais d'échantillonnage et le peu de différences entre les
196 génomes.

197 **Remerciements**

198 Nous Gonché Danesh pour des commentaires et le groupe de modélisation de l'équipe évolution
199 théorique et expérimentale (ETE) pour des discussions. Nous remercions aussi tous les laboratoires
200 qui ont partagé des données de séquence génétique via la base de données GISAID, permettant ainsi
201 des analyses phylogénétiques rapides de l'épidémie.

202 **Références**

- 203 [1] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan,
204 China, of Novel Coronavirus–Infected Pneumonia. *New England Journal of Medicine*. 2020
205 Jan.
- 206 [2] Coronaviridae Study Group of the ICTV. The species Severe acute respiratory syndrome-related
207 coronavirus : classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol*. 2020 :1–9.
- 208 [3] Song HD, Tu CC, Zhang GW, Wang SY, Zheng K, Lei LC, et al. Cross-host evolution of severe
209 acute respiratory syndrome coronavirus in palm civet and human. *Proceedings of the National
210 Academy of Sciences*. 2005 ;102(7) :2430–2435.
- 211 [4] Boni MF, Lemey P, Jiang X, Lam TTY, Perry B, Castoe T, et al. Evolutionary origins of
212 the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *bioRxiv*.
213 2020 :2020.03.30.015008.
- 214 [5] Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-
215 CoV-2. *Nat Med*. 2020 :1–3.
- 216 [6] Zhou H, Chen X, Hu T, Li J, Song H, Liu Y, et al. A Novel Bat Coronavirus Closely Related
217 to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein.
218 *Current Biology*. 2020 ;30(11) :2196–2203.e3.
- 219 [7] Rambaut A. Phylodynamic Analysis | 176 genomes | 6 Mar 2020 ; 2020. Available from :
220 <http://virological.org/t/phylodynamic-analysis-176-genomes-6-mar-2020/356>.
- 221 [8] Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA, et al. Unifying the epi-
222 demiological and evolutionary dynamics of pathogens. *Science*. 2004 ;303(5656) :327–32.
- 223 [9] Volz EM, Koelle K, Bedford T. Viral phylodynamics. *PLoS Comput Biol*. 2013 ;9(3) :e1002947.

- 224 [10] Frost SD, Pybus OG, Gog JR, Viboud C, Bonhoeffer S, Bedford T. Eight challenges in phylo-
225 dynamic inference. *Epidemics*. 2015;10 :88–92.
- 226 [11] Alizon S, Saulnier E. Phylogénie des infections virales. *Virologie*. 2017;21(3) :119–129.
- 227 [12] Volz E, Baguelin M, Bhatia S, Boonyasiri A, Cori A, Cucunubá Z, et al. Report 5 : Phylogenetic
228 analysis of SARS-CoV-2. 2020 :7.
- 229 [13] Scire J, Vaughan TG, Stadler T. Phylodynamic analyses based on 93 genomes. 2020.
- 230 [14] Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots.
231 *PLoS Comput Biol*. 2009;5(9) :e1000520.
- 232 [15] Sagulenko P, Puller V, Neher RA. *TreeTime* : Maximum-likelihood phylodynamic analysis.
233 *Virus Evol*. 2018;4(1).
- 234 [16] Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain : real-time
235 tracking of pathogen evolution. *Bioinformatics*. 2018;34(23) :4121–4123.
- 236 [17] Bell SM, Hodcroft E, Müller N, Wagner C, Hadfield J, Neher R, et al. Genomic analysis of
237 nCoV spread. Situation report 2020-05-15; 2020.
- 238 [18] Rambaut A, Holmes EC, Hill V, O’Toole A, McCrone JT, Ruis C, et al. A dynamic nomenclature
239 proposal for SARS-CoV-2 to assist genomic epidemiology. *bioRxiv*. 2020 :2020.04.17.046086.
- 240 [19] Deng X, Gu W, Federman S, Plessis Ld, Pybus OG, Faria N, et al. Genomic surveillance reveals
241 multiple introductions of SARS-CoV-2 into Northern California. *Science*. 2020.
- 242 [20] Danesh G, Elie B, Michalakakis Y, Sofonea MT, Bal A, Behillil S, et al. Early phylodynamics
243 analysis of the COVID-19 epidemics in France. *medRxiv*. 2020 :2020.06.03.20119925.

- 244 [21] Sofonea MT, Reyné B, Elie B, Djidjou-Demasse R, Selinger C, Michalakis Y, et al. Epidemiolo-
245 gical monitoring and control perspectives : application of a parsimonious modelling framework
246 to the COVID-19 dynamics in France. medRxiv. 2020 :2020.05.22.20110593.
- 247 [22] Anderson RM, May RM. Infectious Diseases of Humans. Dynamics and Control. Oxford :
248 Oxford University Press ; 1991.
- 249 [23] Salje H, Kiem CT, Lefrancq N, Courtejoie N, Bosetti P, Paireau J, et al. Estimating the burden
250 of SARS-CoV-2 in France. HAL. 2020 :pasteur-02548181.
- 251 [24] Stadler T, Kühnert D, Bonhoeffer S, Drummond AJ. Birth-death skyline plot reveals temporal
252 changes of epidemic spread in HIV and hepatitis C virus (HCV). Proc Natl Acad Sci USA.
253 2013;110(1) :228–33.
- 254 [25] Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-
255 CoV-2 transmission suggests epidemic control with digital contact tracing. Science. 2020.
- 256 [26] Gire SK, Goba A, Andersen KG, Sealfon RSG, Park DJ, Kanneh L, et al. Genomic sur-
257 veillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science.
258 2014;345(6202) :1369–72.
- 259 [27] Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable
260 genome sequencing for Ebola surveillance. Nature. 2016;530(7589) :228–232.
- 261 [28] Hartfield M, Murall CL, Alizon S. Clinical applications of pathogen phylogenies. Trends Mol
262 Med. 2014;20(7) :394–404.
- 263 [29] Shen Z, Xiao Y, Kang L, Ma W, Shi L, Zhang L, et al. Genomic Diversity of Severe Acute
264 Respiratory Syndrome–Coronavirus 2 in Patients With Coronavirus Disease 2019. Clin Infect
265 Dis.

- 266 [30] Lythgoe KA, Hall MD, Ferretti L, Cesare Md, MacIntyre-Cockett G, Trebes A, et al. Sha-
267 red SARS-CoV-2 diversity suggests localised transmission of minority variants. bioRxiv.
268 2020 :2020.05.28.118992.
- 269 [31] Schuffenecker I, Iteman I, Michault A, Murri S, Frangeul L, Vaney MC, et al. Genome microevo-
270 lution of chikungunya viruses causing the Indian Ocean outbreak. PLoS Med. 2006 ;3(7) :e263.
- 271 [32] Urbanowicz RA, McClure CP, Sakuntabhai A, Sall AA, Kobinger G, Müller MA, et al. Human
272 Adaptation of Ebola Virus during the West African Outbreak. Cell. 2016 ;167(4) :1079–1087.e5.
- 273 [33] Alizon S, Méthot PO. Reconciling Pasteur and Darwin to control infectious diseases. PLoS
274 Biol. 2018 ;16(1) :e2003815.

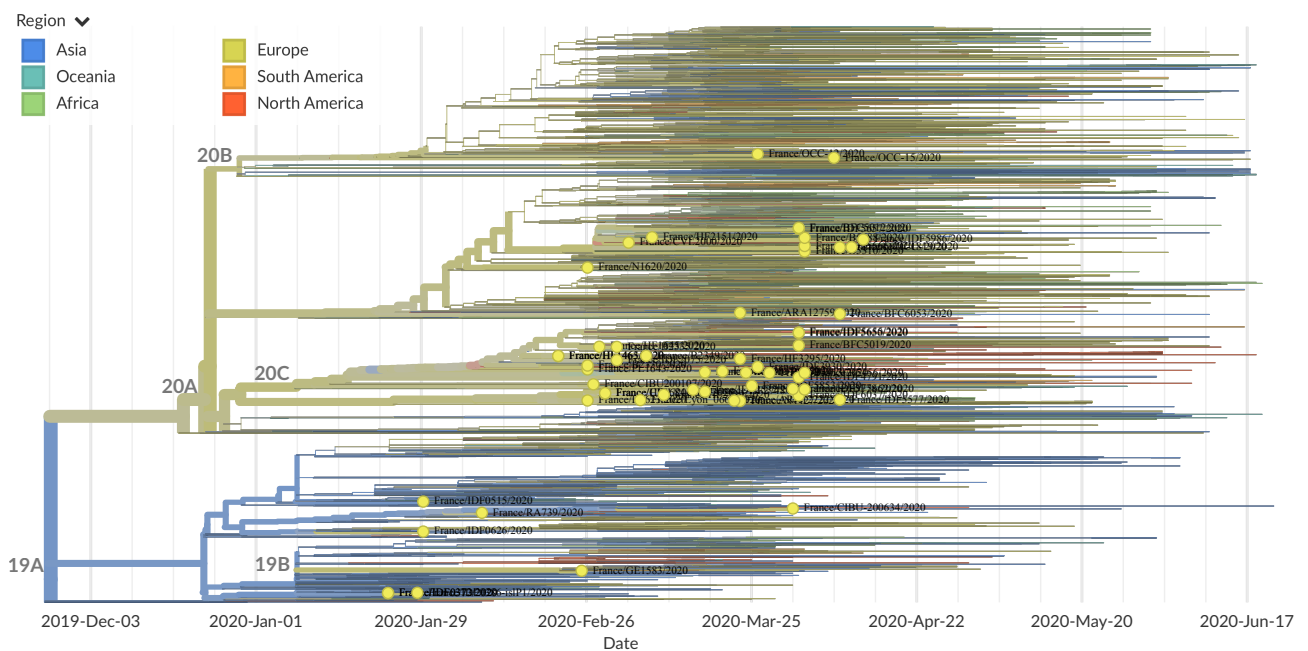


FIGURE 1 – **Phylogénie d’infections de la pandémie de COVID-19 obtenu par Nextstrain.** Cet arbre a été téléchargé le 4 juillet 2020 [15, 16] et obtenu à partir des données de la base GISAID. Les couleurs indiquent les continents sur lesquels ont eu lieu les infections. 57 feuilles correspondant à des séquences issues d’infections en France sont indiquées par des points jaunes. La majorité des séquences, qui appartiennent a priori à la vague épidémique, appartiennent au clade 20A alors que les séquences correspondant aux premiers cas importés sont dans le clade 19A. Cette image est protégée par une licence CC-BY 4.0 de Creative Commons à [nextstrain.org](https://creativecommons.org/licenses/by/4.0/).