médecine/sciences 2025 ; 41 : 291-3



# Chroniques génomiques

## L'ADN comme mémoire informatique (suite)

**Bertrand Jordan**

Biologiste, généticien et immunologiste, Président d'Aprogène (Association pour la promotion de la Génomique), Marseille, France.
brjordan@orange.fr

L'emploi d'ADN comme mémoire informatique [1] (➜) continue de susciter un vif intérêt. *Biomemory*, une start-up française spécialisée dans cette approche, vient de conclure une belle levée de fonds de dix-sept millions d'euros [2]. Fondée en 2021 par des chercheurs du CNRS, l'entreprise[1] a pour but de mettre au point le codage de fichiers dans l'ADN ; elle a déjà réalisé une démonstration de faisabilité en déposant aux Archives Nationales des capsules d'aluminium contenant un ADN qui code pour la déclaration des droits de l'homme et du citoyen [3] (*Figure 1*).

**(➜) Voir *m/s* n° 6-7, 2018, page 622**

Cela constitue un joli coup médiatique, mais il ne s'agit que de courts textes, des fichiers « pesant » quelques kilo-octets (Ko, voir *Encadré*) : on est loin des méga- et giga-octets instantanément accessibles sur nos ordinateurs et nos téléphones. Néanmoins le problème du stockage des données est aigu, et l'ADN offre une solution en principe prometteuse ; d'ailleurs beaucoup d'autres firmes, et non des moindres, s'y intéressent, notamment aux États-Unis. Un consortium d'entreprises dont les membres fondateurs sont *Twist Biosciences* (leader de la synthèse d'ADN), *Illumina* (leader du séquençage), *Western Digital* (spécialiste des disques durs et des mémoires SSD (*Solid State Drive*), et *Microsoft* (qu'on ne présente plus ) a été créé en 2020 ; baptisé *DNA Storage Alliance*[2], il regroupe aujourd'hui une trentaine de partenaires, en majorité des entreprises de biotechnologie et d'informatique mais aussi quelques laboratoires universitaires. L'objectif est bien sûr de faire progresser la technologie (on imagine néanmoins que le secret industriel interviendra pour limiter les échanges), mais aussi de définir des formats et des standards et, plus généralement, de faciliter l'adoption de cette nouvelle technologie. *Biomemory* a naturellement adhéré à ce consortium.


***Figure 1.*** Capsules contenant sous forme d'ADN la déclaration des droits de l'homme (à gauche) et de la femme (à droite), déposées par l'entreprise *Biomemory* aux Archives Nationales (site de *Biomemory*).

### L'étendue du problème

La masse de données enregistrées dans des systèmes informatiques continue à croître de manière exponentielle : on estime qu'elle représente aujourd'hui environ 45 Zo (Zetta octets, voir *Encadré*), soit, pour prendre une image concrète en envisageant le stockage de masse actuel le plus performant, 45 *millions* de bandes magnétiques à haute capacité (1 000 To par bande). Il s'agit par exemple de données de radio astronomie dont l'exploitation dans cinq ou dix ans révèlera peut-être de nouveaux phénomènes cosmiques, ainsi que des très abondantes informations numériques produites par les véhicules autonomes (ou semi-autonomes) et qui sont vitales pour l'amélioration de ces systèmes, et bien sûr de données génétiques, biologiques et médicales essentielles aux progrès de la recherche et des thérapies. Toutes ces données doivent être conservées et rester accessibles à long terme, même si beaucoup d'entre elles ne seront jamais utilisées. Pour ne rien arranger, les systèmes de stockage actuels sont limités dans le temps : si l'on veut sauvegarder durablement

Vignette (© Bertrand Jordan).

[1] https://www.biomemory.com/

[2] https://dnastoragealliance.org/


https://doi.org/10.1051/medsci/2025035

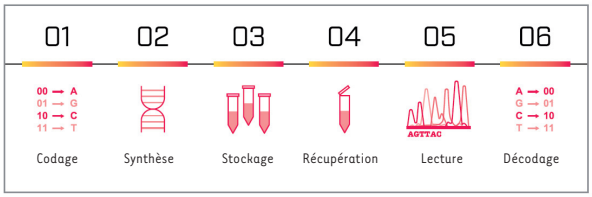


***Figure 2.*** Les étapes du stockage et de la récupération de données en utilisant le support ADN. Le codage indiqué consacre deux bits (pouvant prendre la valeur 0 ou 1) au codage d'une base (T, A, G ou C) ce qui est le minimum nécessaire (voir le texte). Image adaptée à partir du site de la *DNA Storage Alliance*.

les informations, les disques durs ou les mémoires SSD (*Solid State Drive*) doivent être recopiés tous les cinq ans environ, les bandes magnétiques tous les dix. Sans parler des évolutions technologiques : qui est en mesure aujourd'hui de lire une disquette de 1,44 Mo, ou tout bêtement un fichier Word pour MS-Dos de 1997 ? Utiliser l'ADN come un support d'information ultra-compact et quasiment éternel serait susceptible de répondre à ces différentes problématiques – mais pour en arriver là il faut résoudre de sérieuses difficultés techniques.

## La solution ADN

La nature nous révèle un système de stockage ultra-compact : le noyau d'une cellule humaine dont le diamètre est inférieur à 10 µM renferme six milliards de bases d'ADN (deux génomes haploïdes), donc la possibilité de coder au moins autant d'octets (voir plus loin), soit une capacité informatique de six Go. Avec des hypothèses raisonnables sur le mode de codage et la redondance on arrive à une densité possible de l'ordre de 10 To par millimètre cube d'ADN, très supérieure à tous les systèmes de stockage actuels. Cela suppose bien sûr de coder l'information dans le langage de l'ADN (passer de 0 ou 1 à T, A, G ou C) ce qui est trivial du point de vue informatique (*Figure 2*) ; il faut ensuite synthétiser l'ADN correspondant et le stocker dans de bonnes conditions de conservation. Lorsqu'on veut lire l'information, il faut lire (séquencer) l'ADN puis décoder les données pour retrouver le fichier d'origine. Ces étapes sont indiquées dans la *Figure 2*, nous les détaillerons un peu par la suite.

### Les étapes à franchir et l'état de l'art

• Le codage de l'information à stocker peut-être envisagé de manière très simpliste, comme indiqué dans la *Figure 2*. En réalité le problème est plus compliqué compte tenu de la nécessité d'introduire des redondances et des mécanismes de correction d'erreurs. Le codage choisi doit aussi faciliter l'accès direct à n'importe lequel des fichiers enregistrés (accès aléatoire ou *Random Access*) sans qu'il soit nécessaire de lire l'ensemble des informations. Au total, la plupart des systèmes actuellement proposés consacrent un octet (8 bits) au codage d'une base : c'est surabondant puisqu'un octet permet de coder 256 caractères différents ($2^8$) mais cela permet de satisfaire aux besoins. En pratique on peut donc considérer qu'une base dans l'ADN équivaut du point de vue informatique à un octet. Différentes méthodes de codage ont été proposées avec des performances satisfaisantes, et on peut considérer que cette question est « sous contrôle ».

• La synthèse de l'ADN est un des points d'achoppement des systèmes actuels. Il n'est pas possible de synthétiser de longs brins d'ADN, et malgré plus de trente ans d'efforts la taille des oligonucléotides de synthèse ne dépasse toujours guère, en pratique, 100 nucléotides – au-delà le rendement baisse et les erreurs s'accumulent. On ne peut donc pas stocker un fichier sous forme d'une (très) longue molécule d'ADN : il faut le représenter par un ensemble d'oligonucléotides dont la séquence, outre une parcelle de l'information à coder, comporte une adresse permettant l'assemblage du message. C'est faisable – mais à ce jour la synthèse d'ADN reste beaucoup trop chère, au moins mille dollars par Mo alors que l'on souhaite stocker des Go, soit des milliers de Mo. Pour fixer les idées, notons qu'actuellement on estime le coût du stockage d'un To (un million de Mo) sur bande magnétique à 15 dollars [4, 5] ! La synthèse d'ADN est aussi beaucoup trop lente : le record actuel correspondrait à l'écriture de 20 000 Mo en 24h alors qu'un disque dur écrit un tel fichier en quelques minutes.

• Le stockage de l'ADN, en revanche, ne présente guère de difficulté : c'est un des points forts de cette approche. Un ADN lyophilisé stocké à l'abri de l'humidité se conserve *a priori* des milliers d'années :

### Bits, bytes, octets, Tera, Peta et Exa octets

Le bit est l'unité élémentaire d'information, pouvant prendre comme valeur 0 ou 1. Un ensemble de huit bits constitue un octet (noté o) que les anglo-saxons appellent une byte. Les capacités des systèmes de stockage sont généralement exprimées en octets. Un smartphone ou un ordinateur comportent généralement une mémoire de quelques dizaines ou centaines de gigaoctets (Go, un milliard d'octets). Les symboles des quantités correspondent à un saut d'un facteur 1 000[1] : octet, Kilo octet (Ko, une page de texte), méga octet (Mo, une image), Giga octet (Go, un film de 2 heures), Tera octet (To, un disque dur amovible), Peta octet (Po, 1 000 To soit la capacité d'une bande magnétique de stockage) et, en sautant chaque fois d'un facteur 1 000, Exa octet (Eo) puis Zetta octet (Zo). J'essaie de ne pas trop jongler avec ces différentes unités, mais ce n'est pas toujours possible !

[1] En toute rigueur un facteur 2 puissance 10 soit 1024.

on a bien réussi à séquencer l'ADN de Néandertal extrait d'ossements vieux de cent mille ans[3] [6] (➜) ! (➜) Voir *m/s* n° 6-7, 2024, page 556

Et une civilisation future découvrant cet ADN pourra toujours le séquencer, quelle qu'ait été l'évolution des systèmes informatiques. Reste à organiser ce stockage de façon à pouvoir accéder rapidement au fichier recherché : c'est le problème de l'accès aléatoire (*random access*) évoqué ci-dessous.

• La récupération de l'ADN est en principe simplissime : on ouvre la capsule et on prélève l'ADN que l'on va ensuite séquencer. Mais si la capsule contient des milliers de mégabases d'ADN représentant des milliards d'octets, on ne souhaite pas lire l'ensemble pour accéder au fichier recherché. Sur un disque dur ou une mémoire SSD c'est trivial : on accède à la liste des dossiers et fichiers contenus, et on clique sur celui que l'on souhaite utiliser. Sur une bande magnétique, il faut dérouler la bande jusqu'à ce qu'on arrive à ce fichier, c'est beaucoup plus lent[4]. Pour l'ADN, on peut imaginer différents systèmes combinant savamment un codage bien conçu et un stockage sous forme de nombreux petits plots contenant chacun un ensemble différent d'oligonucléotides [7]. Cela reste une complication assez gênante, d'autant plus que les performances actuelles de la lecture sont très médiocres.

• La lecture de l'information passe par le séquençage de l'ADN. On sait que les techniques ont fait des progrès fantastiques : le premier séquençage de l'ADN humain a coûté plus d'un milliard de dollars ; en 2020 ce coût était tombé à mille dollars par génome (3 000 mégabases ou autant de Mo) ; on approche aujourd'hui du « génome à cent dollars » ce qui correspondrait à trente dollars pour lire 1 000 Mo. En imaginant que les prix baissent encore, cela devient envisageable ; mais un autre paramètre est à considérer, celui de la vitesse de lecture. Aujourd'hui il faut environ une journée pour lire un génome humain : superbe progrès puisque la première lecture avait pris dix ans, mais c'est encore beaucoup trop long. Il faudrait arriver à quelques secondes pour égaler les performances d'un disque dur, et on imagine mal comment y arriver.

• Le décodage, enfin, n'est probablement pas un problème, même si l'on adopte des schémas très sophistiqués pour éviter les erreurs et traiter le problème de l'accès aléatoire. Les performances sans cesse croissantes de l'informatique y veilleront…

## De la coupe aux lèvres

On le voit, l'emploi de l'ADN comme stockage informatique est une approche très séduisante mais dont la mise en pratique se heurte encore à de sérieuses difficultés. Le schéma de principe est clair (*Figure 2*), les avantages (compacité extrême, durée presque infinie) sont évidents, mais les technologies sous-jacentes, essentiellement la synthèse d'ADN et son séquençage, sont encore beaucoup trop lentes et trop chères pour une mise en pratique. Il faudrait gagner non pas un, mais plusieurs ordres de grandeur pour que cela change. Certes, on a vu de telles évolutions par le passé (informatique, séquençage), mais rien ne garantit que ce sera possible dans le cas présent. L'attrait médiatique est certain et, comme pour d'autres tentatives assez peu crédibles (la résurrection du mammouth [8] (➜), ou même le dépistage ultra-précoce des cancers [9] (➜), les investisseurs sont séduits. Mais les technologies actuelles de stockage de l'information, perfectionnées au fil de décennies de développement et aujourd'hui ubiquitaires, ne seront pas faciles à détrôner, à moins que d'imprévisibles révolutions scientifiques ne changent la donne. ◊

(➜) Voir *m/s* n° 5, 2022, page 480

(➜) Voir *m/s* n° 10, 2024, page 789


## SUMMARY

### Progress on DNA information storage

Using DNA as a storage medium for digital data promises ultra-high density and safe long-term storage. A number of companies are actively working in this space; they have demonstrated the feasibility of the approach and made progress on its implementation. However, speed and cost improvements spanning several orders of magnitude are needed before storage of information on DNA becomes practical. ◊



### LIENS D'INTÉRÊT

*L'auteur déclare n'avoir aucun lien d'intérêt concernant les données publiées dans cet article.*


## RÉFÉRENCES


1. Jordan B. L'ADN comme mémoire informatique ? *Med Sci (Paris)* 2018 ; 34 : 622-5.
2. Cheminot J. Biomemory, spécialiste du stockage ADN lève 17 M€. *Le Monde informatique*, 10 décembre 2024.
3. Seramour C. Premiers documents encodés sur ADN aux Archives Nationales. *Le Monde informatique*, 24 novembre 2021.
4. Doricchi A, Platnich CM, Gimpel A, *et al.* Emerging Approaches to DNA Data Storage: Challenges and Prospects. *ACS Nano* 2022 ; 16 : 17552-71.
5. Wang S, Mao X, Wang F, Zuo X, Fan C. Data Storage Using DNA. *Adv Mater* 2024 ; 36 : e2307499.
6. Bon C. La paléogénétique ou l'intérêt de l'exploration du passé. *Med Sci* (Paris) 2024 ; 40 :556-9.
7. Organick L, Ang SD, Chen YJ, *et al.* Random access in large-scale DNA data storage. *Nat Biotechnol* 2018 ; 36 : 242-8.
8. Jordan B. La « dé-extinction » du mammouth laineux, un « Colossal » canular ? *Med Sci (Paris)* 2022 ; 38 : 480-3.
9. Jordan B. GRAIL, un rêve de médecine préventive ? *Med Sci (Paris)* 2024 ; 40 : 789-91.


[3] Certes, il était assez dégradé, mais les conditions de conservation n'étaient pas idéales. Cependant on est tout de même parvenu à le lire !

[4] C'est pour cela que les bandes magnétiques sont essentiellement utilisées pour l'archivage.


### TIRÉS À PART

B. Jordan