

## 1. Introduction

La **TPMD** a été proposée pour quantifier l'hétérogénéité du contenu en G+C entre les génomes procaryotes (Sueoka, 1962).

$$\begin{cases} GC \xrightarrow{u} AT \\ AT \xrightarrow{v} GC \end{cases} \Rightarrow \mu_D = \frac{v}{u+v}$$

( $\mu_D$ ) représente la pression de mutation directionnelle = le contenu en G+C d'un ensemble neutre de nucléotides, à l'équilibre)

- Choix d'un **standard de neutralité** : le contenu en G+C des 3<sup>es</sup> positions des codons (Sueoka, 1988).
- Régression linéaire du GC12 contre GC3 : la pente ( $\epsilon \in [0, 1]$ ) représente la **neutralité relative** du GC12.

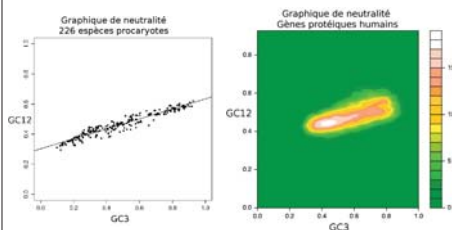


FIG. 1: Graphiques de neutralité pour les génomes bactériens (à gauche) et pour les gènes protéiques humains (à droite). La neutralité relative du GC12 est estimée à 33 % dans le premier cas et à 22 % dans le deuxième cas.

## 2. Le cas particulier de *Leishmania major*

Pour le génome (récemment séquencé) du parasite eucaryote *Leishmania major*, la corrélation entre GC12 et GC3 est significativement **négative**.

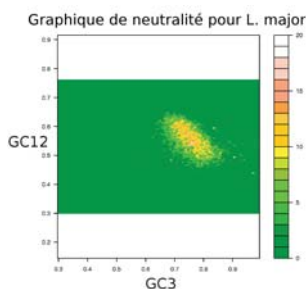


FIG. 2: Graphique de neutralité pour *L. major*. Le coefficient de corrélation de Spearman entre GC12 et GC3 est  $\rho = -0.39$ .

L'interprétation classique en termes de neutralité relative et contraintes sélectives est **impossible**.

## 3. Distribution taxonomique

- La même corrélation négative est observée pour *L. infantum*, mais pas pour *Trypanosoma brucei* et *T. cruzi*.  
⇒ Cette structure génomique est apparue avant la divergence du genre *Leishmania* dans la famille des Trypanosomatidae.
- Seulement deux autres eucaryotes suffisamment documentés (*N. crassa* et *S. japonicum*) présentent une tendance négative, mais moins significative.

## 4. Usage des codons

- La principale source de variabilité pour l'usage des codons synonymes et des acides aminés est liée à la composition en bases.
- Les gènes connus pour être hautement exprimés (tels que les protéines ribosomales, les histones etc.) ont l'usage des codons le plus biaisé.
- Il existe une corrélation négative entre le biais d'usage des codons synonymes et les taux de substitution  $K_a$  et  $K_s$  ( $\rho = -0.51$  et  $\rho = -0.19$ , respectivement).

L'usage des codons semble être influencé aussi bien par le **bias mutationnel** que par la **sélection traductionnelle**.

## 5. Modélisation

La corrélation négative GC12/GC3 n'est pas une conséquence de la variation de l'usage des codons synonymes entre les gènes.

Preuve : construction d'un jeu de données simulé :

- Tous les gènes ont le même usage des codons synonymes, correspondant à la moyenne de tous les CDS de *L. major*.
- La variabilité au niveau de l'usage des acides aminés est inchangée.

Avec ces données, la corrélation GC12/GC3 est  $\rho = -0.70$ .

La corrélation négative est causée par une pression **au niveau des acides aminés**.

- Pour les acides aminés à haut GC12, l'usage des codons synonymes est biaisé vers un GC3 relativement faible, et inversement.

### 5.1 Premier modèle : sélection traductionnelle

**Hypothèses:**

- Les codons préférés sont définis selon le nombre de gènes d'ARNt dans le génome de *L. major*.
- Règle du "Bon choix" (Grosjean & Fiers 1982, Gouy & Gautier 1982) : préférence pour une énergie d'appariement codon/anticodon **moyenne**.
- Probabilités de fixation différentes pour les changements "préféré" ↔ "non-préféré".

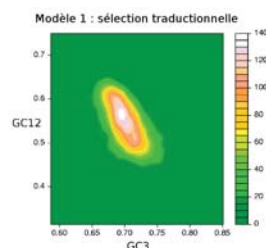


FIG. 3: Premier modèle : prédiction de l'état d'équilibre. Corrélation de Spearman  $\rho = -0.6$

**Results:** A l'état d'équilibre, le modèle prédit une corrélation négative entre GC12 et GC3.

### 5.2. Deuxième modèle : sélection purificatrice

La sélection purificatrice, agissant au niveau des acides aminés, peut influencer l'usage des codons synonymes (Morton, 2001).

**Hypothèses:**

- La sélection purificatrice limite les transversions aux sites deux fois dégénérés.
- Les transitions et les transversions génèrent des contenus en G+C différents :  
 $GC_{eq}(transitions) > GC_{eq}(transversions)$ .

Dans le code génétique standard, le contenu en GC12 des acides aminés codés par 2 codons est inférieur à celui des acides aminés codés par 4 ou 6 codons.

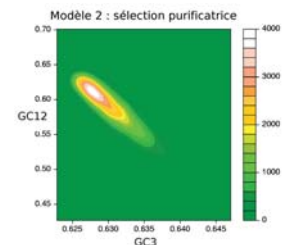


FIG. 4: Deuxième modèle : prédiction de l'état d'équilibre. Corrélation de Spearman  $\rho = -0.85$

Avec ces hypothèses, notre modèle prédit à l'équilibre une forte corrélation négative GC12/GC3.

## 6. Implications pour la TPMD

L'estimation de la neutralité relative et des contraintes sélectives par la régression linéaire GC12/GC3 n'est pas valide dans le cas de *L. major*.

- Choix du standard de neutralité :  
→ Une alternative à l'utilisation du GC3 : le choix des séquences intergéniques comme standard de neutralité, qui supprime l'effet de la sélection traductionnelle.
- Equivalence des transitions et transversions pour le contenu en G+C :  
→ La théorie de la pression de mutation directionnelle devrait prendre en compte les différences potentielles entre les deux types de mutations pour la pression AT/GC.

## 7. Perspectives

La question de l'apparition de la corrélation négative GC12/GC3 au sein du genre *Leishmania* reste ouverte. D'autres explications plausibles :

- Un biais mutationnel dépendant du voisinage, engendrant une sous-représentation des dinucléotides SS.
- Une pression de sélection pour la stabilité des ARNm, qui favoriserait l'homogénéisation du contenu en G+C des séquences transcrites.
- Une variation de la vitesse d'évolution entre les gènes, qui mènerait transitoirement à la formation d'une corrélation négative GC12/GC3.