

Surprise biologique

K. ARP • R. ABBIT

La séquence des bases du chromosome 3 de dipneuste reproduit en base 4 les chiffres de π

Que les lois de la physique fassent intervenir des constantes mathématiques n'étonne plus. En revanche, la présence de principes arithmétiques ou géométriques au cœur des organismes biologiques apparaît plus surprenante. Nous avons tous été fascinés par l'enroulement des pétales des fleurs de tournesol ou des écailles des pommes de pin, où apparaissent les termes de la fameuse suite de Fibonacci 1, 2, 3, 5, 8, 13, 21, 34, 55, etc. Chaque terme de cette suite est la somme des deux précédents, et le rapport de deux termes consécutifs tend à l'infini vers le nombre d'or, $(1 + \sqrt{5})/2$. Moins connu, le cycle de vie des cigales de la moitié orientale des États-Unis qui émergent tous les 13 ans ou tous les 17 ans (13 et 17 sont des nombres premiers) est tout aussi remarquable : S. J. Gould conjecture que cette périodicité non divisible permettrait à ces espèces d'échapper à leurs prédateurs naturels - d'autres insectes -, qui suivent des cycles de deux, trois, quatre ou cinq ans.

La découverte, en janvier de cette année, d'une coïncidence étonnante dans le génome d'un dipneuste *Protopterus aethiopicus* (un poisson osseux d'Afrique qui résiste aux périodes de sécheresse en respirant de l'air) promet d'être plus difficile encore à élucider. L'équipe du Professeur Rod Herring, du Laboratoire de biologie moléculaire de Hammerfest, en Norvège, vient en effet d'identifier sur le chromosome 3 de cet animal une séquence fortement liée à la constante mathématique π . Lorsqu'on écrit la suite des 20 premiers nucléotides du segment du chromosome 3 repéré à Hammerfest, on obtient les lettres GATCAAGGGCTTTTATATAC où, comme c'est l'usage, A indique l'adénine, C la cytosine, T la thymine, G la guanine. Or, en remplaçant A par 0, C par 1, T par 2, G par 3 (cette substitution est suggérée par des considérations biochimiques), on trouve : 3,0210033312222020201, ce qui est l'écriture en base 4 des 20 premiers chiffres du nombre π . Le calcul de : $3 + 2/4^2 + 1/4^3 + 3/4^6 + 3/4^7 + \dots + 1/4^{19}$, donne $863554413089/274877906944 = 3,141592653588304\dots$, ce qui est exact jusqu'à la douzième décimale. En réalité, ce sont 193 nucléotides consécutifs qui coïncident ainsi avec la plus remarquable des constantes mathématiques.

Vu le nombre de plus en plus grand de séquences génétiques déchiffrées, de telles coïncidences ne sont peut-être pas vraiment inattendues. On est loin de connaître les cinq milliards de nucléotides des chromosomes humains, mais la somme cumulée de toutes les séquences contenues dans les banques de données génétiques du monde double à peu près tous les deux ans et atteint déjà plusieurs milliards de nucléotides. Il n'est pas surprenant que, dans une telle masse de données, des singularités sans signification soient présentes. On peut cependant évaluer que la probabilité de tomber par hasard sur une séquence de 20 nucléotides consécutifs correspondant à n ne devient non négligeable que lorsqu'on dispose d'une longueur de 1 000 milliards de nucléotides. Or on en est loin aujourd'hui, et de toute façon c'est 193 et non pas seulement 20 nucléotides consécutifs qu'on a découverts: le hasard ne peut donc en aucun cas expliquer la coïncidence de Rod Herring.

```
GATCAAGGGC TTTTATATAG CTTAGAATAG CAGACAGACT CTATTATGTA
3021003331 2222020201 1220300203 1030103012 1202202320

AAGCGAACGA GCACATTCAA ATCAGTAATA TATTCTCGGA GACGAAAAAT
0031300130 3101022100 0210320020 2022121330 3013100002

AATGTGGTTT CTAGTGACAG TCTGATATCC ACCATTAATA CGTCTAGTAG
0023233222 1203230103 2123020211 0110220020 1321203203

CAAACAGCGC GTGGTCCCAC TCTGAGGAGC AGTTCAAGAC TGA
1000103131 3233211101 2123033031 0322100301 230
```

La suite des 193 nucléotides de l'ADN de dipneuste trouvée sur le chromosome 3 du dipneuste à Hammerfest. La traduction (en remplaçant A par 0, C par 1, T par 2 et G par 3) donne le début du développement en base 4 du nombre π .

Rien n'aurait été possible sans la présence dans le laboratoire d'Hammerfest d'un étudiant italien, Salma Gairdneri, qui est calculateur prodige, et qui en étudiant le chromosome du dipneuste a soudain reconnu la séquence correspondant à n qu'il connaissait par cœur. Lors du premier séquençage de ce morceau de chromosome, le 43e nucléotide ne correspondait pas à celui attendu (on avait trouvé un G), ce qui a conduit l'équipe norvégienne à faire une seconde fois le séquençage et à découvrir qu'il y avait une erreur: le 43e nucléotide était un A, comme le développement de n le laissait prévoir.

Cette découverte devrait susciter un regain de collaboration entre les mathématiciens purs - qui en général éprouvent peu d'intérêt pour les sciences naturelles - et les biologistes, qui de leur côté trouvent peu d'attraits aux mathématiques formelles. Plutôt que de faire tourner pendant des semaines ou des mois des milliers de machines mises en réseaux pour calculer les décimales de n ou pour factoriser des nombres entiers, ne serait-il pas plus judicieux d'accroître les performances de séquençage des machines utilisées aujourd'hui en biologie moléculaire, et d'aller ensuite explorer les bonnes zones ? Une nouvelle façon de faire des mathématiques est sans doute à inventer, qui n'exclurait pas d'aller à la pêche pour trouver d'autres génomes intéressants. Réciproquement, la présence de telles structures mathématiques dans le génome pourrait éclairer la théorie de l'évolution d'une lumière nouvelle.

L'assimilation de l'évolution par sélection naturelle à un processus de calcul, idée déjà exploitée en informatique dans le domaine des algorithmes génétiques, pourrait expliquer la découverte norvégienne. En attendant, on se demande à quoi peut bien servir pour le dipneuste une représentation dans son génome du nombre favori des mathématiciens. Est-ce comme le propose le professeur Ray, d'Édimbourg, parce que la constante l'aide à fabriquer les cellules parfaitement sphériques de son hypophyse, qui avaient déjà frappé les naturalistes au siècle dernier? Ou est-ce, selon la proposition de L. Shol, de Hambourg, parce que les propriétés électriques du dipneuste nécessitent la maîtrise des lois de l'électrostatique qui, chacun le sait, mentionnent souvent n ? Un groupe de linguistes lacaniens de l'évolution prétend même que, d'après leur étymologie, les piafs, pics, pies (surtout), pieuvres et autres pigeons présentent cette particularité du génome.