

DOSSIER

La phylodynamique, l'autre traque du coronavirus

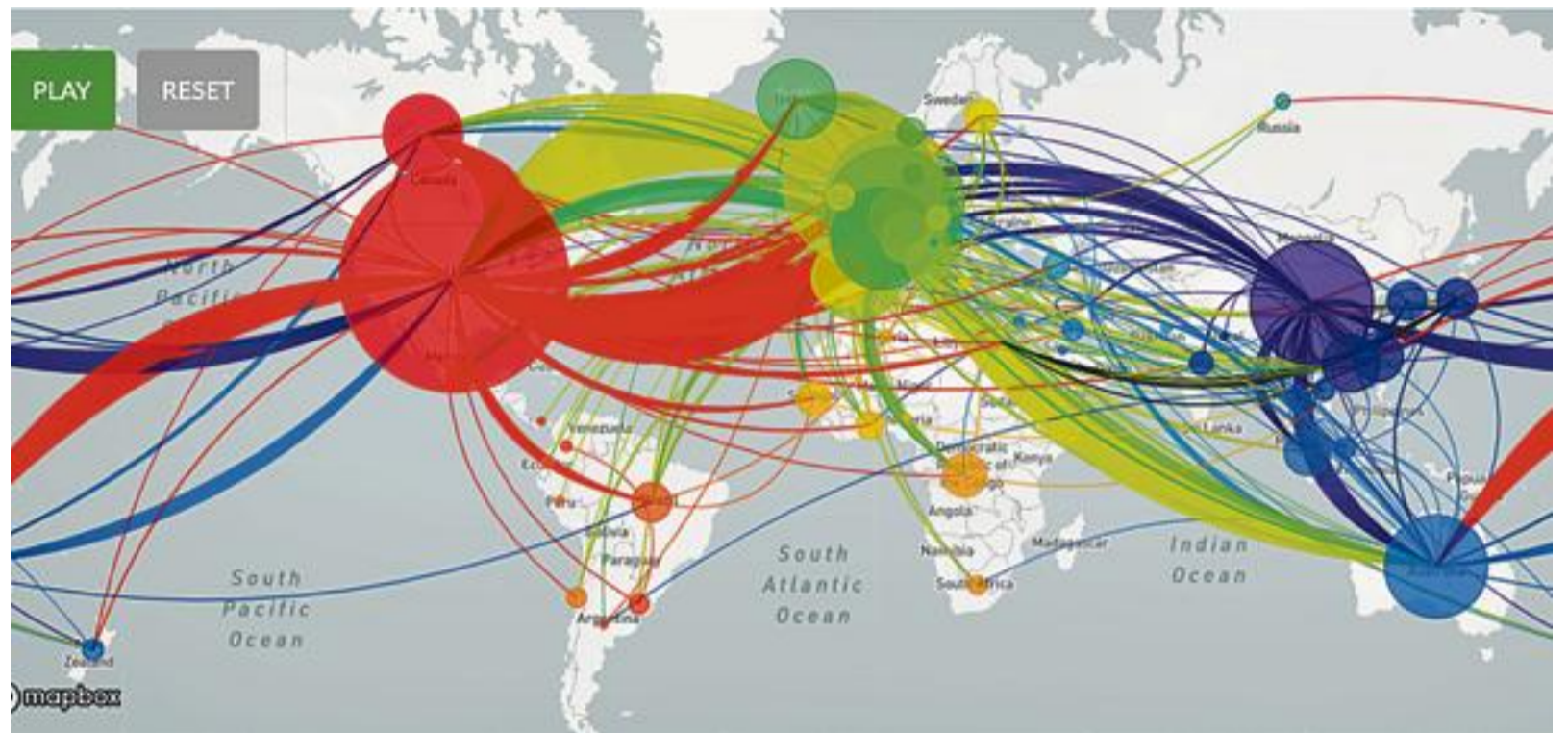
A partir de données de séquences génétiques du virus et à l'aide d'algorithmes, ce nouveau champ d'études explore la chaîne de transmission du Covid-19. Avec des résultats tout à fait comparables à ceux des épidémiologistes de terrain

Le grand public est habitué, grâce aux faits divers ou aux séries télé, à ce que l'ADN aide à confondre les coupables. Mais avec l'épidémie de Covid-19, un autre genre de police génétique est à la manœuvre. S'il ne fait aucun doute que le virus SARS-CoV-2 est bien le responsable de près de 150 000 morts, son patrimoine génétique est en train de révéler aux spécialistes quand il a infecté l'homme, d'où il vient, à quelle vitesse il se répand, combien de gens sont touchés... Ces nouveaux policiers sont des phylodynamiciens, les représentants d'une discipline qui n'a pas vingt ans et qui montre tout son potentiel avec la pandémie en cours. La découverte de son passage probable d'une chauve-souris à l'humain? C'est la phylodynamique. L'origine d'une contamination dès novembre 2019 en Chine? C'est encore elle. Les signes de ralentissement de l'épidémie dans certains pays? Toujours elle. Les origines multiples de l'épidémie en France? Encore et toujours elle...

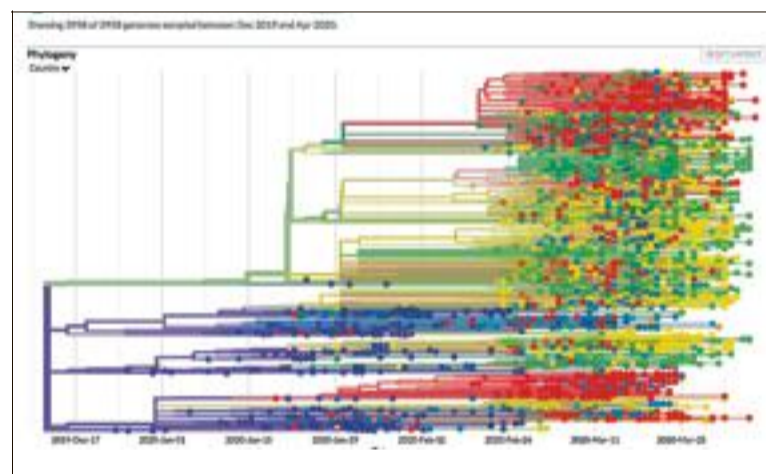
«L'idée de la phylodynamique est que la manière dont les virus se propagent laisse des traces dans leur génome», précise Samuel Alizon, chercheur CNRS de l'équipe Evolution théorique et expérimentale du laboratoire Maladies infectieuses et vecteurs : écologie, génétique, évolution et contrôle, à Montpellier. Ces traces sont si infimes qu'il convient de les examiner avec précaution pour les faire «parler», sous peine de se tromper lourdement. Il s'agit de tout petits changements dans l'enchaînement des quelque 30 000 «lettres» qui constituent le génome de ce virus. Une lettre seulement de différence entre deux génomes est déjà une information précieuse.

A manier avec précaution

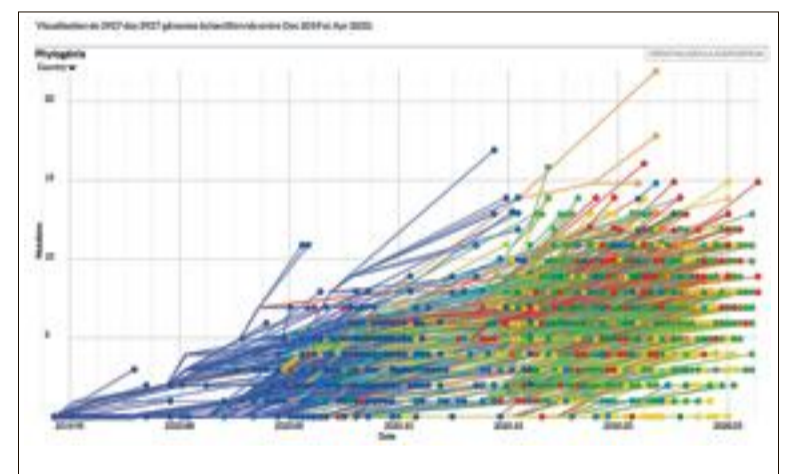
Dès le 20 janvier, une des vedettes du domaine, Trevor Bedford, du centre anticancer Fred Hutchinson, à Seattle (Etats-Unis), acquiert, comme il l'écrit sur son blog, la certitude que ce virus qu'il surveille depuis début janvier est transmissible à l'homme. Une propriété fondamentale qui conditionne la gravité de la maladie à l'échelle planétaire. La Chine ne décrètera une quarantaine dans son premier foyer que le 29 janvier. L'intuition de ce scientifique vient de l'analyse des génomes viraux qui arrivent à partir du 10 janvier de Chine. Ils sont trop semblables pour croire que les malades auraient été contaminés par des animaux. En effet, le virus étant installé depuis longtemps dans ce réservoir, il devrait exister sous des formes assez variées. Or cette diversité est absente des génomes viraux prélevés sur les premiers malades. Sauf à imaginer qu'un même animal ait contaminé autant d'hommes en des endroits distincts, il fallait se rendre à la terrible évidence que le coronavirus avait trouvé un nouvel hôte et qu'il était devenu transmissible.



Mappemonde montrant l'origine des séquences génétiques et les probables importations du virus d'un pays à l'autre. IMAGES EXTRAITES DU SITE NEXTSTRAIN



Vue du nombre de mutations sur les génomes en fonction du temps.



«Arbre» indiquant les «filiations» probables entre les séquences.

Le même spécialiste va très vite mener une seconde enquête pour couper court cette fois à une rumeur. Le 31 janvier, une équipe indienne assure que le génome viral aurait des points communs avec celui du VIH, sous-entendant une manipulation génétique artificielle. Trevor Bedford, dès le lendemain de cette «parution» (l'article a été mis en ligne uniquement sur un site de spécialistes, sans évaluation par une revue scientifique), démonte sur Twitter l'hypothèse. Le genre de

LA DÉCOUVERTE D'ORIGINES MULTIPLES POUR L'ÉPIDÉMIE EN FRANCE? C'EST ELLE...

variations constatées existe aussi naturellement chez un coronavirus de chauve-souris. Et les Indiens auraient mal comparé les séquences entre elles, prenant un artefact dénué de sens pour une similitude riche d'informations. L'épisode montre que la phylodynamique est à manier avec précaution.

Le 10 avril, nouvel exemple des subtilités de la technique. Une autre référence du domaine, Andrew Rambaut, de l'université d'Edimbourg, cloue au pilori des collègues

américains pensant avoir découvert trois variants différents dans les diverses souches de virus, comme ils le prétendent dans une grande revue, PNAS. «Ce qui m'énerve le plus est que ces auteurs ont pris quelques données dans une base, les ont mises dans un logiciel facile à utiliser, ont fait des hypothèses inappropriées et publié ce qu'ils ont trouvé.» Le spécialiste note aussi une grave erreur sur la comparaison avec le coronavirus de la chauve-souris. «Ça me rend un peu triste d'être membre de cette communauté scientifique», déplore sur Twitter un autre spécialiste, François Balloux, professeur de bio-informatique à l'University College de Londres.

NEXTSTRAIN: LE SITE QUI SUIT LE SARS-COV-2 À LA TRACE

L'épidémie de Covid-19 a divisé en deux la population de ceux qui tentent de la suivre. Il y a ceux qui scrutent chaque jour les chiffres du nombre de cas, de décès, d'admis en réanimation... Et il y a ceux qui regardent le site Nextstrain. Si les premiers s'intéressent aux victimes, les seconds sont concentrés sur le coupable, le coronavirus SARS-CoV-2. «Nous avons des centaines de milliers de vues quotidiennes en ce moment», indique James Hadfield, l'une des chevilles ouvrières de cette plate-forme née en 2016. Il travaille dans l'un des deux groupes qui l'a lancée, l'équipe de Trevor Bedford, au centre de recherche contre le cancer Fred-Hutchinson (Seattle, Etats-Unis). L'autre groupe est à l'université de Bâle, en Suisse, dans l'équipe de Richard Neher.

Le site parvient à synthétiser de façon très visuelle et esthétique les propriétés de plus de 3 900 génomes du nouveau virus, prélevés dans une soixantaine de pays (au 16 avril). Une

carte mondiale montre l'origine géographique de ces séquences – surtout en Chine, aux Etats-Unis et en Europe – ainsi que les voies que le virus a suivies. Apparaissent également les régions de la séquence génétique où des mutations (ou substitutions) ont été notées, parmi les 30 000 «lettres» que compte ce génome.

Enfin s'affiche l'arbre généalogique du coronavirus, avec son tronc, ses branches, ses rameaux et toutes les feuilles que constituent les variants. En filtrant par pays, les quelque 200 échantillons français permettent de voir d'un seul coup d'œil qu'il y a eu probablement plusieurs introductions dans le pays, en provenance de Chine, d'Europe voire des Etats-Unis.

L'un des points forts de Nextstrain est justement de présenter ces fameux arbres qui, à la différence de l'origine géographique, sont difficiles à obtenir. Ils sont calculés par des algorithmes proposant les relations les plus probables entre ces séquences qui, parfois,

ne varient que de quelques lettres. A partir de là, les chercheurs traquent l'origine de la contamination, surveillent des mutations dangereuses, voire en déduisent des propriétés de la transmission de l'épidémie. «Nous voulions présenter ces analyses en temps réel de façon à ce qu'elles servent lors d'une épidémie. La manière traditionnelle de publier en science n'est pas adaptée dans ce cas-là, car les résultats peuvent ne pas être à jour. Evidemment, Nextstrain ne souhaite pas remplacer ces articles», explique James Hadfield.

Codes sources ouverts

Pour Nextstrain, tout a commencé en 2016, avec des analyses sur l'épidémie due au virus Ebola en Afrique de l'Ouest et avec d'autres travaux, rétrospectifs ceux-là, sur le virus du Nil occidental. Ont suivi la grippe A (H1N1), la tuberculose, la dengue... Avec le coronavirus, le site a changé de dimension avec plus de 3 500 séquences puisées dans la base de données

Gisaid, gérée par le gouvernement allemand et une association. «Le plus dur du travail est de se tenir à jour!», reconnaît James Hadfield.

L'initiative témoigne aussi d'une plus grande ouverture de la recherche. Les séquences sont partagées, les logiciels utilisés ont des codes sources ouverts, les traductions de Nextstrain (en 17 langues) sont assurées par des volontaires, pas forcément biologistes. Un souci pédagogique confirmé par les tutoriels et les résumés de ce que toutes ces séquences nous apprennent de l'épidémie.

Comme les programmes sont librement accessibles, les chercheurs les utilisent aussi dans leurs laboratoires pour leurs propres calculs et visualisations. «Ma satisfaction est qu'en ce moment nos outils aident l'Institut national de recherche biomédicale de la République démocratique du Congo à faire face à l'épidémie d'Ebola», se réjouit James Hadfield. ■

D. L.

Calculer la vitesse des mutations

L'art phylodynamique est donc subtil. Depuis le début de l'épidémie de Covid-19, l'écart entre les nouveaux génomes et le premier est de moins de vingt lettres de différence sur 30 000 environ. Soit d'un ordre de grandeur analogue au taux d'erreur des techniques de séquençage, d'environ une lettre erronée sur 10 000 lues. A l'inverse, cet écart est de plus de 1 000 lettres entre le probable réservoir du coronavirus, la chauve-souris rhinolophe du Yunnan, et le premier humain touché. «Une centaine de séquences génétiques contient autant d'informations que celles recueillies avec l'ensemble des cas dépistés», assure Samuel Alizon pour défendre l'intérêt de sa discipline dans la compréhension des épidémies. Aux premières observations et controverses ont donc très vite succédé de nouvelles informations, au fur et à mesure qu'arrivaient les séquences du monde entier. Le site Gisaid de dépôt de ces séquences en comptait 9 300 le 16 avril! «Pour un de mes articles dans Science en 2009 sur la grippe A de type H1N1, nous en avions... 11!», se souvient François Balloux, alors à l'Imperial College. En ce moment, mon équipe traite près de 1 000 séquences par jour.»

