



ETUDE DE LA DIVERSITE GENETIQUE DES POPULATIONS A L'AIDE DU POLYMORPHISME ENZYMATIQUE

1. INTRODUCTION

Une **population** est souvent définie comme un ensemble de plantes de la même espèce, de même niveau de ploïdie, de même mode de reproduction et occupant le même habitat. Les populations très étendues géographiquement peuvent être découpées par l'échantillonnage en **sous-populations**. La définition de la population reste complexe. C'est souvent un **échantillon** dont l'expérimentateur définira selon la manière dont il est constitué.

Le polymorphisme génétique d'une population représente la réserve de variabilité nécessaire qui permet à la population de s'adapter à de nouvelles conditions du milieu. La variabilité génétique reste, donc, le garant de la colonisation d'habitats marginaux par les espèces polymorphes. L'un des facteurs du maintien du polymorphisme des populations réside dans l'effet de l'hétérosis qui peut être définie comme l'avantage déterminé par l'état hétérogène de nombreux loci. On parle de superdominance ou overdominance lorsque l'effet hétérosis est produit par l'état hétérozygote d'un seul locus.

L'évaluation des distances génétiques entre les populations nécessite, d'abord, l'étude de leurs polymorphismes, séparément. Ainsi, l'étude de la variabilité intrapopulation précède toujours l'étude de la variation interpopulation.

2. FREQUENCES GENOTYPIQUES ET FREQUENCES ALLELIQUES D'UNE POPULATION POUR UN LOCUS DONNE

Pour un locus donné, on détermine les différents génotypes présents dans une population. Ceci permet de définir les fréquences génotypiques. Si les individus sont diploïdes, la présence de 2 allèles (identiques ou différents) sera notée pour chaque locus. S'ils sont tétraploïdes, 4 allèles seront recensés par génotype. Ainsi, les fréquences alléliques sont déduites des fréquences génotypiques.

Exemple : pour un locus A, les fréquences génotypiques et alléliques d'une population à **N** individus diploïdes sont les suivantes :

Génotypes	Effectifs *	Fréquences génotypiques
A1A1	N11	P11 = N11/N
A1A2	N12	P12 = N12/N
A2A2	N22	P22 = N22/N

- $P1 + P2 = 1$

Allèles	Effectifs *	Fréquences alléliques
A1	2 N11 + N12	P1 = (2 N11 + N12)/2N
A2	2 N22 + N12	P2 = (2 N22 + N12)/2N

* P1 + P2 = 1

Les études isoenzymatiques (méthodes électrophorétiques) sont actuellement plus satisfaisantes pour évaluer les structures génétiques d'une population donnée.

3. EXPRESSION DE LA DIVERSITE GENETIQUE AU NIVEAU INTRAPOPULATION.

- TAUX DE POLYMORPHISME (P)

La proportion de locus polymorphes, encore appelée taux de polymorphisme ou plus simplement polymorphisme (**P**), rend compte de la proportion de protéines variables par rapport à l'ensemble des protéines étudiées. Une population sera dite polymorphe pour un locus donné, si la fréquence allélique de l'allèle le plus fréquent est inférieure à **0,95**. Ne pouvant traiter différemment les gènes très ou peu polymorphes et dépendant du critère utilisé (**95%**, **99%**), ce paramètre présente un intérêt limité.

Les méthodes d'évaluation des états alléliques et des génotypes ne concernent pas tous les loci. On n'étudie une population qu'à travers les loci pour lesquels les génotypes sont faciles à distinguer. Ainsi, les méthodes actuelles ne permettent de révéler les génotypes que d'une fraction d'environ **1%** des loci structuraux qui composent le génome d'une plante.

La signification biologique et adaptative du polymorphisme d'une population est encore un sujet de controverse scientifique. Il pourrait servir pour l'explication des facultés d'adaptation des populations à des milieux hétérogènes, dans le temps et dans l'espace. Le polymorphisme peut être entretenu par des transferts génétiques récurrents entre populations.

- NOMBRE MOYEN D'ALLELES PAR LOCUS (A)

Le nombre moyen d'allèles par locus (**A**), appelé également taux d'allélisme ou richesse allélique, est défini pour n_i allèles au locus i et pour L loci comme :

$$A = \frac{1}{L} \sum_{i=1}^L n_i$$

Exemple : pour 3 loci numéroté 1, 2 et 3 ayant respectivement 2, 3 et 2 allèles,
 $A = (2 + 3 + 2)/3 = 2,33$

Si les paramètres de diversité génétique intra- et interpopulations (diversité de Nei, distance génétique et paramètres de différenciation génétique) sont estimées sur la base des différences entre les fréquences alléliques, le taux d'allélisme tient compte du nombre d'allèles par locus. Par conséquent, la mesure de ce paramètre est particulièrement importante pour les stratégies de conservation. Il était souvent utilisé dans la gestion des collectes et des banques de semences (Asins et Carbonell, 1987; Bataillon et al., 1996).

- TAUX D'HETEROZYGOTIE (h_e et h_o)

L'hétérozygotie (diversité génétique de Nei (1973), h_o (observed heterozygosity) peut être calculée à partir de la fréquence mesurée des hétérozygotes (nombre des individus hétérozygotes divisé par le nombre total des individus de l'échantillon). De même, dans une population panmictique la fréquence théorique des hétérozygotes h_e (expected heterozygosity) à un locus peut être calculée à partir des fréquences alléliques. S'il y a n allèles avec les fréquences $f_1, f_2, f_3, \dots, f_n$, la fréquence théorique des hétérozygotes sera :

$$h_e = 1 - (f_1^2 + f_2^2 + f_3^2 + \dots + f_n^2) = 1 - \sum_{n=1}^n f_n^2$$

Si plusieurs loci sont considérés, l'hétérozygotie moyenne (H_e), représentant la moyenne du taux d'individus hétérozygotes par population, sera la moyenne arithmétique de toutes les valeurs de h_e :

$$H_e = (\sum h_e) / L, \text{ avec } L = \text{nombre de loci.}$$

- INDICE DE FIXATION (F_{IS})

Le paramètre F_{IS} de Wright, dénommé aussi indice de fixation et appelé auparavant coefficient de consanguinité (Wright, 1969) est calculé selon la formule :

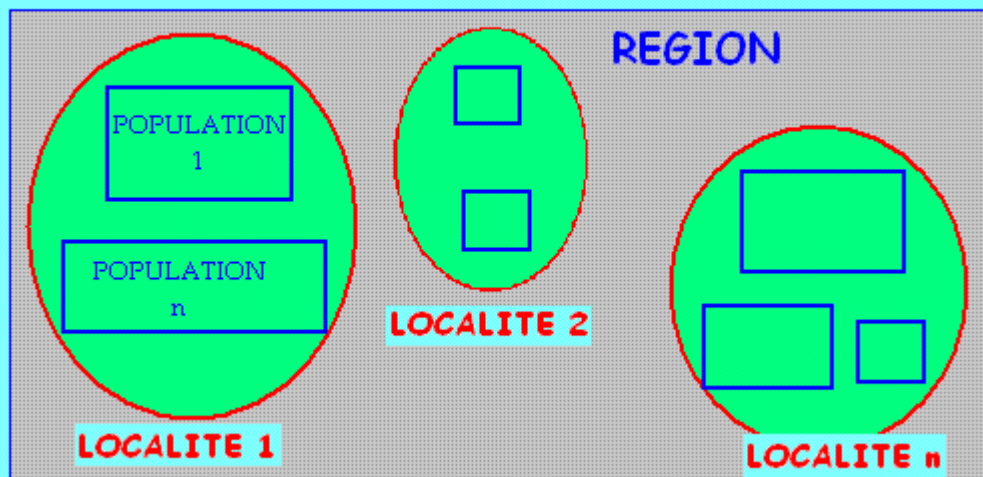
$$F_{IS} = (h_e - h_o) / h_e = 1 - (h_o/h_e)$$

avec h_o , l'hétérozygotie observée et h_e l'hétérozygotie attendue, calculée à partir des fréquences alléliques dans l'hypothèse de Hardy-Weinberg.

Il reflète la différenciation des individus à l'intérieur des populations ($F_{IS} = 1$ signifie fixation complète (cas d'autofécondation), F_{IS} inférieur à 1 : hétérozygotie excédentaire, $F_{IS} = 0$: population en équilibre de Hardy-Weinberg. $F_{IS} < 0$: hétérozygotie excédentaire,

4. EXPRESSION DE LA DIVERSITE GENETIQUE AU NIVEAU INTERPOPULATION ET DISTANCE GENETIQUE.

REPARTITION DE LA DIVERSITE GENETIQUE



- PARAMETRES DE DIFFERENCIATION.

Paramètres F_{ST} , G_{ST}

Wright (1965-1978) a défini l'indice F_{ST} (variance standardisée) comme l'hétérogénéité des fréquences alléliques entre subdivisions d'une population. Il représente la corrélation entre allèles à l'intérieur d'une sous-population par rapport à l'ensemble des sous populations. Ce paramètre est utilisé d'une façon hiérarchique. Soit T , un ensemble formé de S populations dont chacune est composée de I individus. La différenciation des populations par rapport au total (F_{ST}) est calculée en fonction des paramètres F_{IS} (différenciation des individus à l'intérieur des populations) et F_{IT} (différenciation des individus par rapport au total). Ils sont liés par la relation :

$$F_{ST} = 1 - \frac{(1 - F_{IT})}{(1 - F_{IS})}$$

D'après Wright, 1978 : $0 < F_{ST} < 0,05$: différenciation faible
 $0,05 < F_{ST} < 0,15$: différenciation modérée
 $0,15 < F_{ST} < 0,25$: différenciation importante
 $F_{ST} > 0,25$: différenciation très importante

S'il existe une ou plusieurs subdivisions hiérarchiques, elles seront prises en compte dans l'expression mathématique de la différenciation. Ainsi, par exemple, une hiérarchie selon la localité permet de réécrire les formules de différenciation comme suivant :

$$(1 - F_{IT}) = (1 - F_{IS}) (1 - F_{SL}) (1 - F_{LT}).$$
$$(1 - F_{ST}) = (1 - F_{SL}) (1 - F_{LT}).$$

avec : F_{SL} , la différenciation des populations à l'intérieur de la localité et F_{LT} , la différenciation des localités par rapport au total.

Le paramètre F_{ST} est souvent remplacé par un paramètre analogue, le G_{ST} , défini par la formule :

$$G_{ST} = (H_T - H_S) / H_T = 1 - (H_S / H_T).$$

H_S étant la moyenne (sur toutes les populations) des diversités génétiques intrapopulation et H_T , la diversité génétique sur l'ensemble des populations considérées comme une seule population (diversité totale). Lorsqu'il s'agit de plusieurs loci, H_S et H_T deviennent les moyennes (sur l'ensemble des loci) des diversités précédentes. Dans la littérature, $(H_T - H_S)$ est, parfois, remplacée par D_{ST} (diversité entre les populations). Ainsi, on aura :

$$G_{ST} = D_{ST} / H_T.$$

Donc, la diversité totale peut être subdivisée en une composante intrapopulation (H_S) et interpopulation (D_{ST}) : $H_T = H_S + D_{ST}$. Ainsi, G_{ST} représente la part (%) de la diversité interpopulation par rapport à la diversité totale.

Si les populations peuvent être regroupées (en localités, par exemple) la D_{ST} pourra être, de nouveau, subdivisée en composantes intralocalité (D_{SL}) et interlocalité (D_{LT}) ($D_{ST} = D_{SL} + D_{LT}$). On aura : $H_T = H_S + D_{SL} + D_{LT}$.

Chacune des composantes peut être exprimée comme un rapport à la diversité totale :

$$G_S = H_S/H_T, G_{SL} = D_{SL}/H_T, G_{LT} = D_{LT}/H_T \text{ avec, } G_{ST} = G_{SL} + G_{LT}$$

Par ailleurs, Il est possible de quantifier le flux de gènes entre populations, dans un modèle théorique dit 'en île', par le calcul du nombre de migrants par génération (N_m) qui peut être mesuré indirectement à l'aide du G_{ST} . L'expression de N_m devient alors selon Slatkin et Barton (1989) :

$$N_m = \frac{(1 - G_{ST})}{4 G_{ST} \left(\frac{n}{n-1} \right)^2}$$

où n indique le nombre de populations.

Plus la valeur de N_m est supérieure à 1, plus l'échange de gènes est important.

Distance génétique (D).

En se basant sur les fréquences alléliques, les degrés de ressemblance et de dissemblance entre les populations peuvent être évalués. L'indice d'identité de Nei (1972) est d'abord calculé. Soit un locus donné où x_i est la fréquence du $i^{\text{ème}}$ allèle dans la population X et y_i celle du même allèle au même locus dans la population Y. La probabilité d'identité des 2 allèles pris au hasard au sein de la population X est : $P_x = \sum x_i^2$. De même pour la population Y, cette probabilité est : $P_y = \sum y_i^2$. La probabilité d'identité des 2 allèles pris au hasard, l'un dans X et l'autre dans Y, est : $P_{xy} = \sum x_i y_i$. L'indice d'identité de Nei entre les 2 populations (sur la base du locus considéré) est :

$$I = P_{xy} / (P_x P_y)^{1/2}$$

Pour plusieurs loci on calcule la moyenne des probabilités sur l'ensemble des loci étudiés. La distance génétique est :

$$D = - \text{Log } I$$

Si les deux populations possèdent les mêmes allèles à des fréquences identiques on aura :

La détermination des distances génétiques entre populations permet d'évaluer le degré de ressemblance de leurs structures génétiques. Elle peut servir, également, pour montrer si des groupes de plantes prétendus appartenir à des espèces différentes, font partie ou non d'un même complexe d'espèces.

Interprétations de D. Exemples.

1/ Les distances génétiques calculées sur les polymorphismes enzymatiques des populations naturelles variées d'organismes autres que les Mammifères supérieures (primates), sont inférieures à **0,02** (différence d'AA par protéine moyenne) pour des populations appartenant à la même espèce. Les distances entre espèces bien isolées reproductivement, mais du même genre, sont de l'ordre de **0,10 à 1,00** (même espèce).

Si les distances génétiques testées sur des systèmes enzymatiques non compromis par la différenciation liée à la domestication, sont faibles ($D < 0,10$), on pourra suggérer qu'il n'y a pas de barrière reproductive absolue entre les deux populations étudiées.

2/ Une distance génétique entre 2 populations P et Q égale à **0,20** ($D_{PQ} = 0,2$) signifie qu'en moyenne **20%** des protéines d'un individu de la population P diffère par un autre AA des protéines homologues d'un individu de la population Q. Si P et Q étaient complètement isolées depuis leur séparation, on estime que cette séparation a eu lieu **$0,20 \times 10^6 = 200000$ ans** auparavant.

La distance génétique entre deux populations non complètement isolées n'augmente pas indéfiniment au cours du temps.

Selon Nei 1971, le temps de divergence (t) est donné par la formule : **$t = 7,4 \cdot 10^5 \cdot D$**

Selon Nei, 1975, **$t = 7,4 \cdot 10^6 \cdot D$**

3/ Les groupes de gènes dont on peut penser qu'ils sont directement liés au processus d'adaptation, sont à écarter dans les études de distances génétiques, car ils seraient suspectés d'accentuer les distances.

Dendrogramme

Un dendrogramme peut être construit sur la **base des distances génétiques** ou des bandes des profils électrophorétiques des isoenzymes. Il représente donc une classification qui a pour but d'obtenir une représentation schématique simple d'un tableau des données dont les colonnes (variables) caractérisent l'ensemble des lignes (populations ou espèces). Il permet de visualiser les regroupements possibles de ces dernières, et en répartissant des entités en groupes (classes) homogènes, chaque groupe étant bien différencié des autres.

Tous les paramètres de la mesure de la diversité génétique intrapopulation et interpopulation décrits précédemment, peuvent être déterminés à l'aide du logiciel **Biosys-1** (Swofford & Selander, 1981).

5. STRUCTURATION DE LA DIVERSITE ET CONSERVATION DES RESSOURCES GENETIQUES.

Structuration de la diversité génétique.

La décomposition de la diversité totale en 2 composantes liées à la diversité de chaque population et sa différenciation par rapport au reste constitue une approche méthodologique adaptée à la conservation. La particularité de chaque population peut, ainsi, être mise en évidence. Une population ayant une diversité génétique plus élevée n'est pas nécessairement la plus différenciée des autres populations ($H_T = H_S + D_{ST}$).

Dans la littérature, c'est surtout le G_{ST} qui est utilisé pour mesurer le degré de différenciation entre les populations.

Les espèces à large distribution, ayant un flux génétique très élevé, montrent des G_{ST} très faibles, pouvant chuter à moins de 4%. C'est le cas de plusieurs espèces forestières comme les chênes (*Quercus robur* (Zanetto et al., 1994) et *Quercus ilex* (Michaud et al., 1995)).

A l'inverse, de fortes différenciations entre populations ont été trouvées chez des espèces à aire de répartition morcelée ou à distribution discontinue, comme *Pinus torreyana* ($G_{ST} = 1,00$) (Ledig & Conkle, 1983), *Pinus halepensis* ($G_{ST} = 0,30$) (Schiller et al., 1986) et *Pinus brutia* ($G_{ST} = 0,16$) (Conkle et al., 1988).

Une étude synthétique portant sur 322 espèces ligneuses a été faite par Hamrick et al. (1992) qui ont comparé les résultats des diversités intrapopulation (y compris les locus monomorphes) et interpopulation. Il s'est avéré que la distribution géographique de l'espèce est un bon prédicteur de la variation génétique entre les populations (aires endémique,

régionale ou continentale). De plus, les espèces à large distribution présentent un pourcentage de loci polymorphes et une diversité intrapopulation plus élevés.