

**L'information sur les ressources génétiques sous
forme de séquences numériques : un sujet crucial
pour la Convention sur la diversité biologique**

DNA





Construction des capacités
internationales pour
l'évaluation et la gouvernance
de la biologie synthétique



African
centre for biodiversity



etc
GROUP
monitoring power
tracking technology
strengthening diversity

TWN
Third World Network

L'information sur les ressources génétiques sous forme de séquences numériques : un sujet crucial pour la Convention sur la diversité biologique

2018

A propos du Centre africain pour la biodiversité (ACB, African Centre for Biodiversity) : Le Centre Africain pour la Biodiversité (ACB) s'engage à démanteler des inégalités dans les systèmes alimentaires et agricoles en Afrique, à notre conviction du droit des peuples à jouir d'une alimentation saine et culturellement adaptée, produite grâce à des méthodes écologiques et durables, ainsi qu'au droit des peuples à définir leurs propres systèmes alimentaires et agricoles. (Veuillez noter : Le 7 avril 2015, le Centre africain pour la biosécurité a officiellement été rebaptisé pour devenir le Centre africain pour la biodiversité (<http://acbio.org.za/en>). Ce changement a été convenu après consultation au sein de l'ACB pour refléter la façon dont la portée de notre travail a évolué ces dernières années. Nous préserverons notre ancien nom d'African Centre for Biosafety pour toute publication antérieure à cette date ; ces publications doivent être référencées en tant que telles.)

A propos du Groupe ETC : Le Groupe ETC (www.etcgroup.org) travaille pour aborder les problèmes socio-économiques et écologiques liés aux nouvelles technologies qui pourraient avoir un impact sur les populations les plus marginalisées au monde. Il opère au niveau politique mondial et travaille en étroite collaboration avec des organisations de la société civile (OSC) et avec les mouvements sociaux, en particulier en Afrique, en Asie et en Amérique Latine. Le Groupe ETC est basé à Val David, au Canada et à Davao, aux Philippines, avec des bureaux à Mexico et à Guelph, au Canada.

A propos du Réseau Tiers-Monde, TWN, Third World Network : Third World Network (www.twn.my) est un réseau international indépendant à but non-lucratif qui rassemble des organisations et des individus impliqués dans des enjeux relatifs au développement durable, aux Pays du Sud et aux relations Nord-Sud. TWN vise à approfondir la compréhension de dilemmes liés au développement ainsi que des défis auxquels font face les pays en voie de développement afin de contribuer à des changements en matière de politiques publiques pour un développement juste, équitable et écologiquement durable. Le Secrétariat international de TWN est basé à Penang, en Malaisie avec des bureaux à Kuala-Lumpur, en Malaisie et à Genève, en Suisse. Nos chercheurs sont basés à Beijing, Delhi, Jakarta, Manille et New York. Le Secrétariat régional d'Amérique Latine est quant à lui situé à Montevideo, en Uruguay et le Secrétariat régional africain se trouve à Accra, au Ghana.

Ce document a été élaboré dans le cadre du projet BICSBAG de **Construction internationale de capacités pour l'évaluation et la gouvernance de la biologie synthétique** (en anglais Building International Capacity on Synthetic Biology Assessment and Governance). Les partenaires du BICSBAG tiennent à remercier Swedbio du Centre de résilience de Stockholm pour son soutien financier et Frontier Co-op Foundation et CS Fund pour l'élaboration de ces documents.

Pour de plus amples informations, consulter www.synbiogovernance.org

Maquette : Cheri Johnson

Traduction en français : Amandine Semat



Construction des capacités internationales pour l'évaluation et la gouvernance de la biologie synthétique

L'information sur les ressources génétiques sous forme de séquences numériques : un sujet crucial pour la Convention sur la diversité biologique

Introduction

En novembre 2018, en Egypte, l'information sur les ressources génétiques sous forme de séquences numériques sera au programme de la 14^{ème} Conférence des Parties (COP) à la Convention sur la Diversité Biologique (CDB). Les gouvernements ne peuvent se permettre d'ignorer ce sujet. En effet, les nouvelles applications de séquençage numérique transforment actuellement la façon dont les ressources génétiques sont utilisées, et présentent pour la CDB de profondes implications, à bien plus long terme, en particulier en ce qui concerne son troisième objectif : le partage juste et équitable des avantages découlant de l'utilisation des ressources génétiques.

Il est de nos jours de plus en plus facile d'avoir accès à l'information génétique de très nombreux organismes de la biodiversité — sous forme de séquences —, ce qui réduit pour bon nombre d'utilisations commerciales et du domaine de la recherche le besoin d'accéder physiquement aux plantes, aux microbes, aux animaux et autres êtres vivants. Tant que les facteurs qui favorisent ce phénomène (qui inclut le séquençage numérique à faible coût, l'édition de gènes et bien d'autres méthodes de la biotechnologie et de la biologie synthétique) continueront à se développer, cette tendance ne fera que s'accélérer. Chaque jour un plus grand nombre de séquences seront produites

« dans la nature » et partagées via électronique, sans consentement préalable en connaissance de cause de la part des pays fournisseurs ni conditions convenues d'un commun accord.

Ces technologies qui avancent à pas de géants sont en train de bouleverser au sein de la Convention sur la Diversité Biologique les approches traditionnelles quant à l'accès et au partage des avantages découlant des ressources génétiques. Elles couvrent tout le champ de la biodiversité, allant des tout petits organismes — comme les pathogènes viraux — jusqu'à de nombreux génomes, plus conséquents et plus complexes, comme ceux des plantes destinées aux cultures. Il en résulte une situation qui pourrait tout à fait favoriser la biopiraterie, puisque les cadres légaux ne font pas du tout le poids face à ces réalités techniques.

Le partage d'informations sur les ressources génétiques sous forme de séquences numériques est aujourd'hui un aspect fondamental de la recherche, mais tant que ces données seront produites et partagées sans que soient appliquées les contraintes et obligations relatives au partage des avantages découlant des ressources génétiques, les gouvernements des pays en voie de développement, les paysans et peuples autochtones qui ont créé et alimenté cette diversité seront perdants, clairement en désavantage. Les ressources génétiques nationales et les plantes des peuples autochtones se verront accaparées



et exploitées pour produire des séquences génétiques lucratives qui ne profiteront que fort peu, voire pas du tout, à ces peuples.

Ce changement imminent promet d'être si abrupt que le fait de ne pas affronter la question de l'information sous forme de séquences numériques pourrait miner la Convention sur la Diversité Biologique en empêchant la mise en œuvre de son troisième objectif et principe: le partage juste et équitable des avantages découlant de l'utilisation des ressources génétiques. Ce changement pourrait également annuler les efforts conséquents réalisés jusqu'à présent par la CDB et son Protocole de Nagoya sur l'accès et le partage des avantages pour mener à bien cet objectif.

Les gouvernements doivent accepter qu'il n'est absolument pas réaliste d'espérer des résultats de la discussion sur l'information sur les ressources génétiques sous forme de séquences numériques à la CDB s'ils ne mettent pas eux-mêmes ce sujet au centre de leurs préoccupations lors de leur préparation à la COP. Tout ceci est dû au fait que le *statu quo* sert largement les intérêts des pays utilisateurs et ceux des entreprises de biotechnologie. Les pays du Nord ne bougeront

pas de leur propre chef pour répondre efficacement à cette menace à la CDB : ils préfèrent en effet continuer à jouir de l'accès gratuit aux ressources génétiques comme ils le font actuellement grâce à l'énorme *cloud* de données numériques sur les ressources génétiques.

Pour mettre fin à cette avalanche d'informations génétiques, les Parties de la CDB doivent actualiser le texte de la convention pour trouver le moyen de faire appliquer les règles de partage des avantages quant à l'accès et l'utilisation des données génétiques numériques.

Dans le cas contraire, si les collections ex situ continuent à séquencer leurs collections, les chercheurs à développer de petits séquenceurs portables et les bases de données *online* à publier des séquences génétiques en masse, et ce sans aucun égard pour le partage des avantages et sans aucune restriction quant aux demandes de brevets, les Parties finiront par se retrouver devant le fait accompli: les bases d'accès et de partage des avantages sur lesquelles s'est érigée la CDB auront été balayées par un ouragan de *Big Data*.



Des avancées rapides quant au séquençage et à la synthèse

Une synthèse et un transfert de données plus rapides : le meilleur exemple qui permette d'observer la transformation actuellement en cours est celui des tout petits organismes. Dans le secteur de la santé, le séquençage rapide et à faible coût signifie par exemple que le séquençage intégral des virus de la grippe peut être réalisé dans les heures suivant leur isolement. Si cette séquence est téléchargée sur une base de données sur le net, ou encore envoyée par courriel, les technologies de synthèse génétique peuvent alors recréer un virus vivant en moins de trois jours, dans n'importe quel laboratoire dûment équipé, partout dans le monde.

Ainsi il est aujourd'hui possible dans certains cas de déplacer de tout petits organismes d'un endroit à un autre sur la planète, via internet, sous forme de séquences d'informations, et ce beaucoup plus rapidement que s'ils étaient acheminés en tant qu'échantillons physiques par des entreprises telles que DHL ou Federal Express.

Des génomes toujours plus longs : en plus de l'augmentation de la vitesse de transfert et de synthèse des données génétiques, la complexité des constructions géniques (que l'on mesure à la *taille* de leur génome) pouvant être synthétisées à partir d'une seule et unique séquence ne fait qu'augmenter. Le génome

du poliovirus — le premier virus à avoir été complètement synthétisé en laboratoire en 2002 — mesure environ 7 500 nucléotides. En novembre 2016, une équipe de scientifiques américains a annoncé avoir synthétisé entièrement l'adénovirus, dont le génome mesure 34 000 nucléotides, soit quatre fois et demie plus que celui du poliovirus.

Le projet du génome synthétique de la levure — une collaboration internationale de plusieurs laboratoires — prévoit de bientôt synthétiser les 16 chromosomes de *Saccharomyces cerevisiae* — 12 millions de paires de bases d'ADN qui constituent le génome eucaryote dans son ensemble.

Du séquençage à l'édition de gènes : il est actuellement impossible de synthétiser des organismes plus complexes à partir de rien, mais cela n'est absolument pas nécessaire pour que le séquençage de l'information génétique bouleverse les principes d'accès et de partage des avantages établis par la CDB. En alliant des séquences de données génétiques avec des techniques d'édition de gènes de la biologie synthétique telles que le CRISPR, il est possible d'introduire des éléments de la biodiversité appartenant à un endroit précis dans des organismes appartenant à un tout autre endroit sans qu'il n'y ait aucun contact physique, et sans signer quelque Accord type de transfert de matériel (ATTM) que ce soit.



Une entreprise pourrait par exemple trouver sur une base de données la séquence d'un composant alimentaire ou thérapeutique et l'introduire dans une cellule souche cultivée en laboratoire. L'entreprise pourrait ainsi produire un précieux nutriment ou un médicament in-vitro à base de plantes, sans jamais avoir eu accès à la plante source ni avoir souscrit à un ATTM.

Un autre exemple est celui des obtenteurs des sociétés privées attirés par certaines variétés de tomates résistantes à la sécheresse qui pourraient s'intéresser au séquençage du génome de tomates des côtes désertiques de l'Équateur, du Pérou et du Chili. En ayant accès à une quantité suffisante de données sur la caractérisation et le séquençage génétique, ces personnes pourraient alors identifier les mutations qui font que les tomates de ces régions sont plus résistantes à la sécheresse. Elles pourraient alors déposer un brevet pour s'arroger la propriété sur ces mutations et utiliser des techniques d'édition de gènes pour les introduire dans leurs cultures commerciales destinées à la vente en Amérique du Nord et en Europe, et ce, de nouveau, probablement sans se confronter à aucun ATTM ni à aucune contrainte quant à l'accès et au partage des avantages.

Ces médicaments ou variétés végétales adaptées au changement climatique ne sont que de simples exemples, semi-hypothétiques ; le nombre de traits génétiques auxquels on pourrait avoir accès de façon similaire chez d'autres espèces en utilisant des bases de données — et non pas en ayant recours

physiquement aux organismes vivants — est pratiquement infini. Les pays fournisseurs pourraient ainsi permettre sans le vouloir l'accès à leurs ressources génétiques dès lors que les informations numériques sur la biodiversité locale seraient mises en ligne sans les mesures de contrôle adéquates.

L'information sur les ressources génétiques sous forme de séquences numériques à la CDB et autres forums de l'ONU

La gestion de l'information sur les ressources génétiques sous forme de séquences numériques est un problème crucial qui va au delà de la CDB. Elle touche à la biologie lorsqu'il s'agit d'inclure des cultures ou d'autres ressources génétiques dont s'occupent le Traité international sur les ressources phytogénétiques pour l'alimentation et l'agriculture, la Commission des ressources génétiques pour l'alimentation et l'agriculture (CGRFA) de l'Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO) et dans certains cas à l'Organisation mondiale de la Santé (OMS), en particulier (mais pas exclusivement) pour le Cadre de préparation en cas de grippe pandémique (Cadre PIP).

Même si les implications de l'information sur les ressources génétiques sous forme de séquences numériques ont été anticipées depuis plusieurs années maintenant, entre autres par le Traité international sur les ressources phytogénétiques pour l'alimentation et l'agriculture qui les a mentionnées en termes de « dématérialisation », les délibérations de la CDB sur ces séquences numériques sont essentielles en raison de la portée de la Convention, de sa composition et en raison de l'importance de préserver l'approche de la FAO sur ce sujet, approche par ailleurs cohérente avec celles de la Convention.

La présente discussion de la CDB tire ses origines du rapport de 2015 du Groupe spécial d'experts techniques (GSET) sur la biologie synthétique qui a relevé l'impact potentiel sur la Convention que pourrait avoir l'information numérique sur les ressources génétiques quant à « l'accès aux ressources sans avantages partagés » [1]. La Conférence des Parties à la



CDB tenue au Mexique en décembre 2016 a alors adopté une décision sur les informations numériques qui a lancé un plan destiné à aboutir à une décision de fond lors de leur prochaine rencontre en Égypte en 2018 [2].

Lors de la COP 13, de nombreux pays en voie de développement ont pris position pour que la CDB adopte une décision qui insiste sur le fait que l'information sur les ressources génétiques sous forme de séquences numériques doit être traitée exactement de la même façon que les échantillons de matériel naturel pour honorer le partage des avantages résultant des ressources génétiques.

L'organisation Mondiale de la Santé, qui envisage l'information numérique sur les ressources génétiques dans le cadre de la recherche sur les maladies (par exemple l'accès et l'utilisation de séquences génétiques de pathogènes, en particulier du virus de la grippe) a adopté certaines mesures pratiques — encore incomplètes — pour créer des méthodes permettant de gérer l'accès aux séquences génétiques et leur utilisation dans le cadre d'un accord multilatéral sur l'accès et le partage des avantages, notamment avec le Cadre de préparation en cas de grippe pandémique de l'OMS. La question des séquences génétiques de la grippe pandémique devrait être discutée en mai 2018, lors de la 71^{ème} Assemblée mondiale de la Santé [3].

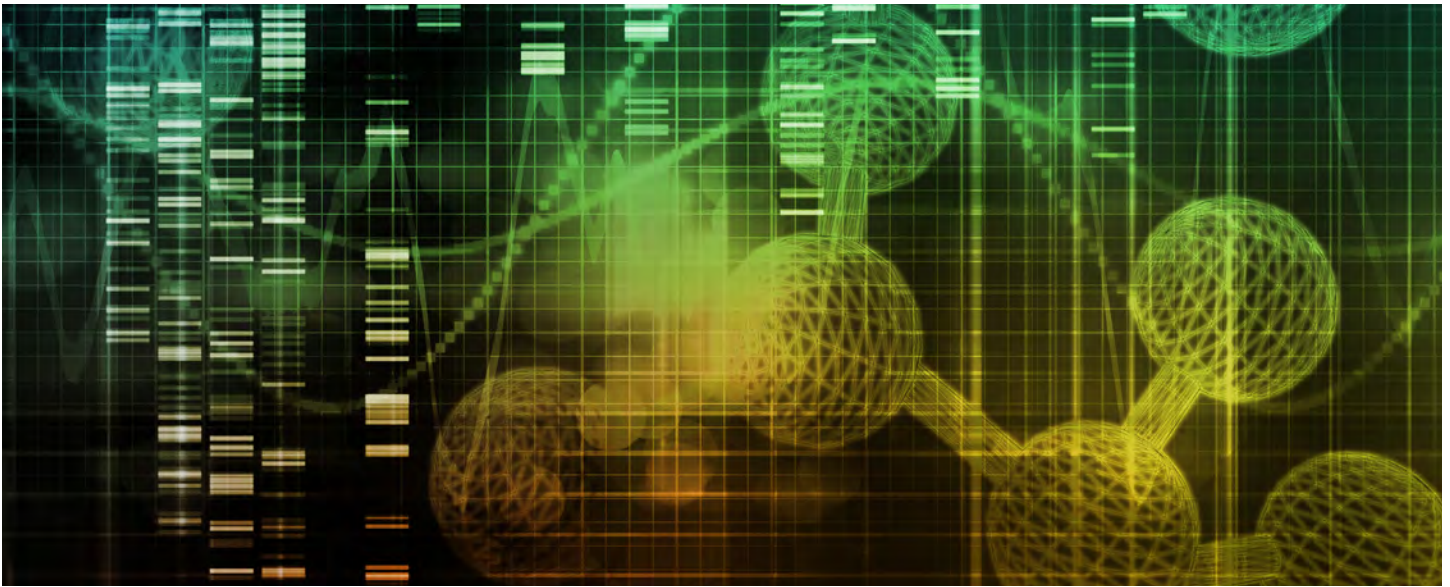
L'OMS a consulté les bases de données

hébergeant des informations sur les séquences génétiques de la grippe et envisage d'élaborer des accords destinés aux utilisateurs de ces bases de données (c'est-à-dire des approches depuis l'interface, ou *front-end*) ainsi que des stratégies qui impliqueraient les serveurs hébergeant ces données génétiques (étiquetage de séquences inclus) et une autre approche, plus complexe et plus difficile : la surveillance, depuis le serveur (*back-end*), de l'apparition de séquences de systèmes multilatéraux (pour la grippe) lors des demandes de droits de propriété intellectuelle.

Termes liés à l'information sur les ressources génétiques sous forme de séquences numériques

Les différents processus internationaux qui traitent l'information sur les ressources génétiques sous forme de séquences numériques utilisent des termes fort divers qui se chevauchent les uns les autres et font parfois double-emploi. La CDB, reconnaissant le besoin de cohérence, a donc décidé de revoir la terminologie avant sa prochaine réunion. On est en droit d'espérer que les termes qu'elle adopte seront également employés partout ailleurs de manière à faciliter la compréhension et la cohérence. Il existe aujourd'hui plusieurs termes en vigueur :

Le terme « **dématérialisation** » a été utilisé durant plusieurs années par le Traité



international sur les ressources phytogénétiques pour l'alimentation et l'agriculture. Il fait référence aux ressources génétiques et au fait que l'information numérique sur ces ressources peut, dans certains cas supplanter le recours aux espèces et spécimens naturels ; en particulier en ce qui concerne l'accès aux avantages. Les ressources génétiques sont déplacées et utilisées sans que cela n'implique le déplacement du matériel d'origine (ou germoplasme). Lors de récentes rencontres, le Traité international sur les ressources phytogénétiques pour l'alimentation et l'agriculture et la Commission des ressources génétiques pour l'alimentation et l'agriculture ont également commencé à employer le terme utilisé dès le début par la CDB : information sur les ressources génétiques sous forme de séquences numériques (DSI en anglais, voir ci-après).

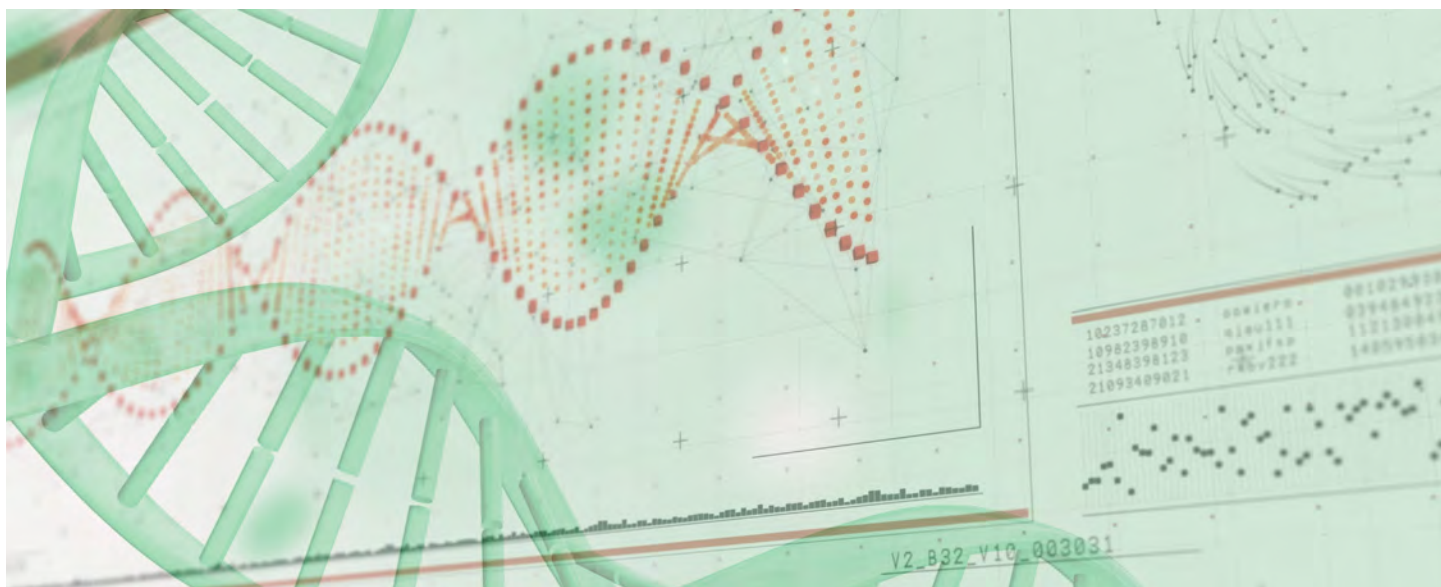
« **Données de séquençage des gènes** » (GSD, d'après son sigle en anglais) est le terme prôné par le Cadre de préparation en cas de grippe pandémique de l'OMS. La discussion sur les GSD comprend ici la séquence génétique de l'ARN du virus de la grippe et cette séquence dans ses formes alternatives, en particulier les modifications pertinentes pour la recherche, les diagnostics, et le domaine thérapeutique comme l'ADN complémentaire (ADNc) et les insertions ou délétions ayant des effets sur la virulence, la croissance en culture cellulaire, etc.

« **Digital sequence information** » (DSI, d'après son sigle en anglais) ou information numérique génétique sur les ressources génétiques est le terme proposé par le Groupe d'experts techniques de biologie synthétique de la CDB ; la portée réelle de cette expression n'a pas encore été définie, bien que beaucoup considèrent qu'elle englobe les séquences numériques d'ADN, d'ARN et d'acides aminés/protéines sous leurs diverses formes, ainsi que l'information épigénétique et autres données de caractérisation des ressources génétiques. Il a été souligné que le mot « numérique » devait être éliminé de cette expression puisqu'il pouvait être limitatif pour différentes raisons comme par exemple le fait que les systèmes d'information à venir (ordinateurs) ne seraient peut être pas numériques, et que les séquences d'information qui ne sont pas aujourd'hui stockées « numériquement » devaient également être un élément à part entière de la discussion.

Concepts-clé à appliquer pour les règles et politiques relatives à l'information sur les ressources génétiques sous forme de séquences numériques

L'information sur les ressources génétiques sous forme de séquences numériques doit être considérée comme l'équivalent des échantillons naturels (comme les semences, par exemple)

L'accès à l'information sous forme de séquences satisfait chaque jour un plus



grand nombre d'objectifs. Ces objectifs sont exactement les mêmes que ceux qui permettraient auparavant l'accès à des matériels physiques et l'utilisation de ces derniers pour créer de nouveaux produits commerciaux qui puissent être brevetés ou placés sous d'autres modalités de droits de propriété intellectuelle. Puisque les séquences de données numériques sont utilisées de cette même façon, et que cela ne fera qu'augmenter dans le futur, il faut appliquer aux séquences numériques les règles pour l'accès et le partage des ressources génétiques que l'on applique aux matériels naturels.

Les accords sur l'accès et le partage des avantages, y compris les accords de transfert de matériel, doivent être actualisés pour couvrir le champ des données de séquences génétiques

La plupart des accords et des lois ayant trait à l'accès et au partage des ressources repose sur l'hypothèse que le matériel génétique est transféré physiquement. Dans l'état actuel des choses, et tels qu'ils sont actuellement rédigés, ces accords ne peuvent être appliqués à l'information sur les séquences génétiques. Un accord sur l'accès et le partage équitable des avantages qui ne contemple pas les séquences génétiques pourrait justifier, en permettant l'accès physique aux ressources génétiques, que le destinataire produise des séquences numériques et tire profit de ces données en dehors du cadre de partage équitable des

avantages résultant des ressources génétiques. Ainsi, si les séquences génétiques ne sont pas traitées exactement de la même manière que les matériaux physiques dans les accords d'accès et de partage des avantages, la biopiraterie ne s'en verra que facilitée, et les moyens par lesquels les utilisateurs pourront échapper à ces obligations de partage des avantages continueront à proliférer.

Les sites qui hébergent des séquences génétiques doivent exiger que leurs utilisateurs acceptent le partage des avantages

Les sites hébergeant des séquences d'information génétique — y compris les bases de données telles que Genbank et l'archive européen de nucléotides (ENA d'après son sigle en anglais) — doivent être sommés d'obliger leurs utilisateurs à accepter le partage des avantages comme condition préalable et obligatoire à l'accès aux données sous forme de séquences génétiques. Les bases de données « en libre accès » ou « accessibles au public » ne signifient et ne peuvent signifier que les utilisateurs sont complètement délivrés de toute obligation. La CDB se doit de développer des règles pour mettre en place de tels accords pour les utilisateurs des données (comme par exemple accepter des Termes et conditions d'utilisation en un *clic*), et les sites hébergeant des informations sous forme de séquences génétiques devraient être sommés de les mettre en œuvre.

L'information sur les séquences génétiques doivent inclure l'ADN, l'ARN, et les séquences d'acides aminés

Les nucléotides génomiques ne sont qu'une partie des séquences qui nous importent. Le matériel héréditaire des organismes ne se limite pas à l'ADN. Il ne s'agit en effet quelquefois de l'ARN. Étant donnée la complémentarité entre les molécules et l'importance de leurs fonctions ces deux séquences génétiques doivent être prises en compte. Les séquences d'acides aminés codées dans les nucléotides sont tout aussi utiles et peuvent être utilisées pour répliquer ou modifier des composants naturels ou encore pour concevoir des systèmes biologiques.

Au delà des séquences génétiques qui servent d'éléments « de base » pour la construction de nouveaux composants ou de nouveaux organismes, il existe une quantité importante d'informations qui font également partie de la séquence génétique ; ces informations comprennent des données sur la caractérisation des ressources génétiques, l'épigénétique, la méthylation et autres informations similaires, ainsi que des données utiles et observables, comme par exemple le nombre de séquences répétées en tandem (STR en anglais).

Conclusion

Il est impossible, pour les objectifs et intentions de la CDB, de retirer l'information d'une séquence génétique une fois qu'elle a été publiée numériquement sans inclure de dispositions quant au partage des avantages. Mettre fin aux publications de ces séquences génétiques *per se* ne devrait évidemment pas être l'objectif des Parties à la Convention. La COP à la CDB doit plutôt agir rapidement pour laisser au clair le fait que les séquences génétiques doivent être accompagnées des mêmes obligations de partage des avantages que les matériels physiques. Ce principe doit également être repris pour les approches d'autres institutions telles que la FAO et l'OMS.

Ainsi, la production et l'utilisation de données et d'informations sur les ressources génétiques ne se verront pas altérées mais auront au contraire lieu chez des acteurs engagés respectant les

objectifs de la Convention quant à l'accès et au partage des avantages. Si les pays développés refusent d'appliquer les règles de partage des avantages à l'information sur les séquences génétiques lors du processus de la Convention, les États qui fournissent le matériel génétique doivent alors envisager d'utiliser les lois sur l'accès et le partage équitable des avantages ainsi que les accords existants pour restreindre le droit des utilisateurs à séquencer leurs ressources génétiques et à les partager.

Une fois que la Convention aura établi le fait que le partage des bénéfices s'applique pleinement à l'information sur les séquences génétiques, le décor sera alors planté pour la prochaine étape du grand défi que devra affronter la CDB quant à ce sujet : comment s'assurer que les capacités des *big data* sont utilisées de manière à véritablement servir les intérêts des peuples autochtones et des communautés locales qui alimentent et préservent la biodiversité, et non ceux des entreprises qui ne feraient qu'utiliser les informations issues de la technologie pour faire avancer la situation actuelle, particulièrement inéquitable.

[1] Voir UNEP/CBD/SYNBIO/AHTEG/2015/1/3. URL: <http://www.cbd.int/doc/meetings/synbio/synbioahteg-2015-01/official/synbioahteg-2015-01-03-en.pdf>

[2] CBD COP Decision XIII/16. Information génétique numérique sur les ressources génétiques. URL: <https://www.cbd.int/doc/decisions/cop-13/cop-13-dec-16-en.pdf>

[3] Voir le site du Cadre du PIP de l'OMS. URL: <http://www.who.int/influenza/pip/en/>