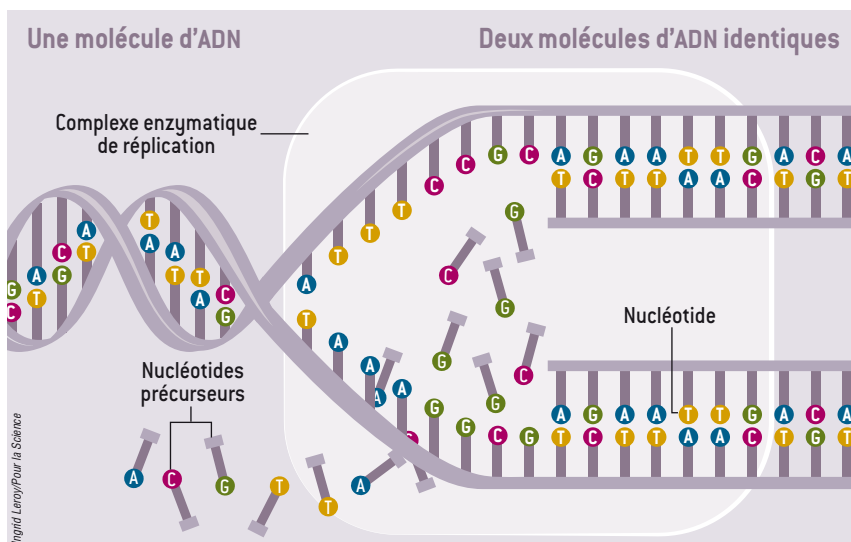




1. LE MÉCANISME qui confère sa pérennité à l'information codée dans le génome est la réplication de l'ADN, orchestrée par un complexe protéique, le complexe enzymatique de réplication. Les deux brins de la molécule d'ADN sont séparés et, sur chaque brin, sont recrutés les nucléotides complémentaires de la séquence lue par le complexe (A et C sont respectivement les complémentaires de T et G). On obtient ainsi deux nouvelles molécules d'ADN identiques.

l'intégralité du futur organisme, déjà formé, mais en plus petit – une sorte de modèle réduit de l'organisme. Mais pour la théorie de l'épigenèse, sa concurrente depuis l'Antiquité jusqu'au XVIII^e siècle, les graines constituaient un plan de l'organisme, que le développement allait réaliser selon des lois d'organisation qui restaient à découvrir. L'épigenèse était ainsi déjà presque une théorie de l'information au sens de Shannon.

Au début du XX^e siècle, la génétique a donné à l'hérédité son caractère numérique: le gène – qui n'était pas encore identifié comme un segment d'ADN – était un objet abstrait que l'on pouvait observer dans un nombre fini d'états disjoints, les allèles, codant les caractéristiques héréditaires des organismes. La découverte, en 1953, de la structure moléculaire du support de l'information héréditaire a clarifié le concept de gène: le plan des organismes contenu dans les graines est un code numérique porté par une molécule, l'ADN.

Cette molécule est un assemblage linéaire de petites molécules – des nucléotides – de quatre types: adénine (A), guanine (G), cytosine (C) et thymine (T). Cet assemblage porte une information numérique: comme les mots dans une écriture alphabétique ou les séquences de 0 et de 1 qui codent les données et les programmes dans les ordinateurs, l'ADN forme un texte dont les mots sont les gènes, écrits dans

un alphabet de quatre lettres A, G, C et T, correspondant aux quatre nucléotides. Ce texte code le plan de l'organisme tel que le prévoyait l'épigenèse.

La réplication, un excellent moyen de conservation

Portée par l'ADN, une information numérique est donc transmise, de génération en génération, lors de la reproduction. Cette information, que l'on nomme le génome, est contenue dans toutes les cellules de tous les êtres vivants. Elle se propage par la réplication de l'ADN, qui le rend pérenne (voir la figure 1). En effet, le support lui-même, comme les multiples supports de l'information numérique en informatique (carte perforée, cassette, cartouche, disquette, disque, mémoire flash), n'est pas conservé indéfiniment. L'ADN d'une cellule qui meurt se dégrade vite, notamment parce qu'il est assimilé par d'autres organismes.

Dans des conditions exceptionnelles de conservation, on peut retrouver une partie de l'information qu'il contient après quelques milliers d'années, difficilement après quelques dizaines de milliers d'années. Ce temps de conservation peut paraître long à l'échelle de nos vies, mais la réplication, elle, permet une préservation de l'information à une tout autre échelle: depuis plus de trois milliards d'années elle permet la copie des génomes et, en quelque sorte, la conservation de cette information dans les organismes vivants.

Grâce à l'ADN et à sa réplication, l'information contenue dans les génomes se propage, se transmet, se copie, se pérennise. Les organismes vivants, en se reproduisant, permettent cette transmission. En informatique aussi, la copie et la transmission de l'information sont un moyen de conservation des données pour faire face aux défaillances possibles des supports. Néanmoins, toute copie, toute transmission *via* un support, provoque des erreurs. Aucun support ni aucun canal de transmission n'est totalement sûr.

Le codage et la transmission de l'information numérique afin de minimiser ces erreurs sont l'objet de la théorie des codes. Le principe est d'introduire de façon méticuleuse et parcimonieuse une « redondance » dans l'information, afin de détecter

Lexique

Aufhebung : Ce terme allemand cher à Hegel désigne, en biologie, la dualité de la mémoire génétique du monde vivant, à la fois préservée et peu à peu détruite par les mutations.

Dérive génétique : Évolution d'une population au hasard de mutations sans effet sur la reproduction, et donc non soumises à une pression de sélection.

Épigenèse : Théorie selon laquelle les graines constituent un plan de l'organisme et non une copie complète miniature de l'organisme parent.

Épigénétique : Ensemble des modifications de l'expression des gènes qui ne sont pas codées dans le génome.

Phylogénie : Généalogie du vivant fondée sur l'étude des caractères anatomiques et moléculaires des organismes vivants.

et de corriger les erreurs. De même, la réplication de l'ADN ne produit jamais des copies parfaites de l'information. La structure de l'ADN et des mécanismes de réparation permettent de corriger une partie des erreurs, sur le principe de la redondance. L'ADN est constitué de deux brins enroulés en une double hélice. Indispensable à la réplication, cette structure en double brin est aussi une façon de conserver l'information en deux copies identiques. Chaque nucléotide a un complémentaire, auquel il est apparié dans la double hélice : adénine est complémentaire de thymine, et guanine de cytosine. En conséquence, l'information est présente deux fois – sur une séquence de nucléotides et sur son complémentaire.

Lors de la réplication, il arrive que les mécanismes dérapent et que dans le brin copié, un nucléotide soit perdu ou inséré, ou pris pour un autre. Il en résulte un « mésappariement », c'est-à-dire la mise en correspondance d'un nucléotide avec un autre qui n'est plus son complémentaire. La plupart du temps, les erreurs ainsi détectées sont réparées en rétablissant le nucléotide complémentaire.

Mais le mécanisme de correction a ses limites, et les génomes répliqués contiennent toujours des différences par rapport à leurs ancêtres. D'abord parce que la fidélité a un coût pour l'organisme, en énergie et en temps passé à fabriquer la génération suivante. Ces erreurs constituent ce que l'on nomme les mutations. Il en existe de multiples possibles, de toutes les formes

et de toutes les tailles : remplacement d'un nucléotide par un autre, insertion ou perte d'un nucléotide ou d'un segment de nucléotides, duplication ou déplacement d'un segment, duplication du génome tout entier. Au moins une centaine de mutations séparent chez l'homme les génomes des parents et des enfants, qui contiennent environ trois milliards de nucléotides. La seconde raison pour laquelle les systèmes vivants ne sont pas parfaitement fidèles, c'est parce que dans ces erreurs figurent parfois les clés de leur survie.

Une information qui évolue est-elle préservée ?

Aufhebung, en allemand, signifie à la fois préservation et destruction. C'est un de ces mots à double sens que le philosophe Georg Wilhelm Friedrich Hegel (1770-1831) affectionnait pour évoquer la dialectique, et c'est le mot qu'ont repris, en 1965, le biologiste austro-américain Emile Zuckerkandl et le chimiste américain Linus Pauling pour décrire la conservation de la mémoire du monde vivant par les molécules : préservation infinie en théorie grâce à la réplication, mais aussi destruction progressive par les mutations.

La plupart des mutations dans les génomes ne se propagent pas aux générations suivantes. D'une part, la sélection naturelle fait un tri. Si une mutation modifie l'efficacité de la reproduction dans un

■ LES AUTEURS



Vincent DAUBIN est directeur de recherche CNRS au Laboratoire Biométrie et biologie évolutive (UMR CNRS 5558) de l'Université Claude Bernard Lyon 1.



Simon PENEL est ingénieur d'étude CNRS dans le même laboratoire.



Éric TANNIER est chargé de recherche de l'Inria dans le même laboratoire.

Ils sont tous trois membres du projet ANCESTRUME qui vise à reconstruire les caractéristiques des organismes disparus en analysant leurs génomes avec des méthodes phylogénétiques (ancestrume.univ-lyon1.fr).

L'ADN et la métaphore informatique, comparaisons et raisons

Voir l'ADN comme une mémoire numérique pousse à l'analogie entre le monde vivant et les ordinateurs. L'ADN est le support d'une information numérique – le génome –, tout comme un disque dur est le support du logiciel ou des données. La métaphore peut être poussée loin, puisque les génomes sont des chaînes de caractères, à l'instar des données informatisées, suites de 0 et de 1. Le génome peut être vu comme un programme contenant les instructions nécessaires à la fabrication et à la reproduction d'un organisme, et le code génétique, qui permet la traduction des séquences génomiques en pro-

téines, est une sorte de moyen d'interpréter ces instructions.

On peut aller encore plus loin. L'évolution a été qualifiée de bricolage par le biologiste François Jacob. Or la conception d'un programme informatique pourrait souvent, sans mettre en cause le professionnalisme des programmeurs, supporter cette qualification : un logiciel évolue toujours au fil du temps, chaque programmeur reprenant des morceaux de code de ses prédécesseurs en les amendant ou en intégrant des modules extérieurs, tels des transferts horizontaux. L'évolution des données numériques est ainsi comparable à celle des génomes.

Le génome a toutefois un statut particulier. Il constitue une sorte de programme autonome, puisqu'il code aussi la fabrication de la machine qui sera capable d'exécuter le programme et de se dupliquer. Là, on dépasse les possibilités des ordinateurs et des programmes informatiques (ce qui sera peut-être démenti sous peu par les imprimantes 3D autorépliquatives *RepRap*). De plus, le génome ne porte pas tout le matériel héréditaire : une part importante d'information est transmise lors de la reproduction, en dehors de la séquence, par « épigénétique ». Faute de bon modèle d'évolution intégrant ces caractères, la métaphore informa-

tique s'arrête à l'information numérique, c'est-à-dire au génome.

On a les métaphores de son temps. Au XVIII^e siècle, les machines les plus perfectionnées étaient les automates, et la métaphore a beaucoup été utilisée pour décrire de façon mécaniste le fonctionnement du corps humain. On a aussi beaucoup comparé l'ADN à un langage, à une carte ou à un atlas. Aujourd'hui, la métaphore informatique est facilitée, d'une part, par l'utilisation constante des ordinateurs pour stocker les génomes et, d'autre part, par l'idée qu'il serait possible d'utiliser l'ADN pour le stockage de données informatiques.