

Projet C++

Charles, ./Ed

April 4, 2016



1 Les lignées S/L

Dans la plus longue expérience d'évolution expérimentale en cours sur *E. coli*¹, 12 populations, provenant d'une même population ancestrale, évoluent indépendamment dans un milieu minimum en glucose depuis 1988 (environ 60000 générations atteintes en 2014). Ces populations subissent un cycle quotidien de croissance dans le milieu, suivi d'un transfert dans un milieu neuf toutes les 24h (soit environ toutes les 7 générations).

Dans une des 12 populations, un événement de diversification a été observé². Deux lignées sont apparues vers 5000 générations et cohabitent depuis lors (c'est-à-dire depuis environ 55000 générations). Ces deux lignées sont appelées L et S en raison de la taille des colonies qu'elles forment sur boîte de Petri (Large et Small). Après analyse des génomes et des interactions écologiques de ces deux lignées, il apparaît qu'un phénomène de construction de niche par cross-feeding explique leur émergence. La lignée L consomme le glucose disponible dans l'environnement et rejette de l'acétate. La lignée S consomme l'acétate produit par la lignée L et rejette de l'éthanol. La saisonnalité de l'environnement (cycle de 24h), impose une contrainte forte puisque la lignée L consomme tout le glucose disponible et entre en phase stationnaire, alors que la souche S se divise plus facilement en phase stationnaire, lorsque l'acétate s'est accumulé. Il semble donc que la saisonnalité et la quantité initiale de glucose soient deux facteurs essentiels du maintien des lignées S et L.

Le but du projet est de simuler les conditions d'émergence et de maintien des deux lignées S et L, en développant un modèle individu-centré d'évolution artificielle.

2 Le modèle

Le modèle est multi-échelle, et inclut 4 niveaux. Les 3 premiers niveaux représentent le "genotype-to-phenotype mapping". Le 4^{ième} niveau représente le paysage de fitness.

- (1) le génotype,
- (2) le réseau métabolique,
- (3) le phénotype et la fitness,
- (4) l'environnement.

¹<http://myxo.css.msu.edu/>

²Ribeck, N., & Lenski, R. E. (2015). Modeling and quantifying frequency-dependent fitness in microbial populations with cross-feeding interactions. *Evolution*, 69(5), 1313-1320.

2.1 Le génotype

Chaque individu possède un génotype $G \in \{G_A, G_B\}$, définissant la capacité de l'individu à consommer le métabolite A (équivalent du glucose) ou le métabolite B (équivalent de l'acétate). Le génotype peut muter à chaque génération en changeant aléatoirement de type avec une probabilité P_{mut} .

2.2 Le réseau métabolique

Les individus croissent dans le milieu virtuel, en consommant les ressources disponibles et en rejetant les produits du métabolisme au moment de leur mort.

4 voies métaboliques sont disponibles :

- La voie $A \rightarrow A$, qui permet aux individus d'importer A dans leur cytoplasme au taux R_{AA} ,
- La voie $A \rightarrow B$, qui permet de convertir A en B au taux R_{AB} ,
- La voie $B \rightarrow B$, qui permet d'importer B au taux R_{BB} ,
- La voie $B \rightarrow C$, qui permet de convertir B en C au taux R_{BC} (C étant équivalent à l'éthanol).

Les individus de génotype G_A ne peuvent réaliser que les voies $A \rightarrow A$ et $A \rightarrow B$ (équivalent lignée L). Les individus de génotype G_B ne peuvent réaliser que les voies $B \rightarrow B$ et $B \rightarrow C$ (équivalent lignée S)³.

Les individus échangent des métabolites avec leur environnement local (la cellule sur laquelle ils se trouvent). Le réseau métabolique des individus de type G_A est régi par le système :

$$\begin{cases} \frac{dA_{out}}{dt} &= -A_{out} * R_{AA} \\ \frac{dA}{dt} &= A_{out} * R_{AA} - A * R_{AB} \\ \frac{dB}{dt} &= A * R_{AB} \end{cases}$$

Celui des individus de type G_B est régi par le système :

$$\begin{cases} \frac{dB_{out}}{dt} &= -B_{out} * R_{BB} \\ \frac{dB}{dt} &= B_{out} * R_{BB} - B * R_{BC} \\ \frac{dC}{dt} &= B * R_{BC} \end{cases}$$

2.3 Le phénotype et la fitness

Chaque individu possède un phénotype $P = (A, B, C)$, qui n'est autre que le vecteur des concentrations métaboliques internes.

³En réalité, les lignées L et S sont toutes deux capables de consommer du glucose et de l'acétate, mais des mutations dans leur réseau de régulation ont conduit chacune d'entre-elles à privilégier une ressource (glucose OU acétate).

La fitness w des individus dépend de leur activité métabolique :

$$w = \begin{cases} B & \text{if } G = G_A \\ C & \text{if } G = G_B \end{cases}$$

Pour éviter le risque que des individus trop mauvais subsistent dans la population, un seuil minimum sera appliqué à la fitness :

$$w = \begin{cases} w & \text{if } w \geq W_{MIN} \\ 0 & \text{if } w < W_{MIN} \end{cases}$$

2.4 L'environnement

L'environnement est représenté par une grille 2D toroïdale de longueur W et de largeur H , sur laquelle diffuse trois métabolites :

- A, qui représente le glucose,
- B, qui représente l'acétate,
- C, qui représente l'éthanol.

Le processus de diffusion avec un taux D peut-être modélisé simplement grâce à l'algorithme suivant. Pour chaque cellule (x, y) de la grille, le voisinage de Moore est parcouru:

Algorithm 1 Diffuse metabolites at cell (x, y)

```

 $A(x, y)_{t+1} \leftarrow A(x, y)_t$ 
 $B(x, y)_{t+1} \leftarrow B(x, y)_t$ 
 $C(x, y)_{t+1} \leftarrow C(x, y)_t$ 
for  $i \in \{-1, 0, 1\}$  do
  for  $j \in \{-1, 0, 1\}$  do
     $A(x, y)_{t+1} \leftarrow A(x, y)_{t+1} + D * A(x + i, y + j)_t$ 
     $B(x, y)_{t+1} \leftarrow B(x, y)_{t+1} + D * B(x + i, y + j)_t$ 
     $C(x, y)_{t+1} \leftarrow C(x, y)_{t+1} + D * C(x + i, y + j)_t$ 
  end for
end for
 $A(x, y)_{t+1} \leftarrow A(x, y)_{t+1} - 9 * D * A(x, y)_t$ 
 $B(x, y)_{t+1} \leftarrow B(x, y)_{t+1} - 9 * D * B(x, y)_t$ 
 $C(x, y)_{t+1} \leftarrow C(x, y)_{t+1} - 9 * D * C(x, y)_t$ 

```

Cet algorithme est efficace, mais devient instable lorsque $D > 0.1$.

Le modèle ainsi décrit couple donc un "genotype-to-phenotype mapping" évolvable (génotype + réseau métabolique + phénotype), avec un paysage de fitness évolvable (l'environnement), puisque les individus modifient leur propre environnement qui en retour modifie les pressions de sélection.

3 L'algorithme évolutif

A chaque pas de temps de simulation, trois événements sont possibles :

- les individus meurent avec une probabilité P_{death} . Dans ce cas, le contenu de leur cytoplasme est rejeté dans l'environnement (à leur emplacement dans la grille),
- les individus survivants entre en compétition pour se diviser dans les espaces libres (gaps). Pour faire simple, l'individu ayant la meilleure fitness dans le voisinage de Moore du gap pourra se diviser. Les deux individus fils héritent d'un génome muté et de la moitié des concentrations internes du parent. Pour éviter les biais, les gaps devront être parcourus aléatoirement,
- tous les individus qui ne sont pas morts réalisent les fonctions de leur réseau métabolique (constante d'intégration = 1, $dt = 0.1$).

En résumé, à chaque pas de temps de simulation:

- les métabolites externes diffusent dans l'environnement,
- les individus meurent aléatoirement,
- les individus survivants entrent en compétition pour la division dans les gaps,
- tous les individus survivants (comprenant les cellules divisées) adaptent leur réseau métabolique.

Afin de simuler l'expérience d'évolution expérimentale décrite en introduction, les individus doivent être transférés périodiquement dans un nouvel environnement. Pour faire simple, seul l'environnement sera réinitialisé. Ainsi, tous les T pas de temps, l'ensemble des concentrations environnementales sont mises à zéro, et une quantité A_{init} de A est introduite dans chaque cellule de la grille.

4 Implémentation

Implémenter le modèle, en utilisant un schéma orienté objet.

5 Cas étudiés

5.1 Dynamique S/L sans mutation

On considère qu'à l'issue d'une simulation, le système peut prendre 3 états :

- **extinction** : la population est éteinte,
- **cohabitation** : les lignées S et L cohabitent,
- **exclusion** : la lignée S s'est éteinte.

Etablir le diagramme de phase du système dans les dimensions T et A_{init} , avec les paramètres suivants :

- $R_{AA} = R_{BB} = R_{AB} = R_{BC} = 0.1$
- $P_{mut} = 0$
- $W = H = 32$ (grille de 1024 cellules)
- $D = 0.1$
- $P_{death} = 0.02$
- $W_{MIN} = 0.001$
- $T \in [1, 1500]$
- $A_{init} \in [0, 50]$
- Temps de simulation = 10000

Au départ de chaque simulation, la population comprend $\frac{W*H}{2}$ individus de génotype G_A , et $\frac{W*H}{2}$ individus de génotype G_B , répartis aléatoirement.

Les transitions de phase devront être représentées avec une précision suffisante (il faudra donc développer une méthode pour affiner localement l'exploration paramétrique au niveau des transitions de phase). De part la nature stochastique des simulations, il est possible que les transitions de phase soient floues.

5.2 Dynamique S/L avec mutations

Même question que dans le cas 5.1, mais avec $P_{mut} = 0.001$.

6 Bonus: exploration du taux de diffusion

Reprendre le cas 5.1 en ajoutant une dimension : le taux de diffusion D ($D \in [0.0, 0.1]$).