

# Introductions et propagation précoce du SRAS-CoV-2 en France, du 24 janvier au 23 mars 2020

Euro Surveill. 2020;25(26):pii=2001200

Fabiana Gámbaro, Sylvie Behillil, Artem Baidaliuk, Flora Donati, Mélanie Albert, Andreea Alexandru, Maud Vanpeene, Méline Bizard, Angela Brisebarre, Marion Barbet, Fawzi Derrar, Sylvie van der Werf, Vincent Enouf, Etienne Simon-Loriere

(traduction post-éditée par N. Bacaër, suggestions d'amélioration : nicolas.bacaer@ird.fr)



## Résumé

Suite à l'émergence du SRAS-CoV-2 en Chine, une surveillance spécifique a été mise en place en France. L'analyse phylogénétique des séquences récupérées grâce à cette surveillance suggère que les introductions initiales détectées, impliquant des virus non clades G, n'ont pas semé la transmission locale. Néanmoins, l'identification des variantes du clade G circulant par la suite dans le pays, la plus précoce provenant d'un patient qui n'a ni voyagé dans des zones à risque ni eu de contact avec des voyageurs, suggère que le SRAS-CoV-2 aurait pu être présent avant les premiers cas locaux enregistrés.

## Texte

Le coronavirus 2 du syndrome respiratoire aigu sévère (SRAS-CoV-2) a été identifié comme la cause d'une épidémie d'infections respiratoires sévères à Wuhan, en Chine, en décembre 2019 [1]. Malgré des mesures de quarantaine strictes à Wuhan et dans les environs, le virus, responsable de la maladie à coronavirus (COVID-19), s'est rapidement propagé à travers le monde, conduisant l'Organisation mondiale de la santé (OMS) à déclarer une pandémie le 11 mars 2020. Peu de temps après l'émergence du virus, une surveillance syndromique spécifique du COVID-19 a été mise en place en France. Parce que la génomique virale, associée à des systèmes de surveillance modernes peuvent aider à comprendre la dynamique des épidémies [2], nous avons séquencé les génomes du SRAS-CoV-2 à partir de cas cliniques échantillonnés dans le cadre de la surveillance.

## Surveillance du COVID-19 dans le nord de la France

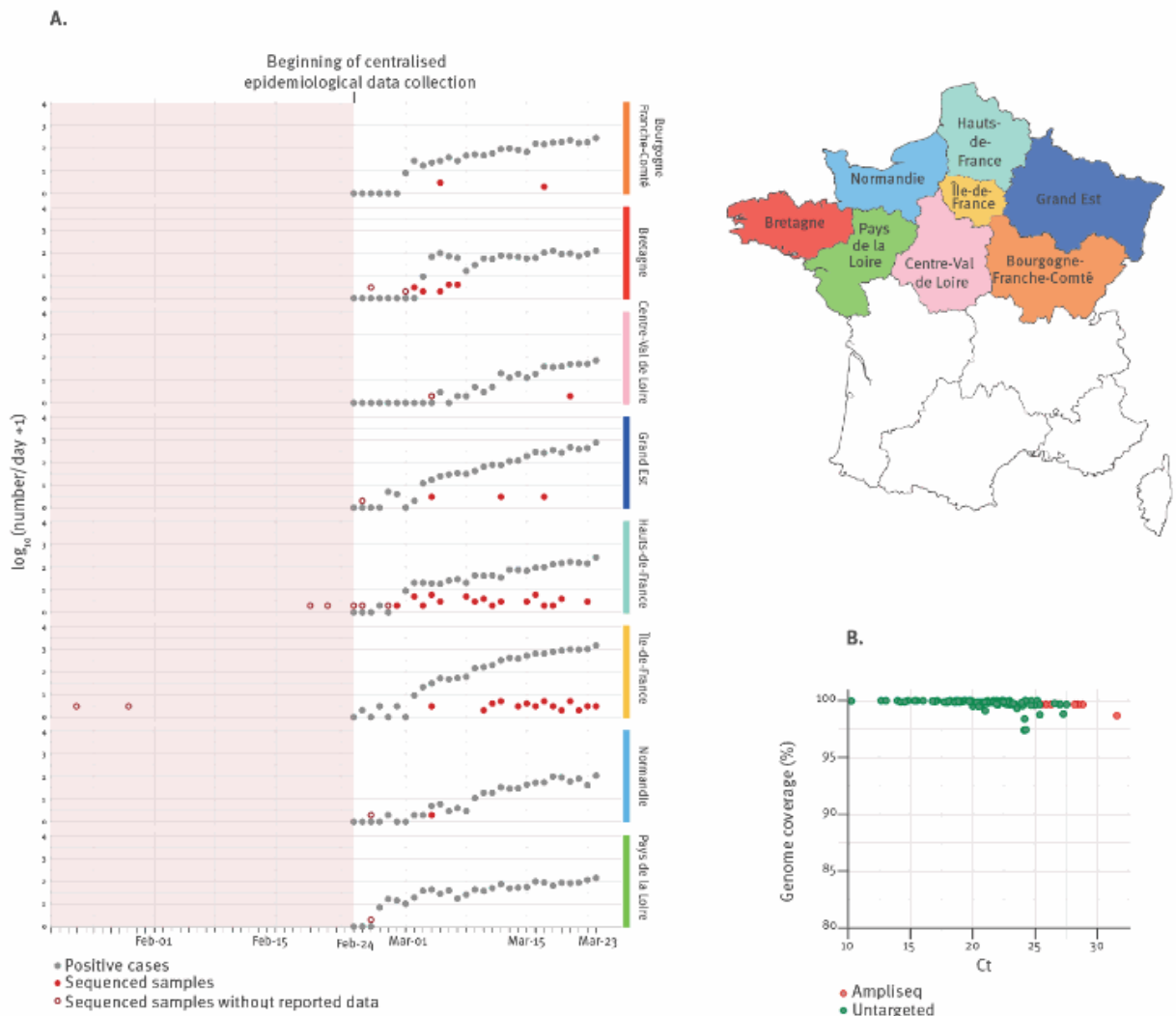
Une surveillance renforcée des cas de COVID-19 a été mise en place en France le 10 janvier 2020, avec pour objectif d'identifier précocement les cas importés pour éviter une transmission secondaire dans la communauté. Dans ce contexte, les premiers cas détectés par le Centre national de référence des virus respiratoires (CNRC) hébergé à l'Institut Pasteur de Paris se sont avérés être les premiers identifiés en Europe. Au fur et à mesure de la progression de l'épidémie de COVID-19 dans le pays, la tâche d'identification des infections au SRAS-CoV-2 a été partagée avec le laboratoire associé au CNRC à Lyon puis étendue aux laboratoires hospitaliers de première ligne dans tout le pays, avec le CNRC à l'Institut Pasteur en se concentrant sur la partie nord de la France, y compris la capitale densément peuplée [3]. Au fur et à mesure que le virus continuait de se propager, il est devenu clair que les patients atteints de COVID-19 pouvaient présenter des caractéristiques cliniques très variables [4], y compris une proportion présentant une infection asymptomatique ou une maladie bénigne [5].

## Échantillonnage des patients et analyse des génomes du SRAS-CoV-2 récupérés

Nous avons généré des séquences complètes du génome du SRAS-CoV-2 à partir d'échantillons de nasopharynx ou de crachats envoyés au CNRC à l'Institut Pasteur dans le cadre de la surveillance continue (Figure 1A). Nous avons combiné les séquences génomiques du SRAS-CoV-2 générées ici, dont 97 du nord de la France et trois d'Algérie avec l'histoire récente de voyages en France, avec 338 séquences publiées et disponibles gratuitement

auprès de l'Initiative mondiale sur le partage de toutes les données sur la grippe (GISAID) EpiCoV base de données et / ou GenBank. Cet ensemble de données a permis de réaliser une analyse phylogénétique pour mieux comprendre les introductions initiales et la propagation du virus en France. Plus de détails sur les méthodes utilisées peuvent être trouvés dans la documentation supplémentaire .

Graphique 1 .Effort de séquençage du génome du SRAS-CoV-2 dans les régions du nord de la France, du 24 janvier au 23 mars 2020



Ampliseq: séquençage basé sur l'amplicon; Ct: seuil de cycle; nombre / jour: nombre de cas confirmés en laboratoire par jour; SRAS-CoV-2: coronavirus du syndrome respiratoire aigu sévère 2.

A. Le graphique représente le nombre de génomes séquencés quotidiennement dans cette étude (cercles pleins ou creux rouges) superposés au nombre de cas positifs rapportés (cercles gris) obtenus auprès de Santé Publique France ( [www.santepubliquefrance.fr](http://www.santepubliquefrance.fr) ). Les cercles creux indiquent les échantillons obtenus à des dates où aucun cas positif pour le SRAS-CoV-2 n'a été signalé. Les données sont présentées séparément pour chaque région du nord de la France comme indiqué sur la carte à droite.

B. Pourcentage de couverture du génome du SRAS-CoV-2 par rapport aux valeurs Ct obtenues à partir de la PCR de transcription inverse en temps réel du SRAS-CoV-2 sur les échantillons originaux, pour les 97 génomes rapportés ici. Pour la fiabilité, le séquençage basé sur l'amplicon a été mis en œuvre pour les échantillons avec des valeurs Ct supérieures à 25. Les couleurs indiquent une approche de séquençage: métagénomique non ciblée (vert) ou séquençage basé sur l'amplicon (rouge).

## Déclaration éthique

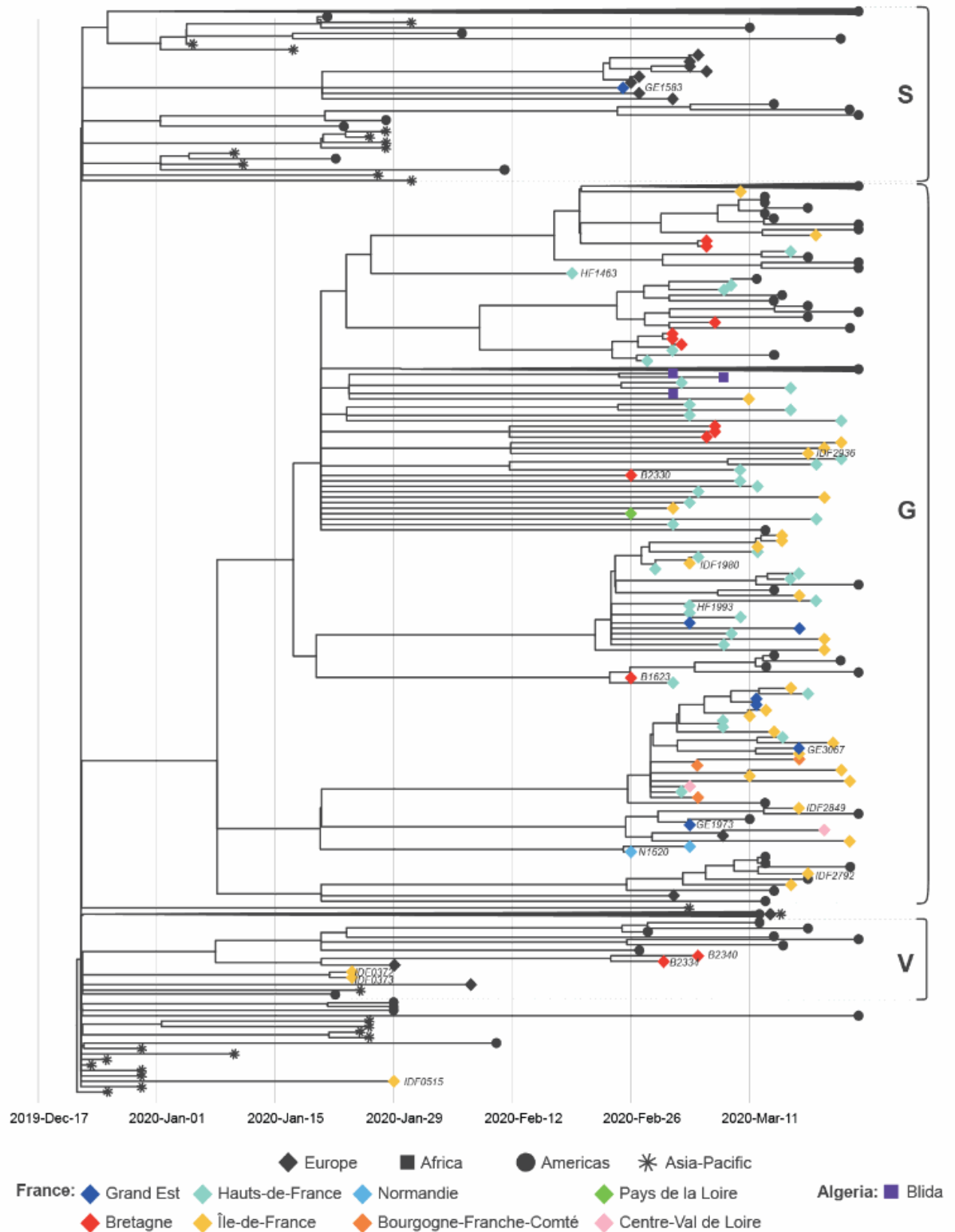
Les échantillons utilisés dans cette étude ont été collectés dans le cadre de la surveillance continue approuvée menée par le CNRC à l'Institut Pasteur (laboratoire de référence de l'OMS fournissant des tests de confirmation

du COVID-19). Les enquêtes ont été menées conformément au règlement général sur la protection des données (règlement (UE) 2016/679 et directive 95/46 / CE) et à la loi française sur la protection des données (loi 78-17 du 06/01/1978 et Décret 2019– 536 du 29/05/2019).

## **Les introductions virales précoces détectées ne semblent pas avoir semé de transmission locale**

Notre analyse indique que la quarantaine imposée aux premiers cas de COVID-19 importés, qui ont été capturés par la surveillance syndromique en France, semble avoir empêché la transmission locale. Les premiers cas européens, originaires d'Île-de-France (IDF) et précédemment décrits ailleurs [6], étaient des importations directes en provenance du Hubei, Chine. Ils ont été échantillonnés le 24 janvier 2020 et les deux génomes viraux respectifs dérivés, IDF0372 et IDF0373, tombent en conséquence près de la base de l'arbre, dans le clade V, selon la nomenclature GISAID ( Figure 2 , Figure 3A). Le clade V est caractérisé par des séquences avec un nucléotide T en position 26144, au lieu d'un G, correspondant à un acide aminé V, plutôt qu'un G, en position 251 de la protéine non structurale 3a. Les génomes IDF0372 et IDF0373 étaient identiques et portaient tous deux une mutation non synonyme G22661T (V367F) dans le domaine de liaison au récepteur de la protéine de pointe, non observée dans d'autres génomes. De même, IDF0515, obtenu à partir d'un échantillon du 29 janvier, correspond à un voyageur du Hubei, Chine. Ce génome basal ne fait pas partie des trois principaux clades proposés par GISAID V, G et S ( figure 2 ), mais porte la mutation G11083T associée à la lignée putative V1 ( figure S2), suggérant une évolution convergente ou une réversion du clade V définissant le changement de G26144T. Des cas précoces subséquents détectés en février dans l'Ouest (Bretagne; B) ou l'Est (Grand Est; GE) de la France (B2334 / B2340, clade V et GE1583, clade S), tous avec des antécédents récents de voyage en Italie, s'ajoutent au diversité génomique des virus du nord de l'Italie, mais ne semblent pas non plus avoir semé de transmission locale dans l'ensemble de séquences échantillonnées actuellement ( figure 2 ).

Graphique 2 .Analyse phylogénétique de séquences d'introductions précoces et de SRAS-CoV-2 en circulation dans le nord de la France, 24 janvier-23 mars 2020

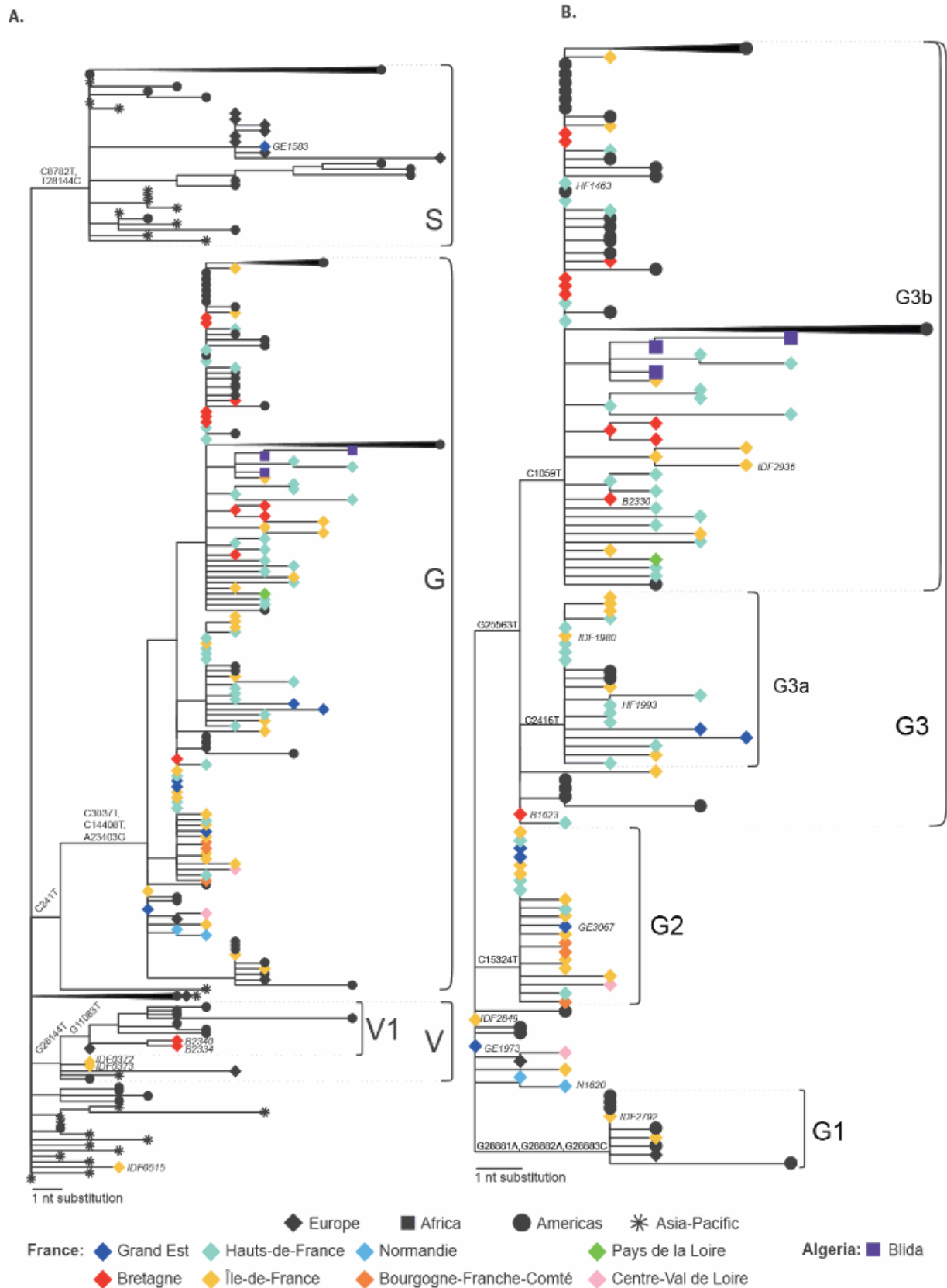


GISAID: Initiative mondiale sur le partage de toutes les données sur la grippe; SRAS-CoV-2: coronavirus du syndrome respiratoire aigu sévère 2.

Arbre calibré dans le temps de 438 séquences de SRAS-CoV-2 comprenant le nord de la France, l'Algérie et des séquences mondiales accessibles au public. L'arbre est enraciné à l'aide de la souche de référence Wuhan / Hu-1 / 2019n (numéro d'accès GenBank: MN908947). Les extrémités de l'arbre sont façonnées et colorées en fonction du lieu d'échantillonnage. La longueur des branches est proportionnelle à l'intervalle de temps entre la date d'échantillonnage et la date déduite de l'ancêtre commun le plus récent. Les trois principaux clades selon la nomenclature GISAID sont indiqués. Les

noms de souches des séquences discutées dans cette étude sont indiqués à côté des conseils correspondants.

Graphique 3 .Arbres phylogénétiques avec séquences de SRAS-CoV-2 montrant les clades (A) S, G, V et (B) du clade G, avec détails sur les lignées correspondantes, nord de la France, 24 janvier-23 mars 2020



GISAID: Initiative mondiale sur le partage de toutes les données sur la grippe; SRAS-CoV-2: coronavirus du syndrome respiratoire aigu sévère 2.

Les pointes des arbres sont façonnées et colorées en fonction du lieu d'échantillonnage. Les longueurs des branches sont proportionnelles au nombre de substitutions de nucléotides de la référence et de la

racine de l'arbre Wuhan / Hu-1/2019 (numéro d'accès GenBank: MN908947). Les clades GISAID et les lignées putatives sont indiqués à droite de chaque panneau. Les noms de souches des séquences discutées dans cette étude sont indiqués à côté des conseils correspondants en italique. Les substitutions de nucléotides partagées entre toutes les séquences de chaque clade ou lignée sont indiquées à côté des nœuds correspondants. Certaines lignées monophylétiques sont regroupées pour faciliter la représentation. Un arbre complet est présenté sur la figure S1 .

## Clades et lignées de SRAS-CoV-2 circulant encore dans le nord de la France

Toutes les autres séquences du nord de la France appartiennent au clade G, défini par deux mutations synonymes (C241T, C3037T) et une substitution non synonyme (A23403G) correspondant à une mutation D614G dans la protéine de pointe ( figure 3 ), et cela inclut les séquences capturées pendant la forte augmentation des cas signalés dans de nombreuses régions fortement touchées ( figure 1 ). Si un échantillonnage plus approfondi sera nécessaire pour confirmer cette hypothèse, l'analyse phylogénétique des séquences récupérées dans la présente étude suggère que l'épidémie française a été principalementensemencée par une ou plusieurs variantes de ce clade, contrairement à ce qui est observé pour de nombreux autres pays européens ( [https://nextstrain.org/ncov/europe?f\\_region=Europe](https://nextstrain.org/ncov/europe?f_region=Europe) ) [7 , 8]. Ce clade peut être davantage classé en lignées (appelées putativement G1, G2, G3, G3a, G3b), bien qu'à nouveau soutenu par seulement une à trois substitutions. Les lignées sont pour la plupart représentées respectivement par des séquences de plusieurs régions. Plusieurs génomes correspondent à des patients en GE, Normandie (N), IDF, Hauts-de-France (HF) et B avec des antécédents récents de voyage en Europe (GE3067, N1620, IDF2792), aux Emirats Arabes Unis (IDF2936), à Madagascar (HF1993 ), Égypte (B1623, B2330) ou reliées aux aéroports parisiens (IDF1980). Ces génomes pourraient représenter des introductions supplémentaires du même clade, puisque les cas respectifs ont été testés positifs pour le virus alors que d'autres infections locales par le virus du clade G avaient déjà été détectées dans le nord de la France. En revanche, dans la lignée G3b,

La surveillance syndromique a permis de capturer l'un des premiers représentants du clade G (HF1463, échantillonné le 19 février) ( Figure 2 ). Surtout, cette séquence porte deux mutations supplémentaires par rapport à la séquence ancestrale reconstruite de ce clade ( figure 3B ). D'autres séquences échantillonnées des semaines plus tard (IDF2849, GE1973) sont plus basales au clade, mettant en évidence la complexité et le risque des inférences basées sur 1 ou 2 substitutions nucléotidiques. Pour cette raison, et la rareté des séquences précoces dans de nombreux pays d'Europe, les estimations phylogéographiques au niveau national et intra-pays ne sont pas fiables avec l'ensemble de données actuel. Il est donc impossible de déduire avec certitude comment le virus a été introduit en France à partir de l'épicentre de l'épidémie, et de multiples voies sont possibles.

## Discussion

Les génomes générés dans cette étude fournissent plus d'informations sur les clades et les variants du SRAS-CoV-2 circulant dans le nord de la France au début de l'épidémie et plus tard pendant la pandémie. Les résultats des analyses semblent indiquer que, au moins pour les premiers cas importés qui ont pu être capturés par la surveillance, ces introductions n'ont pas conduit à une nouvelle transmission du virus dans la communauté. En effet, les séquences de cas importés détectés au début de l'épidémie n'appartenaient pas au clade G, un clade identifiant tous les génomes récupérés plus tard dans l'épidémie. Au sein du clade G, un certain nombre de variantes ont pu être observées. De plus, le premier patient infecté par un représentant du clade G (HF1463) n'avait aucun antécédent de voyage ou de contact avec des voyageurs de retour, ce qui suggère que le virus circulait silencieusement en France en février,5] et observations dans d'autres pays européens [9 , 10]. Bien que cela soit également compatible avec le temps jusqu'à l'estimation de l'ancêtre commun le plus récent pour le clade G ( figure 2 ), l'échantillonnage actuel empêche clairement une inférence fiable pour le moment de l'introduction en France. De plus, si les données actuelles peuvent conduire à émettre l'hypothèse que l'épidémie française aurait pu être principalementensemencée par une ou plusieurs variantes du clade G, des données supplémentaires seront nécessaires pour le confirmer. Une autre explication serait que si l'épidémie a commencé avec des virus appartenant à divers clades, le clade G aurait pu devenir dominant dans le nord de la France à mesure que l'épidémie progressait.

Surtout, alors que tous les premiers échantillons de cas suspects de COVID-19 symptomatiques ont été envoyés au CNRC pour être testés, ce n'était plus le cas à mesure que l'épidémie se développait ( Figure 1A ). De plus, les cas pauci ou asymptomatiques sont à peine représentés dans notre ensemble de données. Cette étude révèle

également des domaines d'amélioration potentielle de la surveillance génomique du SRAS-CoV-2 en France car plusieurs régions sont mal représentées ( Figure 1A). Cela est probablement dû au lourd fardeau pesant sur les hôpitaux, qui, tout en étant en mesure de réaliser des tests locaux grâce au partage rapide des outils de détection moléculaire par le CNRC, auraient pu avoir à réduire le nombre d'échantillons positifs envoyés pour confirmation et séquençage au CNRC. . Pour cette raison, et de la surveillance uniquement syndromique, nous sous-estimons probablement la diversité génétique du SRAS-CoV-2 circulant en France.

En conclusion, notre étude met en lumière l'origine et la diversité de l'épidémie de COVID-19 en France avec des perspectives pour l'Europe, et met en évidence les enjeux des mesures d'endiguement lorsqu'une proportion importante de cas est asymptomatique.

### **Disponibilité des données et des matériaux**

Les génomes du SARS-CoV-2 assemblés générés dans cette étude ont été déposés dans la base de données GISAID ( <https://www.gisaid.org/> ) dès qu'ils ont été générés, les numéros d'accès peuvent être trouvés dans les données S1 (tableau S2) .

### **Remerciements**

Nous tenons à remercier tous les travailleurs de la santé, les employés de la santé publique et les scientifiques impliqués dans la réponse au COVID-19.

Nous remercions les laboratoires hospitaliers du réseau RENAL du nord de la France (liste des noms en Données S1, tableau S4) .

Nous remercions les auteurs, laboratoires d'origine et soumissionnaire des séquences de GISAID et GenBank ( données S1, tableau S2 ). Nous avons évité toute analyse directe des données génomiques non soumises dans le cadre de cet article et avons utilisé ces données génomiques uniquement comme arrière-plan.

Nous remercions Laurence Ma (Plateforme Biomique, C2RT, Institut Pasteur, Paris, France) pour le séquençage MiSeq.

Ces travaux ont utilisé les services de calcul et de stockage (cluster TARS) fournis par le service informatique de l'Institut Pasteur, Paris.

FG fait partie du programme doctoral international Pasteur-Université Paris (PPU), école doctorale BioSPC.

### **Relevé de financement**

Cette étude a été financée par l'Institut Pasteur, CNRS, Université de Paris, Santé publique France, le programme Investissement d'Avenir du gouvernement français, Laboratoire d'Excellence «Biologie intégrative des maladies infectieuses émergentes» (bourse n ° ANR-10-LABX- 62-IBEID), REACTing (Research & Action Emerging Infectious Diseases), France Génomique (ANR-10-INBS-09-09), IBISA et le projet RECOVER financé par le programme de recherche et d'innovation Horizon 2020 de l'Union européenne dans le cadre d'une convention de subvention Non 101003589. ESL reconnaît le financement du programme INCEPTION (bourse Investissements d'Avenir ANR-16-CONV-0005).

### **Conflit d'intérêt**

Aucun déclaré.

### **Contributions des auteurs**

SB, FDo, MA, AA, MV, MBI, ABr, MBa, FG - enquête; FDe - ressources; FG, ABa, ESL - curation et analyse des données, visualisation, rédaction du projet original; SB, VE, ESL, SvdW - rédaction, révision et édition; SvdW, VE, ESL - conceptualisation des études, ressources, supervision; SvdW, ESL - acquisition de financement.

### **Références**

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. , China Novel Coronavirus Investigating and Research Team. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med.* 2020;382(8):727-33. <https://doi.org/10.1056/NEJMoa2001017> PMID: 31978945
2. Grubaugh ND, Ladner JT, Lemey P, Pybus OG, Rambaut A, Holmes EC, et al. Tracking virus outbreaks in the twenty-first century. *Nat Microbiol.* 2019;4(1):10-9. <https://doi.org/10.1038/s41564-018-0296-2> PMID: 30546099
3. Bernard Stoecklin S, Rolland P, Silue Y, Mailles A, Campese C, Simondon A, et al. , Investigation Team. First cases of coronavirus disease 2019 (COVID-19) in France: surveillance, investigations and control measures, January 2020. *Euro Surveill.* 2020;25(6):2000094. <https://doi.org/10.2807/1560-7917.ES.2020.25.6.2000094> PMID: 32070465
4. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. , China Medical Treatment Expert Group for Covid-19. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med.* 2020;382(18):1708-20. <https://doi.org/10.1056/NEJMoa2002032> PMID: 32109013
5. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science.* 2020;368(6490):489-93. <https://doi.org/10.1126/science.abb3221> PMID: 32179701
6. Lescure F-X, Bouadma L, Nguyen D, Parisey M, Wicky P-H, Behillil S, et al. Clinical and virological data of the first cases of COVID-19 in Europe: a case series. *Lancet Infect Dis.* 2020;20(6):697-706. [https://doi.org/10.1016/S1473-3099\(20\)30200-0](https://doi.org/10.1016/S1473-3099(20)30200-0) PMID: 32224310
7. Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl GL, et al. Spread of SARS-CoV-2 in the Icelandic Population. *N Engl J Med.* 2020;382(24):2302-15. <https://doi.org/10.1056/NEJMoa2006100> <https://doi.org/10.1056/NEJMoa2006100> PMID: 32289214
8. Zehender G, Lai A, Bergna A, Meroni L, Riva A, Balotta C, et al. Genomic characterization and phylogenetic analysis of SARS-COV-2 in Italy. *J Med Virol.* 2020. <https://doi.org/10.1002/jmv.25794> PMID: 32222993
9. Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl GL, et al. Spread of SARS-CoV-2 in the Icelandic Population. *N Engl J Med.* 2020;382(24):2302-15. <https://doi.org/10.1056/NEJMoa2006100> PMID: 32289214
10. Onder G, Rezza G, Brusaferro S. Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. *JAMA.* 2020. <https://doi.org/10.1001/jama.2020.4683> PMID: 32203977