



INFORMATIQUE // Les supports numériques actuels sont limités en capacité et vieillissent mal. Pour s'affranchir de ces limites, des chercheurs étudient la possibilité d'avoir recours à l'ADN.

Le stockage sur ADN, disque dur du futur ?

Yann Verdo
yverdo@lesechos.fr

Rien qu'en 2012 l'humanité a produit quelque 3 zettaoctets (3×10^{21} octets) de données numériques. Pour donner une idée du volume que cela représente, le jeune bio-informaticien Christophe Dessimoz, de l'University College London, a calculé que, si on les stockait toutes dans des clefs USB de dernière génération, d'une capacité de 1 téraoctet (10^{12} octets) chacune, celles-ci empliraient un volume équivalent à deux piscines Olympiques. Certes, les industriels produisent à tour de bras quantité de supports pour absorber ce déluge d'information. Mais cette inflation vertigineuse n'est pas le seul problème, ni même le principal. Car, fait méconnu, aucun de ces supports numériques n'est pérenne. Tous se dégradent inexorablement, entraînant la perte irréversible des données stockées. Les plus « résistants » sont les disques optiques, c'est-à-dire nos bons vieux CD et DVD : si les disques réenregistrables se dégradent très vite (moins d'un an pour certains d'entre eux), ceux qui ont été pressés restent intacts au moins une vingtaine d'années. Mais « *personne ne sait exactement combien de temps* », reconnaît le physicien Franck Laloë, rapporteur du rapport « Longévité de l'information numérique », paru en 2010.

Face à ce manque criant de supports pérennes, les biologistes songent depuis longtemps à utiliser l'extraordinaire capacité de l'ADN à stocker et transmettre de génération en génération – et ce, depuis des centaines de millions d'années – l'information génétique des individus. Après tout, l'ADN n'est rien d'autre qu'une très longue chaîne constituée de quatre bases azotées appariées deux par deux – l'adénine (A) avec la thymine (T), la guanine (G) avec la cytosine (C). De même que, depuis la naissance de l'informatique, toute information – texte, image, son – peut être encodée dans une

chaîne de bits prenant deux valeurs (0 ou 1), il devrait être possible de l'encoder avec les quatre nucléotides (A, T, G, C) constituant les maillons de la chaîne ADN.

Aussi fiable qu'un DVD

Voilà pour le principe. De là à transcrire tout le contenu d'une encyclopédie comme Wikipedia dans de l'ADN de blatte – l'une des rares espèces vivantes à pouvoir survivre à un holocauste nucléaire –, comme l'imaginent certains auteurs de science-fiction ou futurologues, il y a cependant un pas. « *Encoder de l'information dans l'ADN "naturel" d'un organisme vivant est un processus extrêmement laborieux pour un résultat instable, du fait des mutations génétiques liées à l'évolution darwinienne* », indique Christophe Dessimoz. Mais l'expérience phare du biologiste américain Craig Venter, qui a réussi l'exploit, en 2010, de synthétiser un génome complet de bactérie, a montré que l'ADN pouvait, comme n'importe quelle autre macromolécule, être fabriqué en laboratoire : les biochimistes produisent en série des nucléotides, briques élémentaires qu'il s'agit ensuite d'assembler dans le bon ordre. Fortes de cette percée technologique, deux équipes de chercheurs, l'une aux Etats-Unis, l'autre en Angleterre, ont transformé l'essai en 2012, en encodant une quantité significative d'informations dans de l'ADN de synthèse. L'équipe américaine s'est constituée à la Harvard Medical School autour de George Church. A cinquante-neuf ans, l'homme n'est pas un inconnu : il a été l'un des artisans du projet Génome humain qui a abouti en 2003 au premier séquençage (c'est-à-dire décryptage) de la totalité de l'information génétique d'un être humain. En génie du marketing, George Church a choisi de coder son propre ouvrage sur la biologie de synthèse, contenant quelque 53.000 mots et 11 images au format JPEG. Soit, une fois numérisée, une masse d'informations lourde de 5,27 millions de bits (mégabits). L'expérience, relatée dans la revue « Science » d'août 2012, a été une réussite. Moins d'un milliardième de gramme d'ADN a été nécessaire pour encoder le contenu du livre. Surtout, après séquençage, cet ADN a pu restituer le texte initial de George Church avec un taux

d'erreur remarquablement faible – moins de 2 bits erronés pour 1 million, soit une fiabilité équivalente à celle offerte par les DVD et bien meilleure que celle des disques durs magnétiques. Après s'être présenté sous la forme d'infimes grains de poussière baignant au fond d'une éprouvette, le livre est redevenu ce qu'il était, de la première à la dernière ligne !

L'expérience menée parallèlement par le Britannique Nick Goldman à l'Institut européen de bioinformatique (EBI) est peut-être encore plus spectaculaire. Publiée dans la revue « Nature » en janvier 2013, elle a consisté à encoder de l'information provenant de supports variés : les 154 sonnets de Shakespeare, un fac-similé au format PDF de l'article fondateur de Watson et Crick sur la structure en double hélice de l'ADN, un enregistrement audio au format MP3 du discours du pasteur King « I have a dream », une photo au format JPEG de tous les membres de l'équipe, etc. Au total 6,06 mégabits de données, soit un volume légèrement supérieur à celui de l'équipe d'Harvard – et correspondant grosso modo au contenu d'une vieille disquette.

Surtout, Nick Goldman et ses collègues (dont Christophe Dessimoz) ont mis au point un système d'encodage – de transformation des 0 et des 1 en A, T, G, C – plus sophistiqué que celui utilisé outre-Atlantique. Résultat : tous les documents cryptés et stockés dans de l'ADN de synthèse ont pu, après séquençage, être récupérés sans aucune erreur !

Une double supériorité

Ces deux prouesses pourraient, à terme, révolutionner le stockage de données numériques. Car l'ADN affiche sur tous les autres supports connus une double et écrasante supériorité. La première est son extraordinaire densité. Après réplification de l'ADN par réaction en chaîne par polymérase (PCR), le livre de George Church se trouvait au fond de l'éprouvette non pas à un mais... 70 milliards d'exemplaires ! Un membre de son équipe a calculé que 4 grammes d'ADN suffiraient à stocker la totalité de l'information numérique produite par l'humanité en une année. Christophe Dessimoz se veut plus prudent : « *Selon nos estimations, cette masse d'informations sous forme d'ADN emplirait un volume équivalent au coffre d'une voiture* », dit-il. Ce qui reste pour le moins acceptable.

L'autre avantage clef de l'ADN est son extraordinaire stabilité dans le temps. A cet égard, le tout récent séquençage du génome d'un hominidé vieux de 400.000 ans, à partir de l'ADN retrouvé dans un fragment d'os au fond d'une grotte espagnole, est particulièrement éloquent. Lyophilisé et conservé dans de bonnes conditions, c'est-à-dire à l'abri de la lumière et de l'humidité, l'ADN conserve son information des centaines de milliers, voire des millions d'années. Voilà qui relègue l'horizon d'une vingtaine d'années offert par nos actuels disques optiques à... la préhistoire. ■