



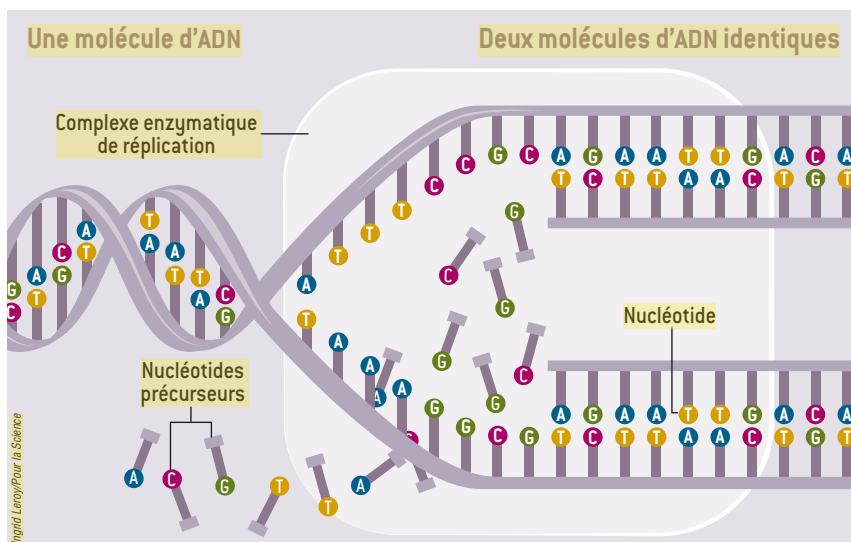
Depuis plus de trois milliards d'années, des informations numériques écrites en langage ADN sont copiées, plus ou moins fidèlement, puis interprétées, conduisant à la formation de nouvelles générations d'organismes. À ce titre, la théorie de l'évolution est une théorie de l'information, analogue à celle qu'on utilise en mathématiques depuis sa formalisation en 1948 par le mathématicien américain Claude Shannon : un ensemble de lois décrivant comment un objet – un texte dans le cas de la théorie mathématique, un organisme dans le cas de la reproduction – peut être reconstitué à partir d'une information sur cet objet, en général codée de façon plus compacte que l'objet lui-même. La biodiversité actuelle résulte de ce processus de duplication – ou réplication – avec erreurs des données génétiques héritées d'une multitude d'ancêtres.

Dans les génomes des organismes actuels figurent les traces de ces milliards d'années de transmission d'informations codées génétiquement. Aussi, chaque lignée ayant enregistré des événements différents depuis sa séparation des autres, il est possible, en comparant ces génomes, de reconstruire l'histoire du vivant. Depuis 50 ans, les phylogénéticiens, qui étudient les relations évolutives entre les organismes, comparent les séquences ADN des espèces actuelles. Avec les progrès du séquençage génomique, ces recherches ont changé d'échelle. Elles permettent aujourd'hui de comprendre à la fois les mécanismes de l'évolution et l'origine des espèces. Quelles informations anciennes peut-on espérer décrypter dans les génomes actuels ? Quelles sont les méthodes employées ? Quelles sont les limites de cette approche ? L'analogie avec la théorie de l'information apporte plusieurs réponses.

L'hérédité : une théorie de l'information

Si, en semant des graines de tomate, on obtient des tomates et non des courgettes, c'est parce que dans les graines figure une information qui dicte la forme de la plante. La nature de cette information, celle qui est transmise des parents à leur progéniture dans tout le monde vivant lors de la reproduction, a été longuement débattue. Selon la théorie de la préformation, par exemple, dont on trouve des traces dès l'Antiquité, chaque graine contenait

© Shutterstock/Tatiana33



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.