

Un algorithme de streaming vidéo aide au stockage sur ADN

De plus en plus de chercheurs se tournent vers l'ADN pour le stockage des données. Une nouvelle étude menée par des chercheurs de l'Université de Columbia et du Centre de Génomique de New York a déterminé qu'un algorithme utilisé pour le streaming vidéo sur les mobiles pouvait débloquer la capacité de stockage de l'ADN en compactant plus d'information sur les 4 nucléotides de base.

Yaniv Erlich et sa collègue Dina Zielinski ont choisi de coder 6 fichiers sur de l'ADN : un système d'exploitation complet, le film français des frères Lumière « L'arrivée d'un train à la Ciotat » (datant de 1895), une carte cadeau Amazon de 50 dollars, un virus informatique, une plaque de la sonde spatiale Pioneer et une étude du théoricien de l'informatique Shannon datant de 1948.

Un algorithme de code fontaine pour compresser les données

Les chercheurs ont compressé les documents dans un fichier maître, puis ont divisé les données dans des chaînes courtes de code binaire (composées de 1 et de 0). Ensuite, ils se sont appuyés sur un algorithme de correction d'erreurs, baptisé code fontaine (utilisé dans le streaming vidéo), pour intégrer les chaînes de manière aléatoire au sein de droplet (gouttelettes). Par la suite, ils ont cartographié les 1 et les 0 dans chaque droplet sur les 4 nucléotides de base de l'ADN : A, G, C et T. L'algorithme cité précédemment permet de supprimer les combinaisons de lettres connues pour créer des erreurs. Il ajoute également un code-barre à chaque droplet pour aider au réassemblage des fichiers *a posteriori*.

Au total, les scientifiques ont généré une liste numérique de 72 000 brins d'ADN, contenant chacun 200 bases longues. Cette liste a été envoyée dans un fichier texte à une start-up de San Francisco, Twist BioScience, spécialisée dans la transformation des données numériques en données biologiques. deux semaines plus tard, les universitaires ont reçu un flacon contenant un brin de molécules d'ADN.

215 Po de données en théorie

Pour récupérer les fichiers à partir de cet ADN, ils ont utilisé un outil de séquençage puis un logiciel spécialisé pour traduire le code génétique en binaire. Et ont récupéré l'ensemble des fichiers avec zéro erreur. Par ailleurs, les spécialistes ont démontré qu'avec leur technique de codage, ils pouvaient créer de manière illimitée des copies de leurs fichiers en multipliant les échantillons d'ADN via la technique PCR (polymerase chain reaction). Et toujours sans erreur lors de la récupération des informations.

Avec ce système de codage, les chercheurs estiment que dans 1 gramme d'ADN, ils peuvent stocker

215 Po de données. Soit 100 fois plus qu'avec les méthodes précédentes. Mais il reste encore un obstacle au développement du stockage de données sur ADN : le coût. Pour mener à bien leur expérience, les scientifiques ont dépensé 7 000 dollars pour synthétiser l'ADN embarquant 2 Mo de données et 2 000 dollars pour le lire.

A lire aussi :

[Stockage sur ADN, Microsoft en route pour le datacenter](#)

[Apple ResearchKit s'intéresse à l'étude de l'ADN](#)

Photo credit: thdoubleu via Visual hunt / CC BY-SA