

Génétique quantitative

Christine Dillmann

christine.dillmann@u-psud.fr

Note : These are my raw course notes. Most references are still missing. I intend to improve it next year.

1 Composantes de la variance phénotypique

1.1 Héritéité et décomposition de la valeur phénotypique

La question de l'héritéité et la transmission des caractères quantitatifs, présentant des variations continues dans une population a été longtemps une source de débats au début du 20ème siècle. La formalisation mathématique a été proposée par Fisher (1918). Il propose un modèle probabiliste de la décomposition de la valeur phénotypique, basé sur une transmission mendélienne des gènes dans une population panmictique (Fig 1).

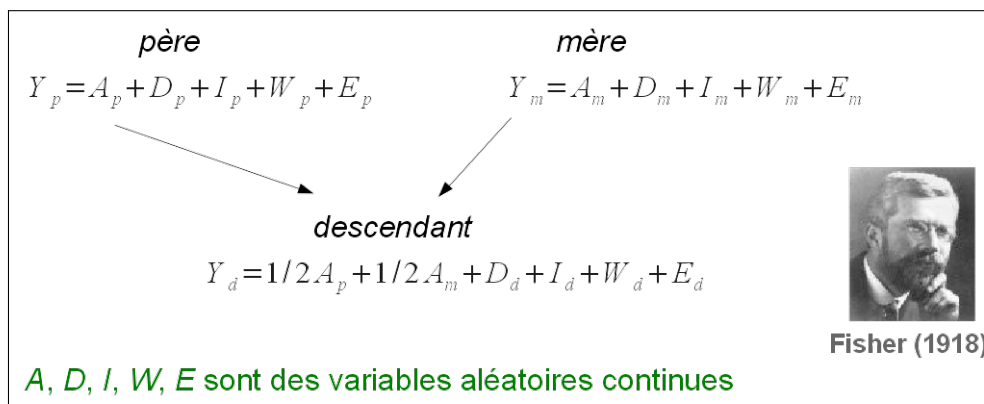


Figure 1: Décomposition de la valeur phénotypique et héritéité.

La valeur phénotypique est décomposée en une somme d'effets qui sont définis par leur mode de transmission d'une génération à l'autre :

- **Effets additifs (A)**: Chaque parent transmet à ses enfants la moitié de ses effets additifs, ce qui correspond au fait qu'un gamète parental, à l'issue de la méiose, contient la moitié de l'information génétique du parent.
- **Effets de dominance (D)** : Les effets de dominance sont des interactions entre allèles à un même locus. Lorsque la rencontre des gamètes se fait au hasard, il n'est pas possible de prédire l'allèle reçu de la mère connaissant l'allèle reçu du père. L'effet de dominance est donc spécifique à chaque individu. Attention, ce ne sera plus le cas lors de croisements **consanguins**.
- **Effets d'épistatie (I)** : Les effets d'épistasie sont des interactions entre allèles à des locus différents. En présence de **liaison génétique**, une partie des effets d'épistasie se transmettent comme des effets additifs.
- **Aléa de méiose (W)** : Un parent transmettant la moitié de ses gènes, il transmet en espérance la moitié de sa valeur génétique additive. Cependant, la transmission résulte du tirage au hasard d'une copie de chaque gène parmi les deux, la valeur du descendant comprend donc un terme aléatoire dit *aléa de méiose*.
- **Effet du micro-environnement (E)** : En l'absence de **corrélations environnementales**, les composantes micro-environnementales des parents sont indépendantes de celles des enfants, et spécifiques à chaque individu. Les effets *E* peuvent cependant se transmettre en partie pour tous les caractères ayant une composante sociale.

1.2 Resemblance entre apparentés

Les individus apparentés se ressemblent car ils partagent des gènes en commun, hérités de leurs ascendants.

1.2.1 Rappels de probabilités

Si Y est une variable aléatoire continue, on peut lui associer une loi de probabilité, c'est à dire une densité $f(y)$ ($\int_{-\infty}^{+\infty} f(y)dy = 1$) et une fonction de répartition $F(y)$ telle que :

$$F(y) = P(Y \leq y)$$

La plupart des caractères quantitatifs suivent des lois normales ou log-normales (loi des grands nombres), définies par deux paramètres, la moyenne ou espérance $\mu = E(Y) = \int_{-\infty}^{+\infty} yf(y)dy$, et la variance

$$var(Y) = E((Y - E(Y))^2)$$

Composition : si a est une constante réelle, alors :

$$E(aY) = aE(Y)$$

$$var(aY) = a^2 var(Y)$$

Si (X, Y) est un couple de variables aléatoires, c'est à dire que l'on peut associer à chaque individu de la population une valeur de X et une valeur de Y , on peut déterminer la loi de probabilité conjointe de X et Y .

Composition : si a est une constante réelle, alors :

$$E(aY + bX) = aE(Y) + bE(X)$$

Indépendance : X et Y sont indépendantes si et seulement si la variance de leur somme est égale à la somme des variances.

Covariance : L'écart à l'indépendance est mesuré par rapport à la variance de la somme $X + Y$:

$$var(X + Y) = var(X) + var(Y) + 2cov(X, Y)$$

La covariance est une mesure de la ressemblance entre les deux variables aléatoires :

$$cov(X, Y) = E((X - E(X))(Y - E(Y)))$$

A noter que $cov(X, X) = var(X)$.

Modèle linéaire et relations de dépendance : Deux variables aléatoires ont une relation linéaire si l'on peut écrire :

$$Y = a + bX + \epsilon$$

où ϵ est une variable aléatoire indépendante de Y et X ($cov(X, \epsilon) = 0$ et $cov(X, \epsilon) = 0$).

On peut montrer alors que $cov(X, Y) = bvar(X)$, où b est appelé le coefficient de régression de Y sur X :

$$b = \frac{cov(X, Y)}{var(X)} \quad (1)$$

Le coefficient de corrélation est une mesure standardisée de la covariance ($-1 \leq r \leq 1$:

$$r = \frac{cov(X, Y)}{\sqrt{var(X)var(Y)}}$$

1.2.2 Composantes de la variance phénotypique

Dans le modèle de Fisher, la valeur phénotypique est décomposée en une somme de composantes aléatoires indépendantes. Pour mesurer l'importance de chaque composante, on peut utiliser la propriété de sommation de la variance des variables aléatoires indépendantes, et mesurer la part de chaque composante dans la variance phénotypique de la population.

Le modèle peut-être réécrit en rassemblant les effets génétiques et non génétiques, ou alors en isolant les effets additifs :

$$P = G + E = A + NA + E = A + D + I + W + E \quad (2)$$

où $NA = D + I + W$ correspond à la somme des effets génétiques non-additifs. Les parts de variance génétique et additive sont appelées *héritabilités*.

Héritabilité au sens large : part de variance génétique dans la variance phénotypique. Mesure aussi le coefficient de la régression génotype-phénotype :

$$H^2 = \frac{\text{var}(G)}{\text{var}(P)}$$

Dans certains cas, lorsque l'on peut disposer de mesures sur des individus différents d'un même génotype (clonage, autofécondation), on peut estimer expérimentalement H^2 ou $\text{var}(G)$.

Héritabilité au sens étroit : part de variance additive dans la variance phénotypique.

$$h^2 = \frac{\text{var}(A)}{\text{var}(P)}$$

Pour estimer l'héritabilité, on peut utiliser des mesures du même caractère sur des individus apparentés.

1.2.3 Régression parent-enfant

On peut utiliser le modèle (Figure 1) pour estimer la covariance entre parents- et enfants. Il suffit de servir des propriétés de sommation pour la covariance entre sommes de termes :

$$\begin{aligned} \text{cov}(Y_p, Y_d) = & \text{cov}(A_p, \frac{1}{2}A_p) + \text{cov}(A_p, \frac{1}{2}A_m) \\ & + \text{cov}(A_p, D_d) + \text{cov}(A_p, I_d) + \text{cov}(A_p, W_d) + \text{cov}(A_p, E_d) \\ & + \text{cov}(D_p, \frac{1}{2}A_p) + \text{cov}(D_p, \frac{1}{2}A_m) + \text{cov}(D_p, D_d) \\ & + \dots + \text{cov}(E_p, E_d) \end{aligned}$$

Si l'on fait les hypothèses suivantes :

- Population panmictique
- Effectif efficace élevé
- L'effet des mutations peut-être négligé
- Pas de sélection
- Pas d'effets de milieu commun

la plupart des termes de covariance s'annulent (effets indépendants). On trouve, $\text{cov}(Y_p, Y_d) = \frac{1}{2}\text{var}(A)$. Doc, la pente de la droite de régression parent-enfant vaut :

$$b_{P/O} = \frac{1}{2} \frac{\text{var}(A)}{\text{var}(P)} = \frac{1}{2} h^2 \quad (3)$$

De façon générale, si l'on néglige les effets d'épistasie, et si X et Y correspond à la mesure d'un même caractère chez deux individus, on a

$$\text{cov}(X, Y) = 2\varphi_A \text{var}(A) + \varphi_D \text{var}(D) \quad (4)$$

où φ_A et φ_D sont les coefficients de simple et double parenté entre X et Y .

1.2.4 Identité par descendance

Gustave Malécot (1945) définit formellement la notion d'identité par descendance.

Coefficient de parenté : Probabilité que deux gènes tirés au hasard dérivent par descendance mendélienne sans mutation d'un même gène ancêtre.

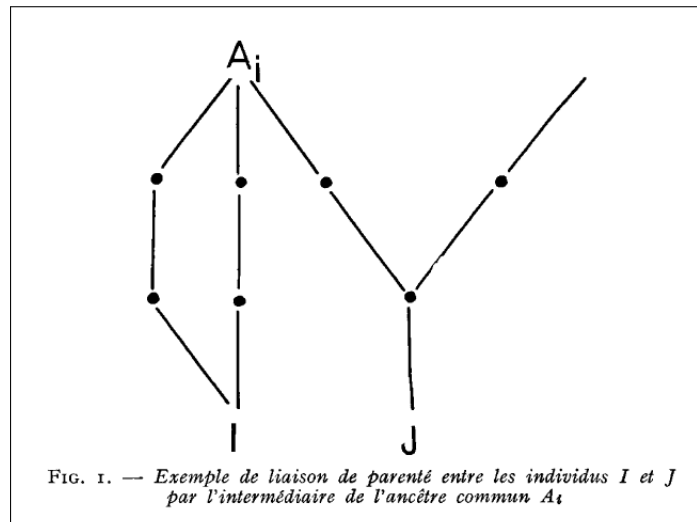


Figure 2: Illustration tirée de l'article de G. Malecot (1972)

A chaque génération, la probabilité de transmettre un gène donné à sa descendance est $1/2$. On peut écrire la relation générale, illustrée Figure 2 :

$$\varphi_A(I, J) = \sum_k \left(\frac{1}{2}\right)^{(nI_k + nJ_k)} \left(\frac{1 + f_{A_i}}{2}\right) \quad (5)$$

où k est le nombre de branches reliant I et J et passant par l'ancêtre commun A_i , et nI_k (nJ_k) le nombre de générations séparant I (J) et A_i .

Exercice 1. Coefficients de parentés

Complétez le tableau ci-dessous en calculant le coefficient d'apparentement φ_A et la valeur des la covariance entre apparentés :

Parenté	φ_A	φ_D	$cov(X, Y)$
Parent-enfant (PO)		0	
Plein-frères		1/4	
Demi-frères		0	
Individu avec lui-même	1/2	1	$\text{var}(A) + \text{var}(D)$

1.2.5 Application : Estimation de l'héritabilité par régression parent/enfant

Exercice 2. Héritabilité de la taille chez l'Homme

Les données du fichier `biom_UPSUD_2012_2016.csv` résultent d'un enquête réalisée chaque année entre 2012 et 2016 auprès de l'ensemble des étudiants de L1 Biologie de l'université Paris-Sud. Des données biométriques ont été récoltées, dont la taille et la taille de chacun des deux parents.

1. Utilisez plusieurs méthodes pour estimer l'héritabilité du caractère, en prenant en compte l'effet du sexe sur la taille.
2. On peut aussi faire une régression entre la valeur moyenne des deux parents et l'enfant. Calculez le coefficient de régression dans ce cas en utilisant (1)
3. Etudiez la relation entre la taille du père et celle de la mère dans cet échantillon.

Que pouvez-vous conclure de cette étude ?

Exercice 3. Pinsons de Darwin

Des données récoltées sur la longueur du bec chez l'un des espèces de pinsons de Darwin (*G. fortis*) ont donné les estimations suivantes pour l'héritabilité de la longueur du bec (Boag, 1983) :

Long. bec	Parent moyen/enfant	père/enfant	mère/enfant
coef. de régression	0.72	0.22	0.39
héritabilité			

Proposez des hypothèses pour expliquer ces résultats.

On sait par ailleurs que les pères ne sont pas toujours les pères biologiques. Une analyse ADN a montré que c'était le cas pour 44 des 223 enfants analysés. En ne conservant que les paires pour lesquelles la relation de paternité biologique est confirmée, on trouve un coefficient de régression père/enfant de $b = 0.36$. Cela change-t-il vos conclusions ?

1.3 Modèle biallélique

On considère le modèle biallélique suivant, en supposant que les effets génétiques sont connus, et que la population est panmictique :

Génotype	A1A1	A1A2	A2A2
G	$G_{11} = a$	$G_{12} = d$	$G_{22} = a$
Fréquence	p^2	$2pq$	q^2

On appelle p (q) la fréquence de l'allèle A1 (A2), et μ la moyenne phénotypique de la population. Sachant qu'un parent transmet, en espérance, la moitié de ses effets additifs, on peut définir l'effet additif d'un allèle α_i comme une espérance conditionnelle

$$\alpha_i = E(P/Ai) - E(P) \quad (6)$$

C'est à dire, comme la moyenne phénotypique des individus, sachant qu'ils portent une copie au moins de l'allèle Ai . Les effets additifs sont centrés par rapport à la moyenne de la population.

Les effets de dominances sont des effets d'interaction entre allèles à un même locus. Ils sont définis comme l'écart entre la valeur génétique et la prédiction additive :

$$\delta_{ij} = E(P/AiAj) - \alpha_i - \alpha_j - E(P) \quad (7)$$

On peut réécrire le tableau suivant, en remplaçant les valeurs génotypiques par leur écriture en fonction des effets statistiques des gènes :

Génotype	A1A1	A1A2	A2A2
$G - \mu$	$2\alpha_1 + \delta_{11}$	$\alpha_1 + \alpha_2 + \delta_{12}$	$2\alpha_2 + \delta_{22}$

Exercice 4. Composantes de la variance :

Calculez les effets additifs et les composantes de la variance dans ce modèle biallélique. Vous étudierez ensuite comment varie la variance additive avec les fréquences alléliques et le coefficient de dominance d/a

Effets additifs La moyenne de la population $\mu = E(P)$ se calcule en remplaçant les G_{ij} par leurs valeurs :

$$\mu = a(p^2 - q^2) + 2pqd = (p - q)a + 2pqd$$

Pour calculer l'effet additif de l'allèle A1, on utilise 6 :

$$\alpha_1 = pa + qd - \mu = q(a - (p - q)d)$$

On fait de même pour α_2 :

$$\alpha_2 = pd - qa - \mu = -p(a - (p - q)d)$$

Enfin, la variance additive vaut :

$$\text{var}(A) = 2E(\alpha_i^2) = 2(p\alpha_1^2 + q\alpha_2^2) = 2pq(a - (p - q)d)^2 \quad (8)$$

Effets de dominance Ici, on utilise 7. Il y a trois termes à calculer, un pour chaque génotype.

$$\delta_{11} = a - 2\alpha_1 - \mu = -2q^2d$$

$$\delta_{12} = d - \alpha_1 - \alpha_2 - \mu = 2pqd$$

$$\delta_{22} = -a - 2\alpha_2 - \mu = -2p^2d$$

Et la variance de dominance vaut :

$$\text{var}(D) = 2E(\delta_{ij}^2) = 4p^2q^2d^2 \quad (9)$$

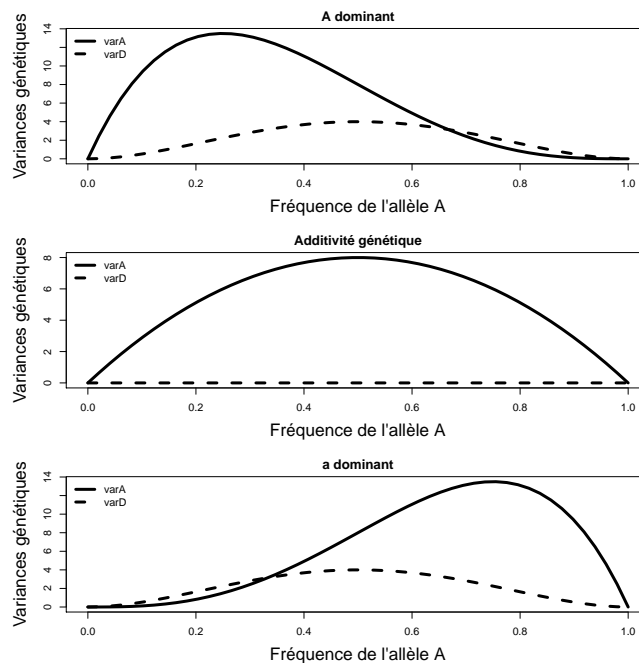


Figure 3: **Composantes de la variance dans un cas biallélique.** La variances additives et de dominances sont calculées en fonction de la fréquence de l'allèle A, qui varie entre zéro et 1, en utilisant (8) et (9).

Bilan :

- Les composantes de la variance varient avec les fréquences des allèles, alors que les phénotypes associés à chaque génotype restent constants.
- Les effets additifs sont additifs au sens de la transmission des effets moyens d'une génération à l'autre. Ils peuvent être importants même en présence de dominance *génétique*, c'est à dire lorsque la moyenne phénotypique des hétérozygotes est différente de la moyenne des homozygotes.
- Dans la plupart des cas, la variance additive est beaucoup plus importante que la variance de dominance.

2 Réponse à la sélection**2.1 The Breeder's equation**

La sélection ne modifie pas l'hérédité des caractères mais les fréquences alléliques. Elle change la probabilité des parents de participer à la génération suivante. Considérons des générations non chevauchantes et une population panmictique. On appelle X_t la variable aléatoire qui décrit la valeur phénotypique d'un individu de la population à la génération t . On peut utiliser le modèle de régression parent-enfant pour décrire la valeur phénotypique des descendants à la génération $t + 1$:

$$X_{t+1} = \mu_t + h^2 \left(\frac{X_t^m + X_t^p}{2} - \mu_t \right) + \epsilon$$

L'espérance conditionnelle de la valeur phénotypique des enfants, sachant la valeur de la moyenne phénotypique des parents vaut :

$$E(X_{t+1} / (X_t^m, X_t^p)) = \mu_t + h^2 \left(\frac{X_t^m + X_t^p}{2} - \mu_t \right)$$

La moyenne phénotypique de la génération suivante s'obtient en intégrant sur l'ensemble des valeurs phénotypiques possibles pour les parents, c'est à dire en intégrant sur les paires $((X_t^m, X_t^p))$, en prenant en compte leur valeur sélective. La valeur phénotypique moyenne des parents qui contribuent à la génération suivante vaut, dans une population panmictique et en supposant l'absence d'effets du sexe sur la valeur phénotypique :

$$\mu_{s,t} = \frac{\int x_t w(x) f_t(x) dx}{\int w(x) f_t(x) dx}$$

où $w(x)$ est la valeur sélective d'un individu de phénotype x . On trouve donc :

$$\mu_{t+1} = \mu_t + h^2(\mu_{s,t} - \mu_t) \quad (10)$$

que l'on appelle **l'équation du sélectionneur**.

2.2 Héritabilité réalisée

On peut utiliser les résultats d'une expérience de sélection pour estimer l'héritabilité d'un caractère en se servant de l'équation du sélectionneur. Si l'on appelle R la réponse à la sélection, et S la différentielle de sélection, on a :

$$R = h^2 S$$

- R est la différence de moyenne de la population d'une génération à l'autre.
- S se calcule comme la différence entre la moyenne des individus sélectionnés (parents de la génération $t + 1$) et la moyenne de la population à la génération t .

- Si l'on dispose de plusieurs générations successives, on peut calculer plusieurs couples de valeurs (R_i, S_i) . L'hérissabilité réalisée se calcule alors comme la pente de la droite de régression de R sur S , en fixant l'ordonnée à l'origine à zéro.

2.3 Gradients de sélection

2.3.1 Un caractère

A une génération donnée, on peut associer une valeur sélective $w_t(x)$ à chaque phénotype x de la population, et supposer qu'il s'agit d'une fonction continue et intégrable sur le domaine de variation de x . En particulier,

$$\bar{w}_t = \int w_t(x) f(x) dx$$

La sélection va conduire à un changement de la fitness moyenne, du fait du changement de la moyenne du caractère. Si l'on suppose que X_t suit une loi normale $\mathcal{N}(\mu_t, \text{var}_t(P))$, on peut calculer $\frac{d\bar{w}_t}{d\mu_t}$ en utilisant la propriété suivante :

$$\frac{df(x_t)}{d\mu_t} \approx f(x_t) \frac{(x_t - \mu_t)}{\text{var}(P)}$$

et le fait que $w_t(x)$ ne dépend pas de la moyenne phénotypique de la population μ_t .

On trouve donc,

$$\frac{d\bar{w}_t}{d\mu_t} = \frac{1}{\text{var}(P)} [\int w_t(x) x_t f(x_t) dx - \mu_t \int w_t(x) f(x_t) dx] = \frac{1}{\text{var}(P)} [\bar{w}_t \mu_{s,t} - \mu_t \bar{w}_t]$$

qui peut également s'écrire

$$\mu_{s,t} - \mu_t = \text{var}_t(P) \frac{d \ln(\bar{w}_t)}{d\mu_t}$$

En utilisant (10), on peut donc écrire sous une autre forme la réponse à une génération de sélection :

$$\boxed{\mu_{t+1} - \mu_t = \text{var}(A) \frac{d \ln(\bar{w}_t)}{d\mu_t}} \quad (11)$$

Cette formulation, et son extension au cas multivarié a été proposée initialement par R. Lande (1976, 1979). Elle montre que la réponse à la sélection sur un caractère dépend de la variance génétique additive du caractère et de la force de la sélection, qui se mesure comme la pente de la fonction reliant la valeur sélective moyenne de la population à sa moyenne phénotypique. *La sélection va déplacer la moyenne phénotypique de la population vers un maximum local de la fonction de fitness.*

2.3.2 Paysage adaptatif

Si l'ensemble des caractères qui déterminent la fitness ont des distributions gaussiennes, on peut utiliser les propriétés de la loi multinormale pour étendre (11) au cas multicaractère. Par exemple pour deux caractères A et B , on aura :

$$\begin{pmatrix} \Delta\mu^A \\ \Delta\mu^B \end{pmatrix} = \begin{pmatrix} \text{var}(A^A) & \text{cov}(A^A, A^B) \\ \text{cov}(A^A, A^B) & \text{var}(A^B) \end{pmatrix} \begin{pmatrix} \beta^A = \frac{\partial \ln(\bar{w})}{\partial \mu^A} \\ \beta^B = \frac{\partial \ln(\bar{w})}{\partial \mu^B} \end{pmatrix}$$

Les coefficients β sont appelés **gradients de sélection** et déterminent le **paysage adaptatif**. On peut condenser l'écriture matricielle en écrivant :

$$\boxed{\Delta\vec{\mu} = G\vec{\beta}} \quad (12)$$

G est la matrice de variance-covariance des effets additifs. Elle est appelée (à tort) **G matrix**. En déterminant les axes de variation génétique les plus forts, la matrice G va déterminer la direction de l'évolution phénotypique dans un paysage adaptatif donné.

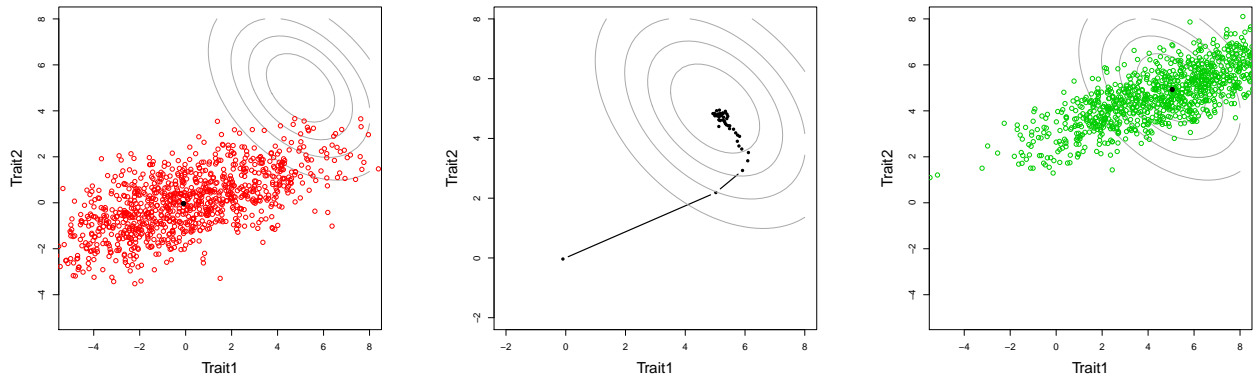


Figure 4: **Evolution dans un paysage adaptatif : Pléiotropie.** Le paysage adaptatif est représenté par des courbes de niveaux dans l'espace à deux dimensions défini par les deux traits d'histoire de vie qui conditionnent la fitness. Les points correspondent aux valeurs phénotypiques des individus. **A gauche.** Situation initiale avec une corrélation génétique positive entre les deux caractères due à de la pléiotropie. Cette corrélation est recréée à chaque génération. **Au milieu.** Evolution de la moyenne phénotypique durant 50 générations. **A droite.** Distribution des valeurs phénotypiques à l'issue des 50 générations.

Les figures 4 et 5 représentent deux simulations de l'évolution d'un couple de caractères dans un paysage adaptatif. Le modèle d'hérédité est basé sur le modèle de Fisher, en supposant l'additivité pour les deux caractères, et en négligeant les effets du milieu. La pléiotropie est simulée en introduisant une corrélation positive entre les deux caractères pour l'aléa de méiose. L'évolution déplace les moyennes phénotypiques vers l'optimum et peut conduire à des modifications des corrélations entre caractères (figure 5).

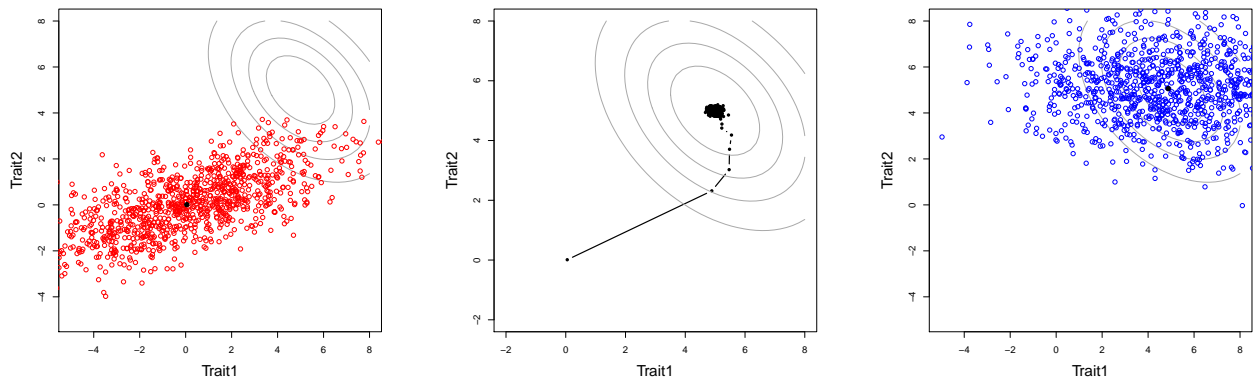


Figure 5: **Evolution dans un paysage adaptatif : Sélection antagoniste.** Le paysage adaptatif est représenté par des courbes de niveaux dans l'espace à deux dimensions défini par les deux traits d'histoire de vie qui conditionnent la fitness. Les points correspondent aux valeurs phénotypiques des individus. **A gauche.** Situation initiale avec une corrélation génétique positive entre les deux caractères due à de la dérive. **Au milieu.** Evolution de la moyenne phénotypique durant 500 générations. **A droite.** Distribution des valeurs phénotypiques à l'issue des 500 générations.

3 Réponse à long terme à la sélection

3.1 Effet des pressions évolutives

S'il est facile de décrire la réponse à une génération de sélection, il n'existe pas d'expression analytique pour la réponse à long terme, à moins de faire des approximations sur l'évolution de la variance

génétique additive. A chaque génération, on a :

$$\Delta\mu_t = \beta_t \text{var}_t(A)$$

- Les pressions évolutives vont faire changer la variance d'une génération à l'autre, à travers les *variations des fréquences des allèles*. Dans un modèle additif, la variance génétique est maximale pour des fréquences intermédiaires des deux allèles (8).
- Les *changements environnementaux* vont modifier à la fois le paysage adaptatif, et donc le gradient de sélection β , et l'effet des allèles (plasticité phénotypique).
- En présence d'*épistasie*, l'effet additif marginal des allèles dépend du contexte génétique. Les variations des fréquences alléliques vont donc induire également des variations des effets des allèles.
- La sélection multilocus crée des associations négatives entre allèles favorables à des locus différents (*déséquilibre de liaison*) qui ne sont pas complètement rompues par la recombinaison (effet Bulmer).

Dans l'ensemble, les pressions évolutives devraient conduire à une diminution de la variance génétique additive sous l'effet de la sélection, jusqu'à un équilibre sélection/dérive/mutation (ref Hill). Par conséquent, la réponse à la sélection devrait diminuer au cours des générations, jusqu'à atteindre asymptotiquement une valeur d'équilibre constante.

3.2 Le modèle infinitésimal

3.2.1 Données empiriques

Les expériences d'évolution expérimentales contredisent la théorie et montrent en général des réponses linéaires à la sélection sur des centaines, voire des milliers de générations. Les deux expériences les plus célèbres sont peut-être l'expérience de sélection pour la teneur en protéines et la teneur en huile des grains de maïs de l'université d'Illinois (<http://mooselab.cropsci.illinois.edu/longterm.html>), et l'expérience *LLTE* de Richard Lenski chez la bactérie *E. coli* (Wiser et al., 2013, *Science* 342:1364).

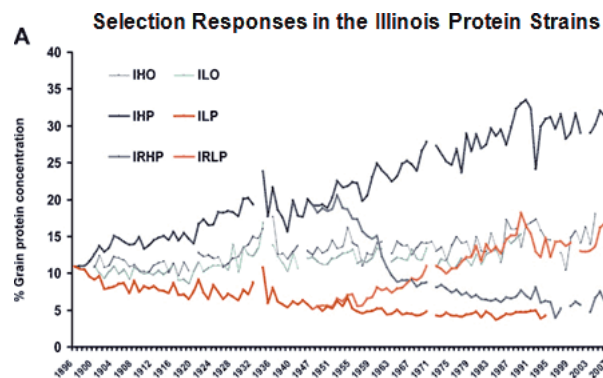


Figure 6: **Illinois Long-term Selection Experiment : évolution de la teneur en protéine des grains de maïs.** La plus longue expérience de sélection chez les plantes a été initiée en 1896 à l'université d'Illinois à partir d'une population de maïs. Au départ, deux couples de populations divergentes ont été créées avec une sélection pour la teneur en huile (IHO,ILO) ou en protéine (IHP,ILP) des grains. Des populations ont aussi été mises en place plus tard, pour tester la réponse à la sélection réverse (IRHP,IRLP). Les populations ILO et IHO représentées ici n'ont pas été sélectionnées pour la teneur en protéines et illustrent les effets de la dérive génétique.

Chez le maïs, la réponse à la sélection est linéaire et ne ralentit pas, même après 200 générations de sélection, en dépit de l'effectif démographique relativement faible (entre 60 et 120 individus phénotypés

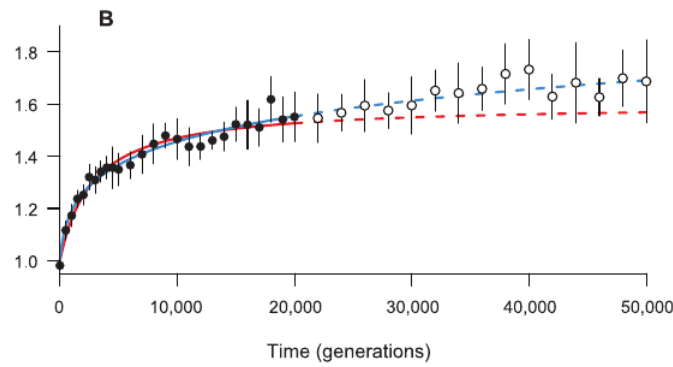


Figure 7: **Lenski's experimental evolution : évolution de la fitness des souches évoluées au cours des 50000 générations de bacth.** L'expérience a démarré en février 1988 avec plusieurs réplicats d'évolution. Depuis, chaque jour et dans chaque réplicats, un échantillon est repiqué sur un milieu de culture neuf. Tous les 75 jours (500 générations), un échantillon de chaque réplicat est congelé. Des expériences de compétition avec la souche ancestrale sont réalisées pour estimer la fitness moyenne de la culture. Les données (ici, la fitness moyenne de l'ensemble des réplicats) s'ajustent mieux avec un modèle prédisant l'augmentation de la fitness selon une loi de puissance (courbe bleue) qu'avec un modèle hyperbolique prédisant une limite à la réponse à la sélection (courbe rouge).

à chaque cycle) et d'une sélection assez forte (20% des individus les plus extrêmes dans chaque population).

Dans l'expérience de Lenski, la fitness continue de croître selon une loi de puissance, donc le logarithme de la fitness croît de façon linéaire.

Tout se passe donc comme si, en milieu constant, la variance génétique additive des caractères sélectionnés restait constante.

3.2.2 Modèle infinitésimal

La vision Mendélienne. On peut expliquer la constance de la variance génétique additive si l'on suppose qu'il existe un nombre infini de gènes gouvernant le ou les caractères sous sélection, contribuant chacun de façon infiniment petite à la variance génétique dans une population de taille infinie. Dans ce cas, la pression de sélection qui s'exerce sur chaque locus va être très petite, conduisant à des variations très faibles des fréquences alléliques.

Pour le montrer, on peut reprendre le modèle biallélique développé plus haut en supposant que l'effet génétique (a) est le même à chaque locus, de même que les fréquences alléliques. Si l'on appelle L le nombre de locus, la variance génétique additive vaut :

$$\text{var}(A) = 2pqLa^2$$

c'est à dire que, pour une variance donnée, l'effet a de chaque locus diminue de façon linéaire avec la racine carrée \sqrt{L} du nombre de locus.

La réponse à la sélection vaut

$$\Delta\mu = \beta\text{var}(A) = 2La\Delta p$$

Le changement de fréquence allélique sous l'effet de la sélection peut alors s'écrire

$$\Delta p = \beta pqa$$

et diminue lui aussi de façon linéaire avec la racine carrée du nombre de locus.

Enfin, le changement de variance génétique additive vaut

$$\Delta\text{var}(A) = \frac{\text{var}(A)}{pq} \frac{\Delta\mu}{L} \left(q - p - \frac{\Delta\mu}{L} \right)$$

La variance génétique additive diminue comme le carré du nombre de locus. Lorsque L devient très grand, $\Delta\text{var}(A)$ tend vers zéro.

La vision Fisherienne. Une autre façon de définir le modèle infinitésimal a récemment été proposée par Nick Barton (Barton et al, 2016 <http://dx.doi.org/10.1101/039768>). Il considère un caractère purement additif dans une grande population panmictique. La valeur phénotypique P_d peut alors s'écrire :

$$P_d = \frac{1}{2}(A_p + A_m) + W$$

Le modèle infinitésimal se traduit par le fait que la variance entre descendants d'un même croisement $V_0 = \text{var}(W)$ est *indépendante des effets génétiques des parents*. La variance de la génération suivante peut donc s'écrire :

$$V_{t+1} = \frac{1}{2}V_t + V_0$$

A l'équilibre, $V_t = 2V_0$, c'est à dire que la moitié de la variance génétique additive est entre-familles, et l'autre moitié intra-famille.

Si la variance de ségrégation V_0 est indépendante des valeurs phénotypiques des parents, elle peut être réduite dans les cas de croisements consanguins, et vaut alors :

$$V_0 \left(1 - \frac{F_{i,i} + F_{j,j}}{2} \right)$$

où $F_{i,i}$ est le coefficient de consanguinité du parent i .

Au total, la variance diminue sous l'effet de la dérive, et augmente sous l'effet des mutations. On a alors :

$$\boxed{\text{var}_{t+1}(A) = \left(1 - \frac{1}{N_e} \right) \text{var}_t(A) + Vm} \quad (13)$$

où Vm est la part de variance apportée à chaque génération par des nouvelles mutations.

A l'équilibre mutation/sélection/dérive, la variance génétique additive tend vers (ref Hill) :

$$\text{var}(A) = 2N_e Vm$$

Ainsi, le modèle infinitésimal, en soi, ne suffit pas à expliquer les réponses à la sélection observées dans les expériences d'évolution expérimentale, à moins de supposer qu'une part important de la variance génétique est générée à chaque génération par de nouvelles mutations.

Le rôle de l'épistasie On voit donc que pour expliquer la constance apparente de la variance génétique additive au cours des générations de sélection dans des populations de petite taille, il faut supposer un mécanisme qui régénère à chaque génération la variabilité phénotypique. L'hypothèse centrale du modèle proposé par Barton (*Heredity*(2017)118, 96109) et Barton et al (2016) est que les phénotypes n'occupent à une génération qu'une partie limitée de la variation possible, étant donnée l'étendue des possibilités de combinaisons multilocus dans un modèle avec un très grand nombre de locus. Ainsi, même des changements très petits de fréquences alléliques peuvent permettre de voir apparaître de nouvelles combinaisons génétiques.

Les simulations présentées Figure 8 montrent qu'il est possible, en présence d'épistasie, d'observer une réponse à la sélection plus importante que dans le modèle additif, avec une diminution de la variance génétique additive au cours du temps à peine inférieure à l'attendu dans le modèle neutre.

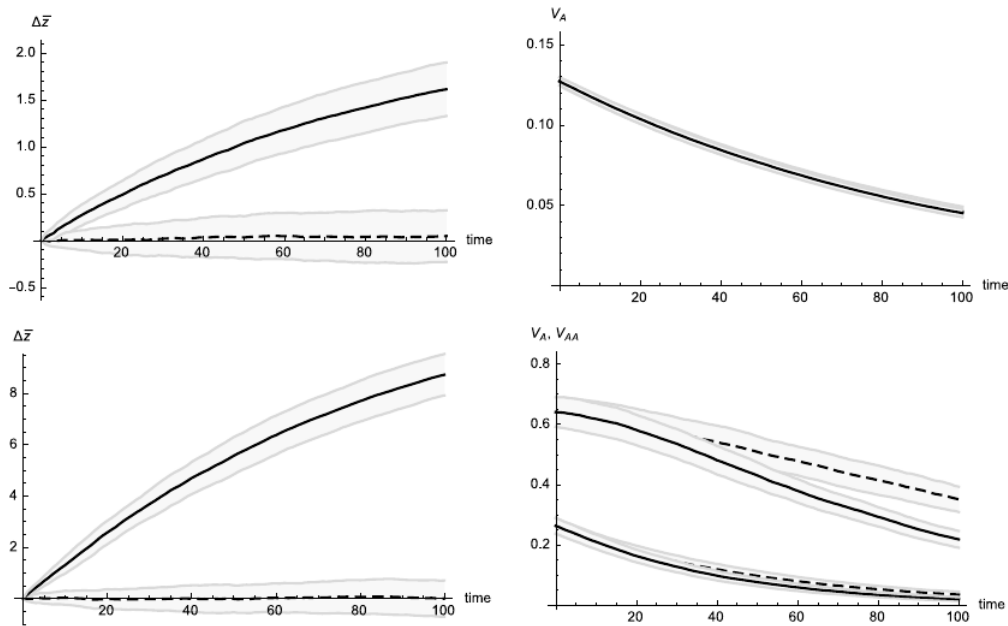


Figure 8: **Réponse à la sélection dans le modèle infinitésimal en présence d'épistasie.** La sélection directionnelle (lignes pleines) est comparée au cas neutre (pointillés), dans un cas additif (**en haut**) ou avec de l'épistasie (**en bas**). Les zones ombrées correspondent à +/- un écart-type. **A gauche** : évolution de la moyenne. **A droite** : évolution de la variance additive (VA) et de la variance d'épistasie additive x additive (VAA). Le modèle considère $M = 1000$ locus dans une population de $N = 100$ individus haploïdes. Les allèles à chaque locus ont le même effet, mais un signe aléatoire $\gamma = +/ - \frac{1}{\sqrt{M}}$. L'épistasie est simulée en choisissant une petite fraction $1/M$ d'interactions entre paires de locus w_{ij} dans une loi normale d'espérance nulle et d'écart-type $\frac{4}{\sqrt{M}}$, en autorisant $w_{ij} \neq w_{ji}$. La valeur phénotypique du caractère est définie comme $z = \delta\gamma + \delta^T w \delta$, avec $\delta = +/ - 1/2$. Les fréquences alléliques de départ sont tirées dans une loi Gamma en forme de U de moyenne $\bar{p} = 0.2$ et de variance $0.2\bar{p}\bar{q}$. Les parents de la génération suivante sont tirés au hasard avec une probabilité proportionnelle à $W = e^{\beta z}$. Dans chaque exemple, trois jeux d'effets additifs et épistatiques sont tirés au hasard, et pour chaque jeu de paramètres, l'évolution de trois populations différentes est simulée (9 répliquats).