HAL open science

Phylodynamique du COVID-19 en France

Gonché Danesh, Baptiste Elie, Samuel Alizon

To cite this version:

Gonché Danesh, Baptiste Elie, Samuel Alizon. Phylodynamique du COVID-19 en France. [Rapport de recherche] 4, Centre National de la Recherche Scientifique (CNRS); Institut de Recherche pour le Développement (IRD); Université de Montpellier (UM), FRA. 2020. hal-02882688

HAL Id: hal-02882688

https://hal.science/hal-02882688v1

Submitted on 9 Jul 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.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Préambule

Phylodynamique

Analyse des données de séquences

Analyse via nextstrain

Signal phylogénétique

Structure de la phylogénie d'infection

Analyses Beast

Clade principal en France

Conclusions

Logiciels utilisés

Sources et remerciements

Phylodynamique du COVID-19 en France

Groupe de modélisation de l'équipe ETE (https://www.mivegec.ird.fr/fr/contact/160-francais/equipes/1205-ete) (Laboratoire MIVEGEC (http://mivegec.ird.fr), CNRS, IRD, Université de Montpellier)

Le 30 mars 2020 (mis à jour le 24 avril).

Préambule

Ce rapport a été produit à des fins académiques et ne constitue pas un support de prise de décision. De plus, les calculs réalisés ici sont faits avec des hypothèses volontairement simplifiées.

L'ensemble de nos rapports sur la pandémie de COVID-19 est disponible sur cette page (http://covid-ete.ouvaton.org).

En matière de santé publique et pour toute question, nous recommandons de consulter et suivre les instructions officielles disponibles sur https://www.gouvernement.fr/info-coronavirus (https://www.gouvernement.fr/info-coronavirus)

L'Organisation Mondiale de la Santé (OMS) dispose aussi d'un site très complet https://www.who.int/fr/emergencies/diseases/novel-coronavirus-2019 (https://www.who.int/fr/emergencies/diseases/novel-coronavirus-2019)

Nous remercions les patients, personnels de santé et les laboratoires de virologie de France qui ont permis de générer les données de séquences de SRAS-Cov-2 à la base de ce travail. Pour plus de détails, voir la liste des données utilisées.

Phylodynamique

Depuis le début des années des 2000, un champ émergent appelé **phylodynamique** propose d'estimer des valeurs de paramètres d'intérêt épidémiologique à partir de données de séquences génétiques et à l'aide de modèle statistiques. Pour plus de détails en français, on peut se référer à cet article (https://hal.archives-ouvertes.fr/hal-01638067) et pour plus de détail en anglais on peut voir une revue plus complète (https://doi.org/10.1371/journal.pcbi.1002947).

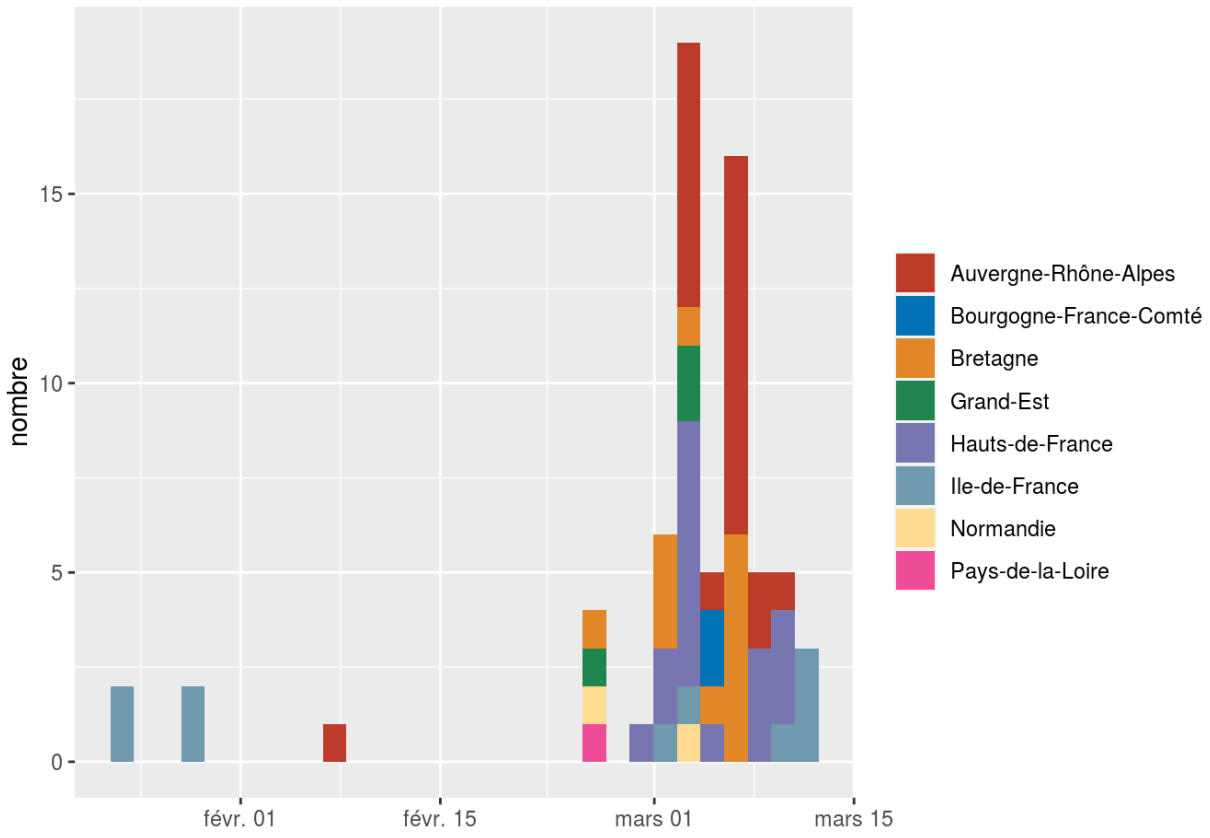
L'idée de la phylodynamique est que la manière dont les virus se propagent laisse des traces dans leur génome. En analysant des génomes issues de plusieurs patient·e·s infecté·e·s, on peut reconstruire un arbre phylogénétique d'infections, qui peut être mis en relation avec une chaîne de transmission. Cet arbre a des feuilles (les extrémités) et des branches. En théorie, les séquences proches dans la phylogénies ont plus de chances d'être issues de patient·e·s proches dans la chaîne de transmission. Toujours en termes d'analogie, si toutes les infections de l'épidémie étaient échantillonnées, on pourrait interpréter chaque feuille de l'arbre comme une fin d'infection et chaque embranchement de l'arbre comme un événement de transmission.

Plusieurs analyses ont déjà été réalisées au niveau mondial, qui sont disponibles sur le site virological.org/ (http://virological.org/). Le site nextstrain.org (https://nextstrain.org/) permet lui de suivre la propagation géographique de plusieurs épidémies virales quasiment en temps réel à l'aide des données de séquences disponibles.

Ce rapport se focalise sur l'épidémie en France en utilisant les données de SRAS-Cov-2 disponibles sur la base de données GISAID (https://www.gisaid.org/) au 28 mars 2020, soit 69 séquences pour la France. Ces séquences ne peuvent pas être publiées dans cette analyse car elles sont pour la plupart sous embargo.

Analyse des données de séquences

Le graphique suivant représente les séquences utilisées, classées en fonction de la date à laquelle le prélèvement a été effectué et de la région de prélèvement.



À noter que 7 séquences ont du être retirées de la base de données GISAID (https://www.gisaid.org/) car elles provenaient des mêmes patient·e·s, ou car elles provenaient de passages en cultures cellulaires.

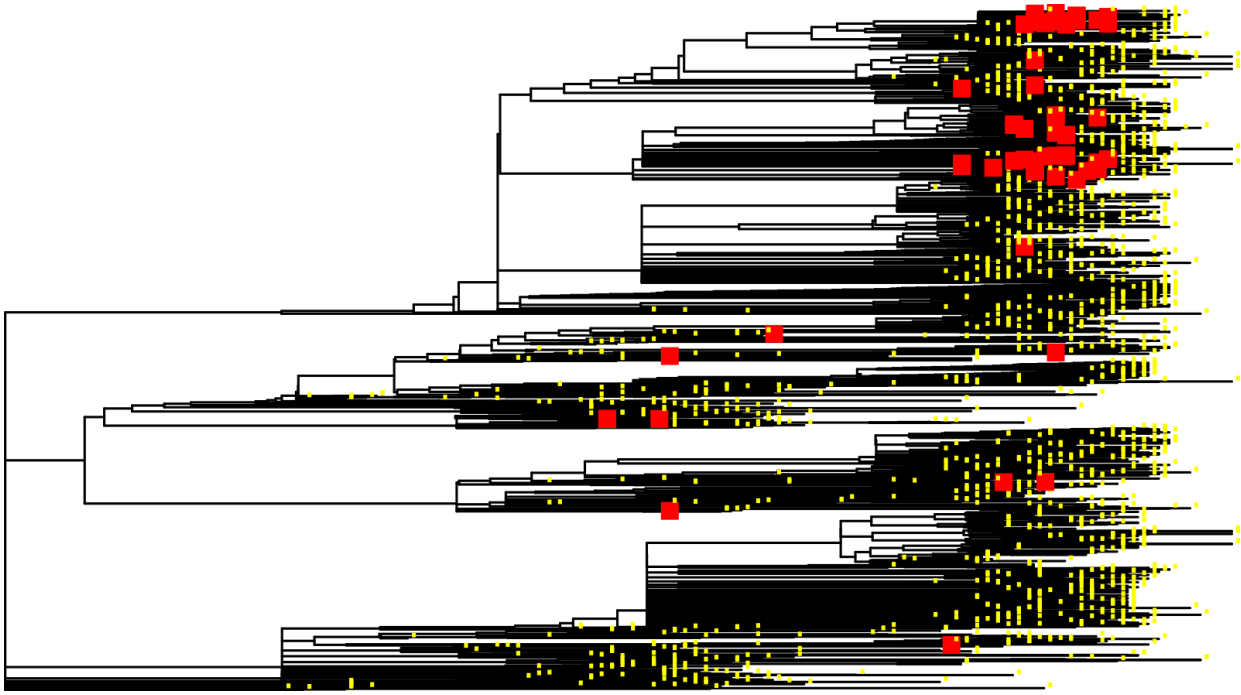
La longueur des séquences exploitable est quasi-parfaite, comme le montre la médiane et l'intervalle de confiance à 95 % du pourcentage de génome couvert par rapport à la séquence de référence (MN908947 (https://www.ncbi.nlm.nih.gov/nuccore/MN908947)) :

```
##  2.5%   50% 97.5%
## 99.67 99.91 99.99
```

Les séquences ont ensuite été alignées et nettoyées à l'aide de la pipeline Augur développée par nextstrain (https://nextstrain.org/docs/bioinformatics/introduction-to-augur). Nous avons pour le moment supposé que les événements de recombinaisons dans les génomes étaient négligeables.

Analyse via nextstrain

La plateforme en ligne de nextstrain (https://nextstrain.org/ncov?f_country=France) permet de visualiser et d'analyser les 68 séquences mentionnées ci-dessus. Quand on les observe sur la phylogénie des séquences mondiales on voit qu'elles sont réparties dans plusieurs clades. On a donc vraisemblablement eu *plusieurs introductions en France*. En effet, si l'épidémie française n'avait été causée que par une seule introduction, alors le dernier ancêtre commun à toutes les séquences françaises n'aurait pas été partagé par des séquences hors de France, ce qui n'est pas le cas dans cet arbre



Séquences françaises (en rouge) parmi la phylogénie mondiale de SRAS-Cov-2 *Phylogénie téléchargée depuis le site nextstrain (https://nextstrain.org/ncov?f_country=France) et annotée en R (http://cran.r-project.org/) grâce au package ape (http://ape-package.ird.fr/).

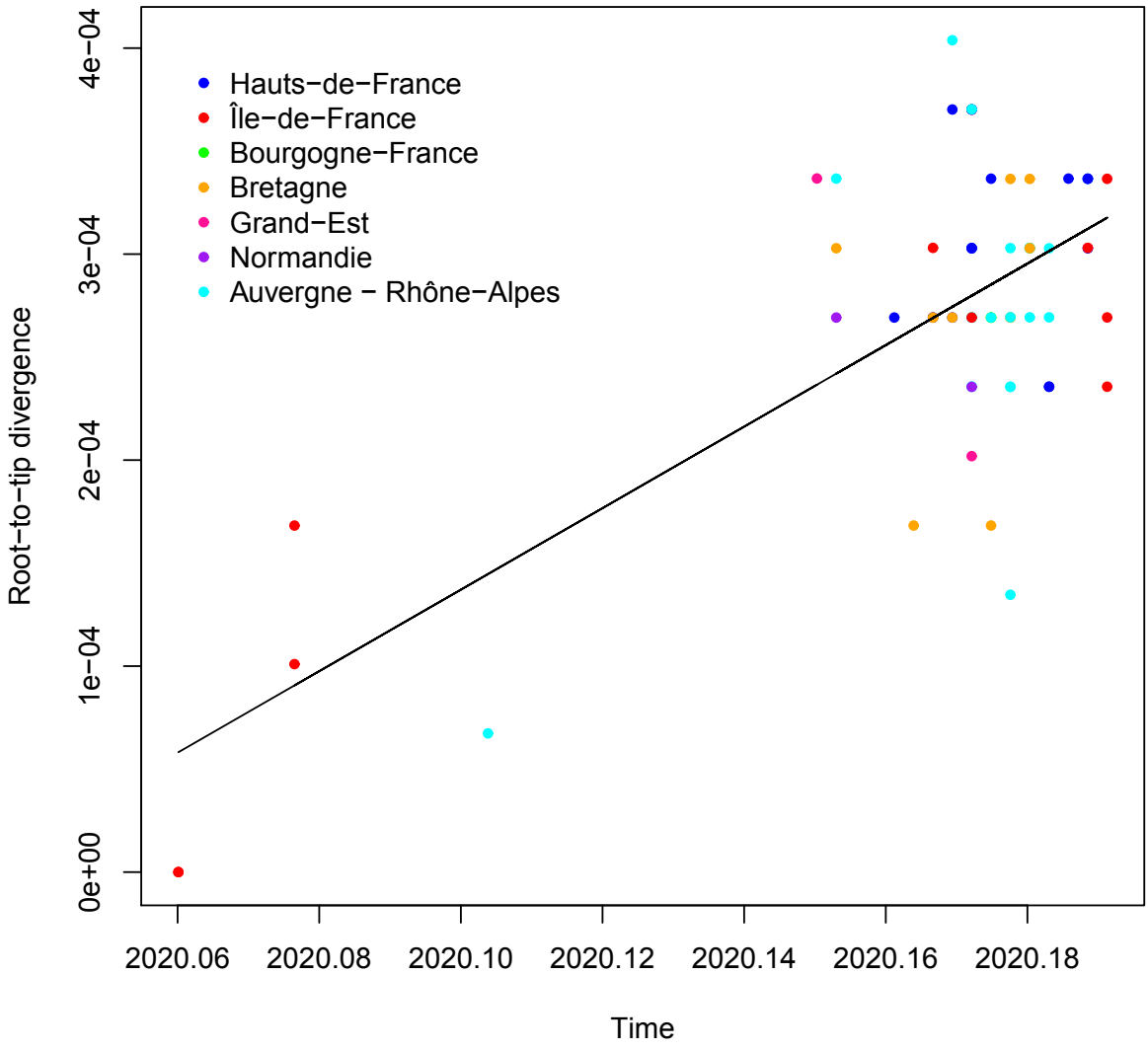
On constate aussi que la majorité des séquences françaises sont regroupées dans le clade supérieur de la phylogénie et que les séquences récoltées le plus tôt dans l'épidémie (donc sur la gauche du graphique) sont dans des clades moins massifs, qui ont donc causé moins d'infections secondaires.

Signal phylogénétique

La phylodynamique permet s'appuyer sur les dates d'échantillonnage pour estimer la *vitesse d'évolution* du virus. Cette dernière s'exprime en substitutions rapportées au temps et à la taille du génome. Une **substitution** désigne une mutation qui s'est fixée dans le génome et il ne faut donc pas confondre taux de mutation et taux de substitution : le premier est une constante biologique liée aux processus biochimiques, tandis que le second correspond aux mutations qui ont survécu (cela exclut donc toutes les mutations délétères ou celles qui pour une autre raison ont disparu). Connaître la vitesse à laquelle les mutations se fixent dans le génome permet de dater des événements tels que l'émergence d'une épidémie (ici celle du COVID-19 en Chine) ou son arrivée dans un lieu précis (ici la France).

Toutefois, il convient d'abord de vérifier que le jeu de données analysé contient suffisamment de signal phylogénétique (donc pour simplifier que suffisamment de mutations ont été accumulées dans les génomes). En effet, si les génomes sont trop petits, ou que le taux de substitution est trop faible, ou que la période d'échantillonnage est trop courte, il se peut qu'il n'y ait pas suffisamment de signal. Dans ce cas, vaut mieux fixer le taux de substitution à une valeur réaliste (par exemple celui estimé sur les données de Chine).

Pour cela, nous avons utilisé le logiciel TempEst (http://tree.bio.ed.ac.uk/software/tempest/) qui effectue une régression linéaire entre la divergence entre deux séquences (en termes de nombre de substitutions) et leur écart dans le temps. Le coefficient de détermination nous indique la fraction de la variance expliquée par la régression et est un proxy du signal phylogénétique. Cet analyse est basée sur une phylogénie inférée en utilisant la méthode de maximum de vraisemblance et le logiciel PhyML (http://www.atgc-montpellier.fr/PhyML/) suivant un modèle de mutation calculé par le logiciel SMS (http://www.atgc-montpellier.fr/SMS/).



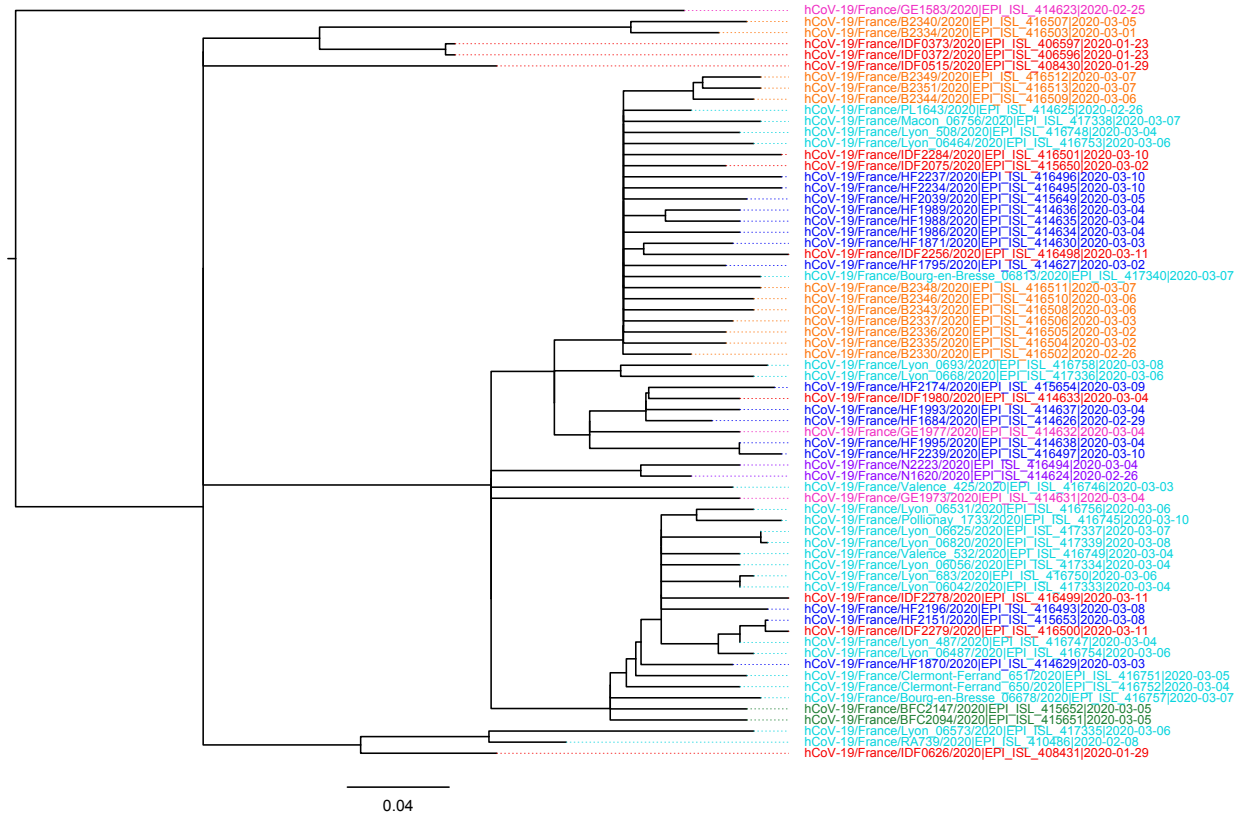
Régression linéaire entre la distance à la racine et la date d'échantillonnage.

La régression aboutit à une relation avec un coefficient de régression de $1,96 \cdot 10^{-3}$ an^{-1}, qui correspond au taux de substitution. Le pourcentage de la variance expliqué est $R^2 = 50\%$.

Ces chiffres sont cohérents avec ceux inférés sur 176 génomes de SRAS-Cov-2 par Andrew Rambaut le 6 mars 2020 (http://virological.org/t/phylodynamic-analysis-176-genomes-6-mar-2020/356) : il trouvait un coefficient de régression de $8,81 \cdot 10^{-4}$ an^{-1} et un coefficient de détermination de $R^2 = 39,4\%$, ainsi qu'une ordonnée à l'origine au 10 Décembre 2019 suggérant une origine récente.

Structure de la phylogénie d'infection

Avant une analyse phylodynamique proprement dite, on peut considérer la structure géographique de l'épidémie échantillonnée. En particulier, pour chaque séquence nous avons deux informations que sont la date et la région d'échantillonnage. Cette analyse est basée sur la même phylogénie que celle utilisée pour estimer le signal phylogénétique et représentée à l'aide du logiciel FigTree (http://beast.community/figtree).



Phylogénie des séquences françaises inférée par maximum de vraisemblance. *La couleur correspond à la région de France où le prélèvement a été échantillonné. La distance est mesurée en années.*

On peut tirer trois enseignements de la structure phylogéographique :

1. Comme on l'a vu sur la phylogénie de nextstrain, il semble y avoir eu de multiples introductions du COVD-19 en France avec des arrivées précoces qui n'ont pas donné de descendance (les séquences provenant a priori de patient·e·s qui ont été isolé·e·s rapidement),

2. On détecte au moins une introduction qui s'est propagée dans toute la France puisque le plus gros sous-arbre (ou clade) contient des séquences de personnes infectées provenant de différentes régions de France,

3. On note toutefois aussi une structure locale de l'épidémie avec de petits groupes (ou petits clades) de personnes dépistées dans la même région.

L'analyse phylodynamique va nous permettre de mieux quantifier la propagation de l'épidémie et de dater les événements épidémiologiques, notamment l'origine de toutes les séquences disponibles en France et l'origine du plus gros clade français.

Analyses Beast

Nous avons réalisé les analyses phylodynamique à partir de notre alignement de séquences datées à l'aide des logiciels Beast (https://beast.community/) et Beast2 (http://www.beast2.org/). Sans entrer dans les détails, ces logiciels utilisent des approches bayésiennes afin d'intégrer des paramètres liés à un modèle d'évolution moléculaire (qui nous serviront à dater les événements) et à un modèle démographique (qui nous serviront à retracer la propagation). Nous avons analysé plusieurs modèles et présentons les résultats de deux d'entre eux :

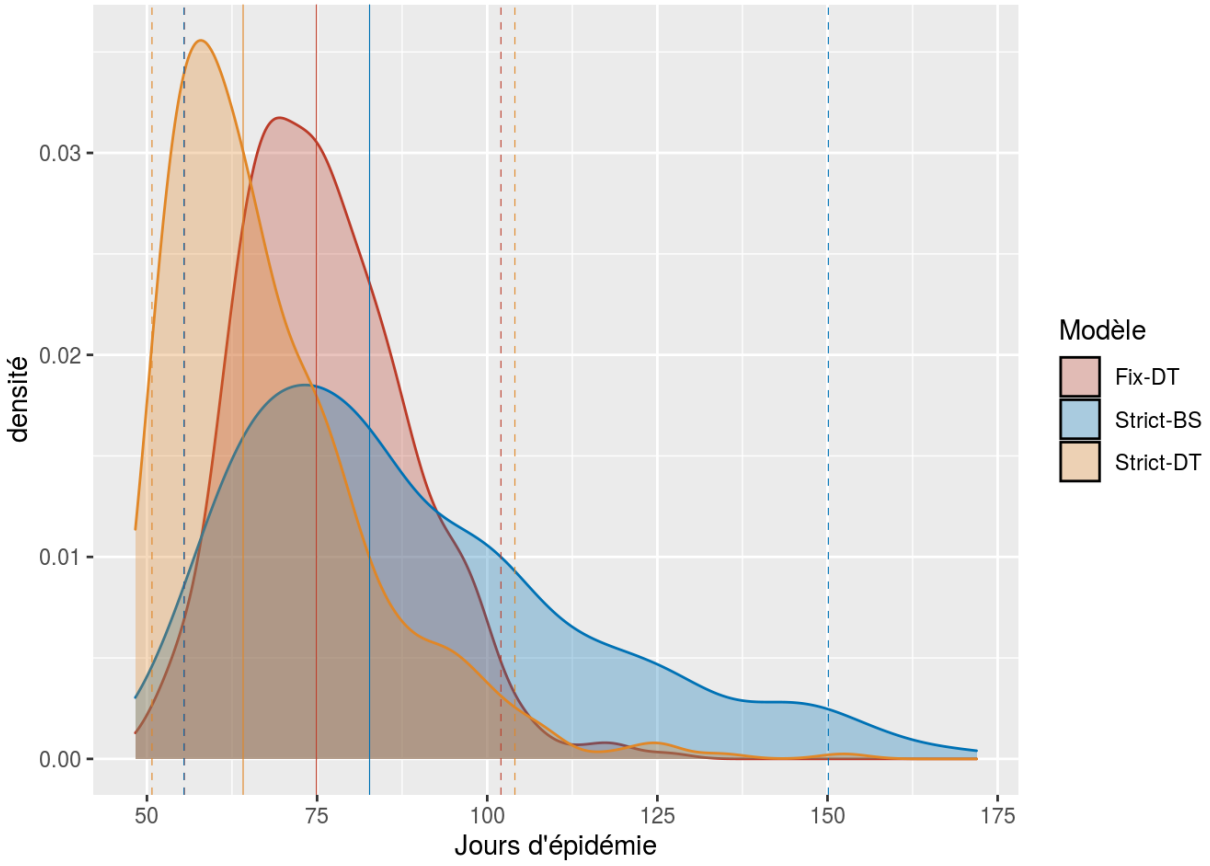
1. Un modèle basé sur un coalescent exponentiel (noté **DT** pour *doubling time*), qui fait l'hypothèse que la population croit de manière exponentielle, ce qui à la date la plus récente d'échantillonnage de séquence (le 11 mars 2020) semble le plus réaliste.

2. un modèle basé sur un coalescent de type “horizon” (noté **BS** pour *bayesian skyline*), qui consiste à diviser le temps en segments et à estimer la taille efficace de l’épidémie (potentiellement proportionnelle au nombre de personnes infectées) pour chaque segment. Ce modèle est a priori moins adapté pour une épidémie en phase de croissance exponentielle mais il est très robuste.

Pour tous ces modèles, nous avons utilisé une horloge moléculaire stricte (les données sont trop limitées pour utiliser une horloge moléculaire relâchée). Nous avons soit inféré la vitesse d’évolution à partir des données (modèle *Strict*), soit imposé la valeur estimée par Andrew Rambaut ((http://virological.org/t/phylodynamic-analysis-176-genomes-6-mar-2020/356)) sur les données de Chine, c’est-à-dire $8,8 \cdot 10^{-4}$ substitutions/position/an (modèle *Fix*).

Date de l’ancêtre commun des séquences analysées

Grâce à l’inférence phylodynamique, on peut dater l’ancêtre commun à toutes les séquences issues de prélèvements effectués en France. Biologiquement, cela peut-être interprété comme la date de l’infection qui a conduit à toutes les infections détectées en France. Vu les introductions multiples du virus en France, cet infection ancestrale se trouvait *a priori* en Chine. Ceci se voit aussi très bien sur la représentation de nextstrain (https://nextstrain.org/ncov) car on trouve des séquences de Chine et du monde entier au sein de l’arbre qui contient toutes les séquences de France.



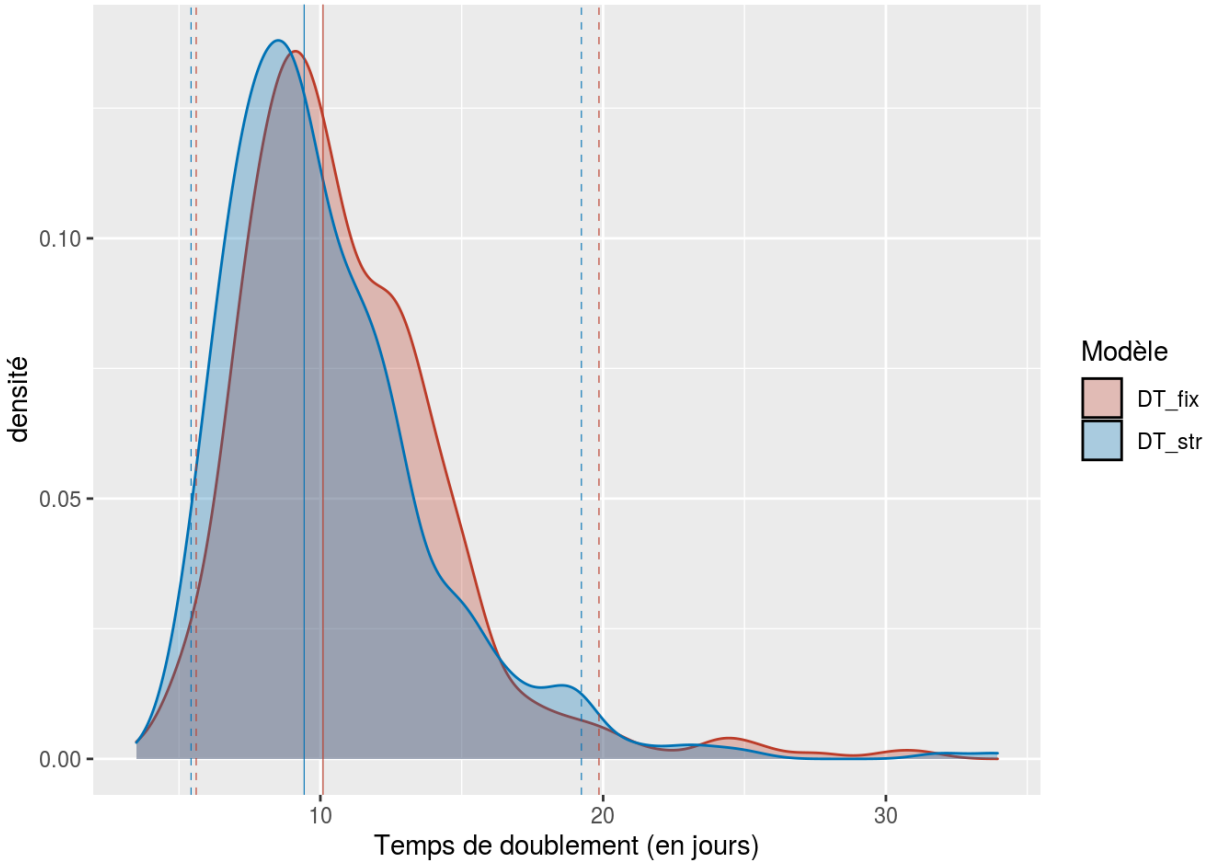
Le graphique montre la distribution des dates d’ancêtres commun selon une horloge moléculaire fixée (“Fix”) ou estimée (“Strict”) et selon un modèle de population (“DT” pour *doubling time* et BS pour *bayesian skyline*). On voit que les médianes (traits verticaux) sont proches quels que soient les modèles d’évolution utilisés (de couleur différentes). La forme exacte des distributions varie, mais les intervalles de confiance sont très similaires pour les trois modèles.

	borne inf. (95 %)	date médiane de l'ancêtre commun	borne sup. (95 %)
Fix-DT	29 nov. 2019	27 déc. 2019	15 janv. 2020
Strict-DT	26 nov. 2019	06 janv. 2020	20 janv. 2020
Strict-BS	10 oct. 2019	19 déc. 2019	15 janv. 2020

Ces dates sont cohérentes avec les plus récentes obtenues par Andrew Rambaut (http://virological.org/t/phylodynamic-analysis-176-genomes-6-mar-2020/356) quant au début de l'épidémie en Chine qui est daté au 17 novembre 2019 avec un intervalle de confiance entre le 27 août et le 19 décembre 2020. Ceci est logique car les séquences françaises sont dispersées dans parmi tous les clades de la phylogénie globale (celle qui inclut toutes les séquences disponibles).

Temps de doublement

Un des avantages de la phylodynamique est que nous pouvons estimer des valeurs de paramètres de dynamique de population. Pour un modèle de coalescent avec croissance exponentielle, le paramètre clé est le temps de doublement, qui correspond au nombre de jours pour que l'épidémie double en taille. Ce paramètre est lié au R_0.



Pour nos deux modèles d'horloge moléculaire (estimée ou fixée), on obtient des résultats très similaires : le temps de doublement de l'épidémie est de 10 jours avec un intervalle de confiance à 95 % assez large (entre 5 et 20 jours).

Ces chiffres sont élevés (donc la croissance de l'épidémie est lente) par rapport aux inférences phylodynamiques réalisées à partir des données de Chine (avec 86 génomes, Andrew Rambaut (http://virological.org/t/phylodynamic-analysis-176-genomes-6-mar-2020/356) trouvait un temps de doublement médian d'environ 7 jours avec un intervalle de confiance entre 4.7 et 16.3 jours) et, surtout, par rapport aux inférences faites à partir des séries temporelles, qui trouvent des temps de

doublement 3 fois plus faibles comme le montrent les analyses de Abbott et alii (https://cmmid.github.io/topics/covid19/current-patterns-transmission/global-time-varying-transmission.html).

	borne inf. (95 %)	temps de doublement médian (en jours)	borne sup. (95 %)
Fix-DT	5.6	10.1	19.9
Strict-DT	5.4	9.4	19.4

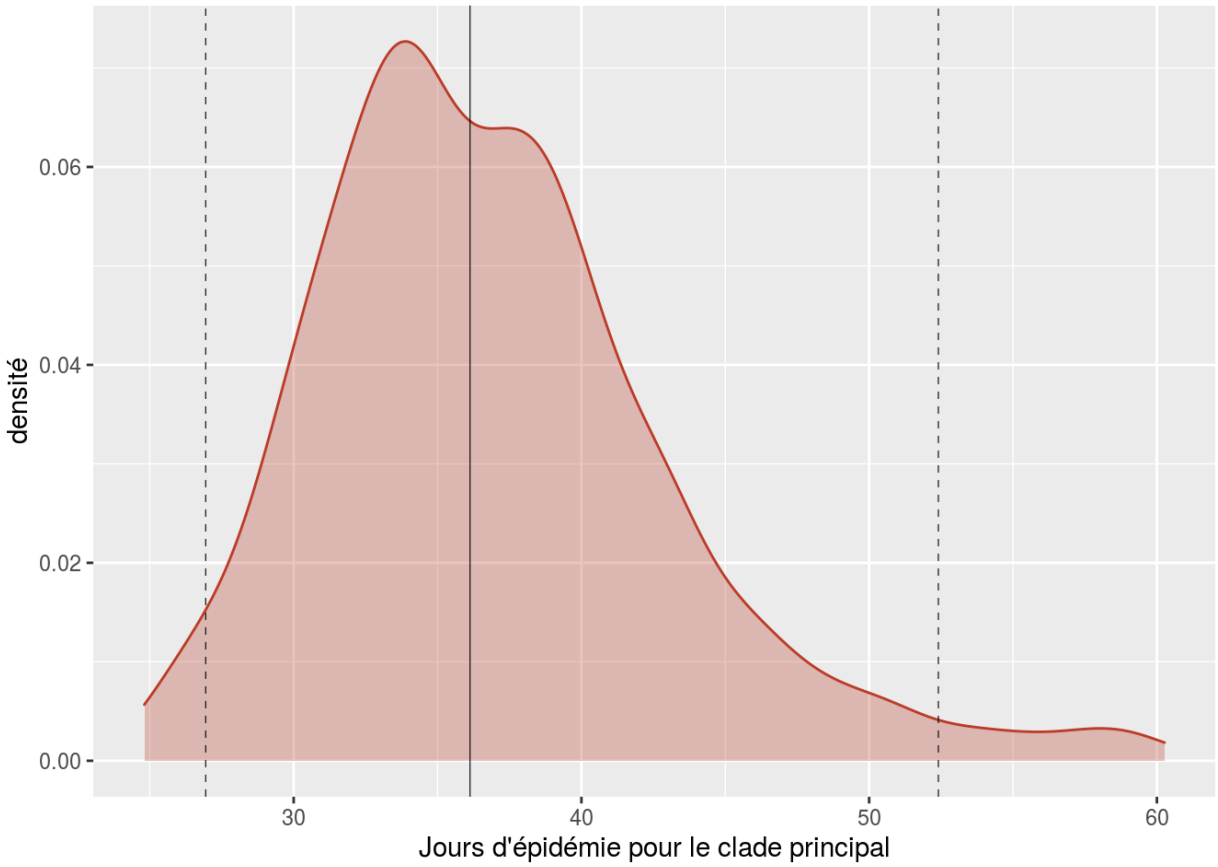
Ces faibles vitesse de croissance de l'épidémie s'expliquent par le fait que la phylogénie utilisée inclut des cas importés en janvier qui n'ont pas généré d'infections secondaires et donc tronquent cette propagation. C'est pourquoi par la suite nous nous concentrons sur le clade principal en France qui exclu ces séquences dans des clades plus petits et isolés.

Clade principal en France

Les données concernant la date de l'ancêtre commun à toutes les infections détectées en France sont peu exploitables car a priori celui-ci se situe en Chine du fait des introductions multiples. Il en va de même pour le temps de doublement et le R_0 qui se trouvent minorés par ces multiples séquences introduites et sans descendance échantillonnée.

Afin d'affiner nos valeurs, nous nous sommes donc focalisés sur les 59 séquences qui forment le clade principal des séquences échantillonnées en France. Ce clade semble assez représentatif car on y trouve des séquences échantillonnées dans 7 régions différentes. Toutefois, le signal phylogénétique n'est plus assez important pour que nous puissions estimer la vitesse d'évolution des données. Nous ne représentons donc ici que le modèle où l'horloge moléculaire est fixée à $8,8 \cdot 10^{-4}$ substitutions/position/an (modèle *Fix* pour l'analyse précédente).

Attention, comme on le voit bien sur la représentation de nextstrain (https://nextstrain.org/ncov?label=clade:A2a), ce clade principal inclut aussi des séquences de nombreux pays. Il se peut donc que son ancêtre commun soit lui aussi en Chine ou dans un autre pays d'Europe (selon nextstrain, son origine la plus vraisemblable est le Royaume-Uni, mais ce résultat est *extrêmement* sensible à l'intensité de l'échantillonnage et la France a quatre fois moins séquencé de virus que le Royaume-Uni).



Concernant le nombre de jours pour remonter à l'ancêtre commun du principal clade de la phylogénie en France, on constate qu'il est bien inférieur à celui pour remonter à l'ancêtre commun à toutes les séquences de France (36 jours contre plus de 75). Cela fait sens étant donné que pour inclure toutes les séquences de France dans une même phylogénie, il faut remonter quasiment au début de l'épidémie.

	borne inf. (95 %)	date médiane du principal clade français	borne sup. (95 %)
Fix-DT	18 janv. 2020	03 févr. 2020	13 févr. 2020

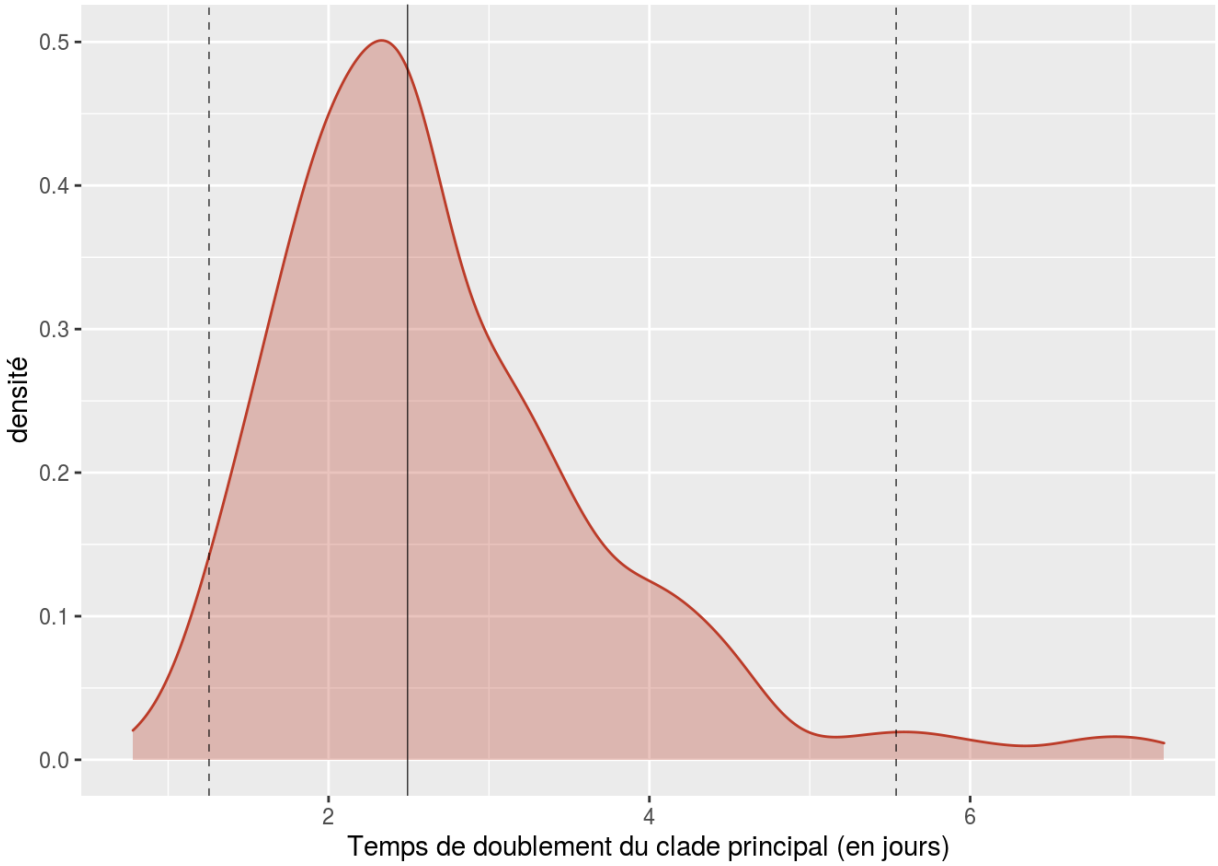
Si l'on considère les dates, les résultats sont assez éloquents. En effet, avec une certitude de 95 % et selon les hypothèse de notre modèle, ce clade a commencé à croître après le 18 janvier et avant le 13 février (nextstrain (https://nextstrain.org/ncov?label=clade:A2a) le date lui au 25 janvier). Pour mémoire, le premier cas importé détecté date du 29 janvier et la courbe d'incidence (nombre de nouveaux cas par jour) n'a commencé à croître de manière continue que le 27 février (comme le montre notre graphique dans notre Rapport 1 (http://alizon.ouvaton.org/Rapport1_R0_France.html)). Autrement dit, avec une erreur minime, on peut dire que l'épidémie était déjà installée en France ou proche de la France (au Royaume-Uni selon nextStrain (https://nextstrain.org/ncov?label=clade:A2a), mais ce résultat est très incertain du fait du peu de séquences disponibles) deux semaines avant la détection des premiers cas en France.

Ces résultats sont sensibles à deux choses. D'une part la vitesse d'évolution, qui a été fixée pour cette analyse. Si le virus a évolué plus vite en France qu'en Chine, alors l'origine du clade de France serait plus récente. D'autre part, à l'échantillonnage qui est limité (59 séquences seulement). Il se peut que tout un pan de l'épidémie installée de manière encore plus précoce n'ait pas été séquencé, ce qui situerait son origine encore plus loin dans le temps.

Enfin, rappelons que le plus gros clade de France inclut aussi des séquences d'autres pays que nous avons retiré pour l'analyse. Il se pourrait donc que le virus ne soit arrivé en France que secondairement en transitant par un autre pays. Toutefois, on voit sur nextstrain

(https://nextstrain.org/ncov?label=clade:A2a) qu'un des sous-clades dont l'origine française est quasi-certaine est datée à 2 jours plus tard.

Voyons maintenant la vitesse à laquelle le noyau de cette épidémie en France a cru.



	borne inf. (95 %)	temps de doublement médian (en jours)	borne sup. (95 %)
Fix-DT	1.3	2.5	5.5

Le temps de doublement de l'épidémie semble bien plus réaliste quand on restreint les données aux clade principal. En effet, pour ce noyau français, l'épidémie double en taille tous les 2,5 jours en moyenne (soit 4 fois plus vite qu'avec l'ensemble des séquences française). Ceci nous confirme qu'une partie des séquences incluses dans l'analyse principale n'ont pas eu de descendances et correspondent à des patient·e·s isolé·e·s avant d'avoir transmis l'infection (vraisemblablement car ils ou elles revenaient de Chine et ont pu être dépisté·e·s avant d'avoir des symptômes). On peut aussi en conclure que les séquences de ces patient·e·s (9 sur les 68) sont sur-représentées par rapport à l'épidémie française.

Conclusions

Malgré le nombre limité de séquences issues de France, il est possible de détecter des introductions multiples de l'épidémie de COVID-19 dans le pays, en particulier via la visualisation de nextstrain (https://nextstrain.org/ncov?f_country=France) pour la France. Le signal phylogénétique inféré est cohérent avec celui obtenu sur des jeux de données plus larges. Les estimations quant à la date de l'ancêtre commun (vers fin 2019) sont cohérentes avec les données déjà publiées sur les séquences disponibles, en particulier celles de Chine.

En nous concentrant sur le clade principal en France, qui regroupe 59 des 68 séquences disponibles, il est possible d'affiner la date de décollage de l'épidémie en France (ou dans sa proximité) ainsi que son temps de doublement. On peut ainsi voir que, selon les hypothèses du

modèle, l'épidémie était déjà installée en France début février, avant d'être détectée dans les relevés d'incidence le 27 février. On voit aussi que le temps de doublement de ce clade principal (tous les 2,5 jours environ) est bien plus réaliste, en cohérence avec les résultats obtenus via les relevés d'incidence et suggère que les patient·e·s isolés avant d'avoir pu transmettre leur infection sont sur-représentés parmi les séquences françaises (environ 13 %).

Attention, pour tout ce qui concerne les origines géographiques du virus, les résultats sont très sensibles à la densité d'échantillonnage, qui varie selon les régions. En particulier, dans le clade où sont la majorité des séquences française (visible sur nextstrain (https://nextstrain.org/ncov?label=clade:A2a)), l'absence de prélèvement avant le 20 février rend toute estimation de la localisation ancestrale du virus très fragile. Pour la France, on sait toutefois que le virus qui était présent le 21 février car la séquence HF1465 (https://nextstrain.org/ncov?label=clade:A2a&s=France/HF1465/2020) issue d'un patient de Compiègne appartient au clade.

En augmentant le nombre de séquences génomiques de SRAS-Cov-2 issues de l'épidémie française, en particulier des séquences prélevée en début d'épidémie, il serait possible de :

- mieux dater l'arrivée de l'épidémie en France,
- mieux estimer les variations du nombre de reproduction au cours du temps (calculé dans notre rapport 1 (Rapport1_R0_France.html)),
- mieux comprendre la propagation entre les différentes régions de France,
- estimer le nombre d'introductions du virus dans le pays,
- détecter de potentielles mutations adaptatives qui modifieraient la propagation du virus.

Pour aller plus loin sur la phylodynamique…

- https://nextstrain.org/ncov (https://nextstrain.org/ncov) : le nec plus ultra de la visualisation phylodynamique au niveau mondial
- Volz *et alii* (2020) (https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-phylogenetics-15-02-2020.pdf) : le rapport du groupe de modélisation partenaire de l'OMS (en anglais).
- http://virological.org/ (http://virological.org/) : site qui regroupe des analyses et des discussions en phylogénétique et phylodynamique avec en particulier l'analyse d'Andrew Rambaut (http://virological.org/t/phylodynamic-analysis-176-genomes-6-mar-2020/356) et celle du groupe de Tanja Stadler (http://virological.org/t/update-2-evolutionary-epidemiological-analysis-of-128-genomes/423) très axée sur la phylodynamique.

Logiciels utilisés

- L'alignement des séquences a été obtenu par la pipeline Augur développée par nextstrain (https://nextstrain.org/docs/bioinformatics/introduction-to-augur).
- L'alignement des séquences peut aussi être réalisé par la version en ligne du logiciel mafft (https://mafft.cbrc.jp/alignment/software/).
- Le choix du modèle génétique de substitution a été réalisé à l'aide du logiciel SMS (http://www.atgc-montpellier.fr/sms/).
- La phylogénie en maximum de vraisemblance a été inférée à l'aide du logiciel PhyML (http://www.atgc-montpellier.fr/PhyML/).

- La recherche de signal temporel d'évolution moléculaire a été réalisée via le logiciel TempEst (http://tree.bio.ed.ac.uk/software/tempest/).
- Les analyses de phylodynamique ont été réalisée grâce à la version 1.8.3 du logiciel Beast (https://beast.community/) et à la version 2.3 du logiciel Beast2 (http://www.beast2.org/).
- La représentation phylogénétique a été effectuée à l'aide du logiciel FigTree (http://beast.community/figtree).
- Des manipulations additionnelles ont été effectuées en R (http://cran.r-project.org/) grâce au package ape (http://ape-package.ird.fr/).

Sources et remerciements

- Les données de séquences génétiques ont été fournies par plusieurs laboratoires internationaux et mis à disposition au travers de GISAID - Global Initiative on Sharing All Influenza Data (https://www.gisaid.org/). La liste des séquences utilisées et les laboratoires les ayant obtenues sont dans le tableau ci-dessous.

accession <fctr>	date <fctr>	Len… <int>	patient <fctr>	Région <fctr>
EPI_ISL_410486	2020-02-08	29876	RA739	Auvergne-Rhône-Alpes
EPI_ISL_406596	2020-01-23	29853	IDF0372	Ile-de-France
EPI_ISL_406597	2020-01-23	29809	IDF0373	Ile-de-France
EPI_ISL_416750	2020-03-06	29879	Lyon_683	Auvergne-Rhône-Alpes
EPI_ISL_416753	2020-03-06	29879	Lyon_06464	Auvergne-Rhône-Alpes

1-5 of 69 rows | 1-5 of 8 columns Previous **1** 2 3 4 5 6 … 14 Next

- Nous remercions les patients, personnels de santé et les laboratoires de virologie de France qui ont permis de générer les données de séquences de SRAS-Cov-2 à la base de ce travail.
- Les auteurs remercient le calculateur à haute performance itrop (plateforme South Green (http://www.southgreen.fr)) de l'IRD de Montpellier pour la fourniture des ressource de calcul à haute performance, qui ont contribué aux résultats présentés dans ce travail (plus de détails sur bioinfo.ird.fr (https://bioinfo.ird.fr/)).
- Gonché Danesh est financée par une bourse doctorale de la Fondation pour la Recherche Médicale (http://www.frm.fr).

- L'équipe de modélisation de l'équipe ETE est composée de Samuel Alizon, Thomas Bénéteau, Marc Choisy, Gonché Danesh, Ramsès Djidjou-Demasse, Baptiste Elie, Yannis Michalakis, Bastien Reyné, Quentin Richard, Christian Selinger, Mircea T. Sofonea.
- Contribution à ce travail :
    - conception du travail : toute l'équipe
    - réalisation des analyses : GD, BE et SA
    - rédaction du rapport : SA

- contacts : samuel.alizon@cnrs.fr (mailto:samuel.alizon@cnrs.fr), gonche.danesh@ird.fr (mailto:gonche.danesh@ird.fr), baptiste.elie@ird.fr (mailto:baptiste.elie@ird.fr)

- approbation : toute l'équipe
- Remerciements à Nicolas Bierne pour les limites liées à l'inférence des états ancestraux.

- (http://creativecommons.org/licenses/by-nc/4.0/)
  Ce(tte) œuvre est mise à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale 4.0 International (http://creativecommons.org/licenses/by-nc/4.0/).