Un modèle de réplication d'un ADN

Aimé Lachal

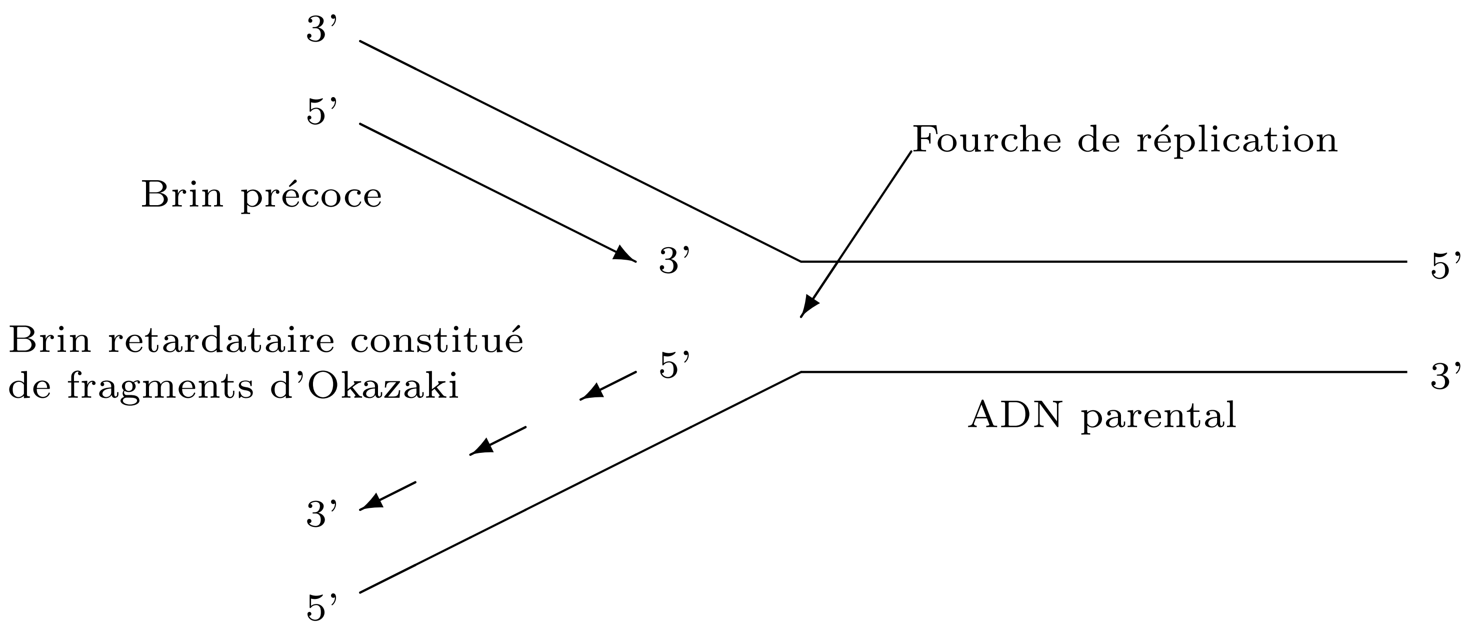
Laboratoire de Mathématiques Appliquées de Lyon
Institut National des Sciences Appliquées de Lyon

1 Composition d'un ADN

- très longue molécule constituée de deux chaînes polynucléotidiques complémentaires ;
- conformation bicaténaire hélicoïdale ;
- bases : adénine (A), cytosine (C), guanine (G), thymine (T) ;
- bases complémentaires : $A \longleftrightarrow T$, $C \longleftrightarrow G$.

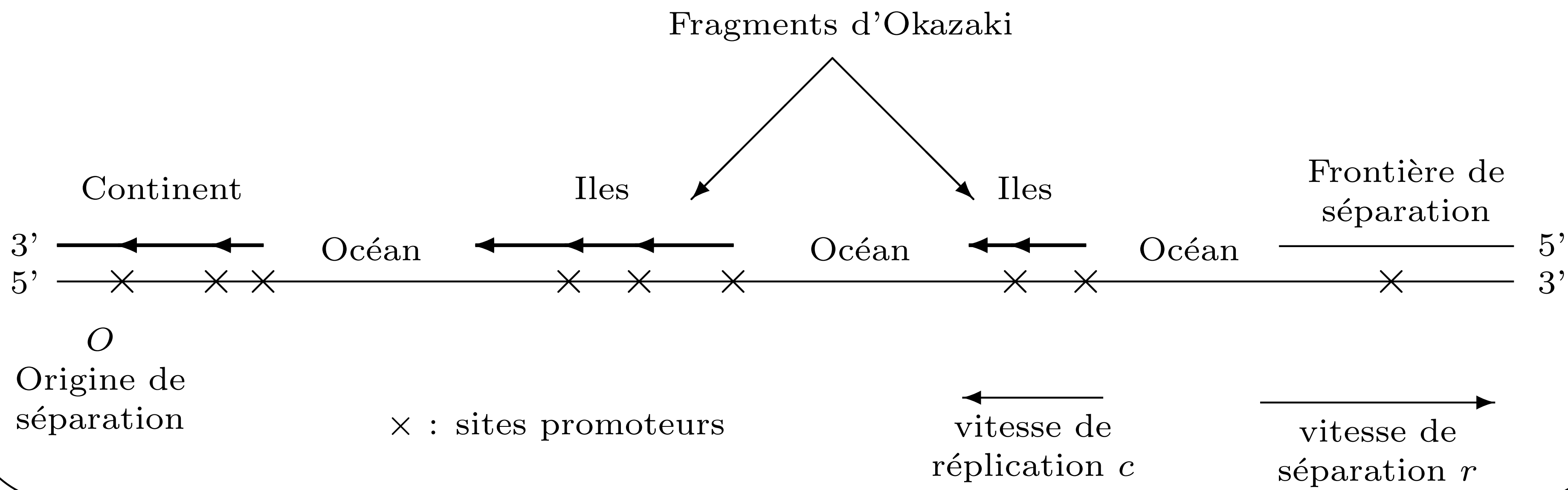
2 Processus de réplication d'un ADN

- *phase de dénaturation :* les deux brins sont séparés après élévation de la température ;
- *phase d'hybridation :* les deux brins sont répliqués à l'aide d'une ADN polymérase ;
  - brin précoce : réplication continue dans le sens de la séparation ;
  - brin retardataire : réplication discontinue dans le sens contraire de la séparation $\longrightarrow$ fragments d'Okazaki (1968).



Le brin retardataire est constitué de deux parties :

- un brin (le *continent*) obtenu par concaténation depuis l'origine de tous les fragments qui sont liés à leurs voisins de gauche ;
- les segments restants qui ne sont pas encore reliés au continent, ce sont les *fragments d'Okazaki* (les *îles*). Les interstices entre ces fragments constituent l'*océan*.



3 Modèle probabiliste de Cowan & Chiu (1994)

Les sites promoteurs sont répartis selon un processus de Poisson spatial d'intensité μ, ou encore un processus de Poisson temporel d'intensité $\lambda = r\mu$.

Références

[1] R. Cowan and S.N. Chiu. Stochastic model of fragment formation when DNA replicates, J. Appl. Prob. 31 (1994), 301–308.

[2] R. Cowan. Stochastic models for DNA replication, in The Handbook of Statistics, Vol. 20, 2001.

[3] D. Piau. Quasi-renewal estimates, J. Appl. Prob. 37 (2000), 269–275.

[4] A. Lachal. Quelques lois de probabilités associées à un modèle de réplication d'un ADN, C.R.A.S. 336 (2003), 175–180.

[5] A. Lachal. Some probability distributions in modelling DNA replication, Ann. Appl. Prob. 13 (2003).

Quelques v.a. intéressantes

- N_t : nombre de fragments d'Okazaki existant à l'instant t ;
- D_t : distance entre la frontière de séparation et l'extrémité droite du continent à l'instant t ;
- P_t : longueur totale des nouveaux morceaux d'ADN à l'instant t (continent inclus) ;
- $\tilde{P}_t = rt - P_t$: longueur totale des interstices entre les fragments d'Okazaki à l'instant t.

Problème : lois des v.a. N_t, D_t *et* $\tilde{P}_t$ *?*

Résultats connus (1994, 2000)

- Cowan & Chiu ont calculé, à l'aide de la théorie du renouvellement :
  - *le nombre moyen asymptotique de fragments d'Okazaki*

$$\lim_{t \to +\infty} \mathbb{E}(N_t) = \lambda\delta + \frac{r}{c}\,;$$

  - *la longueur moyenne asymptotique des interstices*

$$\lim_{t \to +\infty} \mathbb{E}(\tilde{P}_t) = \frac{r(r+c)}{\lambda c}.$$

- Piau a calculé à l'aide d'une équation de quasi-renouvellement :
  - *la distance moyenne asymptotique séparation-continent*

$$\lim_{t \to +\infty} \mathbb{E}(D_t) = r\delta + \frac{r+c}{\lambda} \sum_{n=1}^{\infty} \sigma(n) \Big(\frac{r}{r+c} \Big)^n$$

où $\sigma(n)$ est le nombre de diviseurs de n.

4 Nouveaux résultats (A.L. 2002)

Soit $\varphi(t) = \mathbb{E}(e^{-\alpha N_t})$, $\psi(t) = \mathbb{E}(e^{-\alpha \tilde{P}_t})$ et $\chi(t) = \mathbb{E}(e^{-\alpha D_t})$.

On pose $a = \dfrac{c}{r+c}$, $b = \dfrac{r}{r+c}$ et $Q_n(b) = \displaystyle\prod_{k=1}^{n}(1-b^k)$.

Théorème 1 (Régime transitoire)

$$\varphi(t) \;=\; \begin{cases} e^{-\lambda(1-e^{-\alpha})t} \quad si\ \ 0 \leqslant t \leqslant \delta, \\[2ex] e^{-\lambda\delta(1-e^{-\alpha})} \displaystyle\sum_{n=0}^{\infty} \prod_{k=1}^{n} [1-(1-e^{-\alpha})b^k]\, e^{-\lambda(t-\delta)} \frac{[\lambda(t-\delta)]^n}{n\,!} \ \ si\ \ t \geqslant \delta; \end{cases}$$

$$\chi(t) \;=\; \begin{cases} e^{-r\alpha t} \quad si\ \ 0 \leqslant t \leqslant \delta, \\[2ex] e^{-r\delta\alpha} \displaystyle\sum_{n=0}^{\infty} \Big[\sum_{k=0}^{n} \frac{Q_n(b)}{Q_k(b)} \Big(b - \frac{r\alpha}{\lambda} \Big)^k \Big]\, e^{-\lambda(t-\delta)} \frac{[\lambda(t-\delta)]^n}{n\,!} \ \ si\ \ t \geqslant \delta. \end{cases}$$

Théorème 2 (Régime stationnaire)

$$\begin{aligned}
\lim_{t\to+\infty} \varphi(t) &= e^{-\lambda\delta(1-e^{-\alpha})} \prod_{n=1}^{\infty} [1-(1-e^{-\alpha})b^n], \\
\lim_{t\to+\infty} \psi(t) &= \prod_{n=0}^{\infty} \left[1+\frac{r\alpha}{\lambda}\, b^n\right]^{-1}, \\
\lim_{t\to+\infty} \chi(t) &= e^{-r\delta\alpha} \prod_{n=1}^{\infty} \left[1+\frac{r\alpha}{\lambda}\frac{b^{n-1}}{1-b^n}\right]^{-1}.
\end{aligned}$$

Fonctions de la théorie des partitions : q-séries

Référence : G. Andrews. The theory of partitions, in Encyclopedia of Mathematics and Its Applications, Vol. 2, 1976.

Soit $Q_n(q) = \prod_{k=1}^{n} (1 - q^k)$.

Formules d'Euler :

$$\begin{aligned} \prod_{n=0}^{\infty} (1 - zq^n)^{-1} &= \sum_{n=0}^{\infty} \frac{1}{Q_n(q)}\, z^n, \\ \prod_{n=0}^{\infty} (1 + zq^n) &= \sum_{n=0}^{\infty} \frac{q^{n(n-1)/2}}{Q_n(q)}\, z^n. \end{aligned}$$

Formules de Rothe :

$$\begin{aligned} \prod_{k=0}^{n} (1 - zq^k)^{-1} &= \sum_{k=0}^{\infty} \frac{Q_{n+k}(q)}{Q_n(q) Q_k(q)}\, z^k, \\ \prod_{k=0}^{n-1} (1 + zq^k) &= \sum_{k=0}^{n} \frac{Q_n(q) q^{k(k-1)/2}}{Q_k(q) Q_{n-k}(q)}\, z^k. \end{aligned}$$

Théorème 3 (Décomposition) *Les v.a. N_t, $\tilde{P}_t$ et D_t convergent en loi lorsque $t \to +\infty$ vers les v.a. N_∞, $\tilde{P}_\infty$, D_∞ données par*

$$N_\infty = \sum_{n=0}^{\infty} X_n, \quad \tilde{P}_\infty = \sum_{n=0}^{\infty} Y_n \ \text{ et } \ D_\infty = r\delta + \sum_{n=1}^{\infty} Z_n$$

où

- *X_0 est une v.a. de Poisson $\mathcal{P}(\lambda\delta)$,*
  $(X_n)_{n\geqslant 1}$ est une suite de v.a.i. de Bernoulli $\mathcal{B}(b^n)$ indépendantes de X_0 ;
- *$(Y_n)_{n\geqslant 0}$ est une suite de v.a.i. exponentielles $\mathcal{E}(\lambda/(rb^n))$;*
- *$(Z_n)_{n\geqslant 1}$ est une suite de v.a.i. exponentielles $\mathcal{E}((\lambda b/r) \times (1/b^n - 1))$.*

Corollaire 4 (Moyennes, variances, moments)

- $$\mathbb{E}(N_\infty) = \lambda\delta + \frac{r}{c}, \qquad \operatorname{var}(N_\infty) = \lambda\delta + \frac{r(r+c)}{c(2r+c)};$$

- $$\mathbb{E}(\tilde{P}_\infty) = \frac{r(r+c)}{\lambda c}, \qquad \operatorname{var}(\tilde{P}_\infty) = \frac{r^2(r+c)^2}{\lambda^2 c(2r+c)},$$
  $$\mathbb{E}(\tilde{P}_\infty^n) = \frac{n\,!}{Q_n(b)}\left(\frac{r}{\lambda}\right)^n;$$

- $$\mathbb{E}(D_\infty) = r\delta + \frac{r}{\lambda b}\sum_{n=1}^{\infty}\frac{b^n}{1-b^n} = r\delta + \frac{r}{\lambda b}\sum_{n=1}^{\infty}\sigma_0(n)b^n,$$
  $$\operatorname{var}(D_\infty) = \left(\frac{r}{\lambda b}\right)^2\sum_{n=1}^{\infty}\left(\frac{b^n}{1-b^n}\right)^2 = \left(\frac{r}{\lambda b}\right)^2\sum_{n=1}^{\infty}[\sigma_1(n)-\sigma_0(n)]b^n$$

avec $\sigma_0(n) = \#\{k : k \mid n\}$ *et* $\sigma_1(n) = \sum_{k|n} k$.

5 Indications de preuve

- *équations de renouvellement ;*
- *équations de quasi-renouvellement ;*
- *équations différentielles fonctionnelles.*

Proposition 5

$$\begin{aligned}
\varphi(t) &= \lambda e^{-\alpha}\int_0^t \varphi(t-s)e^{-\lambda s}\,ds + \lambda(1-e^{-\alpha})\int_0^{a(t-\delta)^+}\varphi(t-s)e^{-\lambda s}\,ds + e^{-\lambda t},\\
\psi(t) &= \lambda\int_0^{at}\psi(t-s)e^{-\lambda s}\,ds + \lambda e^{c\alpha t}\int_{at}^t \psi(t-s)e^{-(\lambda+r\alpha/b)s}\,ds + e^{-(\lambda+r\alpha)t},\\
\chi(t) &= \lambda\int_0^{a(t-\delta)^+}\chi(t-s)e^{-\lambda s}\,ds + e^{-(\lambda a+r\alpha)t+\lambda a\delta}.
\end{aligned}$$

Démonstration. — Soit T_1 l'instant auquel apparaît le premier site promoteur.
T_1 suit la loi exponentielle $\mathcal{E}(\lambda)$. En conditionnant par $\{T_1 = s\}$:

- si $t < s, \quad N_t = 0,$
- si $t \geqslant s, \quad N_t \overset{\text{loi}}{=} \begin{cases} N_{t-s} & \text{si} \quad 0 \leqslant s \leqslant a(t-\delta)^+, \\ N_{t-s}+1 & \text{si} \quad s > a(t-\delta)^+. \end{cases}$

Puis

$$\varphi(t) = \int_0^\infty \varphi(t \mid s) \lambda e^{-\lambda s}\, ds$$

avec $\varphi(t \mid s) = \mathbb{E}(e^{-\alpha N_t} | T_1 = s) = \begin{cases} \varphi(t-s) & \text{si} \quad 0 \leqslant s \leqslant a(t-\delta)^+, \\ e^{-\alpha}\varphi(t-s) & \text{si} \quad a(t-\delta)^+ < s \leqslant t, \\ 1 & \text{si} \quad s > t. \end{cases}$

5.1 La v.a. N_t, nombre de fragments d'Okazaki

- Sur $[0, \delta]$, φ vérifie une équation de renouvellement ordinaire ;
- sur $[\delta, +\infty[$, la fonction $\tilde{\varphi}(t) = e^{\lambda t}\varphi(t+\delta)$, $t \geqslant 0$ vérifie une équation différentielle fonctionnelle.

Lemme 6

$$\frac{d\tilde{\varphi}}{dt}(t) - \lambda\tilde{\varphi}(t) + \lambda(1-e^{-\alpha})b\,\tilde{\varphi}(bt) = 0 \quad \textit{avec} \quad \tilde{\varphi}(0) = e^{-\lambda\delta(1-e^{-\alpha})}.$$

Proposition 7 (Solution)

$$\tilde{\varphi}(t) = \sum_{n=0}^{\infty} \varphi_n \frac{(\lambda t)^n}{n\,!} \quad \textit{avec} \quad \varphi_n = e^{-\lambda\delta(1-e^{-\alpha})} \prod_{k=1}^{n} [1-(1-e^{-\alpha})b^k].$$

Corollaire 8

$$\lim_{t\to+\infty} \varphi(t) = \lim_{t\to+\infty} \tilde{\varphi}(t)e^{-\lambda t} = e^{-\lambda\delta(1-e^{-\alpha})} \prod_{k=1}^{\infty} [1-(1-e^{-\alpha})b^k].$$

5.2 La v.a. D_t, distance séparation-continent

- Sur $[0, \delta]$, χ vérifie une équation de renouvellement ordinaire ;
- sur $[\delta, +\infty[$, la fonction $\tilde{\chi}(t) = e^{\lambda t}\chi(t+\delta)$, $t \geqslant 0$ vérifie une équation différentielle fonctionnelle.

Lemme 9

$$\frac{d\tilde{\chi}}{dt}(t) - \lambda\tilde{\chi}(t) + \lambda b\tilde{\chi}(bt) = \lambda\rho e^{(\lambda b - r\alpha)t - r\delta\alpha} \quad \text{avec} \quad \tilde{\chi}(0) = e^{-r\delta\alpha}.$$

Proposition 10 (Solution)

$$\tilde{\chi}(t) = \sum_{n=0}^{\infty} \chi_n \frac{(\lambda t)^n}{n\,!} \quad \text{avec} \quad \chi_n = e^{-r\delta\alpha} \sum_{k=0}^{n} \frac{Q_n(b)}{Q_k(b)} \Big(b - \frac{r\alpha}{\lambda}\Big)^k.$$

Corollaire 11

$$\lim_{t\to+\infty} \chi(t) = \lim_{t\to+\infty} \tilde{\chi}(t)e^{-\lambda t} = e^{-r\delta\alpha} \prod_{n=1}^{\infty} \Big[1 + \frac{r\alpha}{\lambda}\frac{b^{n-1}}{1-b^n}\Big]^{-1}.$$

5.3 La v.a. $\tilde{P}_t$, longueur des interstices

Soit $\tilde{\psi}(t) = \psi(t)e^{\lambda t}$, $t \geqslant 0$.

Lemme 12

$$\begin{cases} \dfrac{d^2\tilde{\psi}}{dt^2}(t) + (r\alpha - \lambda)\,\dfrac{d\tilde{\psi}}{dt}(t) - \lambda r\alpha\,\tilde{\psi}(t) + \lambda b r\alpha\,\tilde{\psi}(bt) = 0, \\ \tilde{\psi}(0) = 1, \quad \dfrac{d\tilde{\psi}}{dt}(0) = \lambda - r\alpha. \end{cases}$$

Proposition 13 (Solution)

$$\tilde{\psi}(t) = \sum_{n=0}^{\infty} \psi_n \,\frac{(\lambda t)^n}{n\,!} \;\text{ avec } \begin{cases} \psi_{n+2} + \left(\dfrac{r\alpha}{\lambda} - 1\right)\psi_{n+1} - \dfrac{r\alpha}{\lambda}\left(1 - b^{n+1}\right)\psi_n = 0, \\ \psi_0 = 1, \quad \psi_1 = \lambda + r\alpha. \end{cases}$$

5.4 La v.a. $\tilde{P}_\infty$

Transformée de Laplace : $L\psi(p) = \displaystyle\int_0^\infty e^{-pt}\psi(t)\,dt.$

On a

$$\lim_{t\to+\infty} \psi(t) = \lim_{p\to 0^+} pL\psi(p).$$

Lemme 14

$$L\psi(p) = \frac{p+\lambda}{p(p+\lambda+r\alpha)} - \frac{\lambda r\alpha}{p(p+\lambda+r\alpha)}\,L\psi\Big(\frac{p+\lambda a}{b}\Big).$$

Corollaire 15

$$\lim_{t\to+\infty} \psi(t) = \frac{\lambda}{\lambda+r\alpha}\,[1 - r\alpha\,L\psi(\lambda a/b)].$$

Proposition 16

$$L\psi(\lambda a/b) = \frac{1}{r\alpha}\Big[1 - \prod_{n=1}^{\infty}\Big(1 + \frac{r\alpha}{\lambda}\,b^n\Big)^{-1}\Big].$$

6 Distance inter-sites promoteurs

Cowan & Chiu introduisent le rapport

$$R = \frac{\text{longueur moyenne des fragments d'Okazaki}}{\text{distance moyenne inter-sites promoteurs}} = \frac{\mathbb{E}(L_\infty)/\mathbb{E}(N_\infty)}{1/\mu}$$

avec

$$L_\infty = D_\infty - \tilde{P}_\infty.$$

Problème : estimation du paramètre μ ?

Proposition 17

$$R = \Big[\lambda\delta - \frac{r}{c} - 1 + \sum_{n=1}^{\infty} \sigma(n)\Big(\frac{r}{r+c}\Big)^{n-1}\Big] / \Big[\lambda\delta + \frac{r}{c}\Big].$$

Expériences

$$\frac{\mathbb{E}(L_\infty)}{\mathbb{E}(N_\infty)} \sim \begin{cases} 150 \text{ bases pour les eucaryotes} \\ 1500 \text{ bases pour les procaryotes} \end{cases}$$

$$\frac{c}{r} \sim \begin{cases} 0,6 \text{ pour les eucaryotes} \\ 6.10^{-3} \text{ pour les procaryotes} \end{cases}$$

Cas $\delta = 0$:

$$R \sim \begin{cases} 1,48 \text{ pour les eucaryotes} \\ 4,77 \text{ pour les procaryotes} \end{cases} \Longrightarrow \frac{1}{\mu} \sim \begin{cases} \frac{150}{1{,}48} = 101 \text{ pour les eucaryotes} \\ \frac{1500}{4{,}77} = 314 \text{ pour les procaryotes} \end{cases}$$

$\Longrightarrow$ *Un site promoteur devrait avoir un motif de 4 bases invariantes.*

Les sites seraient espacés de $4^4 = 256$ bases en moyenne $(256 \in [101; 314])$;
pour un motif de 3 bases, on aurait $4^3 = 64 \notin [101; 314]$;
pour un motif de 5 bases, on aurait $4^5 = 1024 \notin [101; 314]$.