

L'ADN, CLÉ USB DE DEMAIN?

Comment archiver toutes les informations numériques générées chaque année ? Des chercheurs proposent de se servir de l'ADN comme d'une clé USB... éternelle, ou presque.

Par Marine Corniou

Commençons par une devinette. Où peut-on réunir en un seul endroit un film des frères Lumière de 1895, un bon d'achat d'Amazon de 50\$, un virus informatique, une publication sur la théorie de l'information et tout un système d'exploitation d'ordinateur ? Grâce à des chercheurs de l'université Columbia, tout ce contenu est désormais stocké sur un support aussi minuscule qu'improbable : des brins d'ADN.

L'exploit n'est pas si étonnant, quand on sait que l'ADN est une molécule qui porte l'information génétique de tous les êtres vivants, depuis la pâquerette jusqu'au pigeon, en passant par les bactéries et les humains. En piratant ce code universel, les chercheurs peuvent, en théorie, y inscrire n'importe quelle donnée numérique.

Pour comprendre, il faut s'imaginer l'ADN comme une succession de lettres (qui sont en fait des molécules, les nucléotides). Il en existe quatre différentes : A (adénine), C (cytosine), G (guanine), T (thymine). Dans la nature, ces suites de lettres constituent des codes complexes, des sortes de « phrases » que l'on appelle des gènes. Ceux-ci sont décryptés par les cellules : elles s'en servent comme des recettes leur permettant de fabriquer des protéines.

Mais pour stocker des données, ces lettres servent plutôt à traduire un langage numérique. « Les données numériques sont binaires, codées avec des 0 et des 1. C'est assez trivial de traduire un système binaire en base 4 [NDLR, c'est-à-dire avec 4 éléments, comme A, C, G, T]. On peut par exemple dire que les A et les T correspondent à 0 et que les C et les G correspondent à 1 », détaille Christophe Dessimoz, professeur au University



College de Londres, qui a travaillé à l'Institut européen de bio-informatique (EBI) au Royaume-Uni sur le premier cryptage du genre.

Une fois le code déterminé, il suffit de fabriquer les brins d'ADN correspondants, en assemblant les lettres une par une (cette

technique de synthèse d'ADN est déjà bien maîtrisée, et largement utilisée en recherche). Les données sont récupérées en faisant machine arrière, c'est-à-dire en décomposant l'ADN pour lire l'ordre des lettres (ce qu'on appelle le séquençage).

Voilà la théorie. En pratique, il y



a évidemment quelques pièges. Par exemple, il est impossible de fabriquer une immense molécule d'ADN contenant des millions de bases, ou lettres. Il faut donc scinder l'information en de nombreux morceaux. « C'est un peu comme un puzzle. Pour s'y retrouver,

il faut numéroter chaque brin portant l'information numérique codée, et utiliser une sorte de "code postal" qui permet de tout remettre dans l'ordre », ajoute M. Dessimoz qui est aussi chercheur à l'Université de Lausanne, en Suisse.

L'équipe de l'université Columbia a peaufiné ce système de puzzle. Sa technique permet d'encoder 60 % d'information de plus dans la même quantité de molécules qu'auparavant. Mieux encore, elle a copié les brins d'ADN une dizaine de fois sans que cela altère les données, apprend-on dans l'étude publiée par la revue *Science* en mars dernier.

Cette technique permettrait de stocker, dans un seul gramme d'ADN, 215 millions de gigaoctets – 200 000 fois plus que dans un ordinateur portable ! De quoi pulvériser en théorie le record, détenu pour l'instant par Microsoft et des chercheurs de l'université de Washington qui, à l'été 2016, ont stocké dans une éprouvette 200 mégaoctets de vidéoclips et de livres électroniques.

TROP DE DONNÉES !

Incroyablement stable et très compact, l'ADN pourrait bien être la clé de voûte du stockage des données à l'heure du *big data*. « Si toutes les données accessibles sur Internet étaient converties dans de l'ADN, cela tiendrait dans une boîte à chaussures », selon Microsoft.

La question est pressante. La Terre compte 3,5 milliards d'internautes qui génèrent 6 000 tweets, 40 000 recherches Google et 2 millions de courriels par seconde, qui échangent et produisent des monceaux de données, de vidéos, de pages web, etc. Même avec beaucoup d'imagination, il est difficile de se représenter l'immensité de l'univers numérique. D'autant que ce monde virtuel

croît à une vitesse folle. D'ici 2020, sa taille devrait atteindre les 44 zettaoctets (44 000 milliards de gigaoctets, soit 10 fois plus qu'en 2013). Concrètement, c'est l'équivalent de 11 000 milliards de DVD !

En un mot, on produit aujourd'hui plus d'information que l'on peut physiquement en stocker, que ce soit dans les centres de données ou sur des supports de type clés USB, DVD et autres disques durs, dont l'espace est limité, tout comme leur durée de vie – quelques dizaines d'années tout au plus. Voilà qui impose d'effectuer régulièrement des copies de copies.

« Pour beaucoup d'archives des années 1950, 1960 ou 1970, on ne dispose déjà plus des lecteurs qui permettent de les récupérer. L'ADN est universel : tant qu'il y aura de la vie, ce sera pertinent de l'utiliser », explique Christophe Dessimoz.

La technique n'est pas parfaite. Les temps d'accès, c'est-à-dire d'écriture et de lecture, sont loin de pouvoir rivaliser avec ceux d'une clé USB. Quant aux coûts, ils restent prohibitifs à court terme. « La synthèse coûte encore cher, mais on a démontré en 2013 que c'est déjà plus économique d'utiliser de l'ADN que des bandes magnétiques si on veut stocker de l'information à très long terme, par exemple 5 000 ans. Et ça progresse vite : il existe déjà des séquenceurs d'ADN portatifs à moins de 1 000 \$, avec un port USB. Ce n'est plus de la science-fiction », poursuit-il, ajoutant que des chercheurs proposent d'encapsuler l'ADN dans des billes de verre pour qu'il se conserve encore plus longtemps.

Qui sait ? Dans quelques millénaires, nos descendants liront peut-être cet article en fouillant dans des brins d'ADN. 