

## SCIENTES &amp; TECHNIQUES

## ADN : L'ÈRE DE L'ÉCRITURE

Marc CLASTRE\*

*Peut-être l'histoire universelle n'est-elle que  
l'histoire des diverses intonations de  
quelques métaphores.*

J.-L. Borges, *La sphère de Pascal*<sup>1</sup>

**RÉSUMÉ :** Au cours du vingtième siècle, la révolution moléculaire en biologie a eu un impact majeur sur le développement du génie génétique, de la génomique puis de la biologie synthétique. Ces bouleversements dans les sciences de la vie sont abordés à travers le prisme des écrits de l'écrivain Jorge Luis Borges, les références métaphoriques à son œuvre ayant investi le champ de la biologie. Ces analogies trouvent un nouvel écho à l'aune de l'émergence de la biologie synthétique. Après l'ère de la lecture de l'ADN via le séquençage massif et l'annotation des génomes, la biologie entre aujourd'hui dans l'ère de l'écriture en abordant la question de la préservation d'une partie du patrimoine de l'humanité par l'encodage de vastes données numériques dans l'ADN. Pour autant, épuiserons-nous le paradigme de la lecture de l'ADN ?

**ABSTRACT:** In the twentieth century, the molecular revolution in biology has had a major impact on the development of genetic engineering, genomics and finally synthetic biology. These upheavals in the life sciences are examined here through the prism of the writings by Jorge Luis Borges, as metaphorical references to his work have entered the field of biology. These analogies find a new echo in the light of the emerging synthetic biology. After the DNA reading era via massive genome sequencing and annotation, biology has now entered the DNA writing era by tackling the issue of

---

\* Marc Clastre, Professeur de Biologie Moléculaire  
Unité de Recherche 2106 Biomolécules et Biotechnologies Végétales, Faculté de Pharmacie  
Philippe Maupas, Université de Tours. e-mail : marc.clastre@univ-tours.fr  
Communication présentée en séance du 13 octobre 2023.

1. Borges 2010 : 676-679.

preserving parts of humanity's heritage by encoding vast amounts of digital data in DNA. However, the question remains as to whether we will ever exhaust the DNA reading paradigm.

## L'émergence d'une nouvelle discipline

La biologie moléculaire est une discipline scientifique dont l'étude porte sur les structures chimiques et les processus biologiques impliquant les unités de base de la vie, à savoir les molécules. Le domaine de la biologie moléculaire se concentre plus particulièrement sur les acides nucléiques (ADN : acide désoxyribonucléique ; ARN : acide ribonucléique) et les protéines et sur la manière dont ces molécules interagissent et se comportent au sein des cellules.

La biologie moléculaire apparaît dans les années 1940<sup>2</sup>. Elle s'est développée à partir de domaines connexes dont la biochimie, la génétique, la biophysique et la cyberscience (théorie de l'information, cybernétique...) et reste aujourd'hui étroitement liée à ces domaines. L'essor de la biologie moléculaire est un remarquable exemple de révolution scientifique qui s'inscrit dans le sillage de la révolution quantique qui s'est opérée en physique au début du XX<sup>e</sup> siècle.

Dans son ouvrage *Qu'est-ce que la vie ?* paru en 1944, le physicien Erwin Schrödinger, considéré comme l'un des pères de la biologie moléculaire, propose d'aborder le vivant à l'échelle de la molécule, d'étudier le chromosome (qu'il nomme le cristal aperiodique) qui contient les gènes et émet l'hypothèse de l'existence d'un code en miniature qui s'avérera être vingt ans plus tard le code génétique. Schrödinger est le premier à poser le problème central de la biologie en termes d'information codée<sup>3</sup> :

Les atomes de l'hérédité sont organisés selon un ordre imposé du cristal aperiodique qui contient des motifs variables [...] Avec l'image moléculaire du gène, il n'est pas inconcevable que le code en miniature puisse se trouver en correspondance exacte avec un plan de développement très élaboré et contenir en même temps les moyens de les mettre à exécution.

---

2. Morange 1994 ; Thuillier 1975 : 13-36.

3. Schrödinger 1986.

Un autre physicien, Max Delbrück, et le bactériologiste Salvador Luria vont fonder et animer le « groupe du phage ». Il s'agit d'un groupe ouvert qui rassemble l'ensemble des chercheurs qui, entre 1940 et 1960, utilisèrent les phages (ou bactériophages, des virus infectant les bactéries) comme système modèle pour aborder l'étude du fonctionnement des êtres vivants. Peu à peu se développe une génétique bactérienne qui, à travers ses succès, va institutionnaliser la biologie moléculaire à partir des années 1960. La période 1950 à 1970 est particulièrement féconde avec l'élucidation de la nature chimique des gènes, de la structure et de la réplication de l'ADN, la découverte de l'ARN messager (ARNm), le décryptage du code génétique, la découverte de la régulation de l'expression des gènes...

Cette période est marquée par l'idée que le développement d'un organisme trouve son explication à travers une vision déterministe qui repose sur le programme génétique<sup>4</sup>. Le déterminisme génétique est alors interprété comme un programme organisant le vivant à partir de signaux activant la différenciation cellulaire. Cependant, le statut de cause unique de l'information génétique avec son modèle instructif et son caractère omniscient sera régulièrement remis en cause par la suite<sup>5</sup>. On peut rappeler que la vision du tout génétique n'est pas sans conséquences. Ainsi, le concept du prion (PROteinaceous INFectious particles), protéine infectieuse responsable de la maladie de la vache folle et de la maladie de Creutzfeldt-Jacob, a mis beaucoup de temps à s'imposer durant les années 1980 car il était difficilement admissible par la communauté scientifique qu'un agent pathogène puisse être dépourvu d'acide nucléique<sup>6</sup>.

### **L'expansion de la biologie moléculaire**

À partir des années 1970, la biologie moléculaire qui était jusque là une science descriptive, analytique va devenir peu à peu une science pratique. Elle se réinvente à travers le génie génétique, la génomique puis la biologie synthétique, trois domaines qui relèvent du champ des biotechnologies

---

4. Mayr 1961 : 1501-1506.

5. Atlan 1999 ; Fox Keller 1999 ; Kuppiec & Sonigo 2000 ; Maurel & Miquel 2001.

6. Berche 2013 : 53-62.

modernes. Les découvertes réalisées en biologie moléculaire permettent d'élaborer les premiers outils pour manipuler le vivant. Ainsi, le génie génétique émerge au début des années 1970<sup>7</sup>. Il désigne l'ensemble des techniques permettant de modifier le matériel génétique d'un organisme par la manipulation de l'ADN. Ces recherches débouchent sur la mise sur le marché du premier médicament biologique (biomédicament) en 1982. Il s'agit de l'insuline recombinante humaine produite par des bactéries modifiées génétiquement.

Deux techniques vont révolutionner la biologie moléculaire et la faire entrer dans l'ère de la génomique<sup>8</sup>. Cette discipline étudie le fonctionnement d'un organisme non plus à l'échelle d'un gène mais à l'échelle de son génome, c'est-à-dire de l'ensemble du matériel génétique de l'organisme. La première technique, mise au point en 1975, est le séquençage qui permet de lire l'ADN en déterminant l'ordre de l'enchaînement des nucléotides qui le composent<sup>9</sup>. La seconde, apparue en 1986, est la réaction de polymérisation en chaîne (PCR pour Polymerase Chain Reaction). Il s'agit d'une méthode d'amplification des acides nucléiques qui permet d'obtenir un grand nombre de copies identiques d'un fragment d'ADN à partir d'une très faible quantité d'ADN de départ<sup>10</sup>. Ainsi naît en 1988 le projet génome humain (HGP pour Human Genome Project) qui a pour objectif le séquençage du génome humain et le développement de la bioinformatique pour traiter et analyser les larges données issues de la génomique. Une ébauche du génome humain est publiée en 2001 puis la séquence complète, soit trois milliards de nucléotides (l'équivalent d'environ 2000 livres de 500 pages), est publiée en 2004<sup>11</sup>.

Le séquençage massif des génomes d'organismes issus de tous les règnes vivants génère de nouvelles connaissances sur la fonction d'un très grand nombre de gènes. Ces connaissances alliées au développement de puissants outils de bio-ingénierie comme l'édition des génomes (modification localisée des séquences d'ADN) via la méthode CRISPR-Cas9 aboutit à l'émergence de la biologie synthétique<sup>12</sup>. Cette discipline, qui s'inscrit dans la continuité

---

7. Hall 1990.

8. Weissenbach 2016 : 231-239.

9. Sanger & Coulson 1975 : 441-448.

10. Mullis 1990 : 56-65.

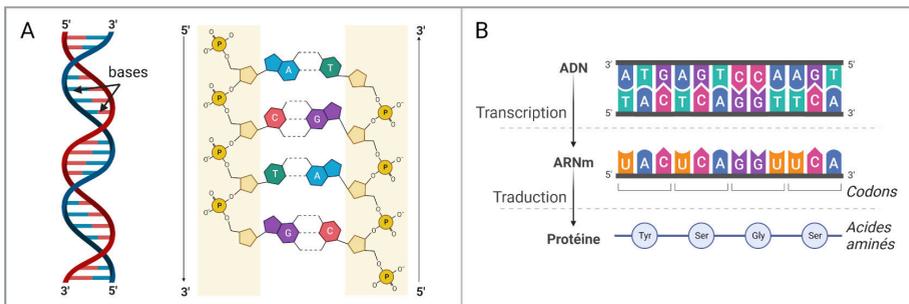
11. International Human Genome Sequencing Consortium 2004 : 931-945.

12. Campos 2013 : 16-25.

du génie génétique, intègre des concepts issus des sciences de l'ingénieur (standardisation, simulation) et vise à rationaliser la bio-ingénierie. Son but est de concevoir et de construire des organismes vivants et des systèmes biologiques artificiels qui accompliront de nouvelles fonctions biologiques répondant à diverses applications dans les domaines de la santé, de l'énergie, des matériaux et de l'environnement. Une avancée majeure de la biologie synthétique est la création de la première bactérie artificielle en 2010<sup>13</sup>.

## Le dogme central

Une fibre chromosomique d'une cellule contient de l'ADN associé à des protéines. Certains fragments de cet ADN comportent les gènes. L'ADN est constitué de briques élémentaires, les nucléotides reliés entre eux par un pont phosphodiester. Chaque nucléotide est composé d'un pentose (le désoxyribose), d'un groupement phosphate et d'une base azotée. Il existe quatre bases dans l'ADN : l'adénine (A), la thymine (T), la cytosine (C) et la guanine (G). La molécule d'ADN est formée de deux brins antiparallèles enroulés l'un autour de l'autre pour former une double hélice (Fig. 1A). Les



**Fig. 1 : Structure de l'ADN et dogme central.** A, à gauche, schéma de la double hélice. Les deux brins antiparallèles se lisent selon l'orientation 5' → 3'; à droite, détail de la structure montrant l'appariement des bases A-T et C-G via les liaisons hydrogènes (en pointillés); B, transcription de l'ADN et traduction de l'ARNm. L'ADN est transcrit en ARNm qui est traduit en protéine. La correspondance entre codons et acides aminés est spécifiée par le code génétique. Dans cet exemple, les codons, présents sur l'ARNm, UAC, UCA et GGU codent respectivement pour la tyrosine (Tyr), la sérine (Ser) et la glycine (Gly).

13. Gibson *et al.* 2010 : 52-56.

deux brins sont reliés entre eux via l'appariement des bases par des liaisons hydrogènes : A est apparié à T et C à G. La séquence d'une molécule d'ADN peut être tout simplement symbolisée par l'enchaînement des bases. Du fait de la complémentarité de ces bases, si l'on connaît la séquence de l'un des deux brins, on peut en déduire la séquence de l'autre<sup>14</sup>.

La structure de l'ADN a été élucidée en 1953<sup>15</sup>. Cette découverte eut un impact énorme sur la communauté scientifique car elle a convaincu l'ensemble des biologistes que l'ADN était le constituant des gènes et par conséquent le support de l'hérédité.

Les gènes sont des fragments d'ADN plus ou moins longs localisés sur le chromosome. Leur structure est assez complexe. Toutefois, ils renferment tous des séquences régulatrices (les promoteurs) qui jouent un rôle dans leur niveau d'expression. L'expression d'un gène regroupe un ensemble de processus biochimiques permettant à l'information stockée dans le gène (l'ordre de l'enchaînement des nucléotides) d'être lue pour diriger la synthèse d'ARNm et de protéines nécessaires au fonctionnement cellulaire. Ainsi, lorsqu'un gène s'exprime, il est tout d'abord converti en ARNm lors du mécanisme de la transcription. L'ARNm, tout comme l'ADN, utilise un alphabet à quatre lettres qui comprend les bases adénine (A), uracile (U), cytosine (C) et guanine (G). Dans un second temps, une partie de l'information portée par cet ARNm est lue par le ribosome. Celui-ci interprète les triplets de nucléotides (appelés codons) de l'ARNm qu'il traduit en acides aminés qui sont assemblés entre eux pour former la protéine (Fig. 1B). Ce mécanisme de synthèse protéique, appelé traduction, permet le passage d'un alphabet de quatre lettres (les nucléotides qui composent l'ADN et l'ARNm) à un alphabet de vingt lettres (les acides aminés qui composent les protéines). La table de correspondance entre les codons et les acides aminés constitue le code génétique. Ainsi, la colinéarité gène-protéine permet, à partir d'une séquence codante d'un gène (lorsque celle-ci est connue), de prédire la

---

14. La longueur d'une molécule d'ADN se compte en paires de bases.

Ainsi, la molécule d'ADN double brin :

- 5'-ATGCTA-3' a une longueur de 6 paires de bases (pb).

- 3'-TACGAT-5'

Tandis que la longueur d'un seul brin :

- 5'-ATGCTA-3' est de 6 bases (ou de 6 nucléotides).

15. Watson & Crick 1953 : 737-738.

séquence en acides aminés de la protéine. Les recherches menées sur l'expression des gènes et de leur régulation ainsi que le craquage du code génétique au cours des années 1960 consolident le principe clé de la biologie moléculaire, « le dogme central » énoncé par F. Crick en 1958<sup>16</sup> :

Cela stipule qu'une fois que l'information est passée dans les protéines, elle ne peut plus en ressortir. Plus précisément, le transfert d'information d'un acide nucléique à un acide nucléique, ou d'un acide nucléique à une protéine, est possible, mais le transfert d'une protéine à une protéine, ou d'une protéine à un acide nucléique, est impossible.

### **Le style de l'ADN**

L'ADN est donc la molécule sémantique dans laquelle l'ordre de l'enchaînement des nucléotides (ou des bases car ce sont elles qui constituent l'alphabet à quatre lettres) le long de la molécule définit l'information génétique portée par l'ADN. Appréhender cette information revient par conséquent à rechercher un ordre dans la cacophonie apparente de l'enchaînement des bases des séquences génomiques. La tâche peut s'avérer parfois ardue dans la mesure où les séquences codantes des gènes ne représentent souvent qu'une infime proportion de la taille totale des génomes (environ 1 % dans le cas du génome humain).

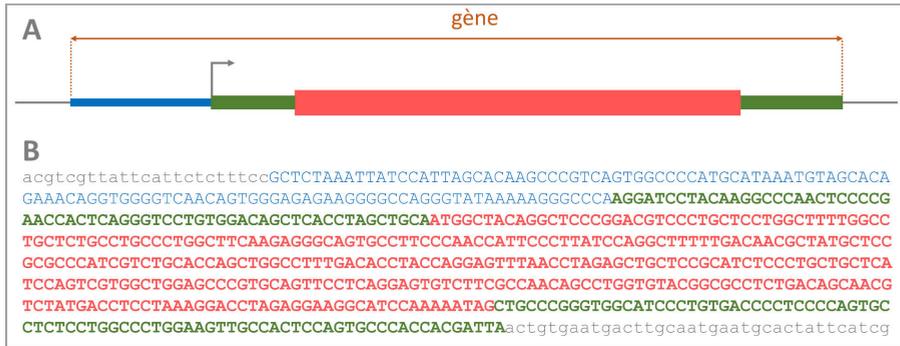
On peut considérer qu'il existe deux niveaux de lecture de l'ADN. Le premier est le séquençage qui va fournir des séquences brutes composées d'un enchaînement des quatre bases. Le second niveau est l'annotation des séquences qui vise à identifier les zones fonctionnelles du génome comme les gènes et les séquences régulatrices (annotation structurale, fig. 2) et à déterminer la fonction des gènes (annotation fonctionnelle). Le développement de logiciels informatiques dédiés à l'annotation et à l'analyse des génomes relève de la bioinformatique<sup>17</sup>.

L'étude de l'organisation des gènes sur les chromosomes nous révèle que l'ordre dans le génome reflète en partie l'organisation cellulaire. A titre d'exemple, on peut évoquer le fait que des gènes impliqués dans les mêmes fonctions physiologiques sont parfois organisés en grappes (gène voisins)

---

16. Crick 1958 : 138-163.

17. Gauthier *et al.* 2019 : 1981-1996.



**Fig. 2 : Annotation structure d'un gène.** A, schéma de la structure d'un gène ; B, séquence nucléotidique (seule les bases de l'un des deux brins de l'ADN sont indiquées) correspondant au schéma. La séquence se lit de gauche à droite et de haut en bas. Promoteur en bleu ; séquence codante en rouge ; régions non traduites en vert. L'unité de transcription correspond aux lettres en gras et son origine est représentée par une flèche sur le schéma.

comme les opérons bactériens et les clusters de gènes dans les génomes des cellules eucaryotes<sup>18</sup>. De manière similaire, chez les bactéries, les gènes possédant un biais de codon similaire sont proches les uns des autres. Par ailleurs, le choix des codons est étroitement corrélé à l'abondance, dans la cellule, des différents ARNt apparentés (ARN de transfert impliqués dans la traduction) et exerce par conséquent un contrôle sur le niveau d'expression des gènes<sup>19</sup>.

Au-delà de l'organisation des gènes sur les chromosomes, existe-t-il des niveaux d'ordre supérieur? La représentation du jeu du chaos (CGR : Chaos Game Representation) permet d'appréhender cette question. La méthode, inspirée des mathématiques du chaos, permet de représenter des fréquences d'oligonucléotides sous forme d'images fractales qui affichent divers motifs locaux et globaux<sup>20</sup>. Dans cette représentation, chacune des quatre lettres de l'ADN est le sommet d'un carré. À partir d'une séquence d'ADN comme celle de la fig. 3A, en commençant au milieu du carré, la

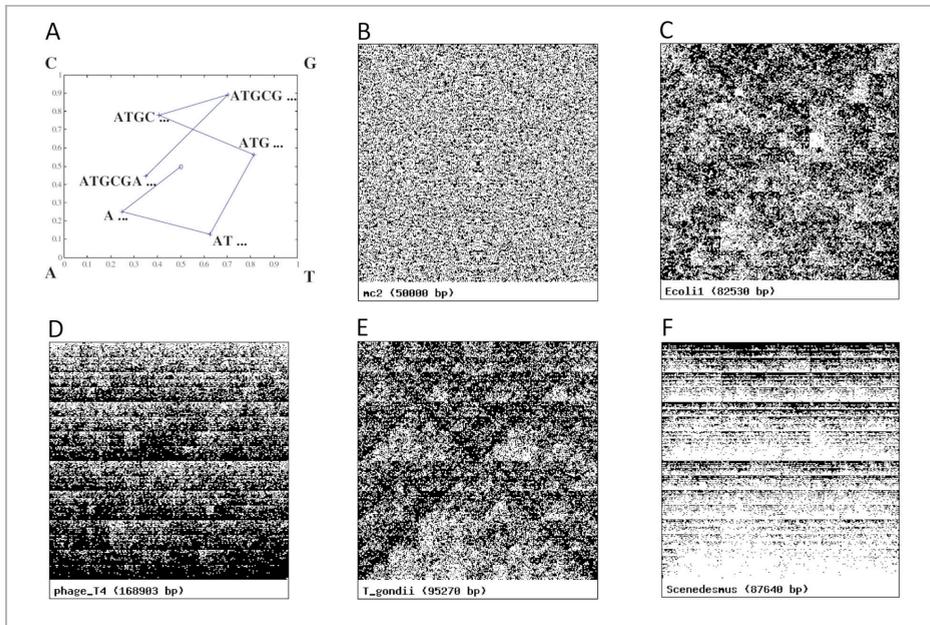
18. Osbourn & Field 2009 : 3755-3775.

19. Bailly-Bechet *et al.* 2006 : e37.

20. Jeffrey 1990 : 2163-2170.

règle à suivre est la suivante : chaque lettre de la séquence d'ADN est représentée par un point qui est le milieu du segment délimité par le point précédent et le sommet correspondant à la lettre en question.

En appliquant cette méthode à des séquences de plusieurs milliers de bases, voire à des chromosomes entiers, on peut observer que des sous-séquences d'un génome présentent les principales caractéristiques du génome entier. Nous obtenons ainsi des signatures génomiques des espèces (Fig. 3C-F). La majorité des segments du génome au sein d'une espèce possèdent des signatures comparables. En conséquence, chaque espèce peut se voir attribuer un style d'ADN qui peut être dérivé de la plupart de ses fragments d'ADN disponibles<sup>21</sup>. Par ailleurs, la distance entre les images peut



**Fig. 3 : Signatures génomiques.** Principe de la CGR pour une séquence d'ADN débutant par ATGCGA (A). Images obtenues pour une séquence aléatoire (B) et pour les génomes de la bactérie *Escherichia coli* (C), du phage T4 (D), du parasite *Toxoplasma gondii* (E) et de l'algue verte *Scenedesmus* (F). Les images B à F ont été générées avec le programme Computation of Chaos Game Representation (CGR) images, disponible en ligne à l'adresse : <http://gscompare.ehu.es/tools/CGR-FCGR/index.php?CGR>

21. Fertil *et al.* 2005 : W512-5.

être considérée comme une mesure de proximité phylogénétique<sup>22</sup>. Ainsi, la CGR nous révèle en partie la complexité intrinsèque des génomes mais sans nous en dire plus sur les lois qui régissent cette complexité.

## Le Livre de la Vie

L'ADN, dès la découverte de la structure en double hélice a acquis un statut tout à fait singulier par rapport aux autres molécules du vivant. On a prêté rapidement des pouvoirs extravagants aux gènes<sup>23</sup> et les chercheurs eux-mêmes, à travers leurs déclarations, ont participé à la sacralisation de l'ADN<sup>24</sup> comme Francis Crick (codécouvreur de la structure de l'ADN) proclamant en 1953, « nous avons découvert le secret de la vie »<sup>25</sup> ou Walter Gilbert (codécouvreur du séquençage de l'ADN) déclarant en 1986 que « la séquence totale de l'homme est le Graal de la génétique humaine »<sup>26</sup>. Lorsqu'une première ébauche du génome humain a été obtenue en 2000, Francis Collins (Directeur du National Human Genome Research Institute) a déclaré « Aujourd'hui, nous célébrons la révélation de la première ébauche du Livre de la Vie. [...] notre propre manuel d'instructions, auparavant connu uniquement de Dieu »<sup>27</sup>.

Utiliser la métaphore du Livre de la Vie n'est pas anecdotique dans la mesure où dans les écritures saintes, et notamment dans *l'Apocalypse de Jean*, Le Livre de Vie est cité à plusieurs reprises. Il s'agit d'un registre céleste dans lequel sont consignés tous nos noms et si nous avons une vie vertueuse sur terre, notre nom ne sera pas effacé du Livre et nous entrerons au Paradis<sup>28</sup>. Ainsi, sommes-nous prédestinés dès la naissance à avoir droit à la vie

---

22. Deschavanne *et al.* 1999 : 1391-1399.

23. Nelkin & Lindee 1998.

24. Une illustration de la sacralisation de l'ADN est la proposition de Kary Mullis (inventeur de la PCR et prix Nobel 1993) de commercialiser, en guise de bijoux, des « amulettes génétiques » contenant des fragments d'ADN amplifiés par PCR provenant de stars du rock.

25. Watson 1984.

26. Roberts 2001 : 1195-1200.

27. White House Event, june 26, 2000 : <https://www.genome.gov/10001356/june-2000-white-house-event>

28. « Il n'entrera que ceux qui sont écrits dans le Livre de Vie de l'agneau » (*Apocalypse* 21:27).

éternelle. Nommer le génome humain le Livre de la Vie cela signifiait que nous sommes avant tout le produit de nos gènes. Dans cet essentialisme génétique, l'ADN incarne dès lors la molécule sémantique qui doit nous dévoiler notre passé et prédire notre futur (études phylogénétiques, gène de prédisposition aux maladies, tests génétiques...). Prédestination génétique et prédestination divine briguent le contrôle de nos vies.

Le Livre de la Vie acquiert ainsi le statut de livre quasi sacré au côté de livres saints des religions monothéistes comme la Thora, la Bible ou le Coran. Ces religions ont leur tradition exotérique qui s'adresse indifféremment à tous les membres d'une communauté religieuse mais également une tradition ésotérique réservée aux initiés. Le Livre de la Vie contient un message chiffré qu'il faut décrypter dans le désordre apparent de l'enchaînement des bases tout comme les livres saints renferment un message secret, ésotérique. C'est le cas par exemple de la Kabbale qui se veut une exégèse ésotérique de la Thora. Déchiffrer le message de la Thora repose sur des méthodes fondées sur la numérologie. Dans l'alphabet hébreu, une valeur numérique est attribuée à chaque lettre. Il existe différents systèmes herméneutiques pour déchiffrer la Thora. L'un d'entre eux, dénommé Gematria, repose sur l'addition des valeurs numériques des lettres et des phrases afin de les interpréter. On établit une identité entre deux mots ou deux versets dont le nombre est identique afin de susciter la révélation des concordances et d'ouvrir de nouveaux horizons dans la compréhension du texte. Ainsi, les kabbalistes développent des algorithmes interrogeant la Thora tout comme les bioinformaticiens le font pour questionner le Livre de la Vie.

### **Métaphores borgésiennes en biologie**

Il est donc désormais question de textes, de lecture et d'interprétation des textes. L'écrivain argentin Jorge Luis Borges (1899-1986) est sans conteste l'un des grands auteurs et lecteurs du XX<sup>e</sup> siècle. Conteur, poète, essayiste, pour qui « la métaphysique n'est qu'une branche de la littérature fantastique »<sup>29</sup>, Borges a pu construire, à travers ses lectures, une œuvre qui révèle une sorte d'universelle analogie. Roger Caillois, qui a largement contribué à

---

29. Borges 2010 : 452-467.

faire connaître l'œuvre de Borges en Europe, expose les thèmes centraux de l'écrivain<sup>30</sup> :

Le principal de ces thèmes clefs me paraît être celui du temps circulaire. Il en entraîne deux autres qui en représentent comme les projections dans les domaines de l'espace et la causalité : celui du labyrinthe et celui de la création récurrente. Ceux-ci, à leur tour, inspirent les symétries et les jeux de miroirs, les systèmes de correspondances et d'équivalences, les compensations et les équilibres secrets qui constituent à la fois la substance et la structure des récits et des poèmes de l'écrivain.

Depuis plusieurs décennies, les références métaphoriques à l'œuvre de Borges ont investi le champ de la biologie. Un éditorial de la revue *Nature*<sup>31</sup> présente un article dans lequel les chercheurs ont réussi à déchiffrer l'origine des motifs de couleur dans le pelage d'une souris rayée africaine en identifiant les gènes impliqués dans la pigmentation. L'éditorial se réfère à la nouvelle *l'Écriture du dieu*<sup>32</sup> qui raconte l'histoire du dernier grand prêtre maya Tzinacán, prisonnier des conquistadores. Il est enfermé dans une prison et dans la cellule voisine se trouve un jaguar. Tzinacán est convaincu que les taches du jaguar ne sont pas des taches aléatoires, mais contiennent un message de son Dieu qui, s'il le déchiffrait, lui offrirait la clé de son évasion.

Au sein de la cellule, une protéine, codée par un gène, peut avoir diverses fonctions. En bioinformatique, cela entraîne des difficultés pour étiqueter les gènes dans des classes fonctionnelles à partir de données d'expression et peut conduire à des biais dans l'apprentissage automatique supervisé<sup>33</sup>. Cet effet, affectant les performances d'apprentissage de l'algorithme, a été nommé « l'effet Borges » en référence à la nouvelle *La langue analytique de John Wilkins*. À travers ce texte, Borges montre qu'il n'y a pas de classification idéale car on y trouvera forcément des ambiguïtés, des redondances et des lacunes : « Il n'existe pas de classification de l'univers qui ne soit arbitraire et conjecturale »<sup>34</sup>.

Dans la nouvelle *Le livre de sable*, Borges a conçu un livre infini dans lequel de nouvelles pages apparaissent constamment à mesure que le lecteur

---

30. Caillois 1981 : 179-192.

31. Deciphering the genes that give mammals their stripes and patterns. 2016 : 5-6.

32. Borges 2010 : 631-635.

33. Mateos *et al.* 2002 : 1703-1713.

34. Borges 2010 : 747-751.

tente d'ouvrir le livre entre la couverture et la première page : «Il me dit que son livre s'appelait le livre de sable, parce que ni ce livre ni le sable n'ont de commencement ni de fin»<sup>35</sup>. En évoquant le livre de sable, Mendes Soares et Valcárcel observent que les pages du livre des génomes semblent également se multiplier au fur et à mesure que nous découvrons de nouveaux gènes et la complexité de leur expression<sup>36</sup>.

La nouvelle qui a grandement inspiré les biologistes est sans conteste *La bibliothèque de Babel*<sup>37</sup>. Divers articles s'y réfèrent, par exemple pour évoquer la diversité des bactériophages<sup>38</sup> ou des enzymes à cytochrome P450<sup>39</sup>. Des outils bioinformatiques s'en réclament explicitement. Ainsi, la suite de logiciels BABELOMICS<sup>40</sup> permet d'intégrer et d'analyser les larges données génomiques et transcriptomiques tandis que le programme BORGES<sup>41</sup> est utilisé dans la détermination de nouvelles structures protéiques.

### De Babel à Mendel

Borges publie en 1941 *La bibliothèque de Babel* dont la première phrase est : «L'univers (que d'autres appellent la Bibliothèque)...». L'auteur décrit une bibliothèque de taille gigantesque formée de salles hexagonales identiques reliées entre elles par des ouvertures. Chaque mur de chaque hexagone porte cinq étagères qui comprend chacune trente-deux livres de quatre cent dix pages. Chaque page contient quarante lignes et chaque ligne environ quatre-vingts caractères. Le nombre total de symboles orthographiques est vingt-cinq (22 lettres de l'alphabet, le point, la virgule et l'espace). Cette bibliothèque contient toutes les combinaisons possibles des vingt-cinq symboles. Si bien que la très grande majorité des livres est de nature chaotique<sup>42</sup> :

Pour une ligne raisonnable, pour un renseignement exact, il y a des lieues et des lieues de cacophonies insensées, de galimatias et d'incohérences [...]

---

35. Borges 2010 : 550-554.

36. Mendes Soares & Valcárcel 2006 : 923-931.

37. Borges 2010 : 491-498.

38. Clark 2012 : 333-335.

39. Ostermeier 2007 : 237-238.

40. Al-Shahrour *et al.* 2005 : W460-4; Alonso *et al.* 2015 : W117-21.

41. Sammito *et al.* 2013 : 1099-1101.

42. Borges 2010 : 491-498.

Il n'y a pas, dans la vaste Bibliothèque, deux livres identiques. De ces prémisses incontrovertibles il déduit que la Bibliothèque est totale, et que ses étagères consignent toutes les combinaisons possibles des vingt et quelques symboles orthographiques, c'est-à-dire tout ce qu'il est possible d'exprimer, dans toutes les langues. Tout : l'histoire minutieuse de l'avenir, les autobiographies des archanges, le catalogue fidèle de la Bibliothèque, des milliers et des milliers de catalogues mensongers, la démonstration de la fausseté de ces catalogues, la démonstration de la fausseté du catalogue véritable, l'évangile gnostique de Basilide, le commentaire de cet évangile, le commentaire du commentaire de cet évangile, le fait véridique de ta mort, la traduction de chaque livre en toutes les langues, les interpolations de chaque livre dans tous les livres.

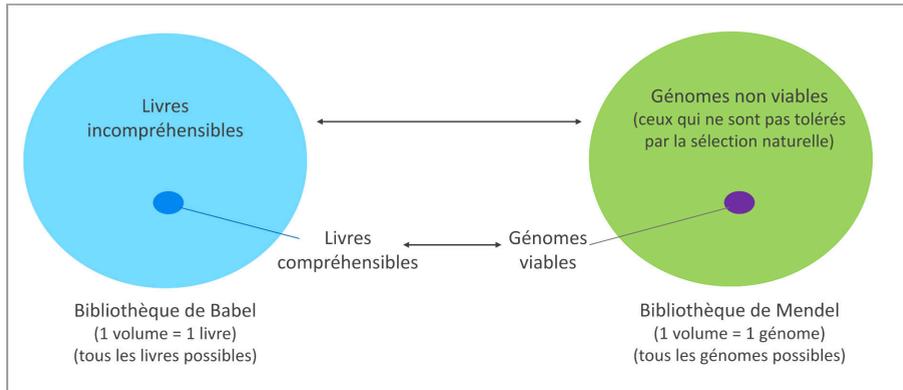
Les bibliothécaires qui parcourent cette bibliothèque essaient vainement de trouver du sens dans des livres cohérents parmi la vaste majorité des livres incohérents. La bibliothèque de Babel est une métaphore de l'univers. Borges suggère que cet univers est conçu pour être interprété. En cela, il rejoint le questionnement des scientifiques : l'univers est-il intelligible ? Quelles sont les lois qui régissent la matière ? Quelles sont celles qui régissent le vivant ? Mais Borges puise sa matière à penser dans la littérature et c'est pourquoi il compare l'univers à une vaste bibliothèque. La bibliothèque de Babel contient  $1,956.10^{1834097}$  livres, soit bien plus que le nombre d'atomes (estimé à  $10^{82}$ ) dans l'univers observable. Cette bibliothèque a été partiellement simulée sur internet<sup>43</sup> par Jonathan Basile qui a conçu un algorithme qui génère des pages aléatoires et les archive.

En 1995, le philosophe Daniell Dennett, dans un ouvrage défendant les thèses de Darwin sur l'évolution<sup>44</sup>, fait une analogie entre ce qu'il nomme la bibliothèque de Mendel et la bibliothèque de Babel. La bibliothèque de Mendel contient tous les génomes possibles, passés, présents et futurs avec toutes les combinaisons possibles des quatre lettres A, T, C et G. Elle comporte donc les génomes viables et non viables. Les génomes viables sont ceux qui sont présents dans les organismes vivants. Les génomes non viables sont ceux qui ne sont pas tolérés par la sélection naturelle. Ils n'existent pas

---

43. Le site internet (<https://libraryofbabel.info/>) utilise un alphabet différent de celui de l'histoire de Borges. Dans la nouvelle, l'alphabet est composé de 22 lettres, de l'espace, du point et de la virgule. Le site internet utilise le même schéma de ponctuation (espace, point, virgule) mais inclut les 26 lettres de l'alphabet anglais.

44. Dennett 1995.



**Fig. 4 : Analogies entre la bibliothèque de Babel et la bibliothèque de Mendel.**

dans la nature. Ainsi, parmi la multitude des génomes possibles, Dennett défend l'idée que la sélection naturelle est un processus aléatoire mais cependant assez puissant pour expliquer l'évolution et l'apparition des génomes viables tout comme la bibliothèque de Babel contient des livres compréhensibles parmi tous les livres possibles (Fig. 4)<sup>45</sup>. Dans les deux cas, même si la proportion de génomes viables et de livres compréhensibles est très vaste, elle reste dérisoire par rapport aux génomes non viables ou aux livres incompréhensibles.

### **Synthia, premier organisme artificiel**

Les analogies dont il est question ci-dessus vont trouver un nouvel écho à l'aune de l'émergence de la biologie synthétique. Bien que le modèle instructif soit aujourd'hui controversé, il reste très prégnant dans le domaine de la biologie synthétique notamment du fait que cette vision du vivant a prouvé son efficacité dans les travaux portant sur la création de bactéries

<sup>45</sup>. Borges ne précise pas dans sa nouvelle quelles sont les 22 lettres de l'alphabet de la bibliothèque de Babel. Si parmi ces 22 lettres se trouvent les lettres a, c, g et t, alors la bibliothèque de Mendel est un sous-ensemble de la bibliothèque de Babel.

artificielles. Ainsi, en 2016, des chercheurs impliqués dans ces travaux déclarent<sup>46</sup> :

Le génome d'une cellule peut être vu comme son système d'exploitation. Il contient le code qui spécifie toutes les fonctions génétiques de la cellule, qui à leur tour déterminent la chimie cellulaire, la structure, la réplication et les autres de ses caractéristiques.

La vision réductionniste de la biologie synthétique où les cellules sont perçues comme des parties qui peuvent être assemblées pour fabriquer un organisme vivant a mené à la question suivante : combien de pièces faut-il pour construire une cellule ? Puis la question s'est peu à peu déplacée pour devenir : peut-on réduire la complexité des génomes naturels en réorganisant et enlevant par approximations les gènes non essentiels et en gardant uniquement les gènes essentiels<sup>47</sup> ? Autrement dit, il s'agit de reformuler la question « Qu'est-ce que la vie ? » en termes génomiques : « Qu'est-ce qu'un ensemble minimal de gènes essentiels ? ». C'est ainsi que les recherches se sont portées sur la création d'un organisme artificiel avec un génome minimal. Outre la justification fondamentale portant sur la compréhension du vivant, la justification majeure est aujourd'hui celle des applications biotechnologiques. La « bactérie minimale » devra présenter un métabolisme et des régulations faciles à appréhender et deviendra une plateforme idéale pour en faire une usine cellulaire en vue de répondre aux problèmes sociétaux liés à la production d'énergie, à la recherche de nouvelles molécules thérapeutiques, à la gestion de l'environnement...

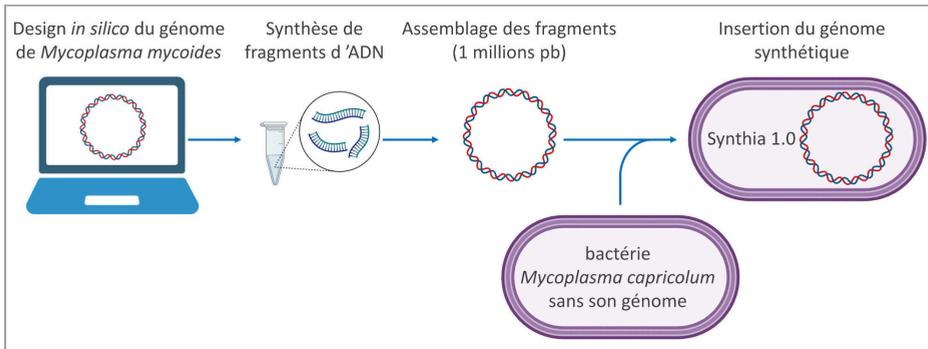
La création de la première bactérie synthétique dénommée Synthia 1.0 a été réalisée en 2010 par le groupe de Craig Venter<sup>48</sup>. L'expérience a consisté, dans un premier temps, à designer *in silico* le chromosome de la bactérie *Mycoplasma mycoides* puis à synthétiser chimiquement des milliers de petits fragments d'une longueur de quelques centaines de paires de bases (pb) qui ont été assemblés séquentiellement pour constituer un chromosome complet de près d'un million de pb comportant 945 gènes (Fig. 5). Ce chromosome a ensuite été inséré dans la bactérie *Mycoplasma capricolum* à laquelle les chercheurs avaient enlevé son chromosome naturel. La nouvelle bactérie

---

46. Hutchison III *et al.* 2016 : aad6253.

47. Un gène essentiel est un gène indispensable à la survie de l'organisme et son inactivation conduit à la léthalité.

48. Gibson *et al.* 2010 : 52-56.



**Fig. 5 : Schéma de construction de la bactérie Synthia 1.0.**

synthétique Synthia 1.0 était viable et présentait un phénotype similaire (même morphologie et même profil protéomique) à *M. mycoides*.

Par ailleurs, lors de la construction de Synthia 1.0, afin de prouver qu'il s'agissait bien d'un chromosome artificiel et non d'une supercherie, les chercheurs ont encodé en langage ADN dans le chromosome quatre filigranes sans préciser la manière dont ils ont été codés. Les messages secrets ont été décryptés par la suite<sup>49</sup>. Ils comportent, entre autres, les noms des co-auteurs de l'article scientifique ainsi qu'une citation du romancier James Joyce : « To live, to err, to fall, to triumph, to recreate life out of life », et une autre du physicien Richard Feynman : « What I cannot build, I cannot understand ».

En 2016, l'équipe de Craig Venter a construit une nouvelle bactérie synthétique sur le même principe que celui décrit pour l'obtention de Synthia 1.0 mais cette fois avec un génome réduit de moitié et ne comportant que 473 gènes<sup>50</sup>. Cette nouvelle version avec son génome minimal a été baptisée Synthia 3.0. Lors de la réduction du génome, les chercheurs se sont rendu compte que la distinction entre gènes essentiels et non-essentiels était trop simpliste. Ils ont ainsi défini une nouvelle catégorie de gènes, à savoir celle des gènes quasi-essentiels qui sont nécessaires pour assurer une croissance robuste de l'organisme minimal.

49. <http://www.righto.com/2010/06/using-arc-to-decode-venters-secret-dna.html>

50. Hutchison III *et al.* 2016 : aad6253.

## L'ère de l'écriture

Avec Synthia 1.0, la biologie est concrètement entrée dans l'ère de l'écriture de l'ADN. Pour un coût d'environ 40 millions \$<sup>51</sup>, la construction de cette bactérie artificielle a considérablement marqué la communauté scientifique dans la mesure où il s'agissait d'une prouesse technologique considérable : convertir une séquence d'ADN numérisée d'une taille d'un million de pb, stockée dans un fichier informatique, en une entité vivante capable de croître et de s'auto-répliquer.

Outre le fait de fabriquer des chromosomes artificiels, l'ADN synthétique fournit également un support pour stocker des données non biologiques. Cette idée a pris de l'ampleur au début des années 2010 pour stocker des données numériques dans l'ADN.

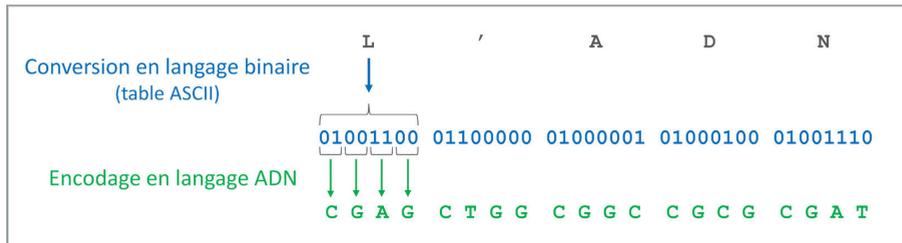
La quantité mondiale de données numériques est estimée à 175 zettaoctets ( $175 \cdot 10^{21}$  octets) d'ici 2025, soit un milliard de disques durs d'un téraoctet. Le stockage de données sur disques durs ou bandes magnétiques nécessite de plus en plus de place et requiert des ressources toujours plus importantes en infrastructures et en énergie. Par ailleurs, le stockage classique est peu durable : les disques durs sont généralement actualisés tous les trois à cinq ans et les bandes magnétiques tous les sept à dix ans. Une partie des données numériques (de 60 % à 80 %) est dite 'froide'. Il s'agit de données que l'on consulte rarement, voire jamais mais elles sont tout de même conservées dans les datacenters.

Il est donc urgent de trouver des solutions alternatives, pérennes et moins énergivores pour stocker les données. L'une des solutions pourrait être l'ADN. C'est une molécule que l'on sait synthétiser et lire par séquençage après l'avoir amplifié par PCR. Elle possède une densité élevée (c'est-à-dire une forte capacité de stockage de l'information pour un volume donné) et présente une longévité élevée qui pourrait nous servir à stocker nos données froides. L'ADN peut être encapsulé pour une conservation à long terme et déposé dans une bibliothèque où sont stockés les pools d'ADN.

Le codage repose sur le principe décrit dans la fig. 6 avec le mot *L'ADN*. Les chiffres, les lettres et la ponctuation d'un texte peuvent être

---

51. Sleator 2010 : 229-230.

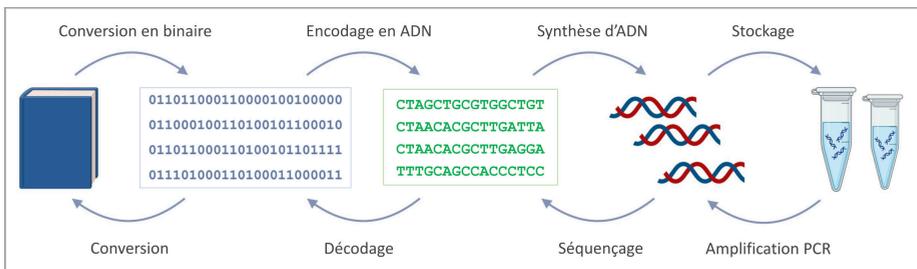


**Fig. 6 : Principe de l'encodage dans l'ADN.**

convertis en langage binaire selon une table comme la table ASCII. Dans celle-ci, chaque signe orthographique est encodé en 8 bits. Ainsi, après avoir converti un texte en séquence binaire, cette dernière est convertie en séquence ADN selon la règle suivante par exemple : 11 : A ; 10 : T ; 01 : C ; 00 : G. Les méthodes de codage ont évolué et sont aujourd'hui plus complexes afin d'augmenter la densité et de réduire le nombre d'erreurs lors de l'encodage puis du décodage.

Une preuve de concept du stockage de larges données dans l'ADN a été apportée en 2012 par le groupe de George Church<sup>52</sup>. Les chercheurs ont montré qu'il est possible d'encoder un livre entier en ADN puis de le relire après séquençage avec un faible pourcentage d'erreurs (Fig. 7).

La méthode a consisté à convertir un ouvrage comportant 53426 mots en binaire selon le code ASCII. La séquence binaire est ensuite traduite en



**Fig. 7 : Principe du stockage de données dans l'ADN et de leur décodage.**

52. Church *et al.* 2012 : 1628.

langage ADN. Puis 54 898 séquences de 159 nucléotides de long ont été synthétisées chimiquement<sup>53</sup>. Chacune de ces séquences ADN correspond à un fragment du texte de l'ouvrage et contient également une adresse afin de pouvoir reconstituer l'ensemble du texte dans le bon ordre lors du décodage. Ces séquences peuvent être stockées pendant un certain temps. Après conservation, si on souhaite les lire, elles sont amplifiées par PCR puis séquencées. Le processus inverse à celui de l'encodage est alors appliqué : les séquences ADN sont décodées pour fournir les séquences binaires qui sont alors converties en texte selon la table ASCII.

Les chercheurs ont calculé la densité en bits/mm<sup>3</sup>. Comparé à d'autres technologies, leur méthode atteint une densité de stockage un million de fois supérieure à celle d'un disque dur ou d'une mémoire flash. Des avancées sont régulièrement réalisées pour augmenter les capacités de stockage dans l'ADN tout en diminuant le nombre d'erreurs. Ainsi, la stratégie de stockage baptisée *DNA Fountain* permet d'atteindre une densité de 215 pétaoctets (215. 10<sup>15</sup> octets) par gramme d'ADN soit l'équivalent de 215 000 disques durs de 1 téraoctets<sup>54</sup>.

En 2017, le groupe de Church réussit à encoder de l'information dans un organisme vivant. En utilisant la technique CRISPR-Cas9, une image animée montrant un cavalier au galop est encodée au sein du génome de la bactérie *Escherichia coli*<sup>55</sup>. Les chercheurs ont ainsi créé un système bactérien capable de capturer, stocker et propager l'information au cours du temps. S'il peut être envisagé dans le futur de stocker de larges données dans des systèmes vivants<sup>56</sup> comme les bactéries (qui seraient stockées dans des banques), ces travaux trouvent également un autre intérêt. Ainsi, la technique

---

53. Afin de minimiser les erreurs, la taille maximale d'un brin d'ADN synthétisé chimiquement est d'environ 200 à 300 nucléotides.

54. Erlich & Zielinski 2017 : 950-954.

55. Shipman *et al.* 2017 : 345-349.

56. Le film *Fahrenheit 451* (2018) est une adaptation cinématographique du roman de l'écrivain de science-fiction Ray Bradbury paru en 1953. Dans un futur dystopique, une dictature entend effacer l'histoire véridique et imposer sa propre version de l'histoire. Les livres sont interdits et les pompiers ont pour mission de les brûler. Il existe toutefois des résistants qui ont la capacité de retenir par cœur le contenu d'un livre à partir du moment où ils l'ont lu une fois. Ils sont les garants des contenus des œuvres. Le film reprend cette idée développée dans le roman mais en y ajoutant une dimension nouvelle à travers le projet Omnis. Il s'agit d'un projet secret des rebelles qui vise à encoder le contenu des livres dans l'ADN de diverses espèces animales, dont les oiseaux qui pourront se reproduire et donc propager ce patrimoine culturel.

mise au point peut être utilisée pour concevoir des bactéries dotées d'une sorte de mémoire cellulaire. Dans ce cas, le système fonctionne de manière similaire à un magnétophone cellulaire : les bactéries écrivent des informations dans leur génome lorsqu'elles sont confrontées à certains événements au cours de leur existence et par la suite il est possible de récupérer ces informations en séquençant leur ADN. Entre autres applications, il peut être envisagé de transformer les cellules en capteurs environnementaux, leur permettant de rendre compte de leur exposition à des produits chimiques particuliers par exemple.

Il reste cependant un obstacle majeur à surmonter : le coût de la synthèse de l'ADN. Le coût du séquençage a considérablement diminué au cours des vingt dernières années. Rappelons que le séquençage du génome humain (3 milliards de pb) publié en 2000 a coûté environ 100 millions \$ alors qu'aujourd'hui ce même séquençage coûte environ 600 \$ (soit  $2.10^{-7}$  \$/base). Le coût de synthèse de l'ADN avec la méthode *DNA Fountain* est de 3500 \$ le mégaoctet<sup>57</sup> ce qui correspond à environ  $9.10^{-4}$  \$/base soit 4500 fois plus que le coût du séquençage. Par conséquent, de nombreuses recherches visent à diminuer drastiquement le coût de synthèse de l'ADN pour l'amener à un coût semblable à celui du séquençage. C'est l'un des objectifs du programme MolecularXiv lancé en 2022, piloté par le CNRS et doté d'un budget de 20 millions € sur 7 ans<sup>58</sup>.

## La construction de la bibliothèque de Mendel

Depuis les travaux pionniers de Church en 2012 sur le stockage d'un ouvrage dans l'ADN (soit 0,66 mégaoctets de données), d'autres travaux ont démontré la faisabilité de stocker dans l'ADN des données plus complexes de plusieurs mégaoctets (images, vidéos, audio, bases de données...). En 2016, Technicolor et l'équipe de Church ont numérisé le film français de 1902 *Un voyage dans la Lune*, réalisé par Georges Méliès, puis l'ont codé en molécules d'ADN synthétique<sup>59</sup>. En 2017, des chercheurs de Microsoft et de

---

57. Erlich & Zielinski 2017 : 950-954.

58. <https://www.ins2i.cnrs.fr/fr/pepr-molecularxiv>

59. <https://www.technicolor.com/news/technicolor-goes-harvard>

l'Université de Washington ont stocké dans l'ADN un volume de 200 méga-octets de données. Il s'agissait des versions numériques d'œuvres d'art, de la déclaration universelle des droits de l'homme dans plus de 100 langues, de 100 livres du Projet Gutenberg et d'une base de données de semences<sup>60</sup>. Pour fêter le vingtième anniversaire de l'album Mezzanine du groupe musical Massive Attack, l'école polytechnique fédérale de Zurich a annoncé en 2018 avoir stocké l'album dans l'ADN<sup>61</sup> pour un coût d'environ 500 000 \$. Le fichier musical a tout d'abord été compressé à 15 méga-octets puis encodé en ADN. Il a fallu synthétiser 920 000 séquences d'ADN. Afin de les conserver, celles-ci ont été encapsulées dans 5 000 billes de verre de taille nanométrique.

En 2020, sous l'impulsion des sociétés Microsoft, Illumina, Twist Bioscience et Western Digital, est créée la DNA Data Storage Alliance<sup>62</sup>. Il s'agit d'une alliance d'organisations industrielles et universitaires dont la mission est de promouvoir un écosystème interopérable basé sur l'ADN comme support de stockage de données. À mesure que les méthodes de stockage de l'ADN commercialement viables seront mieux comprises, l'Alliance envisagera de recommander la création de spécifications et de normes (codage, fiabilité, conservation, systèmes de fichiers...).

Si l'on se replace dans le contexte de la bibliothèque de Mendel, les génomes viables sont ceux qui sont présents dans les organismes vivants sur la planète. Nous découvrons les séquences de ces génomes au fur et à mesure que nous séquençons leur ADN. À ce jour, plusieurs centaines de milliers de génomes sont accessibles dans les bases de données<sup>63</sup>. Quant aux séquences d'ADN synthétiques provenant du stockage de nos données numériques dans l'ADN, elles correspondent à des séquences qui existent dans la bibliothèque de Mendel mais qui font partie des génomes non viables. Par ailleurs, on peut noter une certaine autoréférence pour les données numériques contenant les données d'annotations des génomes viables, puisque dans ce cas, de l'ADN est utilisé pour encoder de l'information portant sur l'ADN.

---

60. <https://blogs.microsoft.com/ai/synthetic-dna-storage-milestone/>

61. <https://ethz.ch/en/news-and-events/eth-news/news/2018/04/entire-music-album-to-be-stored-on-DNA.html>

62. <https://dnastoragealliance.org/>

63. Le site NCBI (National Center for Biotechnology Information) répertorie dans sa base de données 30 530 génomes d'eucaryotes, 567 228 génomes de procaryotes et 66 429 génomes de virus. <https://www.ncbi.nlm.nih.gov/genome/browse#!/overview/>

Ces constatations nous amènent à la conclusion que nous avons déjà commencé à construire la bibliothèque de Mendel.

### **Rationaliser Babel**

À l'aune des développements ci-dessus, nous pouvons relire la bibliothèque de Babel et tenter de conjurer le désordre. Dans la bibliothèque déambulent des bibliothécaires hantés par quelque croyance. Certains sont à la recherche « des livres d'apologie et de prophétie qui justifiaient à jamais les actes de chaque homme »<sup>64</sup>, d'autres sont en quête du livre ultime sensé révéler la clef de la bibliothèque, d'autres encore estiment qu'il faut avant tout éliminer les œuvres inutiles. Habités par leurs convictions, peut-être oublient-ils de regarder avec acuité les autres livres. Nous pouvons faire un parallèle avec la manière dont les génomes étaient perçus il y a quelques décennies. Les biologistes se focalisaient principalement sur les gènes (qui représentent souvent une infime partie du génome notamment chez les eucaryotes) et considéraient le reste du génome comme inutile, le qualifiant d'ADN poubelle. Certaines régions de cet ADN poubelle contiennent de longues séquences répétées qui apparaissaient comme dénuées de signification fonctionnelle. Cette vision a été profondément révisée par la suite lorsqu'on s'est rendu compte que parmi cet ADN poubelle de nombreuses régions jouent un rôle fonctionnel.

De la même façon, si la plupart des livres de la bibliothèque paraissent incohérents, c'est peut-être que les bibliothécaires de Babel ne savent pas les lire. Peut-être contiennent-ils un niveau d'ordre supérieur (Fig. 8A) à l'image de la CGR qui dévoile une complexité des génomes dont les lois intrinsèques nous échappent encore. L'utilisation de la CGR pour déceler un certain ordre dans les livres incohérents pourrait être envisageable dans la mesure où cette méthode alliée à l'intelligence artificielle a été testée pour distinguer divers écrivains sur la base de leur style d'écriture<sup>65</sup>. Pour rationaliser Babel, peut-être faut-il déplacer le problème, le reformuler<sup>66</sup> :

---

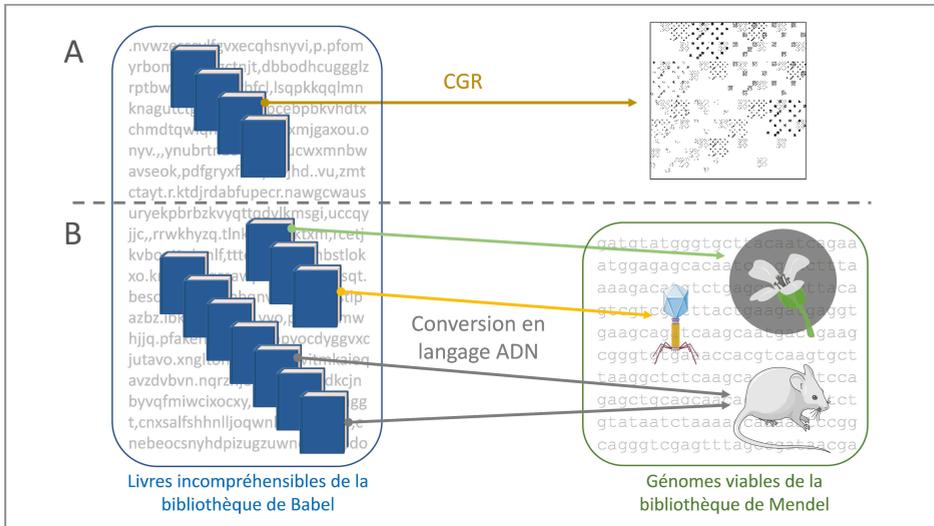
64. Borges 2010 : 491-498.

65. Stoean & Lichtblau 2020 : 1933.

66. Borges 2010 : 491-498.

En effet, la Bibliothèque comporte toutes les structures verbales, toutes les variations que permettent les vingt-cinq symboles orthographiques, mais point un seul non-sens absolu [...]. Ces propositions, incohérentes à première vue, sont indubitablement susceptibles d'une justification cryptographique.

Aussi, parmi la multitude de livres incompréhensibles, certains, ne renfermant que les quatre signes orthographiques a, t, c et g correspondent lettre à lettre à des génomes viables ou des fragments de génome viables de la bibliothèque de Mendel<sup>67</sup>. C'est également le cas pour certains livres incompréhensibles, pouvant renfermer divers signes orthographiques parmi les 25 possibles, qui, une fois numérisés puis convertis en langage ADN



**Fig. 8 : Lire Babel autrement.** A, la méthode CGR, appliquée aux livres incompréhensibles, révèle que certains d'entre eux possèdent un ordre intrinsèque; B, certains livres incompréhensibles, convertis en langage ADN, génèrent des séquences ADN en tout point identiques à celles des génomes viables de la bibliothèque de Mendel.

67. Un livre de la bibliothèque de Babel peut ne correspondre qu'à un fragment de génome de la bibliothèque de Mendel dans la mesure où un livre comporte au maximum 1 312 000 caractères (80 caractères × 40 lignes × 410 pages) et que les génomes contiennent plusieurs millions, voire plusieurs milliards de pb.

trouveront leur correspondance parmi les génomes viables de la bibliothèque de Mendel (Fig. 8B). Ainsi, Babel nous enseigne que tout langage est plausible même sans quelconque intention de signification.

### **Le temps infini de la lecture**

S'il ne fait aucun doute que nous sommes entrés dans l'ère de l'écriture, nous restons toujours fortement ancrés dans l'ère de la lecture. Pour autant, épuiserons-nous un jour le paradigme de la lecture de l'ADN ? Y aura-t-il une fin à l'interprétation c'est-à-dire à l'annotation des génomes ? Cela revient à se poser la question de la finitude de la lecture pour une œuvre, qu'il s'agisse de la bibliothèque de Babel, du Livre de la Vie ou des génomes viables de la bibliothèque de Mendel.

En fait, comme nous le rappelle Gérard Genette<sup>68</sup> :

Le temps des œuvres n'est pas le temps fini de l'écriture mais le temps infini de la lecture [...]. La littérature selon Borges n'est pas un sens tout fait, une révélation que nous avons à subir : c'est une réserve de formes qui attendent leur sens, c'est l'imminence d'une révélation qui ne se produit pas, et que chacun doit produire pour lui-même.

À l'instar du pelage du jaguar où les taches constituent une réserve de formes qui attendent leur sens pour Tzinacán, parmi les formes possibles d'un texte originel, certaines s'incarneront au cours du temps, à chaque fois que l'œuvre sera lue par un lecteur. La bibliothèque de Babel n'a pas été rédigée à l'intention des biologistes mais plusieurs dizaines d'années après sa rédaction Daniel Dennett puis d'autres s'en sont emparés pour y voir des analogies avec le vivant.

La lecture incluant l'annotation des génomes viables de la bibliothèque de Mendel est similaire à la lecture des œuvres littéraires. Si l'on considère le génome humain par exemple, certains résultats issus des recherches scientifiques réalisées depuis 2003 conduisent à le réannoter régulièrement. Ainsi, le nombre de gènes annotés a plus que doublé en vingt ans (de 23 000 en

---

68. Genette 1981 : 364-373.

2003 à environ 62000 en 2023) notamment du fait de la découverte d'un grand nombre de pseudogènes et de gènes transcrits en ARN longs non-codant<sup>69</sup>.

Il semblerait donc que ces observations nous incitent à croire qu'il n'y aurait pas de fin au processus d'annotation des génomes. Mais en fin de compte, nous n'en savons rien. Aussi, laisserons-nous le dernier mot à Borges. L'extrait ci-dessous, tiré d'une *Défense de la Cabale*<sup>70</sup>, se réfère à la Thora mais il n'est pas exclu d'y voir le Livre de la Vie ou encore l'ensemble des génomes viables de la bibliothèque de Mendel :

La conception seule de ce document est un prodige supérieur à tous ceux que ses pages enregistrent. Un Livre impénétrable à la contingence, un mécanisme aux desseins infinis, aux variations infaillibles, aux lumières superposées, dont les révélations vous guettent, comment ne pas l'interroger jusqu'à l'absurde, jusqu'à la prolixité numérique, comme le fit la Cabale ?

## Sitographie

Deciphering the genes that give mammals their stripes and patterns. Nature. 2016, 539 : 5-6. <https://www.nature.com/articles/539005b>

DNA Data Storage Alliance. <https://dnastoragealliance.org/>

Entire music album to be stored on DNA. <https://ethz.ch/en/news-and-events/eth-news/news/2018/04/entire-music-album-to-be-stored-on-DNA.html>

Microsoft and University of Washington researchers set record for DNA storage. <https://blogs.microsoft.com/ai/synthetic-dna-storage-milestone/>

Projet MoleculArXiv. <https://www.ins2i.cnrs.fr/fr/pepr-molecularxiv>

Site internet de la simulation de la bibliothèque de Babel. <https://libraryofbabel.info/>

Technicolor explores role of synthetic DNA for archiving entertainment experiences for thousands of generations. <https://www.technicolor.com/news/technicolor-goes-harvard>

Using Arc to decode Venter's secret DNA watermark. <http://www.righto.com/2010/06/using-arc-to-decode-venters-secret-dna.html>

White House Event, june 26, 2000. <https://www.genome.gov/10001356/june-2000-white-house-event>

---

69. Guigo 2023 : 100375.

70. Borges 2010 : 216-219.

## Bibliographie

- ALONSO R., Salavert F., Garcia-Garcia F., Carbonell-Caballero J., Bleda M., Garcia-Alonso L., Sanchis-Juan A., Perez-Gil D., Marin-Garcia P., Sanchez R., Cubuk C., Hidalgo MR., Amadoz A., Hernansaiz-Ballesteros RD., Aleman A., Tarraga J., Montaner D., Medina I., Dopazo J., Babelomics 5.0: functional interpretation for new generations of genomic data, *Nucleic Acids Res.* 2015, 43 (Web Server issue) : W117-21.
- AL-SHAHROUR F., Minguez P., Vaquerizas JM., Conde L., Dopazo J., BABELOMICS : a suite of web tools for functional annotation and analysis of groups of genes in high-throughput experiments, *Nucleic Acids Res.* 2005, 33 (Web Server issue) : W460-4.
- ATLAN H., *La fin du tout génétique ? Vers de nouveaux paradigmes en biologie.* Ed. Quae, 1999.
- BAILLY-BECHET M., Danchin A., Iqbal M., Marsili M., Vergassola M., Codon usage domains over bacterial chromosomes, *PLoS Comput Biol.* 2006, 2 : e37.
- BERCHE P., Histoire des prions, *Feuillets de biologie*, 2013, 315 : 53-62.
- BORGES J-L., Une défense de la cabale. *Œuvres complètes I*, Paris, éd. Gallimard, Bibl. Pléiade, 2010, p. 216-219.
- BORGES J-L., Tlön, Uqbar, Orbis Tertius. *ID.*, p. 452-467.
- BORGES J-L., La bibliothèque de Babel. *ID.*, p. 491-498.
- BORGES J-L., L'Écriture du dieu. *ID.*, p. 631-635.
- BORGES J-L. La sphère de Pascal. *ID.*, p. 676-679.
- BORGES J-L. La langue analytique de John Wilkins. *ID.*, p. 747-751.
- BORGES J-L. Le livre de sable. *Œuvres complètes II*, Paris, éd. Gallimard, Bibl. Pléiade, 2010, p. 550-554.
- CAILLOIS R., «Les thèmes fondamentaux de J.L. Borges», Ed. de l'Herne, *Cahier de l'Herne*, 1981, p. 179-192.
- CAMPOS L. Un siècle de biologie de synthèse, *Med Sci.* 2013, 29 : 16-25.
- CHURCH GM., Gao Y., Kosuri S., Next-generation digital information storage in DNA, *Science.* 2012, 337 : 1628.
- CLARK JR., Bacteriophages: a biological library of Babel, *Future Virol.* 2012, 7 : 333-335.
- CRICK FH., On protein synthesis, *Symp Soc Exp Biol.* 1958, 12 : 138-163.
- DENNETT DC., *Darwin's dangerous idea: Evolution and the meanings of life.* Ed. Simon & Schuster, 1995.

- DESCHAVANNE PJ., Giron A., Vilain J., Fagot G., Fertil B., Genomic signature: characterization and classification of species assessed by chaos game representation of sequences, *Mol Biol Evol.* 1999, 16 : 1391-1399.
- ERLICH Y, ZIELINSKI D., DNA Fountain enables a robust and efficient storage architecture, *Science.* 2017, 355 : 950-954.
- FERTIL B., Massin M., Lespinats S., Devic C., Dumeé P., Giron A., GENSTYLE : exploration and analysis of DNA sequences with genomic signature, *Nucleic Acids Res.* 2005, 33 (Web Server issue) : W512-5.
- FOX KELLER E., *Le rôle des métaphores dans les progrès de la biologie.* Paris, éd. Les empêcheurs de penser en rond, 1999.
- GAUTHIER J., Vincent AT., Charette SJ., Derome N., A brief history of bioinformatics, *Brief Bioinform.* 2019, 20 : 1981-1996.
- GENETTE G., «La littérature selon Borges», Éd. de l'Herne, *Cahier de l'Herne*, 1981, p. 364-373.
- GIBSON DG., Glass JJ., Lartigue C., Noskov VN., Chuang RY., Algire MA., Benders GA., Montague MG., Ma L., Moodie MM., Merryman C., Vashee S., Krishnakumar R., Assad-Garcia N., Andrews-Pfannkoch C., Denisova EA., Young L., Qi ZQ., Segall-Shapiro TH., Calvey CH., Parmar PP., Hutchison CA III., Smith HO., Venter JC., Creation of a bacterial cell controlled by a chemically synthesized genome, *Science.* 2010, 329 : 52-56.
- GUIGO R., Genome annotation : From human genetics to biodiversity genomics, *Cell Genom.* 2023, 3 : 100375.
- HALL S., *Les débuts du génie génétique.* Paris, éd. Plon, 1990.
- HUTCHISON CA III., Chuang RY, Noskov VN., Assad-Garcia N., Deerinck TJ., Ellisman MH., Gill J, Kannan K., Karas BJ., Ma L., Pelletier JF., Qi ZQ., Richter RA., Strychalski EA., Sun L., Suzuki Y., Tsvetanova B., Wise KS., Smith HO., Glass JJ., Merryman C., Gibson DG., Venter JC., Design and synthesis of a minimal bacterial genome, *Science.* 2016, 351 : aad6253.
- INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM, Finishing the euchromatic sequence of the human genome, *Nature.* 2004, 431 : 931-945.
- JEFFREY HJ., Chaos game representation of gene structure, *Nucleic Acids Res.* 1990, 18 : 2163-2170.
- KUPPIEC JJ., SONIGO P., *Ni dieu ni gène. Pour une autre théorie de l'hérédité.* Paris, éd. Seuil, 2000.
- MATEOS A., Dopazo J., Jansen R., Tu Y., Gerstein M., Stolovitzky G., Systematic learning of gene functional classes from DNA array expression data by using multilayer perceptrons, *Genome Res.* 2002, 12 : 1703-1715.
- MAUREL MC, MIQUEL PA., *Programme génétique : concept biologique ou métaphore ?* Paris, éd. Kimé, 2001.

- MAYR E., Cause and effect in biology, *Science*. 1961, 134 : 1501-1506.
- MENDES SOARES LM., VALCARCEL J., The expanding transcriptome : the genome as the 'Book of Sand', *EMBO J*. 2006, 25 : 923-931.
- MORANGE M., *Histoire de la biologie moléculaire*. Paris, éd. La Découverte, 1994.
- MULLIS KB., The unusual origin of the polymerase chain reaction, *Sci Am*. 1990, 262 : 56-65.
- NELKIN D., LINDEE S., *La mystique de l'ADN*. Paris, éd. Belin, 1998.
- OSBOURN AE., FIELD B., Operons, *Cell Mol Life Sci*. 2009, 66 : 3755-3775.
- OSTERMEIER M., Beyond cataloging the library of Babel, *Chem Biol*. 2007, 14 : 237-238.
- ROBERTS L., Timeline: A History of the Human Genome Project, *Science*. 2001, 291 : 1195-1200.
- SAMMITO M., Millan C., Rodriguez DD., De Ilarduya IM., Meindl K., De Marino I., Petrillo G., Buey RM., De Pereda JM., Zeth K., Sheldrick GM., Uson I., Exploiting tertiary structure through local folds for crystallographic phasing, *Nat Methods*. 2013, 10 : 1099-1101.
- SANGER F., COULSON AR., A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase, *J. Mol. Biol*. 1975, 94 : 441-448.
- SCHRÖDINGER E., *Qu'est-ce que la vie ?* Paris, éd. Christian Bourgois, 1986.
- SHIPMAN SL., Nivala J., Macklis JD., Church GM., CRISPR-Cas encoding of a digital movie into the genomes of a population of living bacteria, *Nature*. 2017, 547 : 345-349.
- SLEATOR RD., The story of *Mycoplasma mycoides* JCVI-syn1.0: the forty million dollar microbe, *Bioeng Bugs*. 2010, 1 : 229-230.
- STOEAN C., LICHTBLAU D., Author identification using chaos game representation and deep learning, *Mathematics*. 2020, 8 : 1933.
- THUILLIER P., « Comment est née la biologie moléculaire », *La recherche en biologie moléculaire*. Paris, éd. du Seuil, 1975, p. 13-36.
- WATSON JD., CRICK FH., Molecular structure of nucleic acid; a structure for deoxy-ribose nucleic acid, *Nature*. 1953, 171 : 737-738.
- WATSON JD., *La double hélice*. Paris, éd. Hachette, 1984.
- WEISSENBACH J., The rise of genomics, *C R Biologies*. 2016, 339 : 231-239.

