

# Projet

Bruno Toupance

FEB 2018

## Objectif

Le processus de dérive génétique a été formalisé dans le cadre mathématique très strict du modèle de Wright-Fisher. Ce modèle établit les relations mathématiques qui décrivent l'influence de la taille  $N$  sur la vitesse d'évolution, par exemple, de l'hétérozygotie de la population.

Ces relations, établies dans le cas d'une population idéale dont les caractéristiques sont très éloignées de celles des populations réelles, peuvent néanmoins être utilisées dans certains cas en remplaçant simplement l'effectif réel de recensement par un effectif fictif, l'effectif efficace ou "effective population size". L'effectif efficace, noté  $N_e$ , est défini comme la taille  $N$  d'une population de Wright-Fisher dérivant à la même vitesse que la population réelle.

Dans ce projet, on se propose de simuler la dérive génétique dans des populations qui ne respectent pas au moins une des hypothèses du modèle de Wright-Fisher pour calculer l'effectif efficace de telles populations.

## Hypothèses du modèle

L'espèce considérée sera supposée diploïde. On s'intéressera à des locus autosomaux. On supposera que les générations sont non-chevauchantes, et qu'il n'y a pas de migration (pas de sous-populations). Pour calculer l'effectif efficace, on se limitera à l'étude de l'évolution de l'hétérozygotie en considérant un locus (neutre et sans mutation) présentant une diversité initiale maximale.

On remettra en cause successivement les hypothèses du modèle de Wright-Fisher suivantes :

1. **Nombre égal de mâles et de femelles (hypothèse contenue dans l'hypothèse d'une espèce monoïque).** On vérifiera dans un premier temps que l'introduction de la dioécie (c'est-à-dire d'une espèce à sexes séparés) avec un sex-ratio équilibré ne modifie pas l'effectif à considérer, puis on fera varier le sex-ratio.
2. **Panmixie.** On simulera une espèce pratiquant un taux  $s$  d'autofécondation. On fera varier ce taux de 0 à 1.
3. **Variance du nombre de gènes transmis par individu = 2.** On simulera différents cas de variance du succès reproducteurs, de 0 (chaque individu laisse exactement 2 copies de gènes) à des valeurs supérieures à 2.
4. **Taille constante de la population.** On simulera dans un premier temps un cycle de quelques générations de croissance suivies d'un goulot d'étranglement démographique restaurant la taille de population initiale. Dans un deuxième temps, on s'intéressera à un unique événement de croissance démographique.

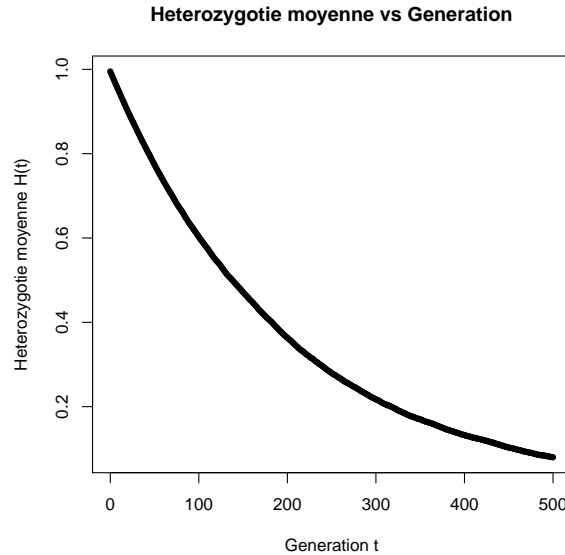


FIGURE 1 – Évolution de l'hétérozygotie moyenne  $\bar{H}_{ET}(t)$  en fonction des générations  $t$  pour une population de Wright-Fisher de taille constante  $N = 100$ . La moyenne de l'hétérozygotie est calculée sur 5000 simulations indépendantes.

## Complément : Estimation de l'effectif efficace $N_e$

L'effectif efficace sera estimé à partir de l'évolution de l'hétérozygotie moyenne  $\mathbb{E}[H_{ET}(t)]$  au cours du temps  $t$  (mesuré en génération). En effet, dans la plupart des cas, on s'attend à trouver la relation suivante :

$$\mathbb{E}[H_{ET}(t)] = H_{ET}(0) \left(1 - \frac{1}{2N_e}\right)^t$$

En pratique, il faudra estimer l'espérance  $\mathbb{E}[H_{ET}(t)]$  par la moyenne arithmétique de l'hétérozygotie  $\bar{H}_{ET}(t)$  pour  $n_{sim}$  simulations indépendantes :

$$\mathbb{E}[H_{ET}(t)] \simeq \bar{H}_{ET}(t) = \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} H_{ETi}(t)$$

où  $H_{ETi}(t)$  est l'hétérozygotie de la simulation  $i$  à la génération  $t$ .

L'hétérozygotie initiale  $H_{ET}(0)$  est parfaitement connue. On peut donc écrire :

$$\ln\left(\frac{\mathbb{E}[H_{ET}(t)]}{H_{ET}(0)}\right) = t \ln\left(1 - \frac{1}{2N_e}\right)$$

On s'attend donc à observer une relation linéaire du type :  $y = \beta x$  dont on peut estimer la pente  $\beta$  par régression linéaire :

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$$

où  $n$  est le nombre de points, c'est-à-dire le nombre de générations simulées.

L'estimation de l'effectif efficace est donc :

$$\hat{N_e} = \frac{1}{2 \left(1 - \exp(\hat{\beta})\right)}$$

Évidemment, cette estimation n'a de sens que si  $\ln\left(\frac{\bar{H}_{ET}(t)}{H_{ET}(0)}\right)$  est effectivement bien linéaire avec le temps  $t$ . Il faudra s'en assurer et adopter une méthode alternative de calcul de  $N_e$  dans le cas contraire.

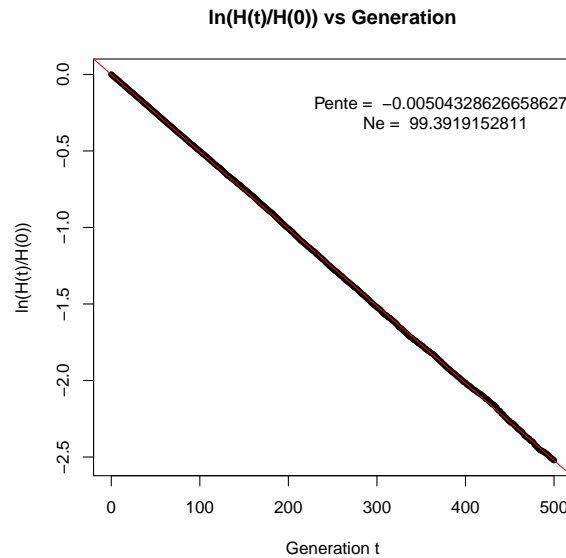


FIGURE 2 – Évolution de  $\ln\left(\frac{\bar{H}_{ET}(t)}{\bar{H}_{ET}(0)}\right)$  en fonction des générations pour une population de Wright-Fisher de taille constante  $N = 100$ . La moyenne de l'hétérozygotie est calculée sur 5000 simulations indépendantes. La relation est effectivement linéaire. L'effectif efficace estimé est  $\hat{N}e = 99.392$ . La population étant une population de Wright-Fisher de taille constante  $N$ , l'effectif efficace attendu est  $Ne = N = 100$ .

## Complément : Variance du succès reproducteur

Dans le modèle de Wright-Fisher, le nombre  $K$  de gènes (ou gamètes) transmis par un individu diploïde à la génération suivante suit une loi de Poisson de paramètre  $\lambda = 2$ . On a donc  $\mathbb{E}[K] = 2$  (population de taille constante) et  $\mathbb{V}[K] = 2$ . Dans une population de Wright-Fisher, environ 13.5% des individus ne se reproduisent pas du tout (aucun gamète transmis), 27.1% ont un seul descendant (un gamète transmis), 27.1% ont deux descendants (deux gamètes transmis), 18.0% ont trois descendants (trois gamètes transmis), 9% ont quatre descendants (quatre gamètes transmis), 3.6% ont cinq descendants (cinq gamètes transmis), etc.<sup>1</sup>

Néanmoins, dans certaines espèces, la variance du succès  $\mathbb{V}[K]$  peut être plus élevée. C'est par exemple le cas des espèces polygynes où un petit nombre de mâles ont beaucoup de descendants et un grand nombre de mâles n'ont aucun descendant. Dans d'autres cas, par exemple chez certaines espèces d'oiseaux qui contrôlent leur nombre d'œufs, la variance du succès  $\mathbb{V}[K]$  peut être plus faible. Dans une population de taille constante ( $\mathbb{E}[K] = 2$ ), l'effectif efficace  $Ne$  est (Wright S, 1938) :

$$Ne = \frac{4N - 2}{\mathbb{V}[K] + 2}$$

## Simulation

Pour faire varier la variance du succès reproducteur  $\mathbb{V}[K]$  tout en maintenant une population de taille constante ( $\mathbb{E}[K] = 2$ ), on peut contraindre la distribution de  $K$  à chaque génération en fixant le nombre d'individus laissant exactement 0, 1, 2, 3, ... gamètes à la génération suivante. Une possibilité est de n'avoir que trois types d'individus :

- $N_0$  individus laissant exactement 0 gamète
- $N_2$  individus laissant exactement 2 gamètes
- $N_4$  individus laissant exactement 4 gamètes

Les nombres d'individus  $N_0$ ,  $N_2$  et  $N_4$  doivent vérifier :

- $N_0 + N_2 + N_4 = N$
- $0 \times N_0 + 2 \times N_2 + 4 \times N_4 = 2 \times N$

1. Ces valeurs peuvent être retrouvées en exécutant la commande `dpois(0:5, lambda=2)` sous R.

En d'autres termes, si on connaît le nombre total d'individus  $N$  et le nombre d'individus ne laissant aucun gamète  $N_0$ , et à condition que  $N_0 \leq N/2$ , on a :

- $N_2 = N - 2N_0$
- $N_4 = N_0$

On peut montrer que :

$$\mathbb{V}[K] = 8 \frac{N_0}{N}$$

En faisant varier  $N_0$  entre sa valeur minimale 0 et sa valeur maximale  $N/2$ , il est donc possible de faire varier la variance  $\mathbb{V}[K]$  entre 0 et 4.

**Remarque :** Le succès reproducteur n'ayant pas de base héréditaire, il conviendra de fixer le nombre de gamètes transmis de chaque individu de manière aléatoire à chaque génération. On pourra éventuellement se servir de la fonction de permutation `d_random__permutation` fournie dans `d_utils.c`.