



3 novembre 2020

Les données de séquence du virus de la COVID-19 doivent être mises en ligne rapidement et être accessibles librement

Lors d'une épidémie, les données permettant de comprendre la biologie du pathogène et son évolution sont essentielles pour la compréhension de sa propagation et son contrôle. Ceci est particulièrement crucial en situation d'urgence sanitaire mondiale telle que la pandémie de SARS-CoV-2.

Les virus changent en permanence, avec l'acquisition de changements mineurs de leur séquence génomique à mesure qu'ils se multiplient et infectent de nouveaux individus. Ces petites modifications peuvent être mises à profit pour tracer la diffusion du virus dans l'espace et le temps, et ainsi aider les autorités de santé publique à comprendre comment le virus se propage et mesurer l'efficacité d'interventions à une échelle très locale (e.g. un service hospitalier) ou plus large (régionale ou internationale). La prévision du développement d'une épidémie passe aussi par l'étude de l'apparition de variants du virus et de potentielles modifications de sa virulence (ou atténuation) et de sa transmissibilité. Enfin, lorsque que des traitements antiviraux ou vaccins sont disponibles, ces données de séquence du virus sont importantes pour s'assurer que ces interventions sont adaptées aux variants du virus en cours de circulation.

Il est donc crucial d'avoir accès à TOUTES les séquences détectées du virus de la COVID-19 accompagnées de « métadonnées » extensives (date, lieu, âge du patient, sexe, ...), le plus rapidement possible, et idéalement en temps réel.

Les espoirs de valorisation financière ou académique de ces séquences ne doivent en aucun cas et pour quiconque, conduire à une rétention d'information, même temporaire, qui est, par ailleurs, condamnée par l'ensemble de la communauté scientifique. Hélas, une telle attitude concerne de très nombreux groupes en France et a été signalé par l'ENA/EBI (Hinxtan).

La structure GISAIID (Global Initiative on Sharing Avian Influenza Data), qui a été créée pour collecter les données génomiques de la grippe, est maintenant impliquée dans la collecte de données sur SARS-CoV-2. **Cette structure (www.gisaid.org) permet une protection de la propriété intellectuelle et est reconnue par la communauté internationale.** Ceci assure



que les auteurs seront crédités ou associés aux publications utilisant leurs séquences. Il faut savoir qu'il est également possible de publier les données de séquences au sein de *l'International Nucleotide Sequence Database Collaboration* (INSDC, dont les trois points d'entrée sont la DNA DataBase of Japan (DDBJ), l'European Nucleotide Archive (ENA/EBI) et GenBank (aux États-Unis) en gardant aussi une protection de la propriété intellectuelle pour un temps donné (de 6 mois à 1 an), tout en permettant une utilisation pour la recherche. Cette possibilité, mal connue, est une alternative solide à GISAID.

Enfin, si des organismes de santé ne sont pas à même de séquencer rapidement une fraction des cas de COVID-19 dans leur région, les deux Centres de Référence sur les Virus Respiratoires sont à même d'aider à ces efforts.

En conclusion, les données de séquence d'un virus qui a fait plus d'un million de morts peuvent grandement contribuer à sauver des vies. Il est lourd de conséquences de ne pas les rendre accessibles le plus rapidement possible sur l'INSDC ou GISAID. Nous en appelons à la responsabilité des chercheurs et à la vigilance de leurs tutelles scientifiques pour que cesse cette rétention d'information qui jette le discrédit sur l'ensemble de notre communauté scientifique et médicale nationale.

Académie des sciences

Pascale Cossart

Secrétaire perpétuel

Pascale.Cossart@academie-sciences.fr

Académie nationale de Médecine

Jean-François Allilaire

Secrétaire perpétuel

jf.allilaire@academie-medecine.fr