

## **Chapitre 02: Mesure de la diversité génétique**

### **10. Mesure de la diversité génétique**

#### **10.1. Fréquences alléliques et fréquences génotypiques**

Lorsqu'une population est polymorphe pour un caractère donné, il est possible de calculer la fréquence des phénotypes observés (Masson et Hove, 2011, 2014).

Par exemple dans une population de  $N$  individus dont  $N_n$  ont le corps noir et  $N_b$  le corps blanc, les fréquences phénotypiques de la population pour le caractère couleur du corps sont les suivantes :

Fréquence du phénotype noir  $f[n] = N_n/N$

Fréquence du phénotype blanc  $f[b] = N_b/N$

Si ce caractère est gouverné par un gène à deux allèles  $A$  et  $a$  autosomaux, avec  $a$  récessif responsable de la couleur blanche, les génotypes  $AA$  et  $Aa$  correspondent au phénotype noir et le génotype  $aa$  au phénotype blanc.

Les fréquences phénotypiques permettent alors uniquement de connaître la fréquence du génotype *aa* puisque parmi les individus noirs, on ne peut pas distinguer les génotypes *AA* des génotypes *Aa*.

La fréquence des allèles *A* et *a* ne peut également pas être calculée.

Si la couleur du corps des individus présente non plus 2 mais 3 phénotypes (noir, jaune et blanc) gouvernés par un couple d'allèles **A1** et **A2** autosomiaux et codominants, les trois génotypes possibles **A1A1**, **A1A2** et **A2A2** peuvent être distingués puisqu'ils correspondent à des phénotypes différents (respectivement noir, jaune et blanc).

La composition phénotypique de la population correspond alors à sa composition génotypique et si on appelle  $N_n$ ,  $N_j$  et  $N_b$  les nombres d'individus présentant les phénotypes noir, jaune et blanc, on peut facilement calculer les fréquences génotypiques dans cette population :

$$f(A1A1) = N_n/N = D$$

$$f(A1A2) = N_j/N = H$$

$$f(A2A2) = N_b/N = R$$

Ainsi, pour un locus donné, une population est complètement décrite si l'on connaît la fréquence de chacune des catégories génétiques.

Dans le cas d'un système diallélique **A** et **a**, la structure d'une population d'effectif **N** est complètement connue si l'on connaît les effectifs

$$N_{AA} \text{ de } AA, N_{Aa} \text{ de } Aa \text{ et } N_{aa} \text{ de } aa \text{ avec } N = N_{AA} + N_{Aa} + N_{aa}$$

à partir desquels on calcule les fréquences relatives des trois génotypes.

Cette caractérisation génétique de la population n'est possible que si l'on sait lire les génotypes individuels, c'est à dire s'il y a codominance.

A partir des fréquences génotypiques, il est facile de calculer les fréquences alléliques dans la population, c'est à dire les fréquences des différents états alléliques du locus considéré.

Dans le cas d'un gène autosomal à deux allèles *A* et *a*, la fréquence de l'allèle *A* est le rapport du nombre d'allèles *A* au nombre total d'allèles à ce locus, soit  $2N$  pour une population de  $N$  individus diploïdes:

- les  $N_{AA}$  individus **AA** sont porteurs de deux allèles **A**

- les  $NAa$  individus  $Aa$  d'un allèle  $A$  et d'un allèle  $a$
- les  $Naa$  individus  $aa$  de deux allèles  $a$ .

Le nombre d'allèles  $A$  dans la population est donc

$$2NAA + NAa .$$

Les fréquences  $p$  et  $q$  des allèles  $A$  et  $a$  sont alors les suivantes:

$$f(A) = p = (2NAA + NAa)/2N \quad f(a) = q = (2Naa + NAa)/2N \quad \text{avec } p + q = 1$$

Autrement dit si  $D$  et  $R$  sont les fréquences des homozygotes  $AA$  et  $aa$ ,  $H$  la fréquence des hétérozygotes  $Aa$ , les fréquences alléliques peuvent aussi être calculées à partir des fréquences génotypiques :

$$f(A) = p = D + H/2 \quad f(a) = q = R + H/2$$

Ces fréquences  $p$  et  $q$  représentent également une estimation de la probabilité qu'un gamète mâle ou femelle porte l'allèle  $A$  ou l'allèle  $a$ .

Il est important de noter que les fréquences alléliques comportent moins d'information que les fréquences génotypiques car on perd la manière dont les allèles sont associés 2 à 2 dans les génotypes individuels.

Exemple du groupe sanguin MN chez l'homme L'examen de 730 aborigènes australiens a donné les résultats suivants :

Groupe sanguin	Génotype	Nombre	Fréquence
[M]	MM	22	0.03
[MN]	MN	216	0.30
[N]	NN	492	0.67

$$f(M) = \frac{2 \times 22 + 216}{2 \times 730} = 0,03 + \frac{1}{2} \times 0,3 = 0,18 = p$$

$$f(N) = \frac{2 \times 492 + 216}{2 \times 730} = 0,67 + \frac{1}{2} \times 0,3 = 0,82 = q$$

## 10.2. Taux de polymorphisme et taux d'hétérozygotie

Pour quantifier la variabilité d'une population étudiée sur plusieurs gènes, différents paramètres peuvent être calculés.

### a. Taux de polymorphisme P

C'est la proportion des gènes polymorphes parmi l'ensemble des gènes étudiés

$$P = \frac{\text{Nbre de gènes polymorphes}}{\text{Nbre total de gènes étudiés}}$$

Par exemple, si 30 loci enzymatiques ont été étudiés par la méthode d'électrophorèse avec 12 loci monomorphes et 18 polymorphes, le taux de polymorphisme P est  $18/30 = 0,6$ .

Ce paramètre présente cependant l'inconvénient de ne pas prendre en compte le nombre d'allèles rencontrés à chacun des loci polymorphes, ni leurs fréquences.

Il est évident par exemple qu'un locus possédant 10 allèles de fréquences voisines apporte plus de variation génétique à la population qu'un locus n'ayant que deux allèles dont un faiblement représenté.

### b. Le taux d'hétérozygotie H

C'est la moyenne des fréquences des hétérozygotes observées à chacun des locus étudiés :

$$H_o = 1/N \sum H_i$$

N étant le nombre total de loci étudiés qu'ils soient monomorphes ou polymorphes.

$H_i$  hétérozygotie au locus i

Le taux d'hétérozygotie fournit une bonne estimation de la variabilité génétique de la population, à condition toutefois que les individus de cette population se reproduisent au hasard.

Des modes de reproduction différents (homogamie, consanguinité, autogamie),  
Cours de Génétique et Dynamique des Populations Dr : bouamous A  
~~conduisent à des situations où  $H_o$  ne donne plus une bonne estimation de la variabilité~~  
génétique.

Les modes de reproduction n'étant pas toujours connus, on calcule également un autre paramètre qui est l'hétérozygotie théorique attendue ( $H_t$ ). Pour un locus A à k allèles  $A_1, A_2, \dots, A_k$  de fréquences  $f_1, f_2, \dots, f_k$ , l'hétérozygotie attendue est la suivante :

$$H_t = 1 - (f_1^2 + f_2^2 + \dots + f_k^2) = 1 - \sum f_i^2$$

C'est une estimation de la fréquence des hétérozygotes si les allèles sont associés au hasard pour former les génotypes.

L'hétérozygotie théorique globale est la moyenne des hétérozygoties attendues à chacun des loci étudiés:

$$H_t = 1/N \sum H_{ti}$$

N étant le nombre total de loci étudiés qu'ils soient monomorphes ou polymorphes  $H_{ti}$  hétérozygotie théorique au locus i

Il est alors possible de comparer la variabilité génétique des populations qui présentent des modes de reproduction différents. Diversité allélique Une autre mesure de la variabilité est la moyenne du nombre d'allèles par locus appelée diversité allélique :

$$A = \text{nbre total d'allèles} / \text{nbre de loci}$$

Pour prendre en compte la fréquence de ces allèles, on peut calculer le nombre d'allèles efficaces  $A_e$ . Pour un locus A à k allèles  $A_1, A_2, \dots, A_k$  de fréquences  $f_1, f_2, \dots, f_k$ , le nombre d'allèles efficaces est :

$$A_e = 1 / (f_1^2 + f_2^2 + \dots + f_k^2) = 1 / \sum f_i^2$$

On calcule alors la moyenne du nombre d'allèles efficaces par Locus