



SCIENCES SUP

Cours et exercices corrigés

Licence • PCEM • CAPES

GÉNÉTIQUE DES POPULATIONS

Jean-Louis Serre

DUNOD

1.3 MESURE DE LA DIVERSITÉ GÉNÉTIQUE ET COMPOSITION GÉNÉTIQUE D'UNE POPULATION

1.3.1 La population

L'espèce est par définition un groupe génétiquement fermé au sein duquel les organismes sont susceptibles, par l'alternance méiose-fécondation de séparer ou de réunir les divers allèles de chacun des gènes et de concevoir des combinaisons génétiques nouvelles par la recombinaison génétique.

Cependant tous les individus d'une même espèce, s'ils sont potentiellement susceptibles de réaliser ce brassage peuvent en être pratiquement empêchés quand des barrières limitent les possibilités de croisements entre certains individus.

Il peut s'agir d'un isolement géographique, lié à l'existence d'une barrière naturelle comme un océan pratiquement infranchissable ou une chaîne montagneuse plus facilement franchissable, ou plus simplement par la distance qui limite la probabilité de croisements entre individus très éloignés.

Il peut s'agir d'un isolement écologique. Par exemple des plantes d'une même espèce occupant un même territoire semblent former une même population. Mais si une disparité dans la composition du sol décale la floraison entre les plantes du sol A et celles du sol B, les échanges génétiques entre les plantes des sols A et B seront limités. Il sera alors nécessaire de définir deux populations A et B. La définition et l'analyse de populations naturelles supposent donc une bonne connaissance de leur biologie et de leur biotope.

Il peut s'agir enfin, chez l'homme, de barrières culturelles, sociales ou ethniques, qui limitent plus ou moins les possibilités d'unions entre individus, même géographiquement proches.

Une espèce peut donc être subdivisée en sous-groupes au sein desquels la possibilité d'échanges génétiques entre individus est effective ; ces sous-groupes sont appelés populations et l'ensemble des allèles qu'ils partagent, pour chacun des gènes de l'espèce, en constitue le patrimoine génétique (pool allélique).

La mesure de la diversité génétique à l'intérieur des populations mais aussi entre les populations, l'origine et le devenir de ces diversités intra- et inter-populationnelles sont un enjeu important de la génétique des populations, notamment chez l'homme, en raison des polémiques qui ont accompagné la définition et l'usage du concept de race, ou les débats sur l'émergence de l'homme moderne à partir de l'*Homo erectus* (voir plus loin).

1.3.2 Variables d'état de la diversité et composition génétique d'une population

Il est possible de définir une chaîne de causalité liant la variabilité phénotypique des caractères et la diversité génétique sous-jacente qui en est la cause (figure 1.6, à gauche). À chacun des niveaux hiérarchiques de la diversité, on peut associer des variables d'état qui mesurent la diversité génétique à ce niveau, les fréquences alléliques, les fréquences génotypiques, les fréquences phénotypiques (figure 1.6, à droite).

Les fréquences phénotypiques sont toujours accessibles directement par le dénombrement des phénotypes présents dans un échantillon, et la question se pose de savoir si il existe des relations mathématiques simples permettant, si on connaît la diversité à un niveau hiérarchique, d'en déduire la diversité à un autre niveau (figure 1.6, flèches doubles à droite). Si de telles relations sont disponibles alors la connaissance de la diversité en un point quelconque des niveaux hiérarchiques permettrait d'avoir une connaissance exhaustive de la diversité génétique de la population en tout autre point.

Or c'est bien le premier but de la génétique des populations que de savoir mesurer la diversité pour définir la composition génétique d'une population ou d'une espèce, au niveau des allèles et des génotypes.

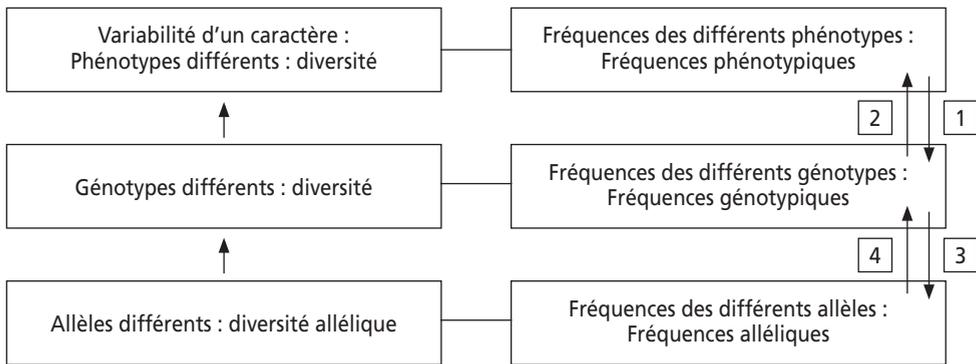


Figure 1.6 Niveaux hiérarchiques de la diversité génétique et variables d'états associés.

Dans un premier temps, on va définir la mesure de la diversité génétique relative à un gène (ou un marqueur) sachant qu'une fraction des problèmes de génétique des populations correspond à une telle situation. Ultérieurement, on généralisera nos résultats à l'étude simultanée de deux gènes ou marqueurs, situation très communément rencontrée en épidémiologie génétique, mais, comme cela a été dit, l'approche analytique devient impossible au-delà et il est nécessaire de recourir à des simulations informatiques.

1.3.3 Codominance et dominance : les limites de la mesure de la diversité génétique

Les phénotypes sont par définition directement accessibles à l'observation et les fréquences phénotypiques peuvent être calculées à partir d'un échantillon d'individus tirés aléatoirement dans la population, mais le calcul des fréquences génotypiques et alléliques n'est possible que si les phénotypes sont codominants, comme le montrent les exemples suivants.

a) Phénotypes codominants : groupe sanguin MN

Le typage est réalisé, comme pour le groupe ABO ou rhésus, par un test d'héماغلutation, les individus appartenant aux groupes [M], [N] ou [MN] selon que leurs

hématies sont respectivement reconnues par l'anticorps anti-M ou l'anticorps anti-N ou les deux (tableau 1.1), parce qu'elles sont porteuses d'une chaîne peptidique spécifiée par l'allèle L^M et/ou l'allèle L^N .

TABLEAU 1.1 ÉCHANTILLON ALÉATOIRE D'UNE POPULATION EUROPÉENNE

Groupes sanguins	[M]	[MN]	[N]
Effectifs observés (total : 1 000)	350	500	150
Fréquences phénotypiques	$350/1\ 000 = 0,35$	$500/1\ 000 = 0,50$	$150/1\ 000 = 0,15$
Génotype	L^M/L^M	L^M/L^N	L^N/L^N

Ici, les phénotypes étudiés sont codominants, il est possible d'associer un phénotype à un seul génotype et réciproquement. Dès lors les fréquences génotypiques sont égales aux fréquences phénotypiques et les fréquences alléliques s'en déduisent aisément. La constitution génétique de la population pour le gène concerné est connue sans ambiguïté.

Deux méthodes de calcul des fréquences alléliques sont possibles, un simple comptage des allèles ou une formule probabiliste.

Méthode des comptages. Elle consiste à dire que les 1 000 individus sont porteurs de 2 000 allèles parmi lesquels les allèles M représentent 350×2 (pour les homozygotes qui possèdent 2) plus 500 (pour les hétérozygotes qui n'en possèdent qu'un). Les fréquences des allèles M et N sont donc respectivement :

$$f(L^M) = (350 \times 2 + 500)/2\ 000 = 1\ 200/2\ 000 = 0,6$$

et $f(L^N) = (150 \times 2 + 500)/2\ 000 = 800/2\ 000 = 0,4$

avec $f(L^M) + f(L^N) = 1$

Méthode probabiliste. On peut calculer les fréquences alléliques par l'application du théorème des probabilités composées.

Considérons plus généralement les trois génotypes possibles résultant des combinaisons diploïdes des deux allèles $A1$ et $A2$ d'un gène et leurs fréquences respectives D , H et R :

– Génotypes :	$A1/A1$	$A1/A2$	$A2/A2$
– Fréquences génotypiques :	D	H	R

La fréquence de l'allèle $A1$ peut être définie comme la probabilité de tirer cet allèle au hasard dans la population, ce qui suppose d'abord de tirer un individu, puis l'un de ses deux allèles :

- l'individu tiré peut être $A1/A1$, avec la probabilité D ; dans ce cas l'allèle tiré au hasard chez cet individu sera $A1$ avec la probabilité 1,
- ou l'individu tiré peut être $A1/A2$, avec la probabilité H ; dans ce cas l'allèle tiré au hasard chez cet individu sera $A1$ avec la probabilité $1/2$, car l'individu est aussi porteur de $A2$.
- ou l'individu tiré peut être $A2/A2$, avec la probabilité R ; dans ce cas l'allèle tiré au hasard chez cet individu sera $A1$ avec la probabilité 0, car il n'en possède pas.

On aura donc $f(A1) = D \times 1 + H \times 1/2 + R \times 0$

D'où la formule générale : $f(A1) = D + H/2$

et $f(A2) = R + H/2$

Dans l'exemple du groupe MN, on retrouve évidemment les mêmes valeurs que celles obtenues par l'autre méthode.

Remarque : si on revient à la figure 1.6, on voit que les relations 1 et 2 sont des relations d'identité et que la relation 3 est le système d'équations qui vient d'être défini, à savoir que la fréquence d'un allèle est égale à la fréquence des homozygotes pour cet allèle plus la moitié de la fréquence des hétérozygotes. Mais la relation 4 est, pour l'instant indéterminée, car avec les fréquences alléliques, il n'est pas possible d'en déduire les fréquences génotypiques.

b) Phénotypes dominants et récessifs : groupe sanguin ABO

Il est toujours possible d'estimer les fréquences phénotypiques (tableau 1.2) mais il est ici impossible d'en déduire les fréquences génotypiques, du moins pour les phénotypes dominants présentant plusieurs possibilités génotypiques sous jacentes ; dans ce cas la fréquence du phénotype, par exemple celle de [A], est égale à la somme des fréquences des deux génotypes, I^A/I^A et I^A/I^O , mais on ne peut connaître la valeur individuelle de la fréquence de chaque génotype, indispensable pour estimer les fréquences alléliques selon les formules vues plus haut. On ne peut mesurer la diversité génétique et, en se rapportant à la figure 1.6, on reste bloqué dans l'ensemble des phénotypes car aucune des quatre relations mathématiques n'est déterminée à ce stade.

TABLEAU 1.2 GROUPES SANGUINS ABO DANS UN ÉCHANTILLON ALÉATOIRE DE LA POPULATION FRANÇAISE.

Groupes sanguins	[A]	[B]	[AB]	[O]
Effectifs observés (total : 500)	213	56	15	216
Fréquences phénotypiques	$213/500 = 0,426$	$56/500 = 0,112$	$15/500 = 0,03$	$216/500 = 0,432$
Génotypes	I^A/I^A ou I^A/I^O	I^B/I^B ou I^B/I^O	I^A/I^B	I^O/I^O

Dans certains cas, une information complémentaire permet de lever l'ambiguïté de la relation phénotype/génotype ; elle est apportée par la connaissance de l'ascendance ou de la descendance des individus étudiés, ou bien, comme dans la drépanocytose, par une étude biochimique donnant accès, pour la mutation étudiée, à des phénotypes codominants (voir exercice).

Cette situation de blocage est commune à tous les caractères présentant des phénotypes mutés récessifs, notamment les maladies récessives, ce qui implique qu'on ne peut pas, pour l'instant, estimer la fréquence d'un allèle pathologique ni celle des porteurs sains.

2.3 LE MODÈLE DE HARDY-WEINBERG

Le modèle théorique général de Hardy-Weinberg sera établi dans le cas le plus simple, le plus général et le plus utile d'un gène autosomique di-allélique, pour une population d'organismes à sexes séparés, présentant des générations séparées.

La généralisation du modèle à un cas multi-allélique, le cas particulier des gènes liés au sexe, le cas des générations chevauchantes ou le cas de deux gènes étudiés simultanément seront exposés au chapitre 3.

Considérons une population constituée des adultes reproducteurs de la génération i .

Dans le cas d'un gène di-allélique (allèles $A1$ et $A2$), la composition génétique de la population est constituée des :

- **trois génotypes possibles :** $A1/A1$ $A1/A2$ $A2/A2$
- **de fréquences génotypiques :** D H R

Ces fréquences D , H et R sont quelconques et seront considérées, en première analyse, égales dans les deux sexes.

On sait (voir chapitre 1) en déduire les fréquences des allèles $A1$ et $A2$, respectivement nommées p et q , soit :

$$p = D + H/2$$

$$q = R + H/2$$

Quelle sera la constitution génétique (fréquences génotypiques et fréquences alléliques), à la génération suivante après un cycle vital ?

Il suffit de se reporter au cycle vital présenté dans le paragraphe précédent pour obtenir la solution en fonction des conditions que nous aurons dû définir et qui forment les « conditions de l'équilibre de Hardy-Weinberg ».

On peut établir ce modèle de deux manières différentes, soit par la formation des couples en suivant pas à pas le cycle vital, soit par le schéma de l'urne gamétique.

2.3.1 Établissement du modèle de Hardy-Weinberg par le cycle vital

a) Formation des couples : condition de panmixie

La première étape du cycle vital (1 dans l'encadré ci-dessus) est la formation de couples reproducteurs. Des règles d'union peuvent exister : on fera l'hypothèse qu'ils se forment au hasard, les couples sont dits **panmictiques (condition de panmixie)**.

Dans ce cas, on peut générer six types possibles de couples (tableau 2.1).

Remarque : il convient de noter que la panmixie ne signifie pas que les six types de couples sont équiprobables ou équi-fréquents (1/6) mais que leurs probabilités respectives sont fonction de la fréquence, dans la population, de chacun des génotypes associés dans le couple. Pour prendre un exemple caricatural les couples noirs \times noirs ne peuvent pas avoir la même probabilité ou la même fréquence dans un pays comme la Suède ou le Sénégal.

b) *Probabilité et fréquences des évènements : condition d'effectif infini de la population*

Dans une population naturelle concrète, ce qui nous importe, et ce qui compte, ce sont les fréquences des couples, les fréquences de leurs descendants, les fréquences des génotypes, les fréquences alléliques, et non les probabilités de ces évènements.

On sait que la fréquence d'un évènement est égale à sa probabilité si le nombre de tirages est très grand (loi des grands nombres) ; par exemple la fréquence des « piles » peut être égale à 0,7 sur dix tirages mais ne peut, sur 100 000 tirages, s'écarter notablement de sa probabilité égale à 0,5.

Afin de pouvoir considérer que les fréquences des couples ou des génotypes sont égales à leurs probabilités respectives, nous considérerons que la population est de taille infinie (concrètement suffisamment grande pour y appliquer la loi des grands nombres aux évènements étudiés). Le seuil grand/petit n'est pas définissable dans l'absolu et ne sera discuté qu'à la fin de l'ouvrage (voir chapitre 8).

Cette condition d'effectif infini s'ajoute à la condition de panmixie. Le tableau de formation des couples et de leurs descendants se présente alors ainsi :

TABLEAU 2.1 FRÉQUENCES DES COUPLES PANMICTIQUES ET DE LEURS DESCENDANTS POUR UN GÈNE DI-ALLÉLIQUE.

Types de couples	Fréquences des couples	Fréquences des descendants A1/A1	Fréquences des descendants A1/A2	Fréquences des descendants A2/A2
A1/A1 x A1/A1	D^2	1	0	0
A1/A1 x A1/A2	$2DH$	1/2	1/2	0
A1/A1 x A2/A2	$2DR$	0	1	0
A1/A2 x A1/A2	H^2	1/4	1/2	1/4
A1/A2 x A2/A2	$2RH$	0	1/2	1/2
A2/A2 x A2/A2	R^2	0	0	1
TOTAL	1	$D^2 + DH + H^2/4$	$DH + 2DR + H^2/2 + RH$	$R^2 + RH + H^2/4$

Maintenant que les couples sont formés et leurs fréquences connues (sous les conditions de panmixie et d'effectif infini), il s'agit de réaliser les fécondations pour obtenir les adultes reproducteurs de la génération suivante de manière à avoir réalisé le cycle vital d'une génération.

c) *Gamétogenèse : condition d'absence de mutations*

Les individus formant les couples produisent des gamètes (étape 2 dans le cycle vital). Pour simplifier, on négligera l'effet des mutations dans le gène considéré. De ce fait on peut considérer comme nulle la fréquence des descendants A1/A2 chez les parents du premier type (première ligne du tableau ci-dessus), et écrire les proportions des lignes 2, 4, 7 du tableau.

d) *Fécondation : condition d'absence de sélection gamétique*

Les deux types de gamètes formés par les hétérozygotes sont théoriquement dans les proportions 50 :50, mais lors de la fécondation (**4** dans le cycle vital), ces proportions peuvent avoir été modifiées si un mécanisme de sélection gamétique (**3** dans le cycle vital) a joué en faveur de l'un et en défaveur de l'autre. En excluant cette éventualité (condition de l'absence de sélection gamétique), on peut écrire les proportions mendéliennes réalisées à la fécondation dans les lignes 3, 5, 6 du tableau.

e) *Développement et croissance des descendants : condition d'absence de sélection zygotique*

Les proportions des génotypes chez les descendants, à la fécondation, ne sont maintenues, au stade final de reproducteur adulte, que si une nouvelle condition est réalisée : l'absence de sélection zygotique (pendant l'embryogenèse ou la croissance ; **5** dans le cycle vital).

Autrement dit les trois génotypes sont supposés avoir la même espérance de vie et ne présentent, du moins pour le gène considéré, aucune mortalité différentielle.

f) *Fréquences des génotypes chez les adultes reproducteurs de la génération suivante : condition d'absence de sélection et de migration*

On peut calculer les fréquences génotypiques des descendants au stade d'adultes reproducteurs (sous toutes les conditions précédentes) en sommant, pour chaque génotype, le produit de la fréquence des couples dont il peut être issu par la fréquence de réalisation de ce génotype au sein de ces couples.

On obtient ainsi les trois sommes figurant en bas du tableau, mais cela suppose encore deux dernières conditions pour qu'on puisse les considérer comme les fréquences génotypiques des adultes reproducteurs de la génération $i + 1$:

- il faut d'abord considérer que les couples sont également fertiles (condition d'absence de sélection en terme de fertilité différentielle). Dans le cas contraire il faudrait pondérer chacun des termes de la somme par la fertilité respective des couples ;
- il faut ensuite considérer que ces sommes, ces fréquences génotypiques, n'ont pas été modifiées dans le temps par l'adjonction à la population d'individus extérieurs (condition d'absence de migrations). En effet un apport extérieur a peu de chance de présenter la même constitution génétique.

Sous toutes ces hypothèses, les nouvelles fréquences génotypiques deviennent :

pour les trois génotypes :	$A1/A1$	$A1/A2$	$A2/A2$
égales à :	p^2	$2pq$	q^2

En effet si $p = D + H/2$ alors $D^2 + DH + H^2/4 = p^2$

et si $q = R + H/2$ alors $R^2 + RH + R^2/4 = q^2$

2.3.2 Établissement du modèle de Hardy-Weinberg par le schéma de l'urne gamétique

a) *Panmixie et pangamie : schéma de l'urne gamétique*

Si on considère que les couples se forment au hasard (étape **1** du cycle vital), c'est-à-dire qu'ils sont panmictiques (condition de panmixie), et que les gamètes eux-mêmes s'unissent au hasard (condition de pangamie) lors de la fécondation (**4** dans l'encadré du cycle vital), on peut alors admettre que tout se passe comme si les couples mettaient en vrac leurs gamètes dans une urne gamétique de spermatozoïdes pour les mâles et d'ovules pour les femelles, et que tout descendant est issu de l'union de deux gamètes tirés au hasard dans chacune des deux urnes. Ce schéma est nommé schéma de l'urne gamétique (figure 2.2). C'est d'ailleurs la réalité biologique pour les espèces végétales, et de nombreuses espèces animales aquatiques, qui émettent leurs gamètes dans le milieu environnant.

Remarque : sauf exception la panmixie s'accompagne toujours de la pangamie et on considère que la condition de panmixie inclue celle de la pangamie.

Si les fréquences alléliques sont les mêmes dans les deux sexes, les urnes sont identiques (on verra plus tard le cas où elles ne le sont pas).

Compte tenu de la composition des urnes il sera alors possible de calculer la probabilité de tirage de chacun des trois génotypes possibles.

D'où le schéma suivant :

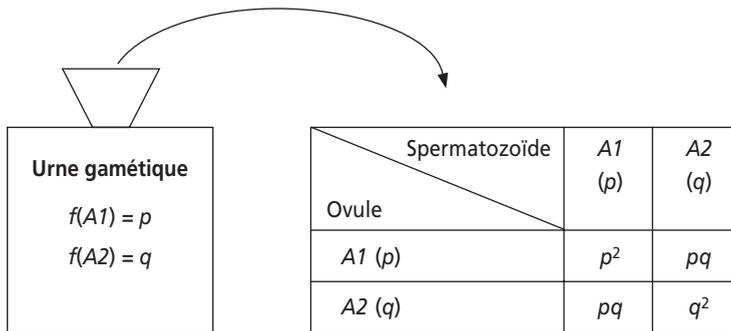


Figure 2.2 Schéma de l'urne gamétique et conséquences mathématiques sur la diversité génétique.

Il faut cependant considérer que les valeurs p^2 , $2pq$, q^2 , obtenues si facilement par ce double tirage, ne sont une réalité biologique, c'est-à-dire les fréquences de génotypes adultes reproducteurs à la génération suivante, que si des conditions additionnelles sont réalisées.

b) *Conditions additionnelles*

La composition génétique de l'urne selon le schéma précédent suppose que chacun des adultes reproducteurs y a placé un nombre égal de gamètes, ce qui signifie qu'il y a **absence de sélection en matière de fertilité**.

Il suppose aussi que les homozygotes A_1A_1 n'y déposent que des gamètes A_1 , ce qui signifie l'**absence de mutations**.

Une fois obtenue par le dépôt des gamètes, la composition de l'urne est stable, ce qui signifie l'**absence de sélection gamétique**.

Le tirage de deux gamètes aboutit alors et seulement alors à la formation d'un génotype :

	A_1/A_1	avec	la probabilité p^2 ;
ou	A_1/A_2	avec	la probabilité $2pq$;
ou	A_2/A_2	avec	la probabilité q^2 .

En fait nous nous intéresserons aux fréquences des génotypes des descendants et non aux seules probabilités de les concevoir. On sait que la fréquence d'un événement est égale à sa probabilité si le nombre de tirages est très grand (loi des grands nombres) ; par exemple la fréquence des « piles » peut être égale à 0,7 sur dix tirages mais ne peut, sur 100 000 tirages, s'écarter notablement de sa probabilité égale à 0,5.

Afin de pouvoir considérer que les fréquences des génotypes obtenus sont égales à leurs probabilités respectives, nous considérerons que la population est de taille infinie (concrètement suffisamment grande pour y appliquer la loi des grands nombres aux événements étudiés). Le seuil grand/petit n'est pas définissable dans l'absolu et ne sera discuté qu'à la fin de l'ouvrage (voir chapitre 8).

Cette condition de taille infinie s'ajoute aux conditions de panmixie, d'absence de mutations, de sélection gamétique et d'égale fertilité des parents.

Les proportions $p^2/2pq/q^2$ ne sont maintenues de la fécondation au stade final d'adultes reproducteurs que si deux autres conditions sont également réalisées :

- l'**absence de sélection zygotique** (pendant l'embryogenèse ou la croissance). Autrement dit les trois génotypes sont supposés avoir la même espérance de vie et ne présentent, du moins pour le gène considéré, aucune mortalité différentielle ;
- l'**absence de migrations**. En effet un apport extérieur a peu de chance de présenter la même constitution génétique ($p^2/2pq/q^2$).

Au bout du compte on aboutit évidemment au même résultat que celui obtenu par la première méthode.

2.3.3 Bilan du modèle de Hardy-Weinberg

a) La relation de Hardy-Weinberg

Les nouvelles fréquences génotypiques correspondent soit au carré des fréquences alléliques pour les homozygotes, soit au double produit des fréquences alléliques pour l'hétérozygote.

Les trois génotypes sont	A_1/A_1	A_1/A_2	A_2/A_2
Leurs fréquences sont égales à	p^2	$2pq$	q^2

La relation ainsi établie entre les fréquences alléliques et les fréquences génotypiques est appelée « relation de Hardy-Weinberg » ou « relation panmictique » car elle découle directement de l'hypothèse panmictique.

Remarque : cette relation mathématique permet, en supposant que l'hypothèse panmictique soit valide, de remonter aux fréquences génotypiques (donc phénotypiques) quand on ne connaît que les fréquences alléliques (voir figure 1.6 et paragraphe 2.3.5.c).

b) L'équilibre de Hardy-Weinberg

Les fréquences alléliques sont inchangées à la génération suivante.

En effet, selon la formule de calcul des fréquences alléliques à partir des fréquences génotypiques (voir chapitre précédent), on a :

$$f(A1) = p^2 + 2pq/2 = p^2 + pq = p(p + q) = p$$

$$f(A2) = q^2 + 2pq/2 = q^2 + pq = q(p + q) = q$$

Cette stabilité de la composition génétique de la population est appelée « équilibre de Hardy-Weinberg ».

c) Les conditions de l'équilibre de Hardy-Weinberg

Les conditions supposées réalisées dans le modèle de l'équilibre de Hardy-Weinberg peuvent se regrouper en trois grands groupes :

- condition 1 :** la population est panmictique ;
- condition 2 :** la population est de taille quasi infinie (loi des grands nombres applicable) ;
- condition 3 :** mutation, sélection, migration sont inexistantes (ou négligeables).

2.3.4 Légitimité des conditions du modèle de Hardy-Weinberg

Le modèle de Hardy-Weinberg est le modèle central de la génétique des populations. Pourtant le nombre et l'importance des conditions sous-jacentes devraient le faire apparaître comme un modèle théorique, abstrait et irréaliste. On sait bien que les mutations existent et sont la source de la variation génétique ayant permis la sélection et l'évolution. Pourtant l'étude de la plupart des gènes dans les populations naturelles donne des résultats compatibles avec ce modèle. Comment expliquer ce paradoxe ?

Tout simplement par le fait que certaines des conditions peuvent parfaitement être réalisées et que les autres, bien qu'illégitimes, n'ont d'effet perceptible que sur une longue échelle de temps. On peut donc les négliger sur l'espace de quelques générations.

Reprenons ces conditions :

a) la panmixie est la condition la plus facilement réalisée. Dans la plupart des espèces les croisements et les fécondations sont réellement aléatoires pour la

plus grande partie des gènes. Il y a évidemment quelques exceptions, comme les allèles d'incompatibilité chez certaines espèces végétales. Mais, dans ce cas, les croisements ne sont pas panmictiques pour ces seuls gènes (et les gènes qui leur sont génétiquement liés, voir chapitre 3), mais sont panmictiques pour le reste du génome.

Chez l'homme les unions ne sont pas panmictiques pour les gènes qui gouvernent la pigmentation, en raison de la discrimination raciale plus ou moins grande existant dans toutes les populations, ou pour les gènes gouvernant la taille parce que les unions tendent à associer préférentiellement un homme de taille égale ou supérieure à la femme. Mais pour tous les autres gènes, ceux qui gouvernent les groupes sanguins, les facteurs sériques ou les maladies, les unions sont panmictiques comme cela a été démontré dans la plupart des études (voir les problèmes d'application) ;

- b)** l'absence de mutation n'est pas une condition légitime dans l'absolu mais elle est acceptable en pratique, sur quelques générations, car, comme on le verra plus loin, l'effet des mutations sur la diversité génétique et la constitution génétique des populations, se mesure sur des centaines ou des milliers de générations ;
- c)** il en est de même pour l'effet de la sélection naturelle. Ce n'est évidemment pas le cas pour la sélection artificielle opérée par les agronomes afin de sélectionner races et variétés animales ou végétales ; mais dans ce cas les croisements ne sont plus panmictiques et les populations en question ne sont plus naturelles ;
- d)** l'absence de migration peut être acceptable pour des populations végétales ou animales réellement isolées, mais c'est une condition dont il est toujours difficile de vérifier la validité. Cette condition est évidemment légitime si on considère l'espèce dans sa globalité.
L'absence de migrations est une condition particulièrement discutable pour ce qui concerne les populations humaines (même dans les isolats). On verra de plus que les migrations, contrairement à la sélection ou aux mutations, peuvent modifier la constitution génétique des populations en peu de générations. Selon la population, le gène étudié, les conditions de l'étude ou la question posée, l'effet des migrations devra être pris en compte ou pourra être négligé ;
- e)** la taille infinie est par essence une condition impossible. On verra que les populations de petite taille sont le siège d'un phénomène spécifique appelé dérive génétique. On verra ensuite que les effets de la dérive et de la sélection peuvent se combiner dans l'histoire des populations. On considérera alors qu'une population est « grande » quand l'effet de la dérive est suffisamment négligeable devant celui de la sélection pour ne faire dépendre son évolution génétique que des effets sélectifs. Inversement une population sera « petite » dès lors que son parcours évolutif, malgré les effets sélectifs, peut aussi dépendre de la dérive.

On peut donc admettre en conclusion que le modèle de Hardy-Weinberg, malgré ses conditions, est acceptable sur le temps de quelques générations, avec cependant une attention particulière à accorder à l'effet des migrations.

2.3.5 L'équilibre de Hardy-Weinberg

a) Mise en évidence des situations d'équilibres allélique et génotypique

Le modèle de Hardy-Weinberg tel qu'il est défini permet de conclure qu'une population panmictique, de grande taille, sans mutations, ni sélection, ni migration, maintient son polymorphisme génétique en l'état.

En effet les fréquences alléliques, p et q à la génération i , n'ont pas varié à la génération suivante. La diversité en termes de fréquences alléliques est donc stable.

Par contre les fréquences génotypiques D , H et R , choisies quelconques à la génération i , ont varié pour prendre des valeurs particulières p^2 , $2pq$, q^2 qui représentent $(p + q)^2$ le développement du carré de la somme des fréquences alléliques.

Cette relation entre les fréquences alléliques p et q d'une part et les fréquences génotypiques p^2 , $2pq$, q^2 d'autre part, est une conséquence directe de la panmixie qui revient à réaliser les fécondations comme deux tirages aléatoires indépendants dans une urne (figure 2.2).

Cette relation de Hardy-Weinberg est très utile car elle va permettre d'estimer les fréquences alléliques dans l'étude d'un gène présentant, selon les allèles, des phénotypes dominants ou récessifs (voir figure 1.6 et tableau 2.2).

Dès que les fréquences génotypiques ont les valeurs p^2 , $2pq$, q^2 , à la génération $i + 1$, elles restent inchangées dans les générations ultérieures tant que les conditions de Hardy-Weinberg sont maintenues. En effet, l'urne gamétique $(p + q)$ restant elle-même inchangée, les fréquences génotypiques qui en sont issues, par panmixie, demeurent égales à p^2 , $2pq$, q^2 .

Cette situation d'équilibre des fréquences alléliques et des fréquences génotypiques appelée « équilibre de Hardy-Weinberg » est atteinte en une génération pour un gène autosomique.

b) Établissement de l'équilibre quand les fréquences alléliques diffèrent entre sexes

On avait supposé, dans l'établissement du modèle, que les fréquences alléliques étaient les mêmes dans les deux sexes. Si tel n'est pas le cas, il est facile de voir, en reprenant le schéma de l'urne gamétique, qu'il faut une première génération de panmixie pour obtenir des fréquences alléliques égales dans les deux sexes, puis une deuxième évidemment, pour obtenir les fréquences génotypiques de l'équilibre de Hardy-Weinberg.

En effet, supposons qu'à la génération g_0 , la composition génétique d'une population soit différente dans chacun des sexes, ce qui peut arriver lors d'une fusion de population, on aura :

Sexe	femelle				mâle		
Génotypes	$A1A1$	$A1A2$	$A2A2$		$A1A1$	$A1A2$	$A2A2$
Fréquences	D	H	R		d	h	r
Fréquences alléliques	$p = D + H/2$			et	$u = d + h/2$		
	$q = R + H/2$				$v = r + h/2$		

À la génération suivante g_1 , après un cycle de panmixie, associée aux autres conditions, on aura dans chacun des sexes la même composition génétique, donnée par le schéma du tirage dans les deux urnes gamétiques parentales de la génération g_0 :

Génotypes	$A1A1$	$A1A2$	$A2A2$	(mâle ou femelle)
Fréquences	$p.u$	$p.v + q.u$	$q.v$	
Fréquences alléliques	$P = p.u + (p.v + q.u)/2$			
	$Q = q.v + (p.v + q.u)/2$			

Les fréquences alléliques ont pris une valeur différente de celles de la génération précédente, mais égales dans les deux sexes, qui ont désormais la même urne gamétique (P et Q).

Par contre les fréquences génotypiques ne vérifient pas la relation de Hardy-Weinberg ($p.u$ n'est pas égal à P^2 , ni $q.v$ à Q^2). Mais dès la génération suivante, g_2 , les fréquences génotypiques vérifieront cette relation en étant égales à P^2 , $2PQ$ et Q^2 , si les conditions de Hardy-Weinberg ont été maintenues.

c) L'équilibre de Hardy-Weinberg n'est pas une situation quelconque

Il convient de noter que pour une même diversité allélique (mêmes fréquences alléliques) il existe une infinité de populations différentes, présentant des fréquences génotypiques différentes, mais qu'il n'en est qu'une seule à l'équilibre de Hardy-Weinberg, présentant entre fréquences alléliques et génotypiques la relation panmixique (tableau 2.2).

TABLEAU 2.2 UNE MÊME DIVERSITÉ ALLÉLIQUE, UNE INFINITÉ DE DIVERSITÉS GÉNOTYPIQUES.

Population	$A1/A1$	$A1/A2$	$A2/A2$	$p = f(A1)$	$q = f(A2)$
1	0,7	0	0,3	0,7	0,3
2	0,6	0,2	0,2	0,7	0,3
3	0,4	0,6	0	0,7	0,3
4	0,5	0,4	0,1	0,7	0,3
5	0,49	0,42	0,09	0,7	0,3

Dans le tableau 2.2, la population 5 est strictement à l'équilibre de Hardy-Weinberg, la population 4 n'en n'est pas éloignée. En pratique on réalisera un test statistique pour statuer sur l'écart entre la diversité génétique observée dans une population et celle attendue sous le modèle théorique de Hardy-Weinberg. Si les écarts ne sont pas significatifs, on admettra que sa constitution génétique est conforme au modèle de Hardy-Weinberg.

d) Signification évolutive du modèle de Hardy-Weinberg

Le modèle de Hardy-Weinberg est fondé sur la conception mendélienne de l'hérédité (ségrégation des allèles à la méiose et réunion de deux allèles de chaque gène à

la fécondation). Il permet, soit de maintenir la diversité génétique (les allèles), notamment en l'absence de sélection, soit de prévoir une variation de la diversité, donc une évolution génétique, notamment en cas de sélection ; toutes choses que ne permettait pas une conception non mendélienne de l'hérédité. On comprend que les jeunes biométriciens darwinistes se soient laissés séduire.

Le développement des modèles de génétique des populations consistera la plupart du temps à partir de ce modèle central de Hardy-Weinberg pour voir comment la composition génétique d'une population peut être modifiée, vers quelle limite, et à quelle vitesse, si telle ou telle des conditions de cet équilibre n'est pas respectée.

2.4 APPLICATION DU MODÈLE DE HARDY-WEINBERG AU CALCUL DES FRÉQUENCES ALLÉLIQUES POUR LES CARACTÈRES PRÉSENTANT DES PHÉNOTYPES RÉCESSIFS

On a vu au chapitre 1 qu'il était facile d'estimer directement les fréquences des différents allèles d'un gène quand les différents génotypes formaient des phénotypes codominants mais que ce calcul se révélait impossible dès qu'un des phénotypes était récessif. C'est alors que le modèle de Hardy-Weinberg présente la grande utilité de fournir le moyen, indirect, d'une telle estimation. Cette utilisation sera illustrée par les quelques exemples qui suivent et dans la plupart des exercices.

2.4.1 Estimation des fréquences alléliques d'un gène responsable de l'albinisme

Il existe plusieurs formes d'albinisme dont chacune est dépendante d'un gène. La forme la plus fréquente résulte d'une mutation touchant le gène de structure de la tyrosinase, la première enzyme de la chaîne de biosynthèse des mélanines. Les individus albinos sont déficients en tyrosinase et sont porteurs de deux copies mutées du gène ; ils sont homozygotes a/a . Ce phénotype est récessif car les hétérozygotes A/a , ayant une copie fonctionnelle A du gène, sont capables d'assurer la synthèse des mélanines.

On observe, dans une population africaine, un phénotype albinos (par déficience de tyrosine) pour 10 000 habitants.

TABLEAU 2.3 DIVERSITÉ GÉNÉTIQUE POUR L'ALBINISME PAR DÉFICIT DE LA TYROSINASE.

Phénotype	Pigmenté		Albinos
Valeurs observées des fréquences phénotypiques	$D + H = 9\,999/10\,000$		$R = 1/10\,000$
Génotypes	A/A	A/a	a/a
Fréquences génotypiques si Hardy-Weinberg	p^2	$2pq$	q^2
Valeurs calculées des fréquences génotypiques	9 801/10 000	198/10 000	1/10 000