

Jean-Pierre Henry Pierre-Henri Gouyon

2^e CYCLE • CAPES • AGRÉGATION

Précis de génétique des populations

Cours, exercices et problèmes résolus



DUNOD

Précis de génétique des populations

Cours, exercices
et problèmes résolus

Jean-Pierre Henry

Maître de conférences à l'Institut national agronomique Paris-Grignon

Pierre-Henri Gouyon

Professeur à l'université de Paris-Sud et à l'INAPG

Maître de conférences à l'École polytechnique

DUNOD

Illustration de couverture : Oxalis, Didier Van de Rostyne, BIOS ©

Ce pictogramme mérite une explication. Son objet est d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit, particulièrement dans le domaine de l'édition technique et universitaire, le développement massif du **photocopillage**.

Le Code de la propriété intellectuelle du 1^{er} juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique s'est généralisée dans les établis-

sements d'enseignement supérieur, provoquant une baisse brutale des achats de livres et de revues, au point que la possibilité même pour les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée.

Nous rappelons donc que toute reproduction, partielle ou totale, de la présente publication est interdite sans autorisation du Centre français d'exploitation du droit de copie (CFC, 20 rue des Grands-Augustins, 75006 Paris).



© Dunod, Paris, 1999

© Masson, Paris, 1998

ISBN 2 10 004752 3

Toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite selon le Code de la propriété intellectuelle (Art L 122-4) et constitue une contrefaçon réprimée par le Code pénal. • Seules sont autorisées (Art L 122-5) les copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective, ainsi que les analyses et courtes citations justifiées par le caractère critique, pédagogique ou d'information de l'œuvre à laquelle elles sont incorporées, sous réserve, toutefois, du respect des dispositions des articles L 122-10 à L 122-12 du même Code, relatives à la reproduction par reprographie.

Table des matières

Avant-propos	VII
Introduction	IX
A. Principales notions de génétique des populations	1
1. Structure génétique des populations	3
2. Étude théorique d'un cas idéal	7
3. Influence du régime de reproduction	11
3.1 La panmixie	11
3.2 Les régimes fermés	13
3.3 Un régime ouvert : l'hétérogamie	19
3.4 Remarques	20
4. Influence des pressions évolutives	25
4.1 Le hasard	25
4.2 Évolution de la fréquence des hétérozygotes	30
4.3 Dérive génétique et consanguinité	31
4.4 La sélection	33
4.5 La mutation	36
4.6 La migration	37
5. Introduction aux modèles à plusieurs locus : le déséquilibre de liaison	45
5.1 La notion de déséquilibre de liaison	45
5.2 Évolution du déséquilibre de liaison sous les hypothèses de Hardy-Weinberg	48
5.3 Causes possibles d'un déséquilibre de liaison	50
Exercices du chapitre A	53
B. Le cryptopolymorphisme	73
1. Généralités	75
2. Fréquence d'un gène létal	79
2.1 Gène létal récessif dans une population panmictique	79
2.2 Gène létal récessif dans une population autogame	80
2.3 Gène létal récessif dans une population se reproduisant végétativement	81
3. Conclusion	83
Exercices du chapitre B	87

C. Le polymorphisme	91
1. Introduction	93
2. Le polymorphisme sélectionné	101
2.1 Polymorphisme équilibré par la superdominance de la valeur sélective	101
2.2 Valeurs sélectives variables	105
2.3 Polymorphisme transitoire	108
2.4 Le problème du fardeau génétique	109
3. Le point de vue neutraliste	111
Exercices du chapitre C	117
D. Problèmes de synthèse corrigés	121
1. Sélection sur le taux d'autofécondation	123
2. Tétraploïdes	127
3. Les chémotypes de thym	129
4. Deux populations : l'effet Wahlund	135
5. Date de floraison	143
6. Trois allèles	149
7. La cyanogénèse du trèfle	153
Corrigés des exercices	159
Chapitre A	161
Chapitre B	171
Chapitre C	173
Lexique	177
Bibliographie	181
Index	183

Avant-propos

Cet ouvrage est destiné à fournir au lecteur les bases de la génétique des populations. Notre but a été ici de présenter les concepts fondamentaux qui permettent de modéliser le devenir des populations et des espèces, et de comprendre la signification d'une partie des débats concernant les questions de biodiversité. Nous nous adressons d'abord aux étudiants en biologie, aux candidats aux concours et aux professeurs du secondaire qui ont vu arriver cette discipline dans les programmes scolaires, mais aussi aux gestionnaires du milieu naturel et à tous ceux qui veulent comprendre au moins le langage dans lequel on peut étudier la variabilité génétique.

La génétique des populations a été élaborée dans les années trente par Fisher, Haldane et Wright sur des bases mathématiques appliquées à l'hérédité mendélienne. Nous avons tenté d'être aussi simples que possible dans ces domaines. Cependant, le lecteur aura besoin, pour pouvoir suivre l'ensemble, de connaître des rudiments de probabilités (lois de probabilité, probabilités conditionnelles, loi binomiale) ainsi, bien sûr, que les bases de la génétique (à un ou deux locus).

Ce précis de génétique des populations est issu du document que les auteurs ont rédigé alors qu'ils étaient enseignants à l'Institut national agronomique de Paris-Grignon sous la direction de **Georges Valdeyron**. Ils tiennent à lui exprimer ici toute leur reconnaissance et leur respect. Ni l'un ni l'autre n'aurions pu nous consacrer à la génétique des populations sans lui : ce livre lui doit donc tout. La rédaction a bénéficié, en outre, de l'aide de ses successeurs, André Gallais et Philippe Brabant ainsi que de celle de Sophie Gerber et d'Irène Till-Bottraud. Qu'ils en soient remerciés.

Jean-Pierre Henry et Pierre-Henri Gouyon

Introduction

Bien que les principes régissant la transmission du matériel héréditaire aient été découverts par Mendel dès 1865, l'importance de ses travaux ne fut pas réalisée à son époque. Ce n'est qu'à partir de 1900 que la redécouverte de ces lois devait donner lieu au développement d'une nouvelle science, la génétique, avec le succès que l'on sait.

À cette époque, beaucoup de généticiens pensèrent que leurs découvertes rendaient caduque la théorie darwinienne de l'évolution par sélection naturelle : pour eux l'évolution s'expliquait par des mutations ; une évolution graduelle de caractères quantitatifs sous l'action de la sélection naturelle, telle que l'avait proposée Darwin (1859), leur paraissait contraire à ce que l'on commençait à connaître de l'hérédité. Pour rapprocher les points de vue, il fallut d'abord éclaircir l'hérédité des caractères quantitatifs et comprendre qu'elle pouvait s'expliquer par la ségrégation de nombreux gènes influençant le même caractère, avec généralement, en plus, une forte influence du milieu (ce sont les travaux de Johannsen sur l'hérédité de la taille des graines de haricot qui l'amènèrent à séparer clairement et à proposer en 1909 les termes « génotype » et « phénotype », ce dernier influencé à la fois par le génotype et par le milieu). D'autre part, il fallut réaliser que si les mutations étaient, bien sûr, un préalable nécessaire à l'apparition de nouveaux gènes, on devait comprendre aussi comment ces nouveaux gènes pouvaient remplacer les anciens, donc raisonner en termes de **fréquence des gènes dans les populations**. C'est là l'objet de la génétique des populations.

La **génétique des populations** est donc née dans les années 1920 de la volonté de concilier la théorie darwinienne de l'évolution et les données de plus en plus précises acquises depuis le début du XX^e siècle sur la transmission du matériel héréditaire. Cette harmonisation a d'abord été faite par des mathématiciens (R.A. Fisher) et des biologistes (J.B.S. Haldane, S. Wright), qui ont construit des modèles décrivant l'évolution des popula-

tions. Le but de la génétique des populations est, par des modèles de plus en plus élaborés, de rendre compte de l'évolution.

Elle repose donc largement sur la construction de modèles mathématiques. Mais ces modèles doivent être confrontés à la réalité. La génétique des populations inclut aussi l'observation de populations naturelles (animales, végétales ou humaines), et des expériences de laboratoire, afin de suivre l'évolution de la fréquence des gènes et de mesurer les paramètres qui interviennent dans les modèles. Un frein a longtemps été la difficulté de trouver dans les populations des gènes « marqueurs » dont on puisse étudier la fréquence, une large part de la variation phénotypique observable dans les populations ayant un déterminisme génétique trop complexe ou trop influencé par le milieu pour être utilisable à cette fin. On était obligé de se limiter à quelques marqueurs tels que les groupes sanguins chez l'homme, des caractères de coloration chez des animaux... Les techniques d'électrophorèse des protéines à partir de 1966, puis plus récemment les techniques d'étude au niveau de l'ADN lui-même, ont mis à la disposition des chercheurs un nombre considérable de nouveaux marqueurs. Les données ainsi acquises ont parfois conduit à remettre en cause certaines idées anciennes, et jouent un rôle important dans les développements en cours.

A

Principales notions de génétique des populations

Structure génétique des populations

Dans une population haploïde, la proportion des individus portant un allèle donné fournit directement la fréquence de cet allèle dans la population. Mais, dans les populations diploïdes, qui sont celles auxquelles nous nous intéresserons principalement, les gènes sont associés par paires dans les individus; il est alors nécessaire de distinguer deux paramètres pour décrire la composition génétique de la population au locus considéré :

- Les **fréquences génotypiques** sont les fréquences des différents génotypes au locus considéré. L'ensemble de ces fréquences donne la **structure génotypique** de la population pour ce locus.

- Les **fréquences alléliques** (dites aussi **fréquences géniques**) sont les fréquences dans la population des différents allèles au locus considéré. Leur connaissance donne la **structure allélique** (ou **génique**) de la population.

C'est l'ensemble de ces paramètres qui permet de décrire la **structure génétique** des populations.

La connaissance de la structure allélique est importante pour le généticien des populations car c'est d'abord l'évolution des fréquences des gènes qui est significative pour l'évolution : la population est vue comme un « pool » de gènes dont la composition est susceptible d'évoluer, les individus n'étant que les porteurs transitoires de ces gènes. D'autre part, il est souvent plus simple dans les modèles de considérer les fréquences des gènes que

celles des génotypes qu'ils engendrent par leur association dans les zygotes. Toutefois, si la connaissance de la structure génotypique fournit facilement celle de la structure allélique (encadré 1), il faut être conscient que ce passage fait perdre de l'information : le passage inverse est impossible sans information ni hypothèse supplémentaire ; en effet, avec les mêmes fréquences alléliques, on peut avoir des structures génotypiques très différentes selon que les gènes homologues sont associés dans les individus plutôt entre gènes semblables (on a surtout des homozygotes) ou plutôt entre gènes différents (on a des hétérozygotes) ; ainsi une population avec 50 % de gènes A et 50 % de gènes a peut être constituée uniquement d'homozygotes AA et aa , ou uniquement d'hétérozygotes Aa , ou de diverses proportions de ces trois génotypes.

Enfin, il peut être nécessaire, pour comprendre l'évolution d'un caractère, de considérer plusieurs locus à la fois. Il faut donc connaître la structure génétique à chaque locus impliqué ; mais cela ne suffit pas. Il faut être conscient, là aussi, que d'autres paramètres seront nécessaires pour décrire la façon dont les gènes de locus *différents* seront associés dans les individus. Ainsi, si nous savons qu'une population contient 50 % d'individus AA et 50 % de aa à un premier locus, 50 % d'individus BB et 50 % de bb à un second locus, la structure génétique est parfaitement définie à chacun de ces locus. Mais cela ne nous dit pas si la population contient des individus $AABB$ et $aabb$, ou $AAbb$ et $aaBB$, ou les quatre et en quelles proportions. Les modèles qui étudient l'évolution de la structure génétique à plusieurs locus à la fois doivent donc manipuler davantage de paramètres et sont souvent d'une grande complexité. Dans la suite de cet ouvrage introductif, nous nous limiterons presque uniquement à des modèles à un seul locus (ce qui permet tout de même des généralisations intéressantes !) et nous nous contenterons de présenter comment on peut décrire l'évolution de la structure génétique à plusieurs locus simultanément par un choix de paramètres assez simple pour que les modèles restent mathématiquement traitables.

ENCADRÉ 1 : STRUCTURE GÉNÉTIQUE DES POPULATIONS

Structure génotypique en un locus

Décrite par l'ensemble des fréquences dans la population des différents génotypes au locus considéré, sa connaissance directe n'est possible que si l'on peut reconnaître les différents génotypes à ce locus, par leur phénotype (variation discontinue, pas de dominance) ou par l'examen de leurs descendance.

$$fr(AA) = \frac{n_{AA}}{N}$$

$fr(AA)$ fréquence des individus ayant le génotype AA .

n_{AA} nombre d'individus ayant ce génotype.

N nombre total d'individus de la population.

Structure allélique

Elle est décrite par l'ensemble des fréquences des allèles au locus considéré.

Cas d'un locus biallélique A, a : 3 génotypes : AA, Aa, aa .

Fréquence de l'allèle A :

$$p = \frac{\text{Nombre de gènes } A}{\text{Nombre total de gènes}} = \frac{2n_{AA} + n_{Aa}}{2N} = fr(AA) + \frac{1}{2}fr(Aa)$$

Fréquence de l'allèle a :

$$q = 1 - p = \frac{\text{Nombre de gènes } a}{\text{Nombre total de gènes}} = \frac{2n_{aa} + n_{Aa}}{2N} = fr(aa) + \frac{1}{2}fr(Aa)$$

Ces formules sont toujours vraies : elles n'impliquent aucune supposition sur le régime de reproduction ou l'existence (ou non) de pressions évolutives.

Cas d'un locus multiallélique : $A_1, A_2, \dots, A_i, \dots, A_n$

Fréquences alléliques : $p_1, p_2, \dots, p_i, \dots, p_n$

$$p_i = \frac{2n_{A_i A_i} + \sum_{j \neq i} n_{A_i A_j}}{2N} = f_{A_i A_i} + \frac{1}{2} \sum_{j \neq i} f_{A_i A_j}$$

Remarque : en général on n'échantillonne qu'une fraction de la population (entité elle-même généralement mal définie dans ses limites). Les fréquences calculées comme ci-dessus à partir d'un échantillon sont des estimateurs sans biais des fréquences de la population.

Étude théorique d'un cas idéal

Le premier modèle publié concernant la structure génotypique d'une population l'a été simultanément et indépendamment par le mathématicien anglais G.H. Hardy et par le médecin allemand W. Weinberg en 1908.

Dans leur modèle, ces auteurs ont adopté des hypothèses permettant de simplifier les calculs et le raisonnement. Ces hypothèses sont les suivantes :

H₁ : les gamètes s'associent *au hasard* par rapport aux gènes considérés (hypothèse de la **panmixie**).

H₂ : la population a une taille *infinie*; la fréquence d'un événement est donc égale à sa probabilité (loi des grands nombres).

H₃ : la fréquence des gènes n'est pas modifiée d'une génération à la suivante par :

- la **mutation**;
- un tri sélectif s'exerçant sur ces gènes (**sélection**);
- un manque d'isolement de la population entraînant l'apport de gènes issus d'autres populations (**migration**).

Considérons une population diploïde dans laquelle coexistent deux allèles en un locus, **A** et **a**, en fréquences respectives p et q ($p + q = 1$). La population gamétique est donc composée de gamètes porteurs de l'allèle **A** en fréquence p , et de gamètes porteurs de **a** en fréquence q . En appliquant les hypothèses H₁ et H₂, la génération diploïde suivante s'obtient par le tableau ci-dessous où la fréquence de chaque case est le produit des fréquences de la ligne et de la colonne correspondantes :

Gamètes mâles Gamètes femelles	A p	a q
A p	AA p^2	Aa pq
a q	Aa pq	aa q^2

La structure génotypique de la nouvelle génération produite est donc :

$$\begin{array}{ll}
 p^2 & \text{individus AA} \\
 2pq & \text{individus Aa} \\
 q^2 & \text{individus aa}
 \end{array}$$

Cette structure génotypique est célèbre sous le nom de **structure de Hardy-Weinberg**.

Selon l'hypothèse H_3 , les gamètes que produira cette nouvelle génération contiendront :

- le gène A en fréquence $p' = p^2 + \frac{1}{2} \cdot (2pq) = p(p+q) = p$
- le gène a en fréquence $q' = \frac{1}{2} \cdot (2pq) + q^2 = q(p+q) = q$

Ainsi, étant donné les hypothèses faites, les fréquences des gènes et des génotypes ne varient pas d'une génération à l'autre.

G.H. Hardy en concluait, dans son article de 1908 : « *En un mot, nous n'avons pas la moindre raison de penser qu'un caractère dominant devrait montrer une tendance à envahir la population, ni qu'un récessif devrait tendre à disparaître.* »

La **loi de Hardy-Weinberg** est actuellement énoncée sous la forme suivante :

Dans une population isolée d'effectif illimité, non soumise à la sélection, et dans laquelle il n'y a pas de mutation, les fréquences alléliques restent constantes.

Si les accouplements sont panmictiques, les fréquences génotypiques se déduisent directement des fréquences alléliques selon la relation ci-dessous; elles restent donc aussi constantes.

$$\begin{array}{lll}
 \text{AA} & \text{Aa} & \text{aa} \\
 p^2 & 2pq & q^2
 \end{array}$$

S'il y a plus de deux allèles au locus considéré ($A_1, A_2, \dots, A_i, \dots, A_n$) en fréquences $p_1, p_2, \dots, p_i, \dots, p_n$, il existe $n(n+1)/2$ génotypes différents et la structure de Hardy-Weinberg devient :

- homozygotes $A_i A_i$ en fréquence p_i^2
- hétérozygotes $A_i A_j$ en fréquence $2 p_i p_j$

Il y a en fait, dans la loi de Hardy-Weinberg, deux aspects tout à fait indépendants.

La première partie de l'énoncé ci-dessus peut être vue comme un truisme : elle dit que s'il n'y a rien pour faire changer les fréquences alléliques (aucune des quatre pressions évolutives : mutation, migration, sélection et hasard lié à un effectif fini), elles ne changent pas. Cela peut être vu aussi comme un principe fondamental, analogue au principe d'inertie en physique. Il est important de réaliser que ce principe est vrai qu'il y ait ou non panmixie. Il est tout aussi vrai, par exemple, pour une population d'organismes haploïdes se multipliant végétativement : dans une population de bactéries avec deux formes différant par leur allèle en un locus (A_1 ou A_2), s'il n'y a pas de mutation transformant un allèle dans l'autre, si le milieu ne favorise pas la multiplication ou la survie d'une des formes, et que la population ne reçoit pas d'éléments extérieurs, si, enfin, cette population peut être considérée comme infinie, les proportions des deux formes ne changeront pas.

La seconde partie de la loi de Hardy-Weinberg établit une relation entre fréquences génotypiques et fréquences alléliques lorsqu'il y a panmixie. Au stade zygotique, cette relation est vraie dès qu'il y a panmixie, même si des pressions évolutives interviennent pour modifier les fréquences alléliques au cours du cycle qui fait passer d'une génération à la suivante. Imaginons par exemple un organisme marin qui émet ses gamètes dans l'eau où ceux-ci s'unissent au hasard ; les zygotes produits seront en fréquences $p^2 AA, 2pq Aa$ et $q^2 aa$, p et q étant les fréquences des gamètes A et a qui se sont unis. Cela reste vrai si, parmi les individus qui ont produit ces gamètes, les individus AA en produisent davantage que les autres (ils ont une meilleure fertilité, c'est une forme de sélection), ou si lors de la production de ces gamètes des gènes A ont muté en a . Simplement, dans ce cas, les fréquences alléliques p et q dans les gamètes et les zygotes ne sont plus les mêmes que les fréquences alléliques de la génération parentale.

Ceci explique que la structure de Hardy-Weinberg décrive beaucoup de populations relativement bien, alors que les hypothèses de Hardy-Weinberg ne sont pratiquement jamais vérifiées. Il suffit qu'il y ait union au hasard vis-à-vis du locus considéré et que rien (par exemple une mortalité différentielle) n'ait modifié sensiblement les fréquences génotypiques entre le stade zygote et celui où est faite l'observation. La conformité d'une population à une structure de Hardy-Weinberg peut être testée à partir de sa composition génotypique (encadré 2).

Le modèle de Hardy-Weinberg constitue le modèle le plus simple possible. Nous allons reprendre une par une les hypothèses de ce modèle de référence, afin de voir ce qui se passe si elles ne sont plus vérifiées; nous allons donc étudier leur influence respective sur les structures et sur l'évolution de la population.

3

Influence du régime de reproduction

Le régime de reproduction décrit la manière dont les gamètes s'assemblent vis-à-vis du (ou des) gène(s) considéré(s), pour former la génération suivante. Chez la plupart des animaux, il s'agit donc de décrire la façon dont se constituent les couples; chez les plantes à fleurs, il peut être plus pertinent de considérer les plantes mères et les grains de pollen qui les fécondent.

3.1 LA PANMIXIE

L'hypothèse H_1 supposait l'association au hasard des gamètes. Ce régime particulier est appelé *panmixie*. Il peut s'agir d'une union au hasard des gamètes eux-mêmes s'ils sont émis dans le milieu, ou bien, ce qui revient au même, d'une formation des couples au hasard relativement aux gènes que portent les conjoints au locus considéré : il ne doit pas y avoir de corrélations entre leurs génotypes, c'est-à-dire que le fait de connaître le génotype d'un des conjoints ne fournit aucune information sur ce que pourrait être le génotype de l'autre.

Ce régime confère à la population une structure génotypique comprenant **$2pq$** hétérozygotes. C'est à cette valeur de $2pq$ que nous allons nous référer pour classer les régimes de reproduction : nous appellerons **régimes fermés** ceux qui donnent un taux d'hétérozygotes inférieur à $2pq$ et **régimes ouverts** ceux qui en donnent plus que $2pq$.

ENCADRÉ 2 : TEST DE LA CONFORMITÉ D'UNE POPULATION À LA STRUCTURE DE HARDY-WEINBERG

Lorsqu'un échantillon issu d'une population est étudié à un locus donné, et que sa composition génotypique peut être établie, il est possible de tester si les fréquences génotypiques observées correspondent ou non à une population ayant la structure de Hardy-Weinberg.

Si l'échantillon présente la composition suivante, pour un locus biallélique A, a :

Génotypes	AA	Aa	aa	Total
Effectifs observés	N_1	N_2	N_3	N

les allèles A et a sont par conséquent représentés en fréquences p et q telles que :

$$p = \frac{2N_1 + N_2}{2N} \quad q = \frac{N_2 + 2N_3}{2N}$$

Ces fréquences sont des estimateurs des fréquences réelles de la population dans laquelle l'échantillon a été tiré. En supposant que la population possède la structure de Hardy-Weinberg, les effectifs attendus dans l'échantillon devraient être :

Génotypes	AA	Aa	aa
Effectifs attendus	Np^2	$2Npq$	Nq^2

Une statistique X , qui mesure l'écart entre les effectifs réellement observés et ceux qui sont attendus si la structure de Hardy-Weinberg est respectée, peut être calculée :

$$X = \sum_{\text{génotypes}} \frac{(\text{effectif attendu} - \text{effectif observé})^2}{\text{effectif attendu}}$$

$$X = \frac{(Np^2 - N_1)^2}{Np^2} + \frac{(2Npq - N_2)^2}{2Npq} + \frac{(Nq^2 - N_3)^2}{Nq^2}$$

Si l'hypothèse est vraie (la population observée a une structure de Hardy-Weinberg), alors la statistique X suit une loi du χ^2 à 1 degré de liberté^a. Cela signifie que si un grand nombre d'échantillons de même taille étaient tirés dans la population et que la valeur de X était calculée pour chacun d'entre eux, l'ensemble de ces valeurs serait reparti selon une loi du χ^2 . Il s'agit alors de comparer la valeur de X obtenue pour l'échantillon étudié avec les valeurs de la loi. Si cette valeur correspond à une probabilité d'être observée inférieure à 5 ou 1 %, on peut rejeter l'hypothèse de conformité. Le risque de rejeter cette hypothèse alors qu'elle est vraie est alors de 5 % ou de 1 %.

a. Si la population présente n allèles au locus étudié, le nombre de degrés de liberté de la loi du χ^2 utilisée sera égal au nombre de génotypes possibles moins n .

3.2 LES RÉGIMES FERMÉS

Ce sont d'une part, la consanguinité, dont la forme la plus extrême est l'autogamie, et d'autre part, l'homogamie.

3.2.1 Exemple de l'autogamie

L'exemple le plus extrême de régime fermé est l'**autogamie**, où chaque individu se reproduit par autofécondation.

Les homozygotes AA et aa ne donnent que des descendants identiques à eux-mêmes alors que les hétérozygotes Aa donnent une disjonction à chaque génération suivant la formule mendélienne :

AA	Aa	aa
$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

La proportion d'individus hétérozygotes est, de ce fait, divisée par deux à chaque génération et la population tend rapidement vers l'homozygotie totale avec seulement des individus de génotype AA et aa . À la génération g , la proportion d'hétérozygotes est donc :

$$H_g = \frac{H_{g-1}}{2} = \frac{H_0}{2^g}$$

Nous avons envisagé ici l'évolution de la structure génotypique. Qu'en est-il des fréquences alléliques dans la population? Pour un locus, avec trois génotypes AA , Aa et aa , si nous appelons D_g , H_g et R_g les fréquences génotypiques respectives à la génération g (D pour « dominant », H pour « hétérozygote » et R pour « récessif », bien que les relations de dominance n'aient aucune importance dans l'affaire), les fréquences alléliques sont :

$$\begin{aligned}
 A \quad p_g &= D_g + \frac{1}{2}H_g = \left(D_{g-1} + \frac{1}{4}H_{g-1}\right) + \frac{1}{2}\left(\frac{1}{2}H_{g-1}\right) \\
 &= D_{g-1} + \frac{1}{2}H_{g-1} = p_{g-1} \\
 a \quad q_g &= 1 - p_g = 1 - p_{g-1} = q_{g-1}
 \end{aligned}$$

Les fréquences alléliques sont donc, ici aussi, constantes. Cela ne doit pas nous surprendre, puisque nous n'avons introduit dans notre modèle aucune des forces susceptibles de faire varier la fréquence des gènes. En l'absence de toute autre force, l'autogamie ne modifie pas les fréquences géniques mais change seulement les fréquences génotypiques en redistribuant les allèles; il en résulte finalement à l'infini la structure génotypique :

$$pAA + qaa$$

Cette disparition des hétérozygotes se produit à la fois pour tous les locus. Une autre façon d'exprimer les choses est donc de dire qu'en autogamie, la proportion des locus hétérozygotes dans le génome de chaque individu est divisée par deux à chaque génération. On tend donc vers des individus homozygotes à tous leurs locus, ce qu'on appelle des **lignées pures**.

3.2.2 La consanguinité

La **consanguinité** est un régime de reproduction où les unions se font entre individus **apparentés**, c'est-à-dire ayant un ou des ancêtres communs. Il s'ensuit que ces individus ont souvent les mêmes gènes hérités de leurs ancêtres communs. Un individu issu d'une union entre deux apparentés est dit **consanguin**. À un locus donné, il aura donc pu recevoir de ses deux parents des copies d'un même gène de l'ancêtre commun; un individu consanguin est, par conséquent, plus souvent homozygote (et moins hétérozygote) qu'un individu issu de l'union de deux individus non apparentés. Dans une population, une propension aux unions consanguines entraîne ainsi un déficit d'hétérozygotes par rapport à $2pq$. Ce régime de reproduction, comme pour l'autogamie, concerne l'ensemble du génome.*

L'autogamie est, en fait, la forme la plus extrême de la consanguinité. En effet, l'individu avec lequel un individu est le plus apparenté (à le plus de gènes en commun), c'est lui-même. Un régime consanguin systématique autre que l'autogamie conduit aussi, bien que plus lentement, à la disparition des hétérozygotes. Ainsi, on peut montrer que lorsque les croisements se font systématiquement entre frère et sœur (ou parent et enfant), un niveau donné d'homozygotie est obtenu après un nombre de générations trois fois plus élevé, environ, qu'avec l'autogamie, mais, là aussi, on aboutit finalement à une disparition des hétérozygotes. Les croisements entre cousins germains, ou entre oncle et nièce, mènent encore plus lentement l'ensemble du génome à l'homozygotie.

Étude détaillée de la consanguinité

À cause de l'importance particulière de la consanguinité en zootechnie, nous détaillons dans l'encadré 3 les méthodes de calcul des paramètres les plus utiles dans les modèles.

Lorsque deux gènes homologues sont deux copies (sans mutation) d'un même gène ancêtre, ils sont dits **identiques**. Noter que ce terme a, en génétique des populations, un sens très précis et restrictif : un homozygote quelconque AA n'a pas, en général, deux gènes identiques; ce ne peut être le cas que s'il est consanguin.

* Ceci ne s'applique pas aux gènes déterminant un régime d'hétérogamie; voir plus loin.

ENCADRÉ 3 : DÉFINITIONS ET CALCUL DES COEFFICIENTS DE CONSANGUINITÉ ET DE PARENTÉ

Définitions

Consanguinité : régime de reproduction où les unions se font entre individus apparentés.

Deux individus sont dits **apparentés** si et seulement si ils ont au moins un ancêtre commun.

Un individu est dit **consanguin** si et seulement si ses deux parents sont apparentés.

Deux gènes sont dits **identiques** si et seulement si ce sont deux copies sans mutation d'un même gène ancêtre.

Principaux paramètres mesurant la consanguinité

Coefficient de parenté de deux individus (Malecot 1948) : le coefficient de parenté ϕ_{IJ} de deux individus I et J est la probabilité que deux gènes homologues tirés au hasard l'un chez I , l'autre chez J , soient identiques.

Coefficient de consanguinité d'un individu : f_I : probabilité que deux gènes homologues de l'individu I soient identiques.

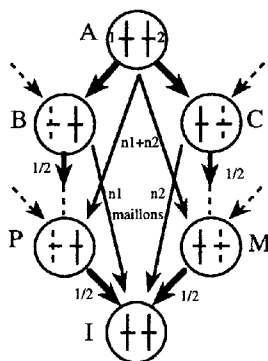
Remarque importante : d'après leurs définitions, le coefficient de consanguinité d'un individu I est égal au coefficient de parenté de ses deux parents (P et M) :

$$f_I = \phi_{PM}$$

Principes du calcul des coefficients de consanguinité et de parenté connaissant le pedigree

Considérons un individu I et ses parents P et M apparentés par l'ancêtre commun A . A a transmis une copie de l'un de ses gènes à chacun de ses enfants B et C . À chaque génération ultérieure, ce gène a une probabilité $1/2$ d'être transmis. La probabilité pour que le gène transmis par P à I provienne de A est finalement $(1/2)^{n1}$, $n1$ étant le nombre de « maillons » de B à I (ou, ce qui revient au même, de A à P). De même, la probabilité pour que le gène transmis par M à I provienne de A est $(1/2)^{n2}$, $n2$ étant le nombre de « maillons » de C à I (ou de A à M).

Ces deux événements sont indépendants. La probabilité pour qu'un gène tiré au hasard chez P et un gène tiré au hasard chez M proviennent de A est donc $(1/2)^{n1 + n2}$.



Cette probabilité assure l'origine commune des deux gènes, mais non leur identité, car chez *A* il y a deux gènes homologues. Les deux gènes transmis à *B* et *C* sont soit des copies d'un même gène de *A* (probabilité $1/2$), soit des copies de chacun des deux gènes homologues de *A* (probabilité $1/2$); dans ce cas, ces deux gènes peuvent être quand même identiques, avec une probabilité f_A (coefficient de consanguinité de *A*). Puisque les deux événements sont disjoints, la probabilité que les deux tirages au sort donnent des gènes identiques est donc $(1/2 + 1/2 f_A)$.

Nous avons donc :

$$f_I = \phi_{PM} = \left(\frac{1}{2}\right)^{n_1 + n_2} \left(\frac{1}{2} + \frac{1}{2} f_A\right)$$

ce qui s'écrit :

$$f_I = \left(\frac{1}{2}\right)^{n+1} (1 + f_A)$$

où $n (n = n_1 + n_2)$ est le nombre de maillons sur la « chaîne de parenté » *P-A-M*.

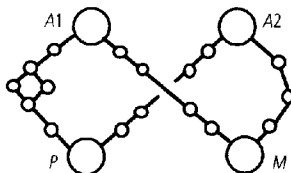
S'il existe plusieurs voies *P-A-M* par lesquelles *P* et *M* ont pu recevoir deux gènes identiques, ou plusieurs ancêtres communs à *P* et *M* dont ils ont pu recevoir deux gènes identiques, la formule se généralise en remarquant que ces situations sont disjointes (système complet d'événements : *P* et *M* peuvent transmettre à *I* deux gènes identiques reçus soit d'un ancêtre, soit d'un autre). On a alors :

$$f_I = \phi_{PM} = \sum_i \left(\frac{1}{2}\right)^{n_i+1} (1 + f_{A_i})$$

Σ sur toutes les chaînes de parenté

n_i : nombre de « maillons » de la i^e chaîne de parenté reliant *P* à *M*

A_i : ancêtre commun à *P* et *M* pour la chaîne *i*



Trois chaînes de parenté de *P* à *M*
(deux par *A1*, une par *A2*)

Il y a autant de chaînes de parenté que de voies par lesquelles deux gènes identiques ont pu être transmis à *P* et *M* (système complet d'événements).

Une chaîne de parenté relie les deux individus *P* et *M* en passant par l'ancêtre commun.

Une chaîne de parenté change une seule fois de direction dans l'échelle des temps : au niveau de l'ancêtre commun.

Une chaîne de parenté ne peut pas passer deux fois par le même individu.

Deux chaînes de parenté sont différentes dès qu'elles ont au moins un maillon différent.

Autres relations permettant de calculer les coefficients de consanguinité et de parenté

Une autre méthode pour calculer les coefficients de parenté, souvent d'un emploi plus simple lorsque les généalogies sont compliquées, consiste à « remonter les générations ». Il existe en effet des relations simples entre les coefficients de parenté de deux individus I et J et ceux de leurs parents; en effet, lorsqu'on tire au hasard un gène chez I , ce peut être celui reçu de son père P ou celui reçu de sa mère M , avec pour chaque cas une probabilité $1/2$. Donc :

$$\phi_{IJ} = 1/2(\phi_{PJ} + \phi_{MJ})$$

De même, si K et L sont les deux parents de J :

$$\phi_{IJ} = 1/2(\phi_{IK} + \phi_{IL})$$

et en prenant à la fois les parents de I et de J (à condition que J ne soit pas le parent de I ou l'inverse) :

$$\phi_{IJ} = 1/4(\phi_{PK} + \phi_{PL} + \phi_{MK} + \phi_{ML})$$

On peut continuer à remonter les générations en prenant les parents de P , M , K , L , jusqu'à tomber sur des coefficients de parenté connus ou des coefficients de parenté d'un individu avec lui-même. En effet, les relations ci-dessus restent valables que P , M , K et L soient différents ou non (par exemple si I et J sont frères, $P = K$ et $M = L$, et : $\phi_{IJ} = 1/4[\phi_{PP} + \phi_{MM} + 2\phi_{PM}]$). En revanche, ces formules ne sont plus valables pour calculer le coefficient de parenté d'un individu avec lui-même. Lorsque l'on tire au hasard un gène chez I et un gène chez I , on peut tirer deux fois le même (probabilité $1/2$) ou les deux gènes de I (probabilité $1/2$); donc, le coefficient de parenté d'un individu I avec lui-même vaut :

$$\phi_{II} = 1/2 + 1/2(f_I)$$

Quelques résultats classiques

Coefficients de parenté :

père/fils	1/4
pleins-frères	1/4
demi-frères	1/8
cousins germains	1/16

Lorsque l'on considère deux individus apparentés, leur degré d'apparentement peut être mesuré par la probabilité qu'ils aient reçu deux gènes identiques de leurs ancêtres communs. Plus précisément, lorsque l'on tire au hasard un gène chez chacun d'eux au même locus, la probabilité que ces gènes soient identiques est appelée le **coefficient de parenté** de ces deux individus. De même, la probabilité qu'un individu consanguin ait, à un locus, deux gènes identiques est appelée le **coefficient de consanguinité** de cet individu. Ces probabilités peuvent être calculées connaissant les relations de parenté, c'est-à-dire la généalogie des individus (encadré 3).

Il peut y avoir consanguinité dans une population d'effectif illimité si les croisements ne se font pas au hasard, mais préférentiellement entre apparentés. En revanche, dans une population d'effectif limité, même si les croisements se font au hasard, la consanguinité est inéluctable. Le nombre d'ancêtres possibles d'un individu étant fini, il suffit de remonter assez haut pour trouver un ancêtre commun à ses deux parents. Si on remonte n générations plus haut, nous avons chacun potentiellement 2^n ancêtres, mais ce nombre finit par être supérieur à la taille de l'humanité; en fait, quelques individus ancêtres apparaîtront de nombreuses fois dans la généalogie.

Comme aucune population réelle n'est illimitée, nous pourrions dire que tous les individus d'une population, d'une espèce, et même tous les êtres vivants, sont apparentés et consanguins; il suffit de remonter assez haut. À vrai dire, cette constatation est d'un intérêt limité : sur des généalogies aussi longues, de nombreuses mutations ont eu lieu; les gènes dérivés d'un même ancêtre commun se sont différenciés et ont cessé d'être identiques.

3.2.3 L'homogamie

On appelle **homogamie** un régime de reproduction où les unions se font entre individus phénotypiquement semblables. Comme en régime de consanguinité, il y a diminution du taux d'hétérozygotes à chaque génération.

Supposons que le phénotype soit gouverné par un couple d'allèles A et a ; il existe, au départ, trois génotypes : AA , Aa et aa .

S'il n'y a pas de dominance, ces génotypes correspondent à trois phénotypes; les accouplements se font donc entre individus ayant le même génotype, et le résultat est identique à celui obtenu par autogamie : les hétérozygotes diminuent de moitié à chaque génération. Mais ici, contrairement à l'autogamie, **seul ce locus est concerné**.

S'il y a dominance de A sur a , les individus aa se reproduisent entre eux et donnent des descendants identiques à eux-mêmes. Le groupe des individus AA et Aa (même phénotype $[A]$) produit trois types de croisements : $AA \times AA$, $AA \times Aa$ et $Aa \times Aa$. Ce dernier type de croisement donne des individus aa qui rejoignent l'autre groupe. La fréquence de a diminue donc dans ce groupe $[A]$; progressivement, il se « purifie » de ses allèles a qui

passent dans l'autre groupe, mais, sur l'ensemble de la population, les fréquences alléliques restent constantes. Le taux d'hétérozygotes décroît, mais plus progressivement que dans le cas précédent (lorsque l'allèle a se fait rare dans le groupe $[A]$, les croisements $Aa \times Aa$ deviennent très rares, de sorte que l'élimination du gène a dans ce groupe devient très lente). Néanmoins, à terme, il ne reste, encore ici, que des homozygotes AA et aa , en fréquences p et q .

- ✓ L'autogamie, la consanguinité et l'homogamie, pourvu qu'elles soient pratiquées de façon absolue par la population, conduisent *progressivement* à l'homozygotie totale. La panmixie, en revanche, amène un taux d'hétérozygotes de $2pq$ en une seule génération (dans le cas d'une population à générations séparées). Un régime mixte (autogamie partielle, homogamie partielle...) comprenant une fraction de croisements en panmixie ne peut donc jamais conduire à l'homozygotie totale de la population, car la fraction panmictique, aussi faible soit-elle, provoque, à chaque génération, l'apparition d'hétérozygotes.

3.3 UN RÉGIME OUVERT : L'HÉTÉROGAMIE

L'**hétérogamie** est un régime de reproduction où les unions se font entre individus phénotypiquement dissemblables (on parle parfois aussi d'« homogamie négative »). Cela entraîne un taux d'hétérozygotes supérieur à $2pq$.

Ce régime concerne notamment les gènes sexuels (ou parasexuels). Ainsi chez les mammifères, le sexe mâle est déterminé par la formule chromosomique XY , et le sexe femelle par la formule XX .

En considérant X et Y comme des allèles, la fréquence de X est égale à $3/4$, celle de Y à $1/4$; le taux d'hétérozygotes en panmixie serait de $2pq = 2 \times (1/4) \times (3/4) = 3/8$. Il est en fait plus élevé : $1/2$ (proportion des mâles). Chez les plantes, les systèmes d'auto-incompatibilité (gamétophytique ou sporophytique) ont aussi pour effet d'interdire tout ou partie des fécondations entre porteurs des mêmes gènes au locus d'incompatibilité, ce qui entraîne une hétérozygotie élevée pour ce locus.

L'hétérogamie a une particularité parmi les régimes de reproduction : à la différence des autres modes de reproduction dont nous avons parlé auparavant, l'hétérogamie modifie non seulement la structure génotypique, mais aussi la fréquence des allèles dans la population. Ainsi, chez les mammifères, la fréquence des chromosomes X et Y est maintenue à $(3/4 : 1/4)$. Cette particularité tient à ceci : si pour une raison ou une autre, les individus reproducteurs d'un sexe sont moins fréquents que ceux de l'autre sexe à une génération, ce sont eux qui, en moyenne, se reproduiront le plus. L'hétérogamie entraîne donc une forme spéciale de sélection qui favorise les gènes de la forme la plus rare, ce qui les maintient à une fréquence d'équilibre précise.

D'autre part, l'hétérogamie « protège » les loci concernés contre l'homozygotie due à la consanguinité. Par exemple, des croisements frère-sœur ne modifient pas la proportion de femelles (XX) et de mâles (XY).

3.4 REMARQUES

1) Les régimes décrits ci-dessus sont des schémas que l'on peut dégager des principales tendances qui existent dans la nature. En fait, les situations réelles ne se conforment jamais qu'approximativement à ces schémas. C'est ainsi qu'il est probable que l'autogamie stricte n'existe pas; que l'on constate une tendance homogamique dans le choix des conjoints pour quelques caractères chez beaucoup d'espèces à panmixie prédominante... Il n'en est que plus remarquable de constater que chaque espèce semble avoir « choisi » une tendance prédominante. Ainsi, les plantes à fleurs se partagent schématiquement en trois grandes catégories : les plantes dites **autogames**, où l'autofécondation prédomine, les **allogames** qui se reproduisent principalement en croisement (on peut considérer alors qu'il y a panmixie pour la plupart des locus) et les plantes à **reproduction asexuée** (multiplication végétative ou apomixie).

2) Le régime de reproduction ne peut être défini, à part pour l'autogamie ou la consanguinité, que pour un locus déterminé, et il peut parfaitement varier d'un locus à l'autre. Dans les populations humaines par exemple, il y a hétérogamie pour le sexe, une certaine homogamie constatée pour les gènes de la couleur de la peau, et à peu près panmixie pour les groupes sanguins.

3) On constate que les régimes de reproduction influencent fortement la structure génotypique, en déterminant si les gènes homologues s'associent dans les zygotes plutôt entre gènes semblables (régimes fermés) ou dissemblables (régimes ouverts); mais, mis à part le cas particulier de l'hétérogamie, ils n'ont pas d'effet par eux-mêmes sur les fréquences alléliques (fig. 1). Ce dernier point doit cependant être nuancé : il est vrai tant que le régime de reproduction est le seul facteur en cause. Mais si une sélection s'exerce qui, par exemple, élimine les homozygotes aa , il est clair que la vitesse d'élimination du gène récessif a de la population dépendra fortement du régime de reproduction : elle sera plus rapide avec un régime fermé qu'en panmixie ou en régime ouvert, où beaucoup de gènes a échapperont à la sélection parce qu'ils sont masqués dans les hétérozygotes.

4) Nous avons classé les régimes de reproduction selon la fréquence d'hétérozygotes par rapport à $2pq$. L'écart à cette fréquence peut être exprimé par un paramètre appelé **indice de fixation** de la population : soit une population ayant, pour un locus à deux allèles (A , a), une structure génotypique absolument quelconque :

Génotypes	AA	Aa	aa
Fréquences génotypiques	D	H	R

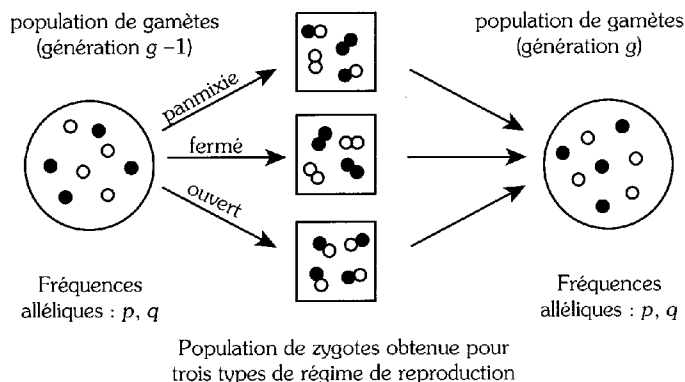


Figure 1 : Le régime de reproduction détermine l'association des gènes homologues dans les zygotes, mais il est sans effet, par lui-même, sur les fréquences alléliques (les deux types de points symbolisent les deux allèles).

On définit l'**indice de fixation** comme :

$$F = 1 - \frac{H}{2pq}$$

$$\Leftrightarrow H = (1 - F) \cdot 2pq$$

$F \in [-1 ; +1]$; $F = 0$ si la structure est celle de Hardy-Weinberg.

$$\begin{aligned} \text{Or } p &= D + \frac{1}{2}H \Rightarrow D = p - \frac{1}{2}H = p - pq(1 - F) \\ &= p(1 - q) + pqF = p^2 + pqF \\ q &= R + \frac{1}{2}H \Rightarrow R = q - \frac{1}{2}H = q - pq(1 - F) \\ &= q(1 - p) + pqF = q^2 + pqF \end{aligned}$$

$D = p^2 + Fpq = (1 - F)p^2 + Fp$	AA
$H = (1 - F)(2pq)$	Aa
$R = q^2 + Fpq = (1 - F)q^2 + Fq$	aa

Cette structure est parfois appelée **structure de Wright**. Quelle que soit la structure génotypique de la population, elle peut toujours s'écrire sous cette forme : connaissant D , H et R , on peut toujours calculer p et F . Inversement, ces deux paramètres suffisent à spécifier entièrement la structure génétique : connaissant p et F , les fréquences génotypiques D , H et R sont totalement déterminées. F représente l'information qui nous manquait pour déduire la structure génotypique de la structure allélique. L'intérêt de l'écriture sous cette forme est qu'elle permet de bien séparer ce qui, dans la structure génotypique, dépend des fréquences alléliques (p et q) et ce qui dépend de la façon dont les gènes sont associés (F , qui mesure un écart à la structure de Hardy-Weinberg). On peut montrer aussi que, si la structure génotypique est observée au stade zygote, F n'est autre que le coefficient de corrélation entre les gamètes qui s'unissent. C'est d'ailleurs ainsi que F a d'abord été défini par Wright, sous le nom de « coefficient de consanguinité de la population ». En fait, ce terme est abusif, car F ainsi défini n'a la signification d'un coefficient de consanguinité que dans la mesure où la consanguinité est bien le seul facteur induisant un écart à la structure de H.W. dans la population. D'autre part, un apparemment généralisé (cas des populations finies) ne provoque pas d'écart à H.W. Même si l'on considère la structure au stade zygote, une corrélation, et donc un écart à la structure de H.W., peuvent être provoqués par d'autres facteurs, l'homogamie par exemple. Si et seulement si la consanguinité est le seul facteur en jeu (toutes les autres conditions de la loi de Hardy-Weinberg sont remplies), F est égal à la moyenne des coefficients de consanguinité f des individus de la population (notion définie à partir du concept d'identité). Autrement dit, F est alors la probabilité de trouver à un locus deux gènes identiques chez un individu tiré au hasard dans la population :

$$F = \bar{f} \in [0;1]$$

et dans l'écriture de la structure génotypique, F peut être remplacé par f . Dans ce cas, on peut d'ailleurs retrouver la structure de Wright par un autre raisonnement : lorsque l'on tire un individu de la population, ses deux gènes sont identiques avec une probabilité f , et ils sont alors AA avec une probabilité p et aa avec une probabilité q ; les deux gènes sont non identiques avec une probabilité $(1 - f)$, et sont alors Aa , Aa ou aa avec les probabilités p^2 , $2pq$, q^2 respectivement. D'où la structure :

$$AA : \quad (1 - f)p^2 + fp$$

$$Aa : \quad (1 - f).2pq$$

$$aa : \quad (1 - f)q^2 + fq$$

Mais ce raisonnement n'est plus valable dès que l'on sort du cadre strict des hypothèses ci-dessus, parce que si d'autres facteurs que la consanguinité interviennent, ou si la consanguinité peut être due à un apparemment généralisé, les gènes non identiques ne sont plus forcément indépendants.

4

Influence des pressions évolutives

Nous appelons **pression évolutive** une action qui s'exerce sur la population en en modifiant la structure allélique.

Il en existe quatre types, correspondant à l'abandon de diverses hypothèses de la loi de Hardy-Weinberg :

- le **hasard** lié à un effectif fini,
- la **sélection**, si certains gènes ont une influence sur la viabilité ou sur l'efficacité de la reproduction,
- la **mutation**,
- la **migration**.

4.1 LE HASARD

Il n'existe pas de populations d'effectif infini (tout au plus existe-t-il des populations suffisamment grandes pour que l'on puisse les considérer comme pratiquement infinies sur un court laps de temps). Or, dans une population finie, les gènes portés par les gamètes qui donnent naissance à une génération représentent un échantillon des gènes de la génération précédente. La loi des grands nombres ne s'applique plus. Même si, en espérance, on attend la même fréquence des allèles qu'à la génération

précédente, cela ne sera réalisé que de façon approximative du fait des effets d'échantillonnage liés à l'effectif fini; ces effets peuvent être importants, surtout si la population est petite. Comme cela se reproduit à chaque génération, et que rien ne tend spécialement à nous ramener aux fréquences d'origine, ces effets cumulés sur de nombreuses générations peuvent devenir considérables. L'effet d'échantillonnage introduit donc une variation aléatoire de la fréquence des gènes au fil des générations. Cette fluctuation est appelée **dérive génétique**.

Considérons une population diploïde à générations séparées, ayant une taille constante de N individus. Mise à part la taille, toutes les autres conditions de la loi de Hardy-Weinberg sont supposées satisfaites (population isolée, sans sélection ni mutation). Soit p_0 la fréquence du gène A dans la génération de départ. La génération diploïde suivante représente un tirage au hasard de $2N$ gènes. Le **nombre** X_1 de gènes A dans cette nouvelle génération est une variable aléatoire dont la loi est une binomiale de paramètres $2N$ et p_0 , d'où :

$$E(X_1) = 2N.p_0$$

$$Var(X_1) = 2N.p_0.(1 - p_0).$$

La **proportion** de gènes A dans la nouvelle génération est :

$$p_1 = X_1/2N$$

Elle a donc pour **espérance** :

$$E(p_1) = p_0$$

et pour **variance** :

$$Var(p_1) = \frac{p_0(1-p_0)}{2N}$$

Si l'on maintient constante la taille N de la population, on peut démontrer par récurrence qu'après g générations, la fréquence p_g de A est une variable aléatoire d'espérance :

$$E(p_g) = p_0$$

et de variance :

$$Var(p_g) = p_0(1 - p_0)[1 - (1 - 1/2N)^g]$$

Cela signifie que si l'on menait une infinité de populations de même taille en parallèle, à partir d'une même population de départ, la fréquence *moyenne* de A sur l'ensemble des populations resterait égale à la fréquence initiale p_0 , mais que la fréquence de A varierait de façon erratique dans chaque population et que les populations divergeraient (fig. 2). La variance ci-dessus représente la variance interpopulations, qui augmente au cours des générations, et ce d'autant plus vite que N est petit. Par ailleurs, la variance mesure, par définition, l'espérance des carrés des écarts à la moyenne; comme cette moyenne est $E(p_g) = p_0$:

$$\text{Var}(p_g) = E(p_g - p_0)^2$$

La variance mesure donc aussi l'ampleur des écarts attendus par rapport à la fréquence initiale p_0 .

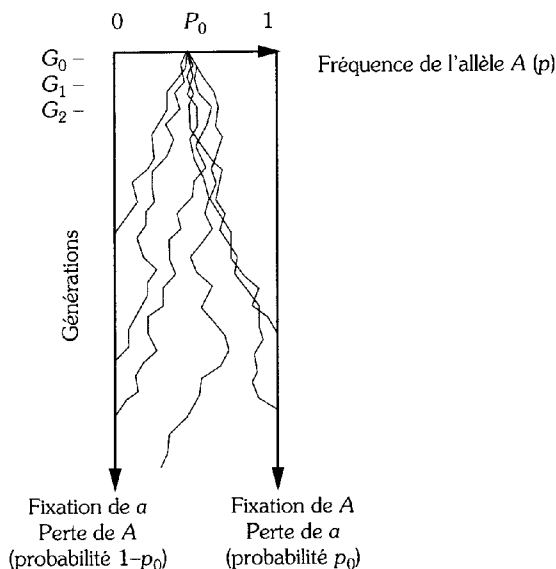


Figure 2 : Fluctuations aléatoires de la fréquence d'un gène dans une série de populations ayant au départ la même fréquence allélique (p_j).

En fait, la fréquence du gène ne va pas fluctuer indéfiniment. Dans une population, elle augmente ou elle diminue, mais comme rien ne tend à ramener la population vers des fréquences moyennes, *tôt ou tard, elle atteindra 1 ou 0*, c'est-à-dire que l'on ne tirera que des gènes A ou que des gènes a . En l'absence de mutations, cette **fixation** aléatoire d'un des allèles

est définitive; si la fréquence p_g prend la valeur 0 ou 1, elle la garde : les valeurs 0 et 1 constituent des « bords absorbants ». Sur l'ensemble des populations envisagées plus haut, on tend vers l'homogénéité de chaque population, mais il y a, comme nous l'avons dit, une différenciation croissante entre les populations, certaines fixant A , d'autres a , de sorte que la diversité allélique est *globalement* maintenue. Plus précisément, comme on sait que $E(p)$ est toujours égale à p_0 , on en déduit qu'à la limite, $p = 1$ avec la probabilité p_0 , et $p = 0$ ($q = 1$) avec la probabilité $1 - p_0$. En d'autres termes, si un allèle est présent dans une population en fréquence p_0 , il va envahir la population sous le seul effet du hasard avec une probabilité p_0 , ou disparaître avec une probabilité $(1 - p_0)$.

Ce phénomène de dérive se produit à une vitesse proportionnelle à la décroissance du terme $(1 - 1/2N)^q$; il est donc d'autant plus rapide que la taille de la population est petite. Une population de petite taille a beaucoup plus de chances d'être rapidement homogénéisée par la dérive qu'une grande population, et ce pour l'ensemble des locus. De plus, une population réelle, composée de N individus, n'évolue généralement pas comme une population idéale de même effectif. Le plus souvent, la population réelle doit être comparée à une population idéale d'effectif N_e , appelé **effectif génétique**, inférieur à l'effectif réel (encadré 4).

Il est important de noter que, même si d'autres forces que le hasard agissent sur les fréquences des gènes, celui-ci agit *toujours*. Comme aucune population n'est infinie, et que dans les populations naturelles, la dérive a eu tout le temps de jouer, on pourrait penser que ces populations sont fixées pour la plupart de leurs locus. En fait, ce n'est généralement pas le cas, parce que la mutation et la migration réintroduisent de nouveaux allèles, et que, pour certains gènes, la sélection opère et peut s'opposer à l'homogénéisation. Nous aurons l'occasion de revenir sur cette question.

Les populations naturelles ont rarement des effectifs constants. La dérive va jouer surtout dans les périodes où les effectifs sont faibles. Si à une époque de son histoire, une population est passée pendant plusieurs générations par un effectif très faible, elle a pu perdre par dérive, pendant cette période de « goulot démographique », une grande part de sa diversité allélique. Elle ne pourra la récupérer ensuite qu'en recevant des gènes d'autres populations, ou par mutation, mais cela est forcément très lent étant donné la faiblesse des taux de mutation. C'est par de tels goulots démographiques que l'on interprète généralement les quelques cas connus d'espèces qui ont une diversité allélique nulle à tous leurs locus. Tel est le cas, par exemple, de l'Éléphant de Mer de la côte ouest de l'Amérique du Nord : on met cela en rapport avec une très forte réduction des effectifs due à une chasse excessive au XIX^e siècle, qui a amené cette espèce au bord de l'extinction : la dérive a dû alors jouer de façon considérable, et là, il n'y a pas d'autres populations qui auraient pu fournir des migrants.

ENCADRÉ 4 : NOTION D'EFFECTIF GÉNÉTIQUE

Dans une population idéale, satisfaisant aux conditions du modèle de Hardy-Weinberg, composée de N individus diploïdes ($2N$ gènes), chaque individu a la même probabilité de participer à la constitution de la génération suivante, ce qui permet de dire par exemple qu'un même gène sera tiré deux fois avec une probabilité $1/2N$. Dans une population réelle de N individus diploïdes il n'est souvent pas possible de tenir le même raisonnement, les formules établies précédemment ne s'appliquent plus.

Pour décrire l'évolution d'une population réelle, il est possible de recalculer les formules en tenant compte des écarts au modèle idéal, mais cela sera souvent compliqué. On choisit plutôt de déterminer l'effectif génétique de la population réelle, noté N_e , défini comme l'effectif d'une population idéale présentant la même évolution, notamment en termes de coefficient de consanguinité f que la population réelle. N_e sera différent de N , et généralement inférieur. Nous allons considérer plus particulièrement deux cas, lorsqu'une population est de taille variable ou si les femelles et les mâles ne sont pas en nombre égal.

- Dans une population idéale de taille N nous avons vu que f_g , coefficient de consanguinité à la génération g , s'exprimait simplement en fonction de f_{g-1} , ou f_0 et N (paragraphe 4.2). Supposons qu'une population, sur deux générations, passe d'une taille N_1 à une taille N_2 , on aurait alors :

$$1 - f_1 = (1 - 1/2N_1)(1 - f_0) \text{ et } 1 - f_2 = (1 - 1/2N_2)(1 - f_1)$$

soit
$$1 - f_2 = (1 - 1/2N_2)(1 - 1/2N_1)(1 - f_0).$$

Dans une population idéale, on aurait $1 - f_2 = (1 - 1/2N_e)^2(1 - f_0).$

Par conséquent : $(1 - 1/2N_e)^2 = (1 - 1/2N_1)(1 - 1/2N_2).$

Cette formule est bien approchée par l'expression : $1/N_e = 1/2(1/N_1 + 1/N_2).$

N_e est la moyenne harmonique entre N_1 et N_2 .

Ceci se généralise à une population qui comprend N_i individus à la génération i . Cette population est alors équivalente, après g générations, à une population idéale d'effectif N_e tel que :

$$\frac{1}{N_e} = \frac{1}{g} \sum_{k=1}^g \frac{1}{N_k}$$

- Considérons le cas d'une population d'individus à sexes séparés. À chaque génération, la moitié des allèles provient de chaque sexe. Par conséquent, dès que l'on s'écarte d'une situation où les deux sexes sont également représentés, le risque de dérive génétique augmente. Si la population comporte N_f femelles et N_m mâles, un indi-

vidu aura forcément deux parents différents, un mâle et une femelle. Il ne pourra être issu de l'union entre deux gamètes provenant du même individu que si celui-ci est l'un de ses grands-parents. Cela se produira lorsque ses deux parents lui transmettront le même gène qu'ils auront eux-mêmes hérité de leur père, avec une probabilité $1/8$, et que ces deux parents auront le même père, avec une probabilité $1/N_m$. Le raisonnement est symétrique avec la mère. La taille réelle de la population sera donc $N = N_f + N_m$ mais sa taille génétique N_e s'écrira (sachant que la probabilité pour que les deux gènes soient des copies du même gène de la génération précédente dans une population idéale est $1/2N$) :

$$\frac{1}{2N_e} = \frac{1}{8N_m} + \frac{1}{8N_f} \Rightarrow N_e = \frac{4N_mN_f}{N_m + N_f}$$

Si une population comporte le même nombre d'individus de chacun des deux sexes, son effectif réel est égal à son effectif génétique. Mais dès que ces nombres sont inégaux, l'effectif génétique est toujours plus faible que le nombre d'individus total de la population.

Un cas particulier de dérive lors de goulot démographique est l'**effet de fondation** : quelques migrants, voire un seul chez une espèce hermaphrodite, sont à l'origine d'une nouvelle population là où l'espèce n'était pas présente. Ils peuvent, par le hasard de l'échantillonnage, représenter un échantillon d'allèles très différent de la composition de la population dont ils sont issus. La nouvelle population qui se développe à partir de ce noyau fondateur aura donc une structure génétique sensiblement différente de la structure d'origine. Selon certains évolutionnistes, ce phénomène pourrait jouer un rôle important dans la formation de nouvelles espèces.

Dans notre modèle, nous avons envisagé la dérive pour une population diploïde, sexuée et panmictique, mais elle se produit de même dans toute population finie où de nouveaux individus naissent tandis que les anciens disparaissent.

La seule chose qui soit propre au type de population que nous avons choisi est l'évolution du taux H d'hétérozygotes.

4.2 ÉVOLUTION DE LA FRÉQUENCE DES HÉTÉROZYGOTES

Dans une population diploïde et panmictique, la fréquence espérée d'hétérozygotes Aa à une génération est $2p_gq_g$, p_g et q_g étant les fréquences alléliques dans les gamètes qui s'unissent pour former les zygotes (en fait, à cause des aléas de l'échantillonnage, cette proportion $2p_gq_g$ ne sera réalisée qu'approximativement).

$$E(H) = E[2p_g(1 - p_g)] = 2[E(p_g) - E(p_g^2)]$$

$$\text{Or} \quad \text{Var}(p_g) = E(p_g^2) - [E(p_g)]^2$$

$$\text{et} \quad E(p_g) = p_0$$

$$\text{Var}(p_g) = p_0(1-p_0)[1 - (1 - 1/2N)^g]$$

$$\text{d'où} \quad E(p_g^2) = p_0(1-p_0)[1 - (1 - 1/2N)^g] + p_0^2$$

$$E(H) = 2p_0(1-p_0)(1 - 1/2N)^g$$

Quand g augmente, $(1 - 1/2N)^g$ tend vers 0, et

$$E(H) \rightarrow 0$$

Les hétérozygotes tendent donc à disparaître. En fait, ils restent à chaque génération en fréquence approximative $2pq$, puisqu'il y a panmixie, mais le produit pq tend vers zéro parce que p tend vers 0 ou q tend vers 0 (fixation), ainsi que nous l'avons vu.

4.3 DÉRIVE GÉNÉTIQUE ET CONSANGUINITÉ

Ces deux phénomènes, que nous avons jusqu'ici séparés, sont liés dans la pratique, car la consanguinité est inéluctable dans les populations d'effectif limité.

On peut illustrer simplement la liaison entre effectif limité et consanguinité : soit une population de N individus diploïdes ($2N$ gènes). Chacun des individus produit une infinité de gamètes. La génération suivante sera constituée par le tirage de $2N$ gamètes dans le pool de gamètes de la génération précédente. Si l'on suppose que les $2N$ gènes des individus de la génération de départ sont non identiques, le coefficient de consanguinité à la génération 0 vaut :

$$f_0 = 0$$

Calculons le coefficient de consanguinité f_g de la génération g . La probabilité que les deux gamètes qui formeront un individu à la génération g portent à un locus donné une copie d'un même gène de la génération précédente est $1/2N$. Ils sont alors identiques avec une probabilité de 1. Si les deux gamètes portent des copies de gènes distincts de la génération $(g-1)$ (probabilité $1 - 1/2N$), ils sont alors identiques avec une probabilité f_{g-1} (coefficient de consanguinité à la génération précédente).

On a alors :

$$f_g = (1/2N) + (1 - 1/2N)f_{g-1}$$

ce qui peut s'écrire :

$$(1 - f_g) = (1 - 1/2N)(1 - f_{g-1}) = (1 - 1/2N)^g(1 - f_0)$$

d'où, puisque f_0 est nul :

$$f_g = 1 - (1 - 1/2N)^g$$

Le coefficient de consanguinité augmente donc d'autant plus vite que la taille de la population est petite. La valeur calculée n'est bien sûr exacte qu'en espérance, puisque le hasard des tirages va jouer.

À terme, on arrive à $f = 1$. Ce calcul nous montre que les choses sont encore pires, en quelque sorte, que ce que nous avait montré le calcul sur la dérive : non seulement on perd la diversité allélique, mais même tous les gènes de la population, au locus considéré, finissent par être des copies d'un seul des gènes de la population de départ. En d'autres termes, si l'on remonte assez haut dans le temps, on trouvera un seul gène ancêtre de tous les gènes actuellement présents à un locus dans une population (fig. 3). Comme nous avons eu l'occasion de le dire à propos de la consanguinité, la

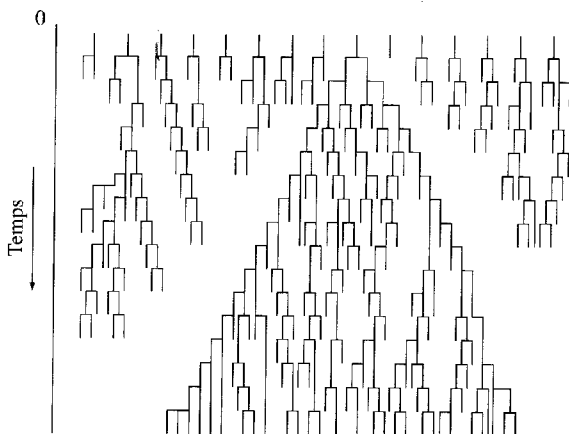


Figure 3 : Le « coalescent » montre que de nombreuses lignées s'éteignent et que tous les allèles présents en fin de processus remontent à un ancêtre unique situé quelque part dans le passé.

mutation peut bien sûr avoir réintroduit une diversité entre ces gènes, mais il n'en demeure pas moins qu'il y a bien un gène ancêtre unique pour chaque locus. Ce phénomène est appelé la **coalescence des gènes** d'une population. Elle existe pour tous les locus, mais le gène ancêtre pour un locus se trouvera généralement dans un autre individu que le gène ancêtre d'un autre locus, et à une époque différente. En fait, comme il peut y avoir eu des recombinaisons à l'intérieur des gènes, même des portions de gènes adjacentes peuvent dériver d'ancêtres différents.

Il s'agit d'un phénomène aléatoire; tant qu'il n'y a pas de sélection, tous les gènes présents à un moment donné ont la même probabilité ($1/2N$) d'être l'heureux gagnant, le futur ancêtre de tous les gènes qui seront dans la population bien des générations plus tard. On ne peut pas non plus dire précisément combien de temps cela prend; ce **temps de coalescence** est une variable aléatoire. Néanmoins, on peut en calculer l'espérance : on montre que, pour k gènes, il faut remonter en moyenne $2k$ générations pour trouver un seul ancêtre commun à tous. Pour une population diploïde ayant un effectif constant de N individus, soit $2N$ gènes par locus, cela nous fait remonter à $4N$ générations, en espérance. De la même façon, il faudra attendre $4N$ générations, en moyenne, pour qu'il ne reste des copies que d'un seul des $2N$ gènes présents à un moment donné.

4.4 LA SÉLECTION

La sélection naturelle et son effet sur l'évolution des populations sont l'hypothèse centrale de la théorie de Darwin. Les généticiens des populations ont introduit cette idée dans leurs modèles à travers le concept de **valeur sélective d'un génotype**.

4.4.1 Notion de valeur sélective

Chaque génotype est affecté d'un paramètre nommé valeur sélective. L'idée est qu'**un génotype est d'autant meilleur que les individus possédant ce génotype laissent en espérance plus de descendants** (puisque c'est vers cela que sont finalisés les êtres vivants).

La valeur sélective d'un génotype, notée ici w , est donc mesurée par le nombre de descendants laissés par les zygotes porteurs de ce génotype à la génération suivante. On peut aussi compter le nombre de gamètes produits qui ont participé à la formation des zygotes de la génération suivante, ce qui revient au même à un facteur 2 près.

La valeur sélective d'un génotype peut être décomposée en deux facteurs :

- la **viabilité** (notée v) représentant la probabilité pour un zygote ayant ce génotype d'arriver à l'âge reproducteur;

• la **fertilité** (notée f) qui est le nombre moyen de zygotes qu'il laisse à la génération suivante (ou de gamètes participant à la génération suivante).

On a alors :

$$w = v \cdot f$$

4.4.2 Évolution d'une population soumise à la sélection

Considérons une population possédant, pour le locus étudié, deux allèles, A et a . Trois génotypes sont possibles : AA , Aa et aa ; on leur affecte des valeurs sélectives : w_1 , w_2 et w_3 respectivement.

Soit p la fréquence de l'allèle A , et $q = 1 - p$ la fréquence de a dans la population gamétique. Si l'on suppose la panmixie, les zygotes produits seront répartis comme suit :

Génotypes	AA	Aa	aa
Fréquences	p^2	$2pq$	q^2

S'il y a des différences de viabilité entre génotypes, ces proportions vont se trouver modifiées chez les adultes reproducteurs. Par ailleurs, s'il y a des différences de fertilité, certains produiront plus de gamètes participant à la génération suivante que d'autres.

En définitive, étant donné les différences de viabilité et de fertilité entre ces trois génotypes, la proportion des gamètes A produits est :

$$p' = \frac{w_1 p^2 + 1/2 \cdot w_2 \cdot 2pq}{\bar{w}}$$

où $\bar{w} = w_1 \cdot p^2 + 2w_2 \cdot pq + w_3 \cdot q^2$ représente la valeur sélective moyenne de la population (les valeurs sélectives de chaque génotype multipliées par leur fréquence), c'est-à-dire la quantité moyenne de zygotes de la génération suivante (ou de gamètes participant aux fécondations) formés par zygote.

On peut alors calculer la variation de la fréquence de l'allèle A d'une génération à la suivante sous l'effet de la sélection :

$$\Delta p = p' - p = \frac{w_1 \cdot p^2 + w_2 \cdot pq}{\bar{w}} - p$$

ce qui se transforme (en réduisant au même dénominateur et en se souvenant que $p + q = 1$) en :

$$\Delta p = pq \frac{(w_1 - w_2)p + (w_2 - w_3)q}{w_1 \cdot p^2 + 2w_2 \cdot pq + w_3 \cdot q^2}$$

On constate que, dans l'évolution de la fréquence allélique, les valeurs absolues des w n'interviennent pas : on pourrait toutes les multiplier ou les diviser par une même quantité. Cela nous confirme qu'il est indifférent de mesurer la valeur sélective par le nombre de zygotes formés ou par le nombre de gamètes utiles. Souvent aussi, par convention, on affecte une valeur arbitraire de 1 à l'une des valeurs sélectives (par exemple la plus grande), et on exprime les autres par rapport à celle-ci (un génotype ayant une valeur sélective de 0,8 sera alors un génotype dont la valeur reproductive est égale à 80 % de celle des meilleurs : quand ceux-ci laissent 1 descendant, le génotype considéré n'en laisse que 0,8). Cela ne change rien pour le calcul de l'évolution des fréquences alléliques; simplement les valeurs sélectives ne représentent plus un nombre absolu de descendants.

Le signe de Δp nous donne le sens de l'évolution : si Δp est positif, cela signifie que la fréquence p de A augmente du fait de la sélection; $\Delta p < 0$ indique que p diminue; $\Delta p = 0$ correspond à un équilibre. Puisque w est toujours positif, que $pq \geq 0$, le signe de Δp est celui du numérateur. Il dépend bien sûr des valeurs sélectives. En écartant le cas où $w_1 = w_2 = w_3$, qui signifie qu'il n'y a pas de sélection (et dans ce cas c'est le hasard qui opère), on peut envisager les situations suivantes :

- $w_1 > w_2 > w_3$ correspond au cas où les génotypes AA se reproduisent davantage (ils sont plus viables et/ou plus fertiles) que les Aa , eux-mêmes meilleurs que les aa . Dans ce cas Δp est toujours positif; le gène A augmente en fréquence jusqu'à $p = 1$ (et $q = 0$; Δp devient alors nul : c'est l'équilibre). Ici, clairement, il y a un gène favorable, A , qui élimine le gène défavorable a . La situation est la même si $w_1 = w_2 > w_3$ ou si $w_1 > w_2 = w_3$.

- Le cas symétrique est $w_1 \leq w_2 \leq w_3$: Δp est négatif; la sélection amène à la fixation de l'allèle favorable a .

- Les cas $w_1 > w_2 < w_3$ (l'hétérozygote est inférieur aux deux homozygotes) et $w_1 < w_2 > w_3$ (l'hétérozygote est supérieur aux deux homozygotes) sont plus complexes : il existe une valeur particulière de p différente de 0 et 1 qui annule le numérateur, c'est-à-dire qu'il peut exister un équilibre où les deux allèles se maintiennent. Nous aurons l'occasion d'explorer plus en détail cette situation.

✓ Ce modèle a l'avantage d'être simple, mais les paramètres qu'il utilise, fréquences alléliques chez les zygotes ou les gamètes, valeurs sélectives, sont très difficiles à estimer dans les populations. Il y a à cela diverses raisons :

- On n'a presque jamais accès aux zygotes ni aux gamètes. On ne peut donc connaître les fréquences à ces stades; on peut, tout au plus, regarder chez les jeunes en espérant qu'il n'y a pas eu de sélection auparavant. Or une éventuelle élimination précoce de zygotes est difficilement détectée, et plus encore une sélection sur les gamètes.

- La connaissance de la structure chez les adultes est généralement insuffisante : à ce stade, les génotypes sont en fréquences attendues :

$$\frac{v_1 p^2}{V} \qquad \frac{2v_2 pq}{V} \qquad \frac{v_3 q^2}{V}$$

mais on ignore ce que valent p et q (valeurs chez les zygotes), puisque les fréquences alléliques chez les adultes ne sont plus les mêmes. En fait, on a trop d'inconnues.

- Par ailleurs, tout cela ne nous dit rien sur d'éventuelles différences de fertilité. S'il y en a, les fréquences chez les adultes ne reflètent pas non plus celles des gamètes et des zygotes qui seront produits.
- Il faut donc des études démographiques précises pour estimer les viabilités et fertilités. Il faut comparer de nombreux individus des divers génotypes pour s'abstraire des effets aléatoires : dans une population, il y a toujours des individus qui se reproduisent plus que d'autres, ce n'est pas forcément dû à la sélection; il faut montrer que les différences constatées sont bien dues au génotype.
- On se heurte à une difficulté encore plus sérieuse : la valeur sélective d'un individu dépend de son génotype sur l'ensemble de son génome. Si l'on décèle des différences, il est difficile d'affirmer qu'elles sont causées par le génotype au locus considéré, et pas à d'autres gènes.
- Enfin, la valeur sélective dépend du milieu : un génotype peut être favorisé dans un milieu et pas dans un autre. Or le milieu naturel est variable...!

4.5 LA MUTATION

La mutation est une **erreur dans la reproduction conforme du message héréditaire**. Elle va transformer un allèle en un autre, nouveau ou déjà présent dans la population. Du point de vue du généticien des populations, les mutations qui importent sont celles qui se produisent dans la lignée germinale. Il s'intéresse donc à des **taux de mutation** par génération.

Si le gène A , en fréquence p , se transforme en a à chaque génération avec une fréquence u (taux de mutation), alors que le taux de réversion est négligeable, la fréquence de A à la génération suivante est :

$$p' = (1 - u) \cdot p$$

La variation de la fréquence de A d'une génération à l'autre, due à la mutation, est :

$$\Delta p = p' - p = -up$$

Après g générations :

$$p_g = (1 - u)^g \cdot p_0$$

Pour que p ait diminué de moitié, le nombre g de générations nécessaire est tel que :

$$p_g = \frac{1}{2} \cdot p_0 = (1-u)^g p_0$$

soit $(1-u)^g = \frac{1}{2}$

d'où $g = \frac{-\ln 2}{\ln(1-u)} \approx \frac{\ln 2}{u} \approx \frac{0,7}{u}$

Pour $u = 10^{-6}$, on trouve $g = 700\,000$. Ce chiffre correspond, chez l'homme et ses ascendants, à une durée approximative de 10 à 20 millions d'années (ce qui nous ramène dans l'ère tertiaire).

Il est clair qu'une force qui prend autant de temps pour diviser simplement la fréquence d'un gène par deux ne peut pas avoir, par ce moyen, joué un rôle important dans l'évolution. Le rôle de la mutation est négligeable sur l'évolution de la fréquence des gènes, comparé à celui des autres pressions évolutives.

En fait le rôle de la mutation dans l'évolution est primordial, mais il est autre : **c'est la seule source de gènes nouveaux**. Mais une fois qu'un nouveau gène est apparu par mutation, ce n'est plus elle qui va déterminer son devenir : si le nouvel allèle est défavorable, ou s'il est plus favorable que les anciens, c'est principalement la sélection qui va déterminer l'évolution ultérieure de sa fréquence. Si le nouvel allèle ne présente ni avantage ni désavantage, c'est la dérive qui interviendra.

- ✓ La mutation n'est pas toujours unilatérale; il peut exister des **mutations reverses**. Ainsi, supposons que le gène A mute en son allèle a avec une fréquence u , et que a mute en A avec une fréquence v . La pression de mutation devient :

$$\Delta p = -up + v(1-p) = v - p(u+v)$$

Cette formule peut évidemment s'annuler (pour $p = v/(u+v)$), ce qui implique que ce système peut à lui seul mener à un équilibre stable. Cependant, l'influence de la pression de mutation sur les fréquences est si faible (voir ci-dessus) que ces fréquences ne peuvent rester constantes sous l'action de ces seules forces; d'autres pressions plus fortes (par exemple la dérive) vont venir modifier la structure de la population et on ne peut pas vraiment considérer qu'il s'agit là d'un équilibre.

4.6 LA MIGRATION

La migration est **le passage de gènes d'une population dans une autre** (par des individus qui passent d'une population à l'autre, par exemple sous forme de graines ou de pollen chez les plantes). Si cet échange se fait entre des populations ayant des fréquences alléliques différentes, elle va tendre à modifier les fréquences alléliques : c'est bien une pression évolutive.

Le modèle le plus simple pour mettre en évidence le rôle de la migration est le modèle dit « en île » : soit p_i , fréquence d'un allèle dans la population étudiée (l'« île »), et p_e , la fréquence de cet allèle chez les immigrants (que nous supposons en proportion m par génération); alors :

$$p'_i = (1 - m) \cdot p_i + m \cdot p_e$$

L'écart des fréquences alléliques entre les deux populations diminue :

$$p'_i - p_e = (1 - m) \cdot (p_i - p_e)$$

et la variation de la fréquence due à la migration est :

$$\Delta p_i = p'_i - p_i = m \cdot (p_e - p_i)$$

L'équilibre est atteint lorsque $p_i = p_e$. Comme on pouvait s'y attendre, la migration tend à homogénéiser les fréquences alléliques entre les populations qui échangent des gènes. Elle s'oppose ainsi à la différenciation qui tend à s'installer entre des populations isolées à partir d'un même stock initial, par dérive, par mutation, et éventuellement par sélection si ces populations sont soumises à des pressions sélectives différentes (milieux différents).

Elle s'oppose aussi, en réintroduisant des allèles, à la perte de variabilité que la dérive provoque dans de petites populations. Il suffit pour cela d'un taux faible de migration. Nous avons vu qu'en l'absence de migration, la consanguinité dans une population finie panmictique augmentait à chaque génération selon la relation :

$$f_g = (1/2N) + (1 - 1/2N)f_{g-1}$$

pour atteindre $f = 1$. Si à chaque génération un taux m de nouveaux immigrants participe au pool de gamètes où sont tirés les gènes qui vont former la nouvelle génération, (pour garder un effectif constant, on suppose qu'ils remplacent des individus de la population initiale), la relation devient :

$$f_g = (1/2N) + (1 - 1/2N) \cdot (1 - m)^2 \cdot f_{g-1}$$

(ou bien on tire deux copies du même gène avec une probabilité $1/2N$, ou bien on tire des copies de deux gènes distincts de $g - 1$, qui peuvent être identiques avec une probabilité f_{g-1} à condition qu'aucun des deux ne provienne d'un nouveau migrant : probabilité $(1 - m)^2$; on suppose que les immigrants ne sont ni apparentés ni consanguins).

On va arriver à une valeur d'équilibre f_e qui ne change plus :

$$f_e = (1/2N) + (1 - 1/2N) \cdot (1 - m)^2 \cdot f_e$$

soit :

$$f_e = \frac{1}{2N[1 - (1 - 1/2N)(1 - m)^2]}$$

Si m est faible, m^2 est négligeable devant m , et :

$$f_e \approx \frac{1}{1 + 4N \cdot m - 2m}$$

soit, puisque m est petit :

$$f_e \approx \frac{1}{1 + 4Nm}$$

N étant le nombre d'individus de la population, et m un taux d'immigrants, Nm est le *nombre* de nouveaux immigrants qui participent à la reproduction de la population à chaque génération. On voit donc qu'un nouveau migrant s'établissant par génération suffit à limiter la consanguinité de la population à une valeur maximale de : $f_e = 1/5 = 0,2$ au lieu de $f_e = 1$ sans migration. Chose remarquable, ce résultat est indépendant de la taille de la population : un immigrant par génération compte autant dans une population de dix individus que dans une de un million d'individus ! La solution de ce paradoxe est la suivante : évidemment, un immigrant dans une population d'un million « pèse » très peu, mais d'un autre côté, l'accroissement de consanguinité dû à l'effectif fini est aussi très faible, de sorte que les deux choses se compensent.

4.6.1 Remarque

La notion de migration traduit la difficulté de définition d'une population. En effet, dans la plupart des cas, il est impossible de réaliser une partition satisfaisante de l'espèce en populations. C'est notamment le cas pour l'espèce humaine, mais également pour beaucoup d'animaux et de végétaux. En fait, chaque individu a plus de chances de se reproduire avec un partenaire originaire d'un lieu géographiquement proche qu'avec un individu éloigné. En fait, le plus souvent, le généticien de terrain délimite la population qu'il étudie de façon arbitraire. Il peut ainsi choisir :

- l'espèce humaine,
- la population européenne,
- la population française,
- la population bretonne,
- la population de Plougastel-Daoulas,
- la famille Kervelec.

Dans cette série de populations, le taux de migration varie de 0 pour la première (à notre connaissance) à un pourcentage élevé. Pour la simplicité des calculs, le premier ensemble peut donc apparaître plus satisfaisant que

les autres. Néanmoins, considérer l'espèce humaine comme une population unique est peu intéressant car :

- l'environnement n'est pas homogène sur la surface du globe;
- les gènes ne sont pas remis en commun sur l'ensemble de l'espèce à chaque génération (un gène porté par un Pygmée a très peu de chances de se retrouver chez un esquimau à la génération suivante); ceci se traduit en disant qu'il y a une certaine endogamie;
- en conséquence de ces deux premiers points, les fréquences géniques sont variables d'un endroit à l'autre. En fait, des études récentes ont montré que ces fréquences varient continûment dans l'espace. Deux populations voisines ont des fréquences proches. Ce n'est qu'en choisissant délibérément des groupes éloignés qu'on peut avoir l'impression qu'une discontinuité existe dans l'espèce humaine. Ce genre de résultat a amené certains biologistes à proposer l'idée que les races humaines n'existent pas.

Dans l'idéal, le généticien devra donc, pour chaque cas, définir une population qui minimisera le taux de migration tout en gardant une certaine homogénéité. En fait, les choses ne se passent pas toujours ainsi dans la pratique, ne serait-ce que parce que l'on ne dispose pas des informations nécessaires. À moins de délimitation évidente (cas d'une petite population isolée), la « population » étudiée sera le plus souvent ce qui peut être échantillonné dans un temps raisonnable. Ce faisant, on risque notamment d'avoir une unité plus petite ou plus grande que celle pour laquelle il y a une remise en commun des gènes à chaque génération. Cela a diverses conséquences, dont la plus notable est l'**effet Wahlund** (encadré 5).

ENCADRÉ 5 : L'EFFET WAHLUND ET LA STRUCTURATION SPATIALE DES POPULATIONS

L'effet Wahlund

Soit un ensemble de k populations où les fréquences d'un allèle A sont respectivement $p_1, p_2, \dots, p_i, \dots, p_k$. La fréquence moyenne de A sur l'ensemble est :

$$P = E(p) = \sum p_i \cdot (n_i / N)$$

où n_i est la taille de la $i^{\text{ème}}$ population, et N l'effectif total. La variance s'écrit :

$$V(p) = \sum (p_i - P)^2 \cdot (n_i / N)$$

Si chaque sous-population a la structure de Hardy-Weinberg :

$$AA : D_i = p_i^2 \qquad Aa : H_i = 2p_i q_i \qquad aa : R_i = q_i^2$$

La structure génotypique observée sur l'ensemble correspond aux fréquences génotypiques moyennes :

$$AA : D_0 = \sum (p_i^2) \cdot (n_i / N) = E(p^2)$$

$$Aa : H_0 = \sum (2p_i q_i) \cdot (n_i / N) = 2E(pq)$$

$$aa : R_0 = \sum (q_i^2) \cdot (n_i / N) = E(q^2)$$

D'un autre côté, si l'ensemble avait été une seule grande unité panmictique, on aurait attendu :

$$AA : D_e = P^2 = [E(p)]^2$$

$$Aa : H_e = 2PQ = 2 \cdot E(p) \cdot [1 - E(p)]$$

$$aa : R_e = Q^2 = [1 - E(p)]^2$$

Or la variance est $V(p) = E(p^2) - [E(p)]^2$, et, puisque $q = 1 - p$, $V(q) = V(p)$.

Il est donc aisé de voir que les fréquences génotypiques observées diffèrent de celles qu'on attendrait si on avait affaire à une seule population panmictique :

Structure observée

$$AA : D_0 = E(p^2) = p^2 + V(p)$$

$$Aa : H_0 = 2E(pq) = 2PQ - 2V(p)$$

$$aa : R_0 = E(q^2) = Q^2 + V(p)$$

Les deux structures ne sont équivalentes que si toutes les sous-populations ont les mêmes fréquences alléliques ($V(p) = 0$, c'est-à-dire tous les p_i égaux à P), évidemment ! Mais dès qu'il y a une certaine différenciation entre populations, $V(p)$ est supérieur à 0, et le fait de considérer le tout comme un seul ensemble conduit à observer un **déficit** d'hétérozygotes.

Cet effet est appelé effet Wahlund : lorsque l'on mélange des sous-populations panmictiques dont les fréquences alléliques sont différentes, on observe sur l'ensemble un déficit d'hétérozygotes par rapport à la structure de Hardy-Weinberg, bien que les sous-unités aient chacune la structure de Hardy-Weinberg.

Dans la pratique, il n'est pas toujours aisé de déceler *a priori* si la population que l'on étudie peut ou non être considérée comme une seule unité panmictique; le généticien devra donc toujours songer à l'effet Wahlund comme explication éventuelle d'un déficit observé d'hétérozygotes : son échantillonnage peut avoir inclus plusieurs sous-populations suffisamment isolées entre elles pour qu'une différenciation génétique ait pu s'établir (cas fréquent chez des plantes ou des animaux peu mobiles, lorsque l'on échantillonne sur une surface étendue), ou il comprend deux populations qui, pour une raison ou une autre, ne se croisent pas librement, ou encore il inclut plusieurs générations de composition génétique différente...

Le déficit d'hétérozygotes est mesuré par l'indice de fixation (voir § 3.4 remarque 4) :

$$F = 1 - \frac{H_0}{2PQ} = \frac{2PQ - 2V(p)}{2PQ} = \frac{V(p)}{PQ}$$

Il est d'autant plus fort que les populations sont plus différenciées.

Structure spatiale des populations et « statistiques F » de Wright

Dans une population, plusieurs causes d'écart à la structure de Hardy-Weinberg peuvent se superposer : la population peut être structurée spatialement, avec plusieurs sous-populations différenciées (effet Wahlund); de plus, chaque sous-population peut s'écarter de la structure de Hardy-Weinberg pour diverses causes (régime de reproduction non panmictique, par exemple). Pour décrire de telles situations, S. Wright a défini le jeu de paramètres suivant :

F_{IS} mesure l'écart à la structure de Hardy-Weinberg à l'intérieur des sous-populations (IS pour « individu dans sous-population ») : dans la $i^{\text{ème}}$ sous-populations la proportion d'hétérozygotes est :

$$H_i = 2p_i q_i \cdot (1 - F_{IS})$$

F_{ST} mesure le déficit dû à la différenciation entre sous-populations (ST pour « sous-population dans total ») : $F_{ST} = V(p)/PQ$

F_{IT} mesure l'écart global à la structure de Hardy-Weinberg dans l'ensemble (IT pour « individu dans total ») : $H_0 = 2PQ \cdot (1 - F_{IT})$

Si toutes les sous-populations ont un même indice de fixation F_{IS} , la fréquence d'hétérozygotes observée sur l'ensemble est :

$$H_0 = \sum_i [2p_i q_i (1 - F_{IS})] (n_i/N) = (1 - F_{IS}) E(2pq)$$

Comme $E(2pq) = 2PQ - 2V(p) = 2PQ \cdot (1 - F_{ST})$

il s'ensuit : $H_0 = 2PQ \cdot (1 - F_{IS}) \cdot (1 - F_{ST})$

Donc : $(1 - F_{IT}) = (1 - F_{IS}) \cdot (1 - F_{ST})$

Cette formule permet donc de séparer ce qui, dans la structure génotypique, revient à la structure spatiale en sous-populations différenciées (F_{ST}), et ce qui revient aux effets internes propres aux sous-populations (F_{IS}). Ainsi, à partir des déficits d'hétérozygotes observés dans un ensemble de populations, on peut estimer F_{IS} , F_{ST} et F_{IT} , par exemple pour décrire la structuration de la variabilité (part intra- et part inter-populations).

En fait, la formule est aussi valable pour n'importe quel découpage dans une population, aussi arbitraire soit-il. Une autre de ses utilisations peut être, par exemple, lorsque l'on a échantillonné uniformément dans une espèce répartie sur un grand territoire, de rechercher pour quelle taille maximale des sous-unités le F_{IS} s'annule, en essayant des découpages plus ou moins fins : cela correspondra à la taille maximale sur laquelle on peut considérer qu'il y a panmixie, ce qu'on appelle un **voisinage**, et ceci peut être relié aux distances moyennes de dispersion des gènes par génération.

Le F_{ST} mesure, lui, le degré de différenciation génétique des sous-populations, et il peut notamment être relié aux taux de migration. Ainsi, si les sous-populations se différencient seulement par dérive, et que la migration s'oppose à cette différenciation, on montre que :

$$F_{ST} = \frac{1}{1 + 4nm}$$

où n est la taille des sous-populations, et m le taux de migration par génération (c'est la même formule que celle que nous avons démontrée pour le coefficient de consanguinité).

(Si les F_{IS} , des sous-populations ne sont pas les mêmes d'une population à l'autre, les formules ci-dessus restent valables en prenant comme F_{IS} la moyenne pondérée :

$$F_{IS} = \frac{\sum_i p_i q_i F_{IS}(n_i/N)}{\sum_i p_i q_i (n_i/N)}$$

En fait, les choses se présentent rarement ainsi dans la pratique : on ignore généralement les tailles relatives des sous-unités et on les suppose égales.)

Introduction aux modèles à plusieurs locus : le déséquilibre de liaison

5.1 LA NOTION DE DÉSÉQUILIBRE DE LIAISON

5.1.1 Signification

Connaissant, dans une population, les fréquences génotypiques à un premier locus (par exemple, fréquence de $AA = x$), et les fréquences génotypiques à un deuxième locus (par exemple, fréquence de $BB = y$), il est tentant de penser que la fréquence des individus de génotype $AABB$ dans cette même population est $x.y$. Ce calcul suppose que les événements « l'individu est AA » et « l'individu est BB » sont indépendants. En fait, il peut très bien en être autrement. Ainsi, tel est le cas, chez l'homme, pour les gènes du groupe sanguin « rhésus » (Rh^+/Rh^-) et les gènes de la couleur de la peau : dans la population noire américaine, la fréquence de l'allèle Rh^- au locus rhésus est d'environ 45 %, alors que ce même allèle a une fréquence très faible dans la population blanche, de l'ordre de 3 %. Par conséquent, si l'on considérait la population américaine dans son ensemble, on se tromperait lourdement en supposant qu'il y a indépendance, *au sens statistique du terme* (comme l'on dit que deux événements sont indépendants), entre le génotype au locus rhésus et celui aux gènes de la couleur de la peau. Par

rapport à ce que prévoirait l'indépendance des événements, certaines combinaisons génotypiques sont « anormalement » fréquentes (ici Rh^- avec gènes de peau sombre). Il est très important de réaliser que cela n'implique pas une liaison génétique entre les deux caractères considérés : **linkage génétique et association non aléatoire de deux gènes dans une population sont deux notions entièrement distinctes** (une population n'est pas une F2 mendélienne!). De façon plutôt trompeuse, l'association non aléatoire est appelée **déséquilibre de liaison**.

L'existence d'un déséquilibre de liaison signifie donc que les gènes de *locus différents* ne sont pas associés au hasard dans la population : certaines combinaisons sont plus fréquentes, d'autres moins fréquentes, que ne le prévoirait une association aléatoire de ces différents gènes.

5.1.2 Expression formalisée du déséquilibre

Considérons deux locus :

- le locus 1 avec deux allèles A et a , en fréquences respectives p_A et q_a ,
- le locus 2 avec deux allèles B et b , en fréquences respectives p_B et q_b .

Plutôt que de s'intéresser aux génotypes des individus diploïdes (il en existe trois pour chaque locus, donc neuf au total), il est plus simple de considérer ce qui se passe au niveau des gamètes ; l'indépendance ou la non-indépendance statistique qui existe dans une génération résulte de celle qu'il y avait dans la population des gamètes qui l'ont formée. Il y a seulement quatre combinaisons gamétiques possibles : AB , Ab , aB et ab .

Considérons l'une d'elles, par exemple AB :

- S'il y a association au hasard dans la population gamétique (*indépendance statistique*), c'est-à-dire pas de déséquilibre, on aura :

$$\text{fréquence du gamète } AB : g_{AB} = p_A \cdot p_B$$

- S'il n'y a pas indépendance statistique, c'est-à-dire s'il y a *déséquilibre*, la fréquence de ce gamète sera différente (supérieure ou inférieure). Elle peut toujours s'écrire :

$$g_{AB} = p_A \cdot p_B + D$$

avec $D > 0$ ou $D < 0$

$$\Leftrightarrow D = g_{AB} - p_A \cdot p_B$$

En remplaçant p_A par $g_{AB} + g_{Ab}$ et p_B par $g_{AB} + g_{aB}$, on peut écrire D de la façon suivante :

$$D = g_{AB} \cdot g_{ab} - g_{Ab} \cdot g_{aB}$$

Le paramètre D mesure le déséquilibre de liaison dans la population pour ces deux gènes : l'écart à l'association aléatoire. Il est lui-même appelé **déséquilibre de liaison**.

La connaissance des fréquences alléliques et de ce paramètre D suffit pour calculer les fréquences des quatre types de gamètes dans la population :

- gamètes AB : $g_{AB} = p_A \cdot p_B + D$

- gamètes Ab : les gamètes portant l'allèle A sont en fréquence p_A , donc : $g_{AB} + g_{Ab} = p_A$

$$\Rightarrow g_{Ab} = p_A - g_{AB} = p_A - p_A p_B - D = p_A q_B - D$$

- gamètes aB : de même $g_{aB} + g_{AB} = p_B$, donc :

$$g_{aB} = q_A p_B - D$$

- gamètes ab : $g_{ab} + g_{aB} = q_a \Rightarrow g_{ab} = q_a q_b + D$

ce que l'on peut résumer dans le tableau de contingence suivant (attention, ce n'est pas un tableau de croisement!).

Locus 2 \ Locus 1	A	a	Somme
B	g_{AB} $p_A p_B + D$	g_{aB} $q_A p_B - D$	p_B
b	g_{Ab} $p_A q_B - D$	g_{ab} $q_A q_B + D$	q_B
Somme	p_A	q_A	1

- ✓ Si au lieu de AB , on était parti d'un autre gamète pour définir D , sa valeur absolue calculée à partir de n'importe laquelle des quatre fréquences gamétiques serait bien sûr la même, mais son signe pourrait être inversé. Ce signe est affaire de convention et sans importance.
- ✓ D étant défini à partir des fréquences des gamètes, il est aussi appelé **déséquilibre gamétique**.
- ✓ Le déséquilibre de liaison est maximal lorsque la connaissance de l'allèle présent à un locus permet de prédire à coup sûr l'allèle de l'autre locus, c'est-à-dire lorsque n'existent que deux types de gamètes sur les quatre possibles (AB et ab , ou Ab et aB).

Dans ce cas, nécessairement :

$$p_A = p_B = p \quad q_A = q_B = q$$

et

$$D = pq$$

Ceci est maximal lorsque

$$p = q = 1/2$$

alors

$$D = 1/4$$

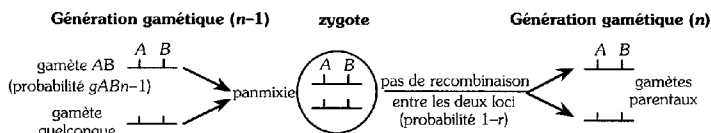
Le déséquilibre est donc compris entre $-1/4$ et $+1/4$ (selon que les deux gamètes existants sont Ab et aB , ou AB et ab).

5.2 ÉVOLUTION DU DÉSÉQUILIBRE DE LIAISON SOUS LES HYPOTHÈSES DE HARDY-WEINBERG

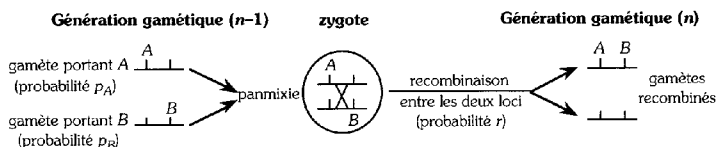
On considère une population où existe un déséquilibre de liaison initial entre deux locus : D_0 à la génération gamétique 0. On suppose que toutes les hypothèses de Hardy-Weinberg sont satisfaites : panmixie et absence de toute pression évolutive (migration, mutation, sélection ou hasard). Les fréquences alléliques vont donc rester constantes. Mais qu'en sera-t-il du déséquilibre de liaison?

Considérons l'un quelconque des gamètes, AB par exemple, à la $n^{\text{ième}}$ génération gamétique. Il est en fréquence g_{ABn} . Un tel gamète AB peut avoir été formé de deux façons différentes (système complet d'événements) :

- soit c'est un gamète parental, c'est-à-dire produit sans recombinaison entre les deux locus à la méiose. Ce gamète existait donc sous la même forme à la génération gamétique $(n-1)$. Si r est le taux de recombinaison entre les deux locus, la probabilité d'obtenir un gamète parental est $(1-r)$ et, sachant qu'il est parental, il est AB avec une probabilité g_{ABn-1} :



- soit c'est un gamète produit par recombinaison (probabilité r). La recombinaison produit un gamète AB à condition qu'elle se fasse entre un gamète portant l'allèle A (probabilité p_A) et un gamète portant l'allèle B (probabilité p_B). Puisque les gamètes s'unissent au hasard (panmixie), la probabilité d'avoir un gamète AB produit par recombinaison est $r \cdot p_A p_B$.



Au total, la fréquence des gamètes AB à la génération n est donc :

$$g_{ABn} = (1-r) \cdot g_{ABn-1} + r \cdot p_A p_B$$

Pour résoudre cette récurrence, on peut écrire :

$$g_{ABn} = (1-r) \cdot g_{ABn-1} - (1-r)p_A p_B + p_A p_B$$

$$(g_{ABn} - p_A p_B) = (1-r) \cdot (g_{ABn-1} - p_A p_B)$$

$$D_n = (1-r) \cdot D_{n-1}$$

$$D_n = (1-r)^n D_0$$

r est compris entre 0 (liaison absolue entre les deux locus) et $\frac{1}{2}$ (indépendance génétique).

- si $r = 0$, la fréquence des gamètes AB ne varie pas d'une génération à l'autre et D est constant; tout se passe comme si on avait un seul locus AB à quatre allèles;

- si r est différent de 0, ce qui est le cas général, $(1-r)$ est strictement inférieur à 1; D ne cesse de diminuer au fil des générations jusqu'à l'équilibre qui est :

$$D = 0 \Leftrightarrow g_{AB} = p_A p_B$$

Ceci justifie l'appellation « déséquilibre » pour ce paramètre : $p_A p_B$ étant la fréquence d'équilibre des gamètes AB , D mesure bien un écart à l'équilibre.

Il diminue d'autant plus vite que les gènes sont génétiquement distants (r grand). En particulier, pour des locus génétiquement indépendants ($r = \frac{1}{2}$), le déséquilibre de liaison diminue de moitié à chaque génération et devient rapidement négligeable.

Néanmoins, cette diminution est toujours progressive. Par conséquent, les fréquences des quatre gamètes évoluent progressivement, de même que celles des génotypes produits par leur union en panmixie, qui se déduisent facilement des fréquences gamétiques. Donc, si on part d'une situation où $D \neq 0$,

l'équilibre des fréquences génotypiques est acquis progressivement, même en panmixie, quand on considère la structure génotypique à plusieurs locus à la fois, alors que, curieusement, pour chaque locus pris séparément, on atteint la structure de Hardy-Weinberg dès la première génération de panmixie.

5.3 CAUSES POSSIBLES D'UN DÉSÉQUILIBRE DE LIAISON

Puisque les déséquilibres tendent à disparaître, pourquoi en observe-t-on dans les populations?

a) Des pressions évolutives peuvent en engendrer :

- Si la population est constituée par un mélange récent de deux populations (ou plus) qui avaient des compositions génétiques différentes, on observera des déséquilibres, qui ne pourront disparaître que progressivement. Un cas évident est celui d'un mélange entre une population n'ayant que les allèles A et B avec une autre n'ayant que les allèles a et b ; mais, en fait, de simples différences de fréquences alléliques entre les populations mélangées suffisent à engendrer un déséquilibre. L'arrivée continue de migrants venant d'une population de constitution génétique différente entretiendra de même un tel déséquilibre. Le mélange de populations est probablement la cause la plus fréquente des déséquilibres de liaison que l'on observe dans les populations humaines, comme celui du couple rhésus-couleur de la peau que nous avons évoqué.

- Certains types de sélection peuvent aussi produire des déséquilibres : on conçoit, intuitivement, que ce puisse être le cas si les combinaisons AB et ab sont sélectivement meilleures que les combinaisons Ab et aB , par exemple (les conditions précises pour que la sélection engendre un déséquilibre sont en fait assez restrictives et complexes à exprimer). Une autre situation où la sélection peut induire un déséquilibre est la suivante : une nouvelle mutation (appelons-la A^*) apparaît à un site chez un individu qui portait un allèle (appelons-le b_1) à un locus situé au voisinage du site muté. Si cette nouvelle mutation est favorable, elle se répand dans la population, entraînant avec elle l'allèle b_1 qui augmente aussi en fréquence. Cet effet est appelé **auto-stop génétique** : le gène b_1 « profite » de sa forte liaison avec le gène sélectionné. Cela engendre donc, au moins transitoirement, un excès de combinaisons A^*b_1 , donc un déséquilibre gamétique, pourvu que la recombinaison qui pourrait réassocier A^* avec un autre allèle du locus b soit assez rare. Ce phénomène est appelé **effet Hill-Robertson**.

- Le simple hasard dans une population finie produit aussi des déséquilibres : le tirage des gamètes qui vont former une génération contient par hasard un excès de combinaisons AB et ab , par exemple.

b) Quelles que soient les causes à l'origine d'un déséquilibre de liaison, celui-ci se maintiendra d'autant plus facilement et longtemps qu'il y aura moins de recombinaisons entre les locus impliqués. Par conséquent :

- Dans les populations, on observe plus souvent des déséquilibres entre des marqueurs très liés qu'entre des marqueurs génétiquement indépendants; ainsi, maintenant qu'on séquence l'ADN, on trouve beaucoup de cas de déséquilibres entre des sites séparés seulement par quelques centaines de paires de bases, alors que les déséquilibres de liaison étaient considérés comme rares dans les mêmes espèces lorsqu'on n'avait que les marqueurs classiques.
- On observe plus souvent des déséquilibres (même entre marqueurs indépendants) chez des espèces à reproduction asexuée et chez des espèces autogames (où la recombinaison existe mais rompt rarement les associations puisqu'elle se fait généralement chez des homozygotes); c'est une des caractéristiques de la structure des populations autogames que les déséquilibres y sont très fréquents, alors qu'il est rare d'en trouver chez les allogames autrement qu'à l'échelle moléculaire. Dans le même ordre d'idées, et pour reprendre l'exemple rhésus-couleur de la peau, il est clair que l'homogamie pour la couleur de la peau aux États-Unis doit ralentir l'effacement du déséquilibre.

Exercices du chapitre A

1 Structure génétique des populations

Exercice A.1.1

Généralement, il n'est pas possible de calculer les fréquences génotypiques à partir des seules fréquences alléliques, à moins de faire des hypothèses supplémentaires (en supposant la panmixie par exemple). Cela est néanmoins possible lorsque l'un des génotypes manque.

Exprimer les fréquences des deux génotypes en présence en fonction des fréquences alléliques p (allèle A) et q (allèle a) dans les cas suivants :

1)	Génotypes	AA	Aa	aa
	Fréquences		0	
2)	Génotypes	AA	Aa	aa
	Fréquences			0
3)	Génotypes	AA	Aa	aa
	Fréquences	0		

Exercice A.1.2

Une étude de la diversité génétique d'une population de Hêtres (*Fagus sylvatica* L.) d'une forêt de Basse-Saxe, en Allemagne, a fourni les résultats suivants, pour deux locus enzymatiques :

- Locus de la 6-phosphogluconate déshydrogénase (6-PGDH, 2 allèles, A_1 et A_2)

Génotypes	A_1A_1	A_1A_2	A_2A_2
Effectifs observés	125	31	4

- Locus de la leucine aminopeptidase (LAP, 3 allèles, B_1 , B_2 et B_3)

Génotypes	B_1B_1	B_1B_2	B_2B_2	B_1B_3	B_3B_3	B_2B_3
Effectifs observés	121	0	2	28	8	1

Calculer les fréquences alléliques à ces deux locus.

(Référence : Gregorius *et al.*, 1986, *Heredity* **57**, pp. 255-262.)

Exercice A.1.3

Chez les chats tigrés, l'un des gènes déterminant la couleur du pelage est situé sur le chromosome X, il s'agit du gène appelé « orange ». Il comprend un allèle O donnant une couleur rousse et un allèle « gris » o . Les femelles possèdent une paire de chromosomes X, les mâles un chromosome X et un

chromosome Y. Dans une population, on observe M mâles parmi lesquels M_o sont roux et M_n sont « non roux ». Parmi les F femelles, F_o sont rousses, F_n sont « non rousses » et F_e ont un pelage « écailles de tortue », mosaïque de gris et de roux.

1) Pourquoi n'observe-t-on pas de mâles ayant un pelage « écailles de tortue » ?

2) Calculer la fréquence des allèles roux parmi les chromosomes X des femelles d'une part ($f_f(O)$), des mâles d'autre part ($f_m(O)$), ainsi que dans la population totale ($f_t(O)$).

2 Étude théorique d'un cas idéal

Exercice A.2.1

Dans une grande population panmictique, deux types de plantes sont observés. Certaines ont des fleurs blanches, d'autre des fleurs rouges. En effectuant des croisements et des autofécondations dans la population, les résultats suivants ont été obtenus :

- tous les descendants issus des plantes à fleurs blanches, autofécondées ou croisées entre elles, présentaient des fleurs blanches,
- les descendants d'un premier type de plantes à fleurs rouges, autofécondées ou croisées avec n'importe quelle autre plante, possédaient des fleurs rouges,
- un deuxième type de plantes à fleurs rouges, autofécondées, donnait environ trois descendants à fleurs rouges pour un descendant à fleurs blanches.

1) Quel serait le résultat du croisement du deuxième type de plantes à fleurs rouges

a) Avec une plante à fleurs blanches?

b) Avec une plante à fleurs rouges?

2) En échantillonnant au hasard 400 plantes dans la population, 335 plantes à fleurs rouges et 65 plantes à fleurs blanches ont été récoltées. À partir de cet échantillon, estimer les fréquences alléliques au locus considéré dans la population. En déduire le nombre attendu de plantes hétérozygotes dans l'échantillon.

3) Les descendants d'une plante à fleurs blanches, fécondée librement dans la population, sont récoltés. En tenant compte des résultats de la question précédente, quelles proportions des différents génotypes sont attendues parmi ces descendants? Quelles seront les proportions des différents phénotypes? Mêmes questions si la récolte se fait, dans les mêmes conditions, sur une plante à fleurs rouges du deuxième type.

Exercice A.2.2

Des daphnies (*Daphnia cucullata*) collectées dans un lac du Vermont, aux États-Unis, ont été étudiées pour leur génotype au locus d'une enzyme, la phosphoglucotase. Deux allèles A_1 et A_2 ont été détectés et les génotypes se répartissaient de la façon suivante :

Génotypes	A_1A_1	A_1A_2	A_2A_2
Effectifs observés	38	91	62

La structure génétique d'une population de coléoptères troglodytes (*Speonomus zophosinus*) du Cloutet, dans les Pyrénées, a été étudiée pour le locus de la leucine aminopeptidase. Deux allèles (B_1 et B_2) étaient présents et la population était structurée de la façon suivante :

Génotypes	B_1B_1	B_1B_2	B_2B_2
Effectifs observés	112	34	28

Dans ces deux exemples, tester si l'hypothèse d'un régime de reproduction panmictique est acceptable.

(Références : Mort et Wolf 1985, *Heredity* **55**, pp. 27-36, Crouau-Roy 1988, *Heredity* **60**, pp. 321-327.)

Exercice A.2.3

Soit, dans une population, un locus à deux allèles, l'allèle A étant dominant sur l'allèle a . Parmi 100 individus, on observe 64 individus de phénotype $[A]$ et 36 de phénotype $[a]$. Quelles sont les fréquences alléliques? Quelle hypothèse faites-vous pour réaliser le calcul? Montrer que l'on ne peut pas tester cette hypothèse avec les données fournies.

Exercice A.2.4

Chez les chats, un locus à trois allèles T^a , T et t^b , est lié à l'aspect des rayures du pelage. L'allèle T^a est responsable du phénotype abyssin qui est caractérisé par un pelage de couleur homogène, sans rayures. L'allèle T conduit au phénotype tigré, caractérisé par deux traits dorsaux et des marques rondes sur les flancs. La tête, la queue et les pattes sont rayées. L'allèle t^b est responsable du phénotype marbré, distingué du précédent par ses flancs tachetés. L'allèle T^a est dominant sur l'allèle T lui-même dominant sur l'allèle t^b . Dans une population, les fréquences de ces trois allèles sont p , q et r respectivement. On peut considérer que la population se reproduit en panmixie.

- 1) Quelles sont les fréquences attendues des différents génotypes?
- 2) Quelles sont les fréquences attendues des différents phénotypes?

La population comporte 12 chats marbrés, 123 chats tigrés et 87 chats abyssins.

- 3) Calculer les fréquences alléliques.

Exercice A.2.5

La mucoviscidose est une maladie génétique due à un allèle muté du gène CFTR. Cet allèle, récessif, est à l'origine d'une diminution du transport des ions chlorure à travers la membrane des cellules des muqueuses. Le mucus

sécrété par ces cellules, déshydraté et visqueux, obture alors les voies aériennes et digestives.

En Europe, un enfant est atteint sur 2 500 naissances.

1) Évaluer la fréquence des individus porteurs d'un allèle muté à l'état hétérozygote. Quelles sont les hypothèses nécessaires pour faire ce calcul?

2) Parmi les allèles mutés, quelle proportion se trouve chez des individus hétérozygotes et quelle proportion chez des individus atteints?

Exercice A.2.6

Chez la *Drosophile*, un locus déterminant la couleur de l'œil est porté par le chromosome X. Les femelles portent deux chromosomes X, les mâles portent un chromosome X et un chromosome Y. À ce locus, l'allèle w (white) donne une couleur blanche à l'œil et est récessif par rapport à l'allèle sauvage w^+ .

Dans une cage à population, on introduit un nombre égal de mâles de race w et de femelles de race pure w^+ . La population se reproduit en panmixie. On admet qu'il n'y a ni mutation, ni sélection et que la population est infiniment grande.

1) Parmi les chromosomes X, calculer la fréquence de l'allèle w^+ chez les femelles (p_f) chez les mâles (p_m) et dans la population totale (p) pour la génération fondatrice G_0 , et pour les générations G_1 et G_2 . Il y a un nombre égal de mâles et de femelles à chaque génération.

2) Calculer $p_m(n)$, $p_f(n)$ et $p(n)$ à la génération n en fonction de $p_m(n-1)$, $p_f(n-1)$ et $p(n-1)$, ainsi que l'écart $E_n = p_f(n) - p_m(n)$ en fonction de E_{n-1} .

3) En déduire les fréquences p_f , p_m et p à l'équilibre.

Exercice A.2.7

Le daltonisme est un dysfonctionnement de la vision des couleurs dû à un gène porté par le chromosome X. Un allèle récessif est responsable de cette anomalie. Waaler, en 1927, étudie la fréquence du daltonisme parmi une population d'enfants de Norvège. Parmi 9049 garçons, 725 sont daltoniens et 8 324 ne le sont pas.

1) Quelle est la fréquence q de l'allèle récessif dans cette population? En déduire la fréquence attendue de filles daltoniennes dans la population.

2) Sur les 9 072 filles de son échantillon, Waaler dénombre 40 daltoniennes. Cette observation est-elle cohérente avec les résultats de la question précédente? Comment expliquer la différence?

3 Influence du régime de reproduction

Exercice A.3.1

1) Démontrer que, pour une population ayant une structure génotypique quelconque :

Génotype	AA	Aa	aa
Fréquences	D	H	R

La statistique du χ^2 calculée dans un test de conformité à la structure de Hardy-Weinberg peut s'écrire :

$$\chi^2 = F^2 N \text{ où } N \text{ est la taille de l'échantillon et } F = 1 - \frac{H}{2pq}$$

2) En déduire la taille de l'échantillon nécessaire pour déceler, par l'examen de la structure génotypique en un locus biallélique, l'existence d'une consanguinité avec $F = 0,1$ (c'est-à-dire pour que le test soit significatif au seuil 5 %).

Exercice A.3.2

Une population est analysée à un locus particulier. Elle présente la composition suivante :

Génotypes	AA	Aa	aa
Fréquences	D_0	H_0	R_0

Cette population est d'effectif illimité et ne fait l'objet d'aucune pression évolutive. Comment évoluent les fréquences alléliques d'une part, les fréquences génotypiques d'autre part si le régime de reproduction de cette population est :

- la panmixie,
- l'autogamie,
- un régime mixte : chaque individu est autogame, sauf pour une proportion t de ses gamètes, qui se reproduisent en panmixie.

Dans le dernier cas, exprimer H_n , fréquence des hétérozygotes à la génération n en fonction de H_{n-1} et chercher la fréquence d'équilibre H_e . Quelle est, à l'équilibre, la valeur de l'indice de fixation F ?

Exercice A.3.3

Dans une population de plantes hermaphrodites existent, à un locus, deux allèles A et a de fréquences $(1-q)$ et q . Pour chaque plante, une

proportion s des fécondations se fait en autogamie. Le reste a lieu en allogamie et il y a alors panmixie pour le couple A, a .

La population comporte des individus de phénotypes $[A]$ et $[a]$. Des graines sont récoltées séparément sur des individus (fécondés librement) des deux phénotypes et sont semées :

- certains individus de phénotype $[A]$ n'ont que des descendants de phénotype $[A]$,
- d'autres individus de phénotype $[A]$ donnent une proportion $(1 - x)$ d'individus de phénotype $[A]$ et une proportion (x) d'individus de phénotype $[a]$,
- les individus de phénotype $[a]$ donnent une proportion $(1 - y)$ d'individus de phénotype $[A]$ et une proportion (y) d'individus de phénotype $[a]$.

Évaluer q et s en fonction de x et de y .

Exercice A.3.4

Chez certaines plantes, comme par exemple le trèfle, existe un locus particulier, le locus S , responsable de ce que l'on appelle auto-incompatibilité. Un grain de pollen possédant un allèle donné ne pourra pas féconder une plante possédant le même allèle.

1) Quelle est l'influence de ce système sur le régime de reproduction? Quelle est la proportion d'hétérozygotes au locus S ? Que penser de cette proportion aux autres locus?

2) Quel est le nombre minimum d'allèles qui peuvent coexister à ce locus dans une population? Calculer l'évolution des fréquences génotypiques d'une génération à l'autre dans ce cas. Quelles seront les valeurs de ces fréquences à l'équilibre?

3) Dans cette population, à l'équilibre, un nouvel allèle apparaît par mutation. Que va-t-il devenir? Que peut-on en déduire sur le nombre d'allèles que peut posséder une population au locus S ?

Exercice A.3.5

Chez une espèce végétale allogame annuelle coexistent deux types d'individus dont les dates de floraison diffèrent, appelés type (1) et (2). En dehors de cette particularité, ces individus sont identiques. L'espèce peut pousser dans deux milieux, au soleil ou à l'ombre, ce qui a une influence sur sa date de floraison. Les dates de floraison observées, en fonction du type de plante et du milieu, sont données dans la partie I du tableau.

Une plante de type (1) a été croisée avec une plante de type (2) l'année 0. Les graines issues de ce croisement ont été semées l'année suivante dans

chacun des milieux et présentaient les caractéristiques données dans la partie II du tableau.

Les plantes obtenues se sont croisées librement, chaque milieu constituant une population totalement isolée, et ont produit la génération F2. Dans cette génération a été dénombré, sur un total de 4000 plantes, le nombre de plantes fleurissant à chaque date (partie III du tableau).

	Milieu	Soleil	Ombre
I	Type de plantes : (1)	1 ^{er} juin	15 juin
	(2)	15 juin	30 juin
II	Croisement F1 : (1) × (2) Année 0	1 ^{er} juin	15 juin
III	Génération F2	1 ^{er} juin : 3000	15 juin : 3000
	Année 1 (4000 plantes)	15 juin : 1000	30 juin : 1000

Les générations suivantes seront constituées par le semis des graines dans le milieu où se trouvait leur mère.

1) Quel pourrait être le déterminisme génétique de la date de floraison dans cette espèce?

2) Quels sont les effectifs attendus des différents génotypes aux différentes dates de floraison parmi les plantes observées l'année 1 dans les deux milieux?

3) Les fréquences génotypiques vont-elles être modifiées? Pour quelle raison? Qu'en est-il des fréquences alléliques?

4) Dans la population située au soleil :

- quelles sont les fréquences alléliques (p_1 et $q_1 = 1 - p_1$) dans le pollen le 1^{er} juin de l'année 1?

- quelles sont, en fonction de p_1 et q_1 , les fréquences des différents génotypes dans les graines des plantes fleurissant le 1^{er} juin de l'année 1?

- quelles seront, en fonction de p_1 et q_1 , les fréquences alléliques p_2 et q_2 parmi les plantes fleurissant au 1^{er} juin de l'année 2?

- calculer q_n en fonction de q_1 en utilisant la suite des inverses $1/q_n$. Comment va évoluer la population? Quelles seraient, à l'équilibre, les fréquences des plantes fleurissant aux différentes dates?

5) Que va-t-il se passer dans la population se trouvant à l'ombre?

6) Que se passerait-il si les deux populations, au lieu d'être isolées, étaient proches et pouvaient librement échanger du pollen? (montrer que si

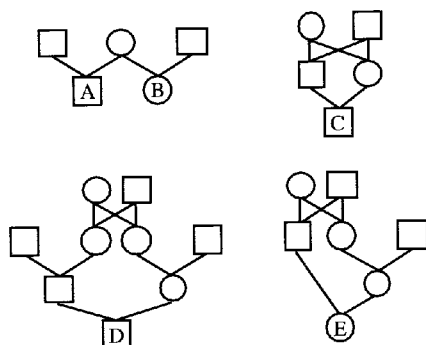
les flux polliniques sont symétriques entre les populations, les flux alléliques ne le sont pas et en déduire que les allèles peuvent être piégés). Quel serait l'état d'équilibre des dates de floraison dans les deux milieux?

Exercice A.3.6

Quel est le coefficient d'apparentement entre deux demi-frères/sœurs (individus A et B)?

Quel est le coefficient de consanguinité d'un individu issu du croisement entre :

- un frère et une sœur (individu C),
- deux cousins germains (individu D),
- un oncle et une nièce (individu E)?



Exercice A.3.7

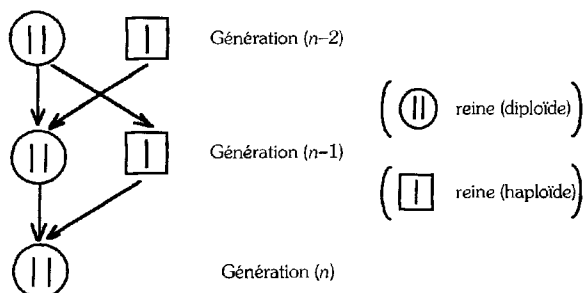
Quel est le coefficient de consanguinité de l'individu K, issu de deux générations successives d'autofécondation, son ancêtre I étant non consanguin?

Quel serait le coefficient de consanguinité d'un individu issu de n générations d'autofécondation?



Exercice A.3.8

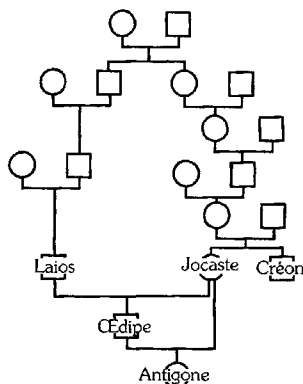
Dans les populations d'abeilles domestiques, à chaque génération, seule une femelle est fertile, la reine. Dans une population isolée, la reine est fécondée par un certain nombre de ses frères. Ses ovules, s'ils sont fécondés, donnent des femelles diploïdes. Les mâles, haploïdes, proviennent d'ovules non fécondés. La généalogie de trois générations de reines est la suivante :



À l'aide de ce schéma, exprimer f_n , coefficient de consanguinité de la reine de la génération n , en fonction de f_{n-1} et f_{n-2} , coefficients de consanguinité des reines des deux générations précédentes.

Exercice A.3.9

Dans la mythologie grecque existe une famille célèbre dont voici une généalogie simplifiée. À partir de ce schéma, calculer le coefficient de consanguinité d'Antigone ainsi que son coefficient de parenté avec son oncle Créon. On supposera qu'en dehors des informations de la figure, il n'existe aucune autre relation de parenté entre les individus.



4 Influence des pressions évolutives

Exercice A.4.1

Soit deux plantes hétérozygotes à un locus, de génotype Aa . Pour constituer la génération suivante, les deux plantes se croisent librement et deux graines sont prélevées au hasard pour obtenir deux plantes.

Quelles sont les probabilités de trouver, dans cette deuxième génération, 0, 1, 2, 3 ou 4 copies de l'allèle A ? Quelle est alors la probabilité d'avoir conservé le polymorphisme?

Exercice A.4.2

Chez une espèce bactérienne, un gène code pour une enzyme permettant la transformation de l'indole en tryptophane. L'allèle A produit une enzyme active, mais les bactéries portant l'allèle a , inactif, ont besoin de tryptophane dans le milieu de culture pour se développer.

1) Une culture mixte de cette espèce est constituée à partir d'une seule bactérie de chaque type (A et a), dans un milieu contenant du tryptophane en quantité légèrement limitante. Dans ces conditions, les bactéries de type a se divisent un peu plus lentement que celles du type A : une division toutes les 30 minutes en moyenne pour les bactéries de type A , toutes les 31 minutes pour les bactéries de type a .

Combien de temps faudra-t-il pour que la proportion de bactéries de type a ne représente plus que 1/1 000 du total?

2) Une culture mixte est constituée de la même façon dans un milieu contenant du tryptophane en quantité largement excédentaire par rapport aux besoins des bactéries. Les bactéries de type A et a poussent alors à la même vitesse (une division toutes les 30 minutes en moyenne). Si l'allèle A mute en a avec une probabilité égale à 10^{-6} à chaque division, combien de temps faudra-t-il pour que la proportion de bactéries de type A soit réduite à 1/1 000 du total?

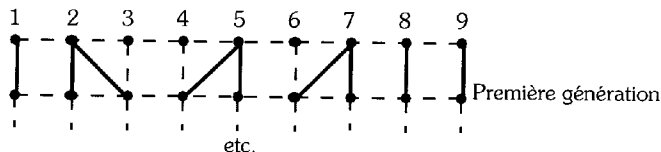
3) En fait, la culture est réalisée dans des flacons de 100 ml, et la croissance s'arrête quand les bactéries ont atteint une densité de 10^8 cellules par ml (phase stationnaire). Au départ, on utilise un nombre égal de bactéries de chaque type. Quand la phase stationnaire est atteinte, on repique un échantillon de la culture dans un milieu neuf de même composition, et on recommence indéfiniment. Comment la constitution de la culture va-t-elle évoluer :

a) s'il n'y a pas de mutation?

b) s'il y a une mutation de A en a avec une fréquence de 10^{-6} , sans mutation de a en A ?

Exercice A.4.3

Soit une population comportant neuf individus haploïdes. À chaque génération, chaque descendant a pour unique parent un individu tiré au hasard parmi les neuf individus de la génération précédente. La taille de la population est constante. L'évolution de cette population peut être simulée en utilisant une table de nombres tirés au hasard. Le tableau comprend, dans chaque colonne, neuf chiffres tirés aléatoirement de 1 à 9. On choisit de représenter chaque génération par une colonne de ce tableau, elle indique alors quel individu a laissé un descendant à la génération suivante. (Si un individu est cité deux fois, c'est qu'il a laissé deux descendants; s'il est cité n fois, c'est qu'il en a laissé n .) Ainsi, la génération engendrée par la première colonne est constituée d'un descendant des individus 1, 8 et 9 et de deux descendants des individus 2, 5 et 7. Cette situation peut être représentée de la façon suivante :



La deuxième génération peut être engendrée de la même façon à partir de la deuxième colonne du tableau (ou d'une autre) et ainsi de suite.

1) En partant d'une colonne quelconque du tableau, tracer l'évolution d'une population pendant une vingtaine de générations.

Quel peut être le devenir d'un allèle, assimilé à un individu de la population?

Trouver l'ancêtre commun le plus récent de deux individus quelconques de la dernière génération.

Trouver l'ancêtre commun le plus récent des 9 individus de la dernière génération.

2) Soit une population de N individus haploïdes. On va supposer maintenant que chaque individu d'une génération est le descendant d'un des N parents de la génération précédente au hasard, chaque descendant étant tiré indépendamment des autres. Calculer la probabilité pour que deux individus d'une même génération soient issus du même parent. Calculer la probabilité $P(X = 2)$ pour que deux individus d'une même génération possèdent un ancêtre commun deux générations avant, trois générations puis g générations auparavant ($P(X = g)$). Calculer l'espérance de cette probabilité :

$$E(P(X)) = \sum_{k=1}^{\infty} k P(X = k)$$

Tableau 1 : Nombres au hasard

7	1	3	9	2	3	3	8	4	4	9	9	9	7	2	7	4	2
5	9	2	5	3	5	6	8	7	9	2	7	5	9	2	8	3	6
9	2	8	3	2	4	6	1	8	5	7	6	3	1	2	7	2	4
8	5	2	7	9	4	1	2	1	5	3	2	7	2	4	6	4	3
1	1	5	5	9	6	7	4	4	1	6	1	7	5	5	8	8	9
2	5	8	8	8	4	5	4	1	6	2	7	7	9	3	7	1	7
7	8	5	3	9	1	4	8	8	3	5	3	5	3	5	6	3	6
2	6	8	8	2	6	3	4	1	6	3	9	4	9	4	8	7	5
5	6	2	2	4	3	8	1	5	8	6	3	4	4	8	2	3	3

7	1	3	9	2	3	3	8	4	4	9	9	9	7	2	7	4	2
5	9	2	5	3	5	6	8	7	9	2	7	5	9	2	8	3	6
9	2	8	3	2	4	6	1	8	5	7	6	3	1	2	7	2	4
8	5	2	7	9	4	1	2	1	5	3	2	7	2	4	6	4	3
1	1	5	5	9	6	7	4	4	1	6	1	7	5	5	8	8	9
2	5	8	8	8	4	5	4	1	6	2	7	7	9	3	7	1	7
7	8	5	3	9	1	4	8	8	3	5	3	5	3	5	6	3	6
2	6	8	8	2	6	3	4	1	6	3	9	4	9	4	8	7	5
5	6	2	2	4	3	8	1	5	8	6	3	4	4	8	2	3	3

c'est-à-dire le nombre de générations qu'il faut remonter en moyenne pour trouver un ancêtre commun à deux individus quelconques d'une génération, dans une population d'individus haploïdes.

On rappelle que :

$$\sum_{k=0}^n x^k = \frac{1-x^{n+1}}{1-x}$$

3) Écrire un programme en basic simulant ce processus.

Exercice A.4.4

Un troupeau bovin laitier comprend 15 vaches et 2 taureaux. Quel est son effectif génétique?

Exercice A.4.5

Soit une population d'individus haploïdes. À un locus donné, on observe deux allèles A et a dans la population. Les individus portant l'allèle A (respectivement l'allèle a) ont une valeur sélective w_A (respectivement w_a). En l'absence de toute autre pression évolutive, si p est la fréquence de l'allèle A à une génération, exprimer la valeur de p' , fréquence de A à la génération suivante. En déduire l'expression de $\Delta p = p' - p$. Comment la population va-t-elle évoluer? Y a-t-il un équilibre polymorphe possible?

Exercice A.4.6

Dans une population de maïs, espèce allogame, existe un locus pour lequel les valeurs sélectives relatives sont les suivantes :

Génotypes	AA	Aa	aa
Valeurs sélectives	$3/4$	1	0

1) Donner en fonction de p , fréquence de l'allèle A , les variations de cette fréquence à chaque génération.

2) Existe-t-il un état d'équilibre polymorphe $pq \neq 0$?

Exercice A.4.7

Un aviculteur possède deux populations de poules, l'une est composée de poules blanches à pattes jaunes (souche A), l'autre est composée de poules rousses à pattes blanches (souche B). Ces deux souches sont destinées à produire en croisement ($A \times B$) des poulets commercialisables. Dans un premier temps, l'aviculteur laisse, à l'intérieur de chaque population, les individus se reproduire entre eux. La souche A donne une descendance A' d'individus à pattes jaunes, parmi eux 99 % ont des plumes de couleur blanche et 1 % des plumes de couleur rousse. La descendance B' de la souche B est constituée d'individus à plumes rousses parmi lesquels 97 % présentent des pattes blanches et 3 % des pattes jaunes.

1) Quel est le déterminisme génétique de la couleur des pattes et du plumage? (On choisira la solution la plus simple, un seul locus pour chaque caractère.)

2) Dans le produit du croisement C , quelle est la proportion s de poules répondant au standard (plumes et pattes blanches)?

On élimine dans la descendance A' les individus à plumes rousses et dans la descendance B' les individus à pattes jaunes. On réalise alors le croisement $C' = (A' \times B')$.

3) Quelle sera dans ce cas la proportion s' d'individus répondant au standard? Quel est le gain de l'aviculteur du fait de cette sélection?

Exercice A.4.8

L'anémie falciforme est une maladie génétique humaine liée à un allèle létal récessif. En Afrique, les individus hétérozygotes à ce locus sont protégés de la malaria. Leur valeur sélective vaut par conséquent 1,18, celle des homozygotes étant de 0,97.

1) Ce locus est dans une situation d'équilibre en Afrique. Quelle est la fréquence q_0 de l'allèle récessif? Quelles sont les fréquences génotypiques?

2) En Amérique du Nord, la malaria est absente : les hétérozygotes ne sont plus favorisés, leur valeur sélective est égale à celle des homozygotes. Dans la population noire américaine, on observe 12 % d'hétérozygotes et 88 % d'homozygotes. Exprimer l'évolution de la fréquence de l'allèle récessif d'une génération à l'autre ($1/q_n$ en fonction de $1/q_{n-1}$ puis de $1/q_0$). En déduire le nombre de générations écoulées depuis l'arrivée de la population africaine en Amérique, liée à l'esclavage.

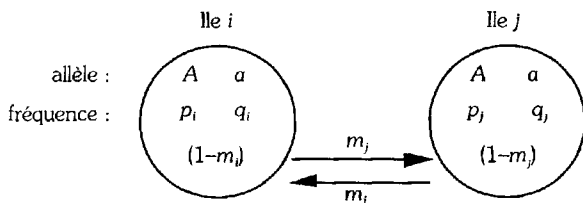
Exercice A.4.9

Dans une espèce végétale annuelle aux graines disséminées par le vent, on considère un locus à deux allèles A et a , non soumis à la sélection. Cette espèce se trouve sur deux îles voisines. Les fréquences alléliques sont différentes dans les deux îles. À la génération n , la population de l'île i contient une proportion m_i d'individus issus de graines venues de l'autre île, et une proportion $1 - m_i$ d'individus descendant des plantes déjà en place. La situation est parallèle dans l'île j . On suppose que la migration est indépendante du génotype au locus considéré, qu'il n'y a pas de mutation, et que les populations sont suffisamment grandes pour que la dérive puisse être négligée.

1) Calculer les fréquences $P_i(n)$ et $P_j(n)$ de l'allèle A à la génération n en fonction des proportions m_i , m_j et des fréquences $P_i(n-1)$, $P_j(n-1)$.

2) En supposant que les taux de migration sont constants dans le temps, déterminer la situation d'équilibre.

3) Soit E_n l'écart entre les fréquences de l'allèle A dans les deux populations à la génération n . Combien de générations seront nécessaires pour que cet écart soit réduit de moitié si $m_i = 1\%$ et $m_j = 2\%$?



Exercice A.4.10

Le coquelicot est une espèce auto-incompatible. Un ensemble de plantes a été étudié à un locus présentant deux allèles, A_1 et A_2 . Les effectifs suivants ont été observés :

Génotype	A_1A_1	A_1A_2	A_2A_2
Effectifs	80	80	40

1) Calculer les fréquences alléliques. Cette population a-t-elle la structure de Hardy-Weinberg?

La population étudiée résulte en fait du mélange de plantes prélevées en deux lieux distincts. En respectant l'origine des plantes, les effectifs suivants sont observés :

	Lieu 1			Lieu 2		
Génotype	A_1A_1	A_1A_2	A_2A_2	A_1A_1	A_1A_2	A_2A_2
Effectifs	62	33	5	18	47	35

2) Qu'en est-il des structures génotypiques à ces deux endroits? Qu'en déduisez-vous sur la cause du déficit trouvé sur l'ensemble?

5 Introduction aux modèles à plusieurs locus : le déséquilibre de liaison

Exercice A.5.1

La composition d'un échantillon de gamètes d'une population est connue pour deux locus, l'un à trois allèles, A_1 , A_2 et A_3 et l'autre à deux allèles B_1 et B_2 . Les gamètes se répartissent de la façon suivante :

Locus 1 Locus 2	A_1	A_2	A_3	Total
B_1	25	9	18	52
B_2	3	39	6	48
Total	28	48	24	100

Combien de paramètres sont nécessaires pour décrire les déséquilibres gamétiques dans ce cas? Quelles sont leurs valeurs?

Exercice A.5.2

Les graines de deux lignées de maïs (plante allogame dont la reproduction peut être considérée comme panmictique) sont mélangées en proportions p et q . Ces deux lignées diffèrent notamment pour deux locus bi-alléliques (A , a et B , b) :

Lignée 1 : génotype $AA BB$

Lignée 2 : génotype $aa bb$

Ce mélange est semé au champ, il constitue la génération G_0 . Les graines sont ressemées d'année en année, avec des effectifs assez grands pour que la dérive puisse être négligée. On suppose qu'aucune sélection ne s'exerce, que le champ est isolé et que le taux de mutation est négligeable.

Quelle sera la composition de la population dans les générations G_0 , G_1 , G_n :

- au locus A/a ?
- au locus B/b ?
- aux deux locus simultanément?

En déduire la valeur du déséquilibre gamétique dans ces générations.

Application numérique : $p = q = 1/2$; $r = 1/2$ (les deux locus sont indépendants).

Suivre l'évolution de la fréquence des génotypes $aa bb$ et $aa BB$ (calcul pour G_1 à G_5 , G_∞ et graphe).

B

Le cryptopolymorphisme

Généralités

Une des causes du maintien de la variabilité dans les populations est le fait que la mutation produit constamment des gènes nouveaux. Le plus souvent, cependant, le résultat des mutations est défavorable : la protéine produite par le gène muté fonctionne moins bien, ou même le gène n'est plus du tout fonctionnel (cas d'une délétion ou d'une mutation « non-sens », par exemple). Si le gène muté apporte un désavantage sélectif aux individus qui le portent, la sélection tend à l'éliminer. Toutefois ce genre de mutation se produit sans cesse, réintroduisant ces gènes défavorables dans les populations. Ce doit être un cas fréquent : la plupart des maladies génétiques chez les humains correspondent sans doute à une telle situation, et plus généralement toutes les mutations entraînant la perte de fonction des gènes. Tous les locus sont concernés : il ne doit guère y avoir de gènes dont la perte de fonction ne soit plus ou moins défavorable. On peut penser que si de tels gènes existaient, ils cesseraient bientôt d'être fonctionnels en accumulant des mutations diverses ou en subissant des délétions. Tel est peut-être d'ailleurs le cas de certains « pseudogènes » ; ce sont des séquences de l'ADN ayant plus ou moins l'organisation et la séquence de gènes fonctionnels, dont ils seraient à l'origine des copies redondantes apparues par duplication, mais avec des anomalies diverses, telles que l'insertion de codons « stop », qui les rendent non fonctionnels.

Si l'allèle muté a est récessif (ce qui est le cas le plus fréquent), les valeurs sélectives relatives que l'on peut attribuer aux trois génotypes sont :

AA	Aa	aa
$w_1 = 1$	$w_2 = 1$	$w_3 = 1 - s$

s représentant le désavantage sélectif des individus *aa* qui expriment la mutation par rapport aux individus normaux (pour $s = 0$, il n'y a pas de désavantage; pour $s = 1$, *a* est **létal**).

Dans une population panmictique, la pression de sélection a pour valeur :

$$\Delta_s p = pq \frac{sq}{1 - sq^2}$$

La pression de mutation est, quant à elle :

$$\Delta_m p = -up$$

avec **u** le taux de mutation de *A* en *a* par génération.

$\Delta_s p$ est positive : la sélection tend à faire augmenter la fréquence de l'allèle « normal » et à éliminer l'allèle muté. $\Delta_m p$ est négative : la mutation fait diminuer la fréquence de *A*. On atteint un équilibre où la fréquence des gènes ne change pas d'une génération à l'autre lorsque les deux pressions s'équilibrent, c'est-à-dire :

$$\Delta_s p + \Delta_m p = 0$$

soit :

$$\frac{spq^2}{1 - sq^2} = up$$

$$\frac{sq^2}{1 - sq^2} = u$$

d'où :

$$sq^2 = \frac{u}{1 + u} \approx u$$

Le taux de mutation u étant faible (de l'ordre de 10^{-5} ou 10^{-6}), il apparaît que sq^2 doit être faible, ce qui veut dire que :

- soit **s est très faible**, auquel cas :

$$w_1 = w_2 \approx w_3$$

Il n'y a donc pratiquement pas de sélection, et dans ce cas, nous avons eu tort de négliger le hasard, qui doit être la pression évolutive principale : il n'y a pas vraiment d'équilibre et la dérive va jouer (éventuellement en fixant l'allèle défavorable);

- soit **q est faible**, auquel cas nous avons bien un équilibre entre l'effet de la mutation, qui tend à accroître la fréquence du gène défavorable, et celui de la sélection qui tend à l'éliminer, mais cet équilibre s'établit à une fréquence de *a* très faible; la coexistence des deux allèles est un élément de variabilité peu important dans la population, presque tous les individus exprimant le phénotype « normal » [*A*].

Cette situation est appelée **cryptopolymorphisme**; c'est celle qu'on observe pour la mucoviscidose, pour l'hémophilie, la galactosémie et, en général, pour presque toutes les maladies génétiques graves.

Pour un gène autosomal récessif défavorable dans une population panmictique :

- la fréquence des individus atteints aa est :

$$q^2 = \frac{u}{s}$$

- la fréquence de l'allèle défavorable est :

$$q = \sqrt{\frac{u}{s}}$$

2

Fréquence d'un gène létal

Un gène létal est un gène qui confère aux individus qui l'expriment une valeur sélective nulle, soit qu'ils soient non viables, soit qu'ils soient stériles.

2.1 GÈNE LÉTAL RÉCESSIF DANS UNE POPULATION PANMICTIQUE

C'est le cas vu ci-dessus avec $s = 1$; alors à l'équilibre :

- la fréquence d'apparition à chaque génération des individus létaux aa est égale au taux de mutation :

$$q^2 = u$$

- la fréquence de l'allèle défavorable dans la population est :

$$q = \sqrt{u}$$

- la fréquence des porteurs du gène à l'état hétérozygote est :

$$2pq \approx 2q = 2\sqrt{u}$$

Si $u \approx 10^{-6}$, elle est donc de l'ordre de 1 individu sur 500.

2.2 GÈNE LÉTAL RÉCESSIF DANS UNE POPULATION AUTOGAME

Comme précédemment, nous allons rechercher l'influence de la sélection seule, celle de la mutation seule, et calculer la fréquence pour laquelle elles s'équilibrent.

La pression de sélection dépend de la fréquence d'apparition des individus atteints. Soit q la fréquence de l'allèle a dans la population de reproducteurs. La population qui se reproduit est composée de deux génotypes (les aa sont morts ou stériles) :

	AA	Aa
en fréquences	$(1 - 2q)$	$2q$
Lorsque ces individus se reproduisent en autogamie, ils produisent :		
Adultes :	AA $(1 - 2q)$	Aa $2q$
	↓	↓
Zygotes :	AA	$\frac{1}{4}AA \quad \frac{1}{2}Aa \quad \frac{1}{4}aa$

Les individus aa (en fréquence $1/4 \cdot 2q = 1/2q$) disparaissent. Parmi les adultes reproducteurs de la génération suivante (nous avons ainsi bouclé une génération, où la sélection a joué), la fréquence des hétérozygotes est donc :

$$2q' = \frac{1/2 \cdot 2q}{1 - 1/2q} = \frac{2q}{2 - q}$$

soit :

$$q' = \frac{q}{2 - q}$$

La pression de sélection est :

$$\Delta_s q = q' - q = \frac{-pq}{2 - q}$$

La pression de mutation est toujours :

$$\Delta_m q = -uq$$

A l'équilibre :

$$\Delta_s q + \Delta_m q = 0$$

soit

$$\frac{q}{2 - q} = u$$

$$q = \frac{2u}{1 + u} \approx 2u$$

- La fréquence de l'allèle défavorable dans la population adulte est donc :

$$q = 2u$$

• La fréquence d'apparition à chaque génération des individus létaux aa est :

$$\frac{1}{4} \cdot 2q = \frac{1}{2} q = u$$

Ici encore, elle est donc égale au taux de mutation.

2.3 GÈNE LÉTAL RÉCESSIF DANS UNE POPULATION SE REPRODUISANT VÉGÉTATIVEMENT

Dans la population qui se reproduit,
soit x la proportion de génotypes AA ,
et $(1-x)$ la proportion de Aa .

Ces individus se reproduisent identiques à eux-mêmes, à la mutation près :

Adultes	AA	Aa
Fréquences	x	$(1-x)$
	↓	↓
Descendance (propagules avant sélection)	$(1-u)^2 \mathbf{AA} + 2u(1-u)\mathbf{Aa} + u^2 \mathbf{aa}$	$(1-u)\mathbf{Aa} + u \mathbf{aa}$

Les individus létaux aa sont donc produits en fréquence :

$$x \cdot u^2 + (1-x) \cdot u$$

Dans les adultes produits (après sélection qui élimine les aa), la proportion x' de **AA** est donc :

$$x' = \frac{x(1-u)^2}{1-xu^2-(1-x)u}$$

$$x' = \frac{x(1-u)}{1+ux}$$

$(1-u)/(1+ux)$ est manifestement inférieur à 1, donc x' est inférieur à x ; la fréquence des **AA** ne cesse de décroître et tend vers 0. A l'équilibre, la population reproductrice est composée exclusivement de Aa , et la fréquence du gène létal est :

$$q = \frac{1}{2}$$

Dans cette population à l'équilibre, la fréquence d'apparition du génotype aa est encore u , taux de mutation.

- ✓ Ce calcul donne une idée de ce que devient la population; la fréquence q ne peut que croître. Mais l'équilibre (100 % d'hétérozygotes) est rarement atteint, vue l'extrême lenteur du phénomène. En effet, le terme $(1 - u)/(1 + ux)$ est équivalent à $1 - u(1 + x)$, et très peu différent de 1. Tout ce que nous avons dit sur la lenteur d'action de la mutation est vrai ici.

2.3.1 Gène létal dominant

Quel que soit le régime de reproduction, un gène létal dominant ne peut provenir que d'une mutation récente. La fréquence du gène létal A est donc $p = u$, mais le phénotype létal Aa apparaît avec une fréquence $2u$. La population est composée de :

$$2u \text{ } Aa \text{ et } (1 - 2u)aa$$

Conclusion

Selon le régime de reproduction, la fréquence d'équilibre d'un gène récessif létal peut donc varier dans un intervalle très large (de $2u$ à $1/2$), mais elle s'ajuste toujours de telle sorte qu'à l'équilibre, la fréquence d'apparition du phénotype létal est toujours égale à la fréquence de la mutation qui introduit le gène létal. En d'autres termes, à l'équilibre, la fréquence d'apparition d'une tare génétique dans la population ne dépend pas de la manière dont la population se reproduit. En particulier, et contrairement à une idée reçue, un régime consanguin, s'il est pratiqué depuis suffisamment longtemps pour que l'équilibre soit atteint, n'entraîne aucune augmentation de la fréquence d'apparition des tares. Ce qui est vrai, en revanche, c'est que l'arrivée d'un système de reproduction consanguin dans une population se reproduisant auparavant en panmixie démasque instantanément un certain nombre de gènes récessifs délétères. C'est d'ailleurs un moyen de reconnaître une espèce autogame d'une espèce allogame. En autofécondation, la première fournit des individus normaux, alors que la seconde fournit des individus tarés en grand nombre.

L'indépendance entre le régime de reproduction et la fréquence d'apparition des individus exprimant le phénotype défavorable reste vraie pour un gène non létal, simplement défavorable : si la valeur sélective des individus exprimant le gène est $(1 - s)$, la fréquence d'apparition des individus exprimant le caractère est toujours à l'équilibre u/s lorsque le gène est récessif (ainsi que nous l'avons calculé pour une population panmictique), et environ $2u/s$ si le gène est dominant.

Cela peut être montré assez simplement, au moins dans le cas d'un gène récessif. Considérons une population au stade zygotes (ou propagules), *avant sélection* :

Génotypes	AA	Aa	aa
Fréquences	D	H	R
Valeurs sélectives	$w_1 = 1$	$w_2 = 1$	$w_3 = 1 - s$

La fréquence de l'allèle A à ce stade est : $p = D + \frac{1}{2}H$.

Au stade adulte (après sélection, qui élimine une fraction s des aa), les fréquences génotypiques deviennent :

$$\frac{D}{1-sR} \quad \frac{H}{1-sR} \quad \frac{(1-s)R}{1-sR}$$

et la fréquence de A est :

$$p_{ad} = \frac{D + 1/2H}{1-sR} = \frac{p}{1-sR}$$

Lors de la reproduction, des mutations se produisent. La nouvelle fréquence de A devient :

$$p' = (1-u) \cdot p_{ad} = \frac{(1-u) \cdot p}{1-sR}$$

L'équilibre est réalisé lorsque $p' = p$, c'est-à-dire lorsque :

$$1-sR = 1-u$$

$R = u/s$

Nous n'avons fait ici aucune hypothèse sur le régime de reproduction. Ce calcul a une valeur générale.

Une valeur sélective de $(1-s)$ signifie qu'une proportion s des individus exprimant le caractère défavorable est éliminée par la sélection. Cela représente donc une proportion de la population égale à :

$$L = s \cdot R = s \cdot \frac{u}{s}$$

$L = u$

Cette proportion éliminée est appelée le **fardeau de mutation** (L pour *load*). Il représente la diminution de la valeur sélective moyenne de la population due aux mutations à ce locus. Cette valeur sélective moyenne est ici de :

$$\bar{w} = 1 - u$$

(il s'agit ici d'une valeur sélective relative, 1 étant celle des meilleurs individus, ceux qui ne portent pas la mutation).

Ce résultat remarquable, obtenu par Haldane, montre qu'à l'équilibre, la fraction de la population éliminée à cause des mutations délétères récessives en un locus est toujours u , quel que soit le régime de reproduction, et surtout *quelle que soit l'importance de l'effet délétère de la mutation* : un gène faiblement délétère (s faible) diminue autant la valeur sélective moyenne de la population, est à l'origine d'autant de « morts génétiques », qu'un gène létal ($s = 1$). En effet, si la proportion des individus aa éliminés est beaucoup plus faible dans le premier cas que dans le second, la fréquence à l'équilibre de ces individus dans la population est aussi beaucoup plus élevée, de sorte que les deux effets se compensent exactement.

De même pour un gène dominant (même incomplètement) :

$$L \approx 2u$$

indépendamment, ici encore, de l'importance de l'effet délétère et du régime de reproduction.

Exercices du chapitre B

Exercice B.1

La galactosémie est une maladie héréditaire. Elle s'explique par la présence d'allèles récessifs (à différents gènes), responsables d'un déficit d'enzymes impliquées dans le métabolisme du galactose (sucre présent dans le lait). La dégradation hépatique du galactose est perturbée et il y a alors accumulation de métabolites toxiques du galactose. Les nourrissons galactosémiques ne survivaient pas autrefois. Ils sont à présent sauvés par une suppression totale du galactose dans l'alimentation

1) Quel(s) type(s) de descendant(s) peut donner un couple de galactosémiques?

2) Quelle était la fréquence des allèles responsables de la galactosémie à l'époque où cette maladie était incurable? (La fréquence de la mutation directe sera prise égale à 10^{-6} , la mutation reverse sera jugée négligeable. La population peut être considérée comme panmictique aux loci concernés.)

3) Désormais, ce caractère est vraisemblablement indifférent à la sélection. Quelle sera la fréquence d'équilibre des allèles inactifs?

Exercice B.2

La maladie de Tay-Sachs est une maladie génétique due à une déficience de l'hexosaminase A, une enzyme lysosomale qui dans les conditions normales catalyse l'hydrolyse du ganglioside GM2, un glycolipide abondant dans les membranes des neurones. Le nouveau-né est normal mais les premiers signes d'une dégénérescence progressive du tissu nerveux apparaissent vers l'âge de 6 mois. Cette dégénérescence est fatale et les enfants atteints ne survivent que jusqu'à 4 ou 5 ans. Cette maladie affecte autant de filles que de garçons.

1) En Europe occidentale, il naît environ un enfant atteint pour 250 000 naissances. Évaluer la fréquence de l'allèle responsable de la maladie, en supposant qu'il est récessif par rapport à l'allèle normal puis en supposant qu'il est dominant. Quelles hypothèses supplémentaires faites-vous pour réaliser les calculs?

2) Évaluer le taux de mutation qui produit l'allèle déficient dans les deux hypothèses (récessivité ou dominance).

3) Une étude a montré que parmi les enfants atteints, un grand nombre (15 % environ) était né d'unions entre cousins germains, alors que la fréquence de ce type d'union n'était que de 5 pour 1 000 dans la population globale.

Cette observation vous renseigne-t-elle quant au caractère récessif ou dominant de l'allèle?

Exprimer, en fonction de f , coefficient de consanguinité, et q , fréquence de l'allèle déficient, la probabilité pour un enfant issu de l'union entre deux

cousins germains d'être atteint. Calculer cette probabilité grâce aux probabilités conditionnelles et en déduire la valeur de q .

Exercice B.3

Soit un allèle récessif létal introduit par mutation dans une population panmictique. Sachant que le taux de mutation est égal à u :

1) Quelle sera la fréquence de cet allèle à l'équilibre? Quelle est la proportion d'individus hétérozygotes à ce locus?

2) En supposant qu'il existe dans le génome N locus de ce type soumis au même taux de mutation, quel sera le nombre moyen d'allèles létaux portés à l'état hétérozygote par individu?

3) Quelle sera la valeur du fardeau génétique correspondant? (On négligera u devant 1 pour tous les calculs.)

4) Sachant que $(1-x)^n \approx e^{-nx}$ quand x est très faible, et en prenant $N = 10^4$ et $u = 10^{-5}$ calculer le nombre moyen d'allèles létaux portés par individu et calculer la valeur du fardeau. Chez l'Homme, le nombre moyen d'allèles récessifs létaux présents chez chaque individu est évalué à 4. Quel(s) écart(s) au modèle utilisé pour les calculs pourrai(en)t expliquer la différence entre les valeurs attendues et observées?

C

Le polymorphisme

1

Introduction

Nous avons vu que l'équilibre sélection-mutation aboutit à une situation de « cryptopolymorphisme » avec la forme « normale » très prédominante, et un ou quelques variants rares défavorables. Cela n'explique pas la majeure partie de la variation couramment observée dans le monde vivant, c'est-à-dire tous les cas où les différents phénotypes sont en proportions comparables. C'est à de telles situations que Ford (1940) a réservé le terme de **polymorphisme**. Il définit le polymorphisme comme « *la coexistence dans une population de deux formes discontinues ou davantage, dans des proportions telles que la plus rare ne peut être maintenue par la mutation récurrente* ». Par cette précision, il entendait écarter la situation de cryptopolymorphisme.

Cette variabilité phénotypique est la conséquence d'un polymorphisme sous-jacent au niveau des gènes. Plus précisément, le **polymorphisme génétique** est la *coexistence de plusieurs allèles en un locus*. Par opposition, un locus pour lequel on ne trouve qu'un allèle est dit **monomorphe**. Pour écarter le cas du cryptopolymorphisme (si on l'incluait, tous les locus pourraient être dits polymorphes, puisque tous peuvent muter en des formes défavorables), il faut ajouter que les allèles doivent être *en fréquences du même ordre de grandeur*, qu'il n'y ait pas un allèle très prédominant. Dans la pratique, les généticiens se fixent un seuil, un « critère de polymorphisme » : le polymorphisme génétique est la *coexistence dans une même population de deux allèles ou plus en un locus, le plus fréquent d'entre eux ne dépassant pas la fréquence de 0,99*. Cette définition est plus opérationnelle que la précédente, mais il ne faut pas se dissimuler ce que le seuil

de 0,99 à d'arbitraire (certains utilisent d'ailleurs d'autres critères, par exemple un critère de 0,95 au lieu de 0,99). Ces critères sont choisis aussi pour être compatibles avec l'étude d'un nombre raisonnable d'individus (de l'ordre de la centaine).

En toute rigueur, et conformément à l'énoncé de Ford, la définition du polymorphisme suppose qu'il y a coexistence d'allèles différents *au sein de la même population*. C'est en tout cas toujours dans ce sens que nous l'emploierons. L'existence de formes différentes dans des populations différentes ne constitue pas, à proprement parler, un cas de polymorphisme (on doit parler plutôt, dans ce cas, de *polytypisme*).

Le polymorphisme est un fait d'observation courante : la couleur des yeux, des cheveux, les groupes sanguins chez l'homme ainsi que la taille ou la musculature, ou encore la couleur du pelage chez les chats, sont des exemples parmi bien d'autres. La plupart des différences que nous constatons entre les individus se ramènent à des cas de polymorphisme génétique, même s'il n'est pas toujours facile de préciser les unités génétiques en cause.

Or ce fait banal constitue, à bien y réfléchir, un paradoxe flagrant : si le moteur de l'évolution est bien la sélection exercée par le milieu, comme Darwin l'a proposé, nous devrions nous attendre que la structure héréditaire qui se reproduit le plus efficacement soit la seule à subsister à terme. Une population ne devrait donc comporter généralement qu'un allèle à chaque locus, à la mutation près. Aussi une explication possible, lorsque nous nous trouvons en face d'allèles polymorphes, est-elle que la sélection n'agisse pas : il s'agirait d'un polymorphisme **neutre** (*une polymorphisme et des allèles sont dits neutres sélectivement si le remplacement d'un allèle par un autre ne modifie pas les valeurs sélectives des individus qui les possèdent*). Toutefois, dans ce cas, la dérive doit opérer, et nous savons qu'elle aboutit, elle aussi, à l'homogénéité.

Pour expliquer l'existence au sein des populations d'une diversité sous contrôle génétique, nous ne pouvons donc échapper au dilemme suivant :

- soit il existe un mécanisme sélectif responsable du maintien du polymorphisme : c'est l'**hypothèse sélectionniste**;
- soit les allèles polymorphes sont neutres mais alors la population n'est pas à l'équilibre : c'est l'**hypothèse neutraliste**.

Jusqu'à une époque assez récente, les exemples de polymorphisme que l'on pouvait donner portaient le plus souvent sur des caractéristiques morphologiques ou de coloration. Peu de gènes avaient été identifiés avec précision et il y avait généralement, entre les gènes analysés et leur expression, une grande distance, en termes de complexité des mécanismes biologiques mis en jeu. Quelques rares exemples étaient cependant bien étudiés : ainsi, les groupes sanguins, connus en nombre croissant depuis le début du

siècle ont fait l'objet de nombreux travaux de génétique des populations humaines. Il était toutefois difficile d'estimer le nombre de locus polymorphes au sein d'une espèce ou d'une population : l'existence même d'un locus non polymorphe ne pouvait être détectée que lorsqu'on avait eu la chance de trouver un variant rare (cryptopolymorphisme). La plupart des généticiens évaluaient le nombre de locus polymorphes dans une population à quelques dizaines, peut-être quelques centaines : avec seulement 25 locus polymorphes, et même si ceux-ci ne comportent que deux allèles par locus, il peut exister 3^{25} génotypes différents, soit près de 10^{12} : c'est bien plus qu'il n'en faut pour rendre compte de la variabilité phénotypique observée.

On dispose, depuis les années 1960, d'un moyen de quantifier l'importance du polymorphisme au niveau même des produits des gènes grâce à la technique de l'*électrophorèse des protéines* sur gel d'amidon ou d'acrylamide, que Lewontin et Hubby (1966) ont eu l'idée d'appliquer à ce problème. Une goutte d'un extrait protéique, souvent obtenue par simple écrasement des tissus, est déposée sur une plaque de gel imbibée d'une solution d'électrolytes, et dans laquelle on établit un champ électrique (fig. 4). Chaque protéine se déplace dans ce champ avec une vitesse qui est fonction de sa charge; ces molécules prennent donc, après un temps défini, une position caractéristique par rapport au front de migration. Si l'on arrête alors la migration et qu'on met en contact avec le gel un produit pouvant servir de substrat à une réaction enzymatique et donnant un produit coloré, on observera une bande à l'endroit où se sont arrêtées les molécules de l'enzyme correspondante. On peut alors distinguer des **isozymes**, enzymes ayant des fonctions voisines puisqu'elles sont actives sur le même substrat artificiel.

Lorsque le gène subit une mutation, l'enzyme qu'il code peut perdre totalement son activité. Dans ce cas, la bande correspondante sur le « zymogramme » disparaît : on est en présence d'un « allèle nul ». En fait, on trouve peu d'allèles nuls car ceux-ci sont le plus souvent contre-sélectionnés et ne subsistent qu'en fréquence très faible dans les populations (cryptopolymorphisme). Il peut arriver, au contraire, que la mutation respecte l'activité enzymatique, au moins qualitativement, et donne une « enzyme allèle » de l'enzyme d'origine (on parle d'**allozyme**). C'est le cas notamment pour certaines mutations ponctuelles qui remplacent un acide aminé par un autre. L'examen du code génétique montre qu'une substitution d'une base dans l'ADN entraîne environ une fois sur trois un remplacement d'un acide aminé par un acide aminé chargé différemment; il en résulte une modification de la charge globale de la protéine, donc de sa mobilité électrophorétique. Dès lors, la bande, au lieu d'être supprimée, se trouve légèrement décalée. Ainsi, en un locus déterminé, l'étude d'un nombre suffisant d'individus révèle souvent plusieurs allèles.

Un intérêt particulier de cette méthode est qu'elle révèle aussi bien les locus non polymorphes, contrairement aux méthodes génétiques classiques.

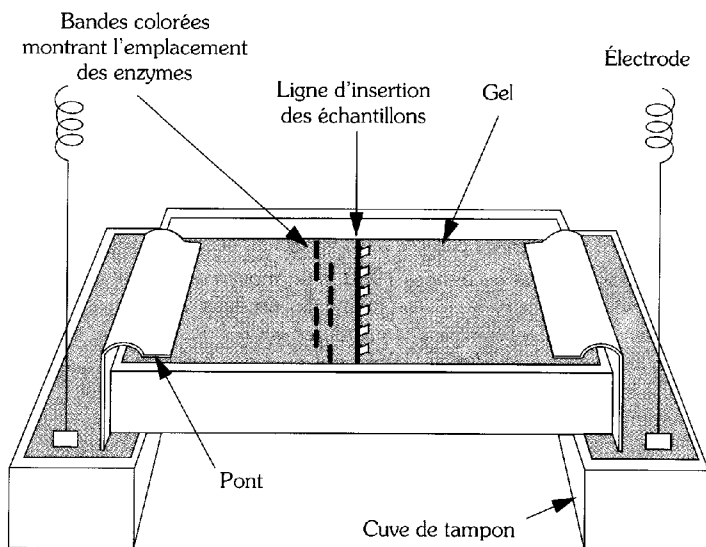


Figure 4a : L'ÉLECTROPHORÈSE SUR GEL est une méthode qui permet d'évaluer l'importance de la variabilité génétique dans les populations naturelles grâce à l'examen des variants protéiques des divers individus. Dans un premier temps, on prélève un échantillon de tissu de chaque organisme et on en extrait les protéines; chaque extrait protéique est ensuite disposé sur un gel d'amidon, d'agar ou de polyacrylamide. Puis on fait passer un courant dans le gel pendant quelques heures; chacune des protéines de l'échantillon migre alors à travers le gel selon une direction et à une vitesse qui dépendent de sa charge électrique et de son poids moléculaire. Au terme de la migration, on recouvre l'ensemble du gel d'une solution contenant le substrat spécifique de l'enzyme étudiée et un réactif coloré. L'enzyme présente dans le gel catalyse la conversion du substrat en un produit qui se combine avec le réactif pour donner une bande de couleur à l'endroit où la protéine a migré. Comme les divers allèles d'une même enzyme diffèrent par leur structure moléculaire et par leur charge (et ont par conséquent des mobilités différentes dans un champ électrique), il est possible de visualiser, pour chaque individu, la composition génétique du locus qui code pour une enzyme donnée, d'après le nombre et la position des bandes électrophorétiques.

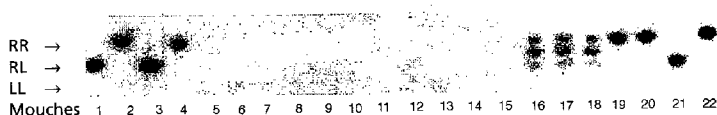


Figure 4b : CE GEL ÉLECTROPHORÉTIQUE a été coloré pour la malatedéhydrogénase, enzyme qui intervient dans le métabolisme respiratoire. Ce gel représente les migrations respectives de cette enzyme chez 22 mouches appartenant à l'espèce *Drosophila equinoxialis*. Deux variants polypeptidiques (c'est-à-dire deux chaînes protéiques codées par deux allèles différents) apparaissent dans cette expérience : un polypeptide à migration rapide (designé par R) et un autre à migration lente (designé par L). La malate-déshydrogénase est constituée de deux chaînes polypeptidiques qui se combinent spontanément après leur synthèse; les individus homozygotes ne fabriquent qu'une forme de cette enzyme (soit RR, soit LL), alors que les hétérozygotes en fabriquent trois : R, LL, et RL (la dernière forme est de mobilité électrophorétique intermédiaire). À l'électrophorèse, les homozygotes se présentent donc sous la forme d'une seule bande, alors que les hétérozygotes en ont trois. Un tel cas, qui ne fait intervenir que deux allèles, représente une situation simplifiée : en fait, un certain nombre de locus codant pour les protéines ont jusqu'à cinq allèles ou plus dans les populations.

On a donc pu estimer, par ce procédé, la diversité génétique dans de nombreuses populations. Les individus sont souvent hétérozygotes pour quelque 10 % de leurs locus, et sur la totalité d'une espèce, une très forte proportion des locus (30 à 80 %) présente de la variabilité, soit plusieurs milliers en extrapolant à l'ensemble du génome.

Plus récemment, le développement des techniques de biologie moléculaire a permis d'aborder l'étude du polymorphisme au niveau de l'ADN lui-même. Le séquençage est une méthode trop lourde pour être couramment pratiquée à l'échelle des populations, mais d'autres techniques, comme la mise en évidence du polymorphisme dans les sites de restriction de l'ADN, peuvent être utilisées. On a ainsi révélé une large variabilité entre individus pour le nombre et la position de divers sites de restriction (on parle de **polymorphisme de longueur des fragments de restriction**, ou plus couramment, selon l'abréviation anglaise, de RFLP) (fig. 5). Ce polymorphisme concerne tout le génome, y compris et surtout les portions non codantes qui n'étaient pas accessibles par l'étude des protéines.

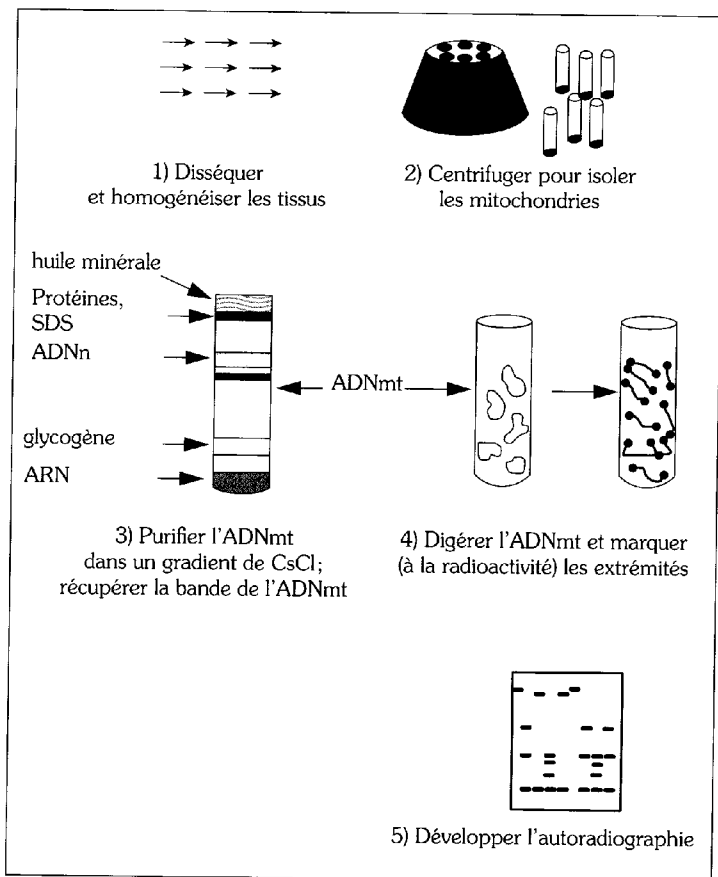


Figure 5a : Protocole général pour l'analyse de RFLP d'ADN mitochondrial (ADNmt).

La découverte de cette quantité de polymorphisme a conduit les généticiens des populations à revoir leurs hypothèses sur les mécanismes de maintien de la variabilité, que nous allons maintenant considérer.

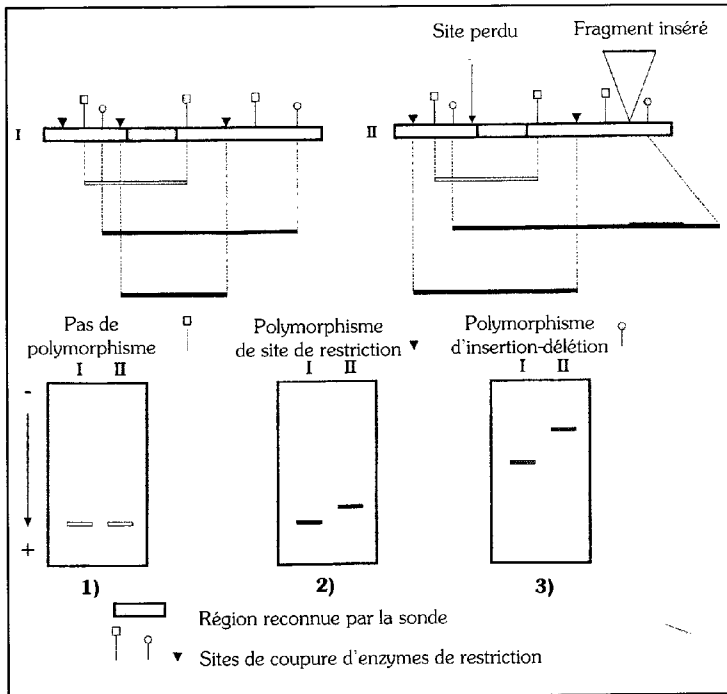


Figure 5b : RFLP

Lorsque deux individus, que l'on va supposer homozygotes pour des raisons de simplicité, sont comparés après digestion de leur ADN par un enzyme donné et hybridation avec une sonde donnée, trois situations élémentaires peuvent se présenter :

- au voisinage de la région du génome reconnue par la sonde, les sites de restriction sont les mêmes, et il n'y a pas de différence d'insertion-délétion : les profils sont identiques (1),
- il y a des différences au niveau des sites de restriction, sans différences d'insertion-délétion : les profils sont différents (2),
- il y a des différences d'insertion-délétion, sans différences au niveau des sites de restriction : les profils sont différents (3). (D'après de VIENNE, D., 1997. Les marqueurs moléculaires et leurs applications, série « Biotechnologies végétales » dirigée par Y. DEMARLY. Coordinateur scientifique Emmanuel PICARD. UNISAT Université audio-visuelle francophone. CNED/AVPELF-UREF).

2

Le polymorphisme sélectionné

2.1 POLYMORPHISME ÉQUILIBRÉ

PAR LA SUPERDOMINANCE DE LA VALEUR SÉLECTIVE

La formule de la variation des fréquences géniques du fait de la sélection :

$$\Delta p = pq \frac{(w_1 - w_2)p + (w_2 - w_3)q}{w_1 \cdot p^2 + 2w_2pq + w_3 \cdot q^2} \quad (\text{cf. § 4.4.2})$$

montre que la sélection seule doit être capable d'expliquer l'existence d'un état d'équilibre polymorphe.

L'état d'équilibre se définit par $\Delta p = 0$.

L'état polymorphe se traduit par $pq \neq 0$.

Si $\Delta p = 0$ et que $pq \neq 0$, la seule solution est :

$$(w_1 - w_2)p + (w_2 - w_3)q = 0$$

Or $p > 0$ et $q > 0$, donc il faut que :

$$w_2 \notin [w_1, w_3]$$

Autrement dit, si $w_1 \geq w_2 \geq w_3$ ou si $w_1 \leq w_2 \leq w_3$ (ce qui représente tous les cas de dominance complète ou intermédiaire du caractère valeur sélective), aucun équilibre polymorphe n'est possible avec les hypothèses faites. Nous avons vu que ces situations aboutissent à la fixation de l'allèle favorable (A et a respectivement). Restent deux possibilités : $w_1 < w_2 > w_3$ ou

$w_1 > w_2 < w_3$. Dans ces cas, il existe une fréquence d'équilibre p_e qui appartient à $]0 ; 1[$, telle que :

$$(w_1 - w_2)p_e + (w_2 - w_3)(1 - p_e) = 0$$

soit :

$$p_e = \frac{w_2 - w_3}{2w_2 - w_1 - w_3}$$

Nous avons trouvé un équilibre (en plus des deux équilibres non polymorphes qui existent toujours). Il reste à chercher si, à partir d'une fréquence quelconque, on converge bien vers cette valeur, c'est-à-dire si cet équilibre est stable. Pour ce faire, cherchons dans quel sens évolue la fréquence p si elle s'écarte de l'équilibre d'une petite quantité k ; la fréquence de A est alors :

$$p = p_e + k \quad (\text{avec } k < 0 \text{ ou } k > 0)$$

$$\text{d'où :} \quad \Delta p = pq \frac{(w_1 - w_2)(p_e + k) + (w_2 - w_3)(1 - p_e - k)}{w_1 \cdot p^2 + 2w_2pq + w_3 \cdot q^2}$$

Δp a le signe du numérateur, qui permet d'écrire, en remplaçant p_e par sa valeur :

$$k \cdot (w_1 - 2 \cdot w_2 + w_3)$$

- Quand $k > 0$, c'est-à-dire $p > p_e$, ceci est positif (donc p augmente) si $w_1 > w_2 < w_3$, et négatif (donc p diminue) si $w_1 < w_2 > w_3$.

- Quand $k < 0$, c'est-à-dire $p < p_e$, ceci est positif (donc p augmente) si $w_1 < w_2 > w_3$, et négatif (donc p diminue) si $w_1 > w_2 < w_3$.

En définitive, lorsque la valeur sélective de l'hétérozygote est inférieure à celle des deux homozygotes ($w_1 > w_2 < w_3$), si p est supérieur à p_e , p aura tendance à augmenter encore, et à s'écarter de plus en plus de p_e à chaque génération. Inversement, si p est inférieur à p_e , il diminuera au fil des générations. Il est donc clair que l'« équilibre » p_e n'est pas stable : même si on partait de $p = p_e$, la moindre perturbation (mutation, fluctuation aléatoire) entraînerait la perte de cet équilibre. La fréquence du gène dans la population va évoluer vers les équilibres stables 0 ou 1 selon sa valeur de départ (inférieure ou supérieure à p_e).

À l'opposé, si $w_1 < w_2 > w_3$, alors l'équilibre p_e est stable : quelles que soient les fréquences de départ, la fréquence du gène A tend vers p_e .

Il suffit donc que la valeur sélective des hétérozygotes soit supérieure à celle des homozygotes pour obtenir un polymorphisme stable.

Cette situation s'appelle la **superdominance** de la valeur sélective (pour un caractère mesurable, on dit qu'il y a dominance intermédiaire si l'hétérozygote est quelque part entre les deux homozygotes, dominance complète

s'il est égal au meilleur homozygote [ou au moins bon], et superdominance s'il est au-delà).

Le rôle de la superdominance dans l'existence du polymorphisme a longtemps été considéré comme primordial; il est aujourd'hui assez controversé.

En effet, si l'on connaît de très nombreux cas où les individus issus de croisements entres souches différentes sont supérieurs à leurs parents (supériorité de taille, de vitesse de croissance, de valeur sélective... : c'est ce que l'on appelle la **vigueur hybride** ou **hétérosis**), il est souvent possible d'interpréter cette supériorité comme résultant de l'addition des effets favorables des gènes apportés par les parents, mais on n'a pu qu'exceptionnellement démontrer qu'il s'agissait d'allèles; il peut s'agir, le plus souvent, de complémentation. L'hétérosis peut donc très bien s'expliquer, dans la plupart des cas, sans faire appel à la superdominance.

Toutefois, il existe au moins un exemple de superdominance pour lequel une analyse assez complète de la situation existe : c'est l'**anémie falciforme**.

L'hémoglobine humaine est une protéine formée de deux monomères polypeptidiques, α et β . La molécule β comporte 148 acides aminés, dont le sixième est un acide glutamique. Une mutation ponctuelle du sixième codon peut transformer l'acide glutamique en valine; la molécule ainsi transformée manifeste des propriétés aberrantes, se traduisant par différentes anomalies sanguines chez les homozygotes : les hématies prennent, à pression d'O₂ normale, une forme contournée évoquant la forme d'une faucille (d'où le nom de globules falciformes) provoquant un blocage des capillaires et une anémie généralement mortelle pour le nouveau-né. Cette maladie génétique est appelée drépanocytose ou anémie falciforme.

Le gène de la drépanocytose peut exister chez les populations européennes, en situation de cryptopolymorphisme; il y a dominance du gène normal, en ce sens que les parents d'anémique(s), obligatoirement hétérozygotes, sont sensiblement normaux, bien que leurs hématies contiennent les deux types d'hémoglobine; l'homozygote récessif est, quant à lui, létal. Le gène est maintenu en fréquence très faible.

La situation est très différente dans les zones impaludées d'Afrique noire, où la fréquence du gène muté peut dépasser 10 %; il y a alors un réel polymorphisme, qui s'explique par une situation de superdominance. On a démontré, en effet, que les globules des hétérozygotes qui contiennent le mélange des deux hémoglobines étaient lysés plus rapidement que les hématies normales, si bien que le parasite n'avait pas le temps d'y terminer son cycle de développement. Les hétérozygotes sont donc indemnes de la malaria (malaria et paludisme sont synonymes); leur longévité moyenne et leur taux de reproduction sont donc plus grands que ceux des homozygotes normaux.

Quelques autres cas de superdominance ont été rapportés dans la littérature. L'un d'eux est directement issu de l'étude du polymorphisme enzymatique

par électrophorèse : il s'agit de l'alcool-déshydrogénase, ou Adh, de la *Drosophile* ; pour cet insecte, la désalcoolisation du milieu est une fonction importante. Il n'est donc pas surprenant que l'Adh représente 1/100 du poids total de l'animal. Les études d'allozymes ont révélé l'existence d'un polymorphisme du gène de l'Adh, avec deux allèles. La fréquence de ceux-ci varie en corrélation avec les doses d'éthanol auxquelles sont soumises les populations. Les deux protéines correspondantes ont été analysées, et se sont révélées avoir des conditions d'activité différentes : sensibilité à la température, pH optimum... Il semble que l'hétérozygote possédant les deux protéines soit capable de faire face à des conditions plus diverses que les homozygotes. Il serait donc plus apte qu'eux à utiliser un milieu diversifié, dans l'espace ou dans le temps ; mais dans une condition de milieu déterminée, il ne se montre pas plus performant que l'un ou l'autre des homozygotes. Dans une telle situation, où l'hétérozygote n'est jamais supérieur aux deux homozygotes à un instant donné, mais seulement en moyenne sur une gamme de milieux, on parle plutôt de *superdominance marginale*.

Cet exemple, quoique moins bien démontré, est peut-être plus intéressant que celui de l'anémie falciforme, car plus facilement généralisable à d'autres locus.

Le polymorphisme ne peut donc s'expliquer, dans le cadre de notre modèle, que par la superdominance. Jusque dans les années soixante, c'est généralement elle qui était invoquée pour expliquer la plupart des cas de polymorphisme alors connus. Avec l'avènement des techniques d'électrophorèse qui révélaient une quantité impressionnante de polymorphisme et permettaient de reconnaître facilement les hétérozygotes (contrairement à la plupart des marqueurs morphologiques étudiés auparavant), on s'attendait à découvrir de nombreux autres exemples de superdominance. Mais cet espoir ne s'est pas réalisé : les cas démontrés de superdominance, et même de sélection agissant sur des polymorphismes enzymatiques restent rares ou équivoques (même pour l'Adh de la *Drosophile*, l'intervention de la sélection est très probable, mais l'existence de la superdominance reste douteuse).

Il faut souligner que la mise en évidence d'une sélection naturelle s'exerçant sur un locus particulier est souvent difficile. Pour des raisons pratiques d'abord : dans beaucoup de cas, les fonctions exactes *in vivo* des enzymes révélées par électrophorèse sont inconnues ; tester tous les facteurs sélectifs envisageables est une tâche sans fin ; les différences de valeurs sélectives entre individus sont souvent faibles, et obtenir des données sur la survie et la fertilité des différents génotypes dans différentes conditions de milieu est un travail colossal. À cela s'ajoute une difficulté théorique : les individus diffèrent inévitablement par de nombreux gènes, et on ne peut jamais être sûr qu'une différence sélective observée est due au gène étudié, et non pas aux gènes voisins qui seraient en déséquilibre de liaison avec lui.

Néanmoins, la rareté apparente des cas de superdominance, le fait que le polymorphisme enzymatique soit aussi considérable chez les organismes haploïdes que chez les diploïdes (les plus hauts niveaux de polymorphisme observés le sont chez des bactéries!) ont jeté un fort doute sur l'importance de ce phénomène dans le maintien de la variabilité, et si la part de la superdominance reste encore un sujet débattu parmi les généticiens, force a été de rechercher d'autres mécanismes pouvant expliquer le maintien du polymorphisme.

2.2 VALEURS SÉLECTIVES VARIABLES

Nous avons supposé, jusqu'ici, que les valeurs sélectives sont constantes; dans ce cas, la superdominance est le seul mécanisme pouvant maintenir un polymorphisme à l'équilibre. Mais le polymorphisme peut aussi être maintenu par la sélection si les valeurs sélectives sont variables.

- Elles peuvent être variables *en fonction des fréquences géniques* elles-mêmes : $w = f(p)$; c'est la **sélection fréquence-dépendante**; cette dénomination résulte de ce que la valeur sélective d'un génotype n'est pas une constante, mais varie avec la fréquence des génotypes dans la population. Plus précisément, et sans entrer dans la formalisation mathématique de ce type de sélection, on conçoit que le polymorphisme est maintenu si *un allèle est avantageé lorsqu'il devient rare* (encadré 6); il va donc remonter en fréquence : la sélection s'oppose à sa disparition (pour ce type de sélection, on a proposé aussi les appellations « avantage du rare » ou « sélection apostatique », allusion plaisante au fait que c'est celui qui ne fait pas comme les autres, l'apostat, qui est favorisé). Bien qu'à première vue, une telle sélection puisse paraître peu plausible, il y a plusieurs situations où cela peut se produire. On a montré, par exemple, que les prédateurs s'attaquent de préférence aux proies qu'ils reconnaissent, donc à celles dont l'aspect correspond au type le plus fréquent, qui est de ce fait contre-sélectionné, jusqu'à ce qu'il ne soit plus le plus fréquent. De même, les pathogènes évoluent en développant des virulences contre les génotypes les plus communs dans les populations d'hôtes, de sorte que, parmi ceux-ci, les génotypes rares peuvent être favorisés. Par ailleurs, si différents génotypes ont des façons différentes d'exploiter les ressources du milieu (des « niches écologiques » différentes) ou des besoins légèrement différents, cela favorise les types les plus rares qui ont moins à souffrir de la compétition avec leurs congénères et, ayant ainsi une valeur sélective supérieure, augmenteront, ici encore, jusqu'à ne plus être les plus rares; cette situation peut avoir une valeur très générale. Enfin l'hétérogamie entraîne aussi, comme nous l'avons vu, une sélection qui avantage la forme rare.

- Les valeurs sélectives peuvent être *variables dans l'espace ou dans le temps*. En effet, le milieu n'est pas homogène : il peut comporter des

ENCADRÉ 6 : LE CONCEPT DE POLYMORPHISME PROTÉGÉ

Une condition suffisante du maintien d'un polymorphisme est que, si un allèle est en fréquence très faible, il doit toujours tendre à augmenter en fréquence à la génération suivante : si cela est vrai pour chacun des allèles en présence, il est clair que le polymorphisme ne peut pas être perdu. Le polymorphisme est dit protégé. Pour voir si un polymorphisme est protégé ou non, il suffit donc d'étudier ce qui se passe lorsque la fréquence de chaque allèle, successivement, tend vers 0. Cela revient à étudier la stabilité de chacun des points $p_i = 0$, p_i étant la fréquence du i^{e} allèle : le polymorphisme est protégé si tous ces points sont des équilibres instables.

L'intérêt de cette approche est que les situations correspondant à ces bornes sont généralement plus simples à analyser : les équations se simplifient. Cette méthode est particulièrement utile lorsque l'on doit analyser des situations complexes (allèles multiples, sélections variables...) où la recherche explicite des points d'équilibre « intérieurs » (p_i tous loin de 0 ou 1) s'avère difficile (on peut d'ailleurs avoir un polymorphisme protégé sans qu'il existe d'équilibre interne, les fréquences oscillant ou parcourant des cycles, ou avec plusieurs équilibres possibles).

Pour illustrer la simplicité de cette méthode, nous nous contenterons de reprendre l'analyse du cas de sélection à valeurs sélectives constantes avec deux allèles A , a dans une population panmictique : si l'allèle a est rare, il n'existe pratiquement que les génotypes AA et Aa ; on voit que a ne sera pas perdu (sera protégé) si les Aa ont une valeur sélective supérieure à celle des AA (si $w_2 > w_1$). Inversement, quand A est rare, les seuls génotypes présents sont Aa et aa , et pour éviter la perte de A il faut que w_2 soit supérieur à w_3 . On retrouve ainsi la condition de la superdominance : $w_1 < w_2 > w_3$.

Ceci peut aussi s'exprimer mathématiquement : nous avons vu que :

$$p' = \frac{w_1 \cdot p^2 + 1/2 \cdot w_2 \cdot 2pq}{w_1 \cdot p^2 + 2 \cdot w_2 \cdot pq + w_3 \cdot q^2}$$

Si p tend vers 0 et que q tend vers 1, ceci se simplifie beaucoup en :

$$p' = \frac{w_2 \cdot p}{w_3}$$

(les autres termes sont négligeables). Pour que p augmente ($p' > p$), il faut donc que $w_2 > w_3$. L'étude symétrique de q' à l'autre borne ($p \approx 1$, $q \approx 0$) fournit le reste de la condition du maintien du polymorphisme : $w_2 > w_1$.

« micro-habitats », dont chacun convient mieux à un génotype particulier. On conçoit qu'une telle situation puisse faciliter le maintien d'un polymorphisme génétique. Il y a à cela certaines conditions cependant. Si la « maille » des variations spatiales de l'environnement est fine par rapport au déplacement des individus, de sorte que chaque individu expérimente au cours de sa vie l'ensemble des micro-habitats, c'est le génotype ayant en moyenne la meilleure valeur sélective sur l'ensemble des micro-habitats qui est avantagé. Ce meilleur génotype peut éventuellement être l'hétérozygote : on retombe sur la superdominance marginale ; mais si c'est un homozygote, le polymorphisme n'est pas maintenu. En revanche, si l'environnement est à « mailles larges », chaque individu n'expérimente généralement qu'un seul habitat dans sa vie, ce sont les génotypes spécialistes de chaque habitat qui sont favorisés et on peut, sous certaines conditions, maintenir un polymorphisme sans qu'il y ait superdominance, même marginale. Il s'agit bien de polymorphisme au sens où nous l'avons défini, et non de polytypisme, si la reproduction a lieu entre individus vivant dans les différents habitats, de sorte qu'on a bien affaire à une seule population. Ce peut être assez facilement le cas chez les plantes, où les individus sont inféodés à un habitat, mais où la dispersion du pollen et des graines assure les échanges génétiques entre habitats.

Un des cas les mieux étudiés de polymorphisme lié à un environnement en mosaïque est fourni par des populations d'Avoines sauvages introduites en Californie depuis la colonisation espagnole. Dans certaines populations de cette espèce autogame coexistent, à quelques décimètres de distance, des individus génétiquement différents, les uns montrant une adaptation à des micro-habitats plus secs, les autres plus humides, selon le micro-relief du terrain et la nature du sol. Le passage d'un micro-habitat à l'autre s'accompagne quasi automatiquement du remplacement d'une forme par l'autre. Or, ces deux types d'individus s'avèrent différer par leurs allèles à de nombreux locus, y compris des locus enzymatiques ou morphologiques sans relation apparente avec l'adaptation.

Dans le même ordre d'idées, des fluctuations temporelles de l'environnement favorisant alternativement différents génotypes peuvent permettre le maintien du polymorphisme. Ici encore, il faut que les fluctuations soient assez lentes pour que chaque individu n'expérimente pas l'ensemble des conditions (ainsi, les variations saisonnières ne peuvent pas favoriser le polymorphisme chez les Chênes, mais elles peuvent jouer un rôle pour des populations de Daphnies, de Drosophiles, ou de micro-organismes ayant de nombreuses générations par an). Elles doivent par ailleurs être assez rapides pour qu'aucun des allèles n'ait le temps de se fixer dans la période où il est favorisé.

Tous ces cas nous amènent à constater que les mécanismes conservant le polymorphisme par une variation des valeurs sélectives ressemblent beaucoup à ceux qui sont invoqués pour expliquer la coexistence de diverses

espèces dans un même milieu. Ce sont fondamentalement les mêmes. Ils ne maintiennent pas exclusivement la variabilité intraspécifique qu'est le polymorphisme, mais entrent dans un ensemble plus vaste qui est celui des causes de la diversité du monde vivant dans son ensemble, de la biodiversité. C'est d'ailleurs cette question de la coexistence d'espèces différentes dans un même milieu qui a amené l'élaboration du concept de niche écologique. La superdominance est le seul mécanisme qui soit strictement réservé à la variabilité intraspécifique, à la génétique des populations des organismes diploïdes, objets de cette étude. Tous les autres facteurs de maintien de la biodiversité sont du ressort de l'écologie générale.

2.3 POLYMORPHISME TRANSITOIRE

Si un allèle nouvellement apparu dans une population par mutation ou par migration confère à ses porteurs une valeur sélective supérieure, il va remplacer l'ancien allèle, à moins qu'il ne soit perdu par dérive aléatoire lorsqu'il est encore très rare : en effet, un gène apparu par mutation n'est présent au départ qu'en un exemplaire dans un individu, et, même s'il confère une valeur sélective supérieure à son porteur, sa probabilité de perte dans les premières générations est forte. En revanche, dès qu'il atteint une fréquence appréciable, la probabilité de perte d'un gène favorable devient presque nulle, sauf dans une population vraiment très petite.

Si le nouvel allèle n'est pas perdu, il se substitue à l'ancien, mais cette substitution prend un certain temps, d'autant plus grand que la différence de valeur sélective entre les individus ayant l'ancien allèle et ceux ayant le nouveau est faible. Pendant toute la période où le nouvel allèle augmente en fréquence aux dépens de l'ancien, on observe un polymorphisme, mais qui n'est que **transitoire**.

La même situation s'observe si le milieu se modifie de manière qu'un allèle auparavant défavorable, qui ne pouvait être maintenu qu'à l'état cryptopolymorphe, devient favorable. L'exemple le plus célèbre d'une telle situation est celui du mélanisme industriel, étudié en particulier en Angleterre chez la Phalène du bouleau : à partir de 1850, une forme noirâtre de ce Papillon a remplacé la forme claire dans les régions industrielles où la pollution a assombri les paysages par les dépôts de suie et la destruction des lichens. Cela s'est fait dans certaines zones en moins de cinquante ans ; sachant que cette espèce n'a qu'une génération par an, un remplacement aussi rapide est le signe d'une forte sélection. C'est la prédation par les oiseaux, plus forte sur la forme la plus visible, qui est le facteur sélectif.

Une telle inversion du sens de la sélection, de même que l'apparition d'un nouveau mutant favorable, doivent être des éventualités plutôt rares. La part du polymorphisme qu'elles expliquent est sans doute limitée.

2.4 LE PROBLÈME DU FARDEAU GÉNÉTIQUE

L'hypothèse de la superdominance pour expliquer le polymorphisme a longtemps paru suffisante à la plupart des généticiens, tant que la quantité de polymorphisme était inconnue.

La découverte à partir des années 1960 du polymorphisme intense des populations naturelles grâce à l'électrophorèse a posé aux évolutionnistes un certain nombre de problèmes. Nous en avons déjà évoqué certains, mais le plus important est le problème du **fardeau génétique** (*genetic load*).

De façon générale, on appelle **fardeau génétique** la **diminution de valeur sélective d'une population due au maintien d'individus ayant des valeurs sélectives différentes**. Il peut être mesuré par :

$$L = \frac{w_{\max} - \bar{w}}{w_{\max}}$$

où w_{\max} est la valeur sélective des meilleurs individus de la population, et \bar{w} la valeur sélective moyenne. Il représente donc la proportion d'individus de la population éliminés par la sélection, comparée à une population qui ne contiendrait que les meilleurs génotypes.

La mutation, qui produit sans cesse des individus tarés, est responsable d'un **fardeau de mutation** existant dans les populations. Nous en avons parlé à propos du cryptopolymorphisme.

Le polymorphisme transitoire crée, pendant la période de remplacement, un **fardeau de substitution**.

Dans le cas d'un polymorphisme équilibré, c'est le maintien du polymorphisme sur lequel s'exerce la sélection qui produit un fardéau génétique; on parle de **fardeau de ségrégation**, puisque c'est la ségrégation mendélienne, à chaque génération, qui refabrique des individus de valeur sélective moindre (par exemple, dans le cas de la superdominance, on ne peut maintenir 100 % d'hétérozygotes en reproduction sexuée; des homozygotes défavorables sont ségrégués à chaque génération). Paradoxalement, c'est la sélection elle-même, responsable du maintien du polymorphisme, qui, ici, maintient le fardéau!

On peut essayer d'appréhender ce problème sur un exemple simple (mais irréaliste).

Prenons un cas de superdominance extrême : $w_2 = 1$; $w_1 = w_3 = 0$. Le polymorphisme est maintenu à $p_e = q_e = \frac{1}{2}$.

Les zygotes produits à chaque génération ont la structure :

$$\frac{1}{4} AA \qquad \frac{1}{2} Aa \qquad \frac{1}{4} aa$$

À chaque génération, la moitié des individus de la population (les homozygotes) ne se reproduit pas. Ceci impose un certain coût à la population, un **fardeau génétique**. Ainsi, dans le cas de l'anémie falciforme, ce fardeau est élevé : il contient la somme des individus anémiques *aa* et les homozygotes *AA* morts par malaria.

Ici, il ne s'agit de maintenir polymorphe par sélection qu'un seul locus, mais si cela se passe de la même façon sur un autre locus, cela signifie l'élimination d'autres individus, homozygotes à ce second locus. Il est évident que de tels mécanismes ne peuvent maintenir beaucoup de locus à l'état polymorphe, car alors, une trop faible fraction de la population se reproduirait. Si pour chaque locus les homozygotes étaient létaux, pour maintenir 100 locus à l'état polymorphe par ce système, il faudrait supposer que la proportion d'individus se reproduisant pour former la génération suivante soit de l'ordre de $(1/2)^{100}$, soit environ 10^{-30} , ce qui est évidemment impossible. Pour en maintenir seulement 10, il faut déjà que la proportion de reproducteurs soit de $1/1000$, ce qui est inconcevable dans certaines espèces telles que l'espèce humaine (par exemple). Nous nous sommes placés dans le cas le plus défavorable : bien entendu, on peut supposer des différences de valeurs sélectives beaucoup plus faibles entre homozygotes et hétérozygotes que dans notre exemple, mais même alors, un calcul analogue montre qu'il paraît impossible qu'une sélection avec superdominance maintienne le polymorphisme de plusieurs milliers de locus.

Le problème est d'ailleurs le même avec tous les autres types de sélection pouvant contribuer au maintien du polymorphisme. En effet, toute sélection entraîne un coût, c'est-à-dire la non-reproduction de certains individus. Plus on veut sélectionner de gènes à la fois, plus ce coût est élevé car on sacrifie beaucoup d'individus dans la population, et plus l'effectif reproducteur réel est faible, ce qui a pour conséquence une dérive accélérée pour les autres gènes. Ce problème est bien connu des sélectionneurs, aussi bien de végétaux que d'animaux : ils se gardent bien de sélectionner d'un coup tous les caractères recherchés dans une nouvelle variété ou race, de peur de perdre par dérive une grande partie des gènes sur lesquels ils n'ont pas de prise directe, y compris ceux qui peuvent être intéressants.

Le point de vue neutraliste

Une solution radicale au problème du fardeau est de supposer que le polymorphisme intense observé dans les populations naturelles ne s'accompagne pas de différences de valeurs sélectives, autrement dit qu'il est **sélectivement neutre**. Il n'y a donc plus aucun coût.

En revanche, il doit y avoir dérive, de sorte qu'ici le polymorphisme ne peut être que transitoire. Il faut donc supposer que les populations sont assez grandes pour que la fixation aléatoire d'un gène ne se fasse que très lentement, et d'autre part que la mutation réintroduit sans cesse de nouveaux allèles neutres. Il en résulte une sorte d'**équilibre dynamique** où de nouveaux allèles se substituent progressivement aux anciens, avant d'être remplacés à leur tour, mais, globalement, la variabilité est maintenue parce que la mutation en recrée autant que le hasard en détruit.

La vision neutraliste de l'évolution est donc la suivante : il se produit sans cesse des mutations. Celles qui sont défavorables tendent à être éliminées; exceptionnellement, le nouveau gène est plus favorable que l'ancien et, s'il n'est pas perdu d'emblée, il le remplace assez rapidement (ce sont ces rares mutations favorables qui sont significatives pour l'évolution). L'originalité des « neutralistes » est d'affirmer que de nombreuses mutations ne sont ni favorables ni défavorables, et c'est donc le hasard qui va déterminer leur devenir : beaucoup de ces mutations neutres sont perdues plus ou moins vite par dérive, mais comme il se produit continuellement, l'une ou l'autre finira par se fixer.

Dans une population de N individus ($2N$ gènes), si nous appelons η le taux de mutation neutre, $2N\eta$ nouveaux gènes neutres sont produits à

chaque génération. Chacun des $2N$ gènes de la population, muté ou non, aura la même probabilité de se fixer dans le futur (c'est précisément ce que signifie le mot « neutre »), soit une probabilité $1/2N$. Par conséquent, c'est un nouveau gène muté de cette génération qui se fixera avec une probabilité $1/2N \times 2N\eta = \eta$. Comme ceci est vrai à chaque génération, on peut dire qu'il apparaît un nouvel allèle neutre destiné à se substituer à l'ancien toutes les $1/\eta$ générations en moyenne. Kimura, qui a développé la théorie neutraliste, a montré aussi que pour une mutation qui doit se fixer, il s'écoule en moyenne $4N$ générations entre son apparition et sa fixation (N étant la taille de la population diploïde) : ce n'est autre que le temps de coalescence.

C'est évidemment une durée considérable dans la plupart des espèces. Pendant tout ce temps, on observe du polymorphisme. La fixation est d'ailleurs toute provisoire puisqu'une autre mutation neutre la remplacera à son tour. Si la population est assez grande ($4N > 1/\eta$), l'allèle destiné à remplacer celui en cours de fixation est déjà apparu et en train d'augmenter à son tour en fréquence, de sorte qu'à tout moment on observe un polymorphisme avec généralement plus de deux allèles par locus (fig. 6).

Il apparaît donc aussi que le rythme de **substitution** d'un allèle neutre par un autre est indépendant de la taille de la population, puisqu'il est égal à $1/\eta$ (c'est-à-dire que si le taux de mutation neutre est de 10^{-6} , il y a substitution d'un allèle par un autre allèle en moyenne toutes les 10^6 générations).

La comparaison des séquences de protéines de divers organismes a montré que le rythme d'évolution de certaines séquences protéiques était essentiellement constant. Ainsi, chez toutes les lignées de Vertébrés, la séquence de l'hémoglobine α a évolué au rythme d'environ une substitution tous les 600 000 ans, et ce, aussi bien dans les lignées qui ont connu une forte et rapide évolution morphologique ou écologique, comme les hominidés ou les ongulés parmi les Mammifères, ou les amniotes parmi les Vertébrés, que dans les lignées n'ayant peu ou pas évolué morphologiquement depuis très longtemps (Sélaciens, Salamandres ou, parmi les Mammifères, les Opossums). Cette apparente constance dans le rythme d'évolution d'une protéine ou d'une séquence d'ADN donnée, ce qu'on a appelé l'**horloge moléculaire**, est un argument puissant en faveur de la théorie neutraliste (bien que la réelle constance de l'horloge moléculaire, comme ses causes, restent encore controversées).

Le rythme d'évolution est sensiblement constant pour chaque protéine, mais il varie considérablement d'une protéine à l'autre : il est de 1,2 substitution par acide aminé et par milliard d'années pour l'hémoglobine α , de 0,3 pour le cytochrome c (un transporteur d'électrons de la chaîne respiratoire), de 0,01 pour l'histone H_4 (protéine qui participe à la structure du chromosome eucaryote), mais de 8,3 pour le fibrinopeptide (segment de chaîne polypeptidique qui, dans le processus de coagulation du sang, est séparé de la fibrine lors du clivage enzymatique du fibrinogène sous l'action de la

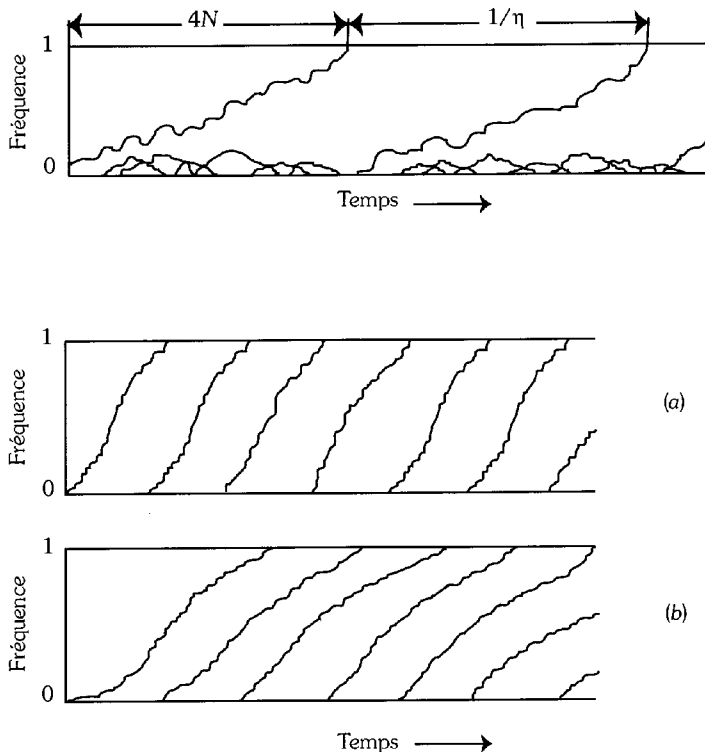


Figure 6 – Processus de la substitution des allèles neutres (d'après Kimura).

En haut : schématisation du processus. De nouveaux allèles neutres apparaissent sans cesse par mutation. La plupart sont perdus, mais certains finissent par se fixer et remplacent les anciens. Cela arrive en moyenne toutes les $1/\eta$ générations, η étant le taux de mutation neutre au locus. Le temps moyen entre l'apparition d'un nouvel allèle neutre destiné à remplacer l'ancien et le moment où il se fixe est de $4N$ générations, N étant la taille de la population.

En bas : comparaison d'une population de petite taille (a) et de grande taille (b). Le taux de mutation neutre est le même, donc le taux de substitution aussi mais comme $4N$ est différent, on trouve à tout moment plus de polymorphisme neutre dans la population de grande taille que dans celle de petite taille.

thrombine). Cela ne signifie pas que le fibrinopeptide subisse plus de mutations que l'histone H₄, mais que la proportion des mutations qui sont neutres est très différente entre ces deux protéines : l'histone H₄ a pratiquement la même séquence chez tous les eucaryotes, de l'homme au petit pois. Ce doit être le signe que toute mutation modifiant sa structure primaire est contre-sélectionnée; au contraire, le fibrinopeptide n'a pour fonction que d'empêcher, par sa présence à l'extrémité N-terminale du fibrinogène plasmatique, la polymérisation spontanée en réseau de fibrine. La séquence précise des acides aminés qui le constituent n'aurait guère d'importance, aussi beaucoup des substitutions qui s'y produisent sont neutres. De même, au niveau de l'ADN, certaines séquences évoluent plus vite que d'autres : ainsi, les introns évoluent plus rapidement que les exons, signe que les changements de séquence des premiers sont plus souvent sans importance fonctionnelle que pour les seconds; dans les exons, quand on compare les séquences d'espèces voisines, les troisièmes bases des codons ont un taux de substitution supérieur aux deux premières, pour la même raison. De façon générale, les portions non codantes du génome évoluent plus vite. Tout cela est le contraire de ce qu'on attendrait sur une base sélectionniste : si une majorité des évolutions observées étaient adaptatives, les molécules et séquences les plus importantes dans l'adaptation des organismes seraient celles que l'on voit évoluer le plus vite.

Il est important de comprendre que la théorie neutraliste ne conteste pas le rôle de la sélection naturelle dans l'évolution des organismes, pas plus qu'ils ne contestent que le polymorphisme de l'anémie falciforme soit sélectionné. Simplement, dans ce modèle, on considère que, dans la quantité de polymorphismes révélés par l'électrophorèse ou l'étude de l'ADN, la plus grande partie est neutre, les cas de polymorphismes sélectionnés étant une toute petite minorité. De même, dans l'évolution des protéines, quelques substitutions ont certainement une signification adaptative (on ne pourrait peut-être pas échanger impunément l'hémoglobine α de la carpe et de l'homme, différentes pour 68 acides aminés), mais ces substitutions adaptatives sont noyées dans une masse de substitutions neutres, sans signification au niveau de l'adaptation.

Le débat entre « sélectionnistes » et « neutralistes », né à la fin des années 1960 des nouvelles données apportées par l'électrophorèse et des travaux de Kimura, a été très vif dans les années 1970-1980, et a donné lieu à une quantité considérable de recherches. Mais il est impossible de prouver qu'un polymorphisme est neutre (on ne peut jamais être sûr d'avoir éliminé tous les facteurs sélectifs) et souvent très difficile de prouver qu'il ne l'est pas. C'est sans doute ce qui a permis à la discussion de devenir très animée. Les calculs des neutralistes ont en tout cas montré que leurs hypothèses pouvaient rendre compte de la plupart des faits observés et permettaient des prédictions apparemment justes sur l'évolution moléculaire et la structura-

tion de la variabilité. Aussi, alors que dans les années 1950-1960 bien des évolutionnistes refusaient l'idée même qu'une mutation neutre puisse exister, la plupart admettent aujourd'hui qu'une bonne part du polymorphisme enzymatique et de l'ADN pourrait être neutre, même si des désaccords subsistent sur la part respective du hasard et de la sélection.

Exercices du chapitre C

Exercice C.1

Chez trois espèces cultivées, on a procédé à la même expérience : on a autofécondé différentes plantes appartenant à la même variété cultivée et on a semé les graines issues de cette autofécondation. Certaines plantes étaient albinos, c'est-à-dire dépourvues de chlorophylle, un phénotype létal, la plantule mourant dès qu'elle a épuisé les ressources de la graine. La proportion de plantules albinos, sur 10 000 graines semées, a été :

chez le millet	0
chez le mil	12
chez la vigne	2522

Comment expliquez-vous ces résultats? Quelles sont les caractéristiques des variétés qu'on peut espérer commercialiser chez ces différentes espèces?

Exercice C.2

Dans une population de plantes qu'on supposera infinie (pourquoi?) coexistent deux allèles A_1 et A_2 . À la génération g , on étudie les individus adultes de cette population. Les fréquences de ces deux allèles sont alors respectivement p et q ($= 1 - p$). La population se reproduit en panmixie.

Quelles seront les proportions des différents génotypes dans la population adulte à la génération $g + 1$ et comment ces proportions évolueront-elles dans le futur :

1) si les deux allèles n'ont aucun effet sur la valeur sélective des individus qui les portent?

2) si les viabilités (probabilités de survivre du stade zygote au stade adulte) des trois génotypes sont égales mais que les fertilités (nombre de descendants fabriqués par un adulte en espérance) sont telles que :

$$w(A_1A_2) = w(A_2A_2) = w(A_1A_1)/2$$

3) Comment la réponse à la question précédente serait-elle modifiée si la différence portait non plus sur la fertilité mais sur la viabilité?

4) Application numérique : la population de départ contient :

10 individus A_1A_1 , 40 individus A_1A_2 et 40 individus A_2A_2 .

D

**Problèmes
de synthèse corrigés**

Sélection sur le taux d'autofécondation

Dans une population de plantes, deux types d'individus se trouvent réunis. Ces deux types sont semblables en tout sauf en ce qui concerne le taux d'autofécondation.

Les individus de type (1) autofécondent une fraction s de leurs gamètes femelles; les individus de type (2) autofécondent une fraction $s + x$ de leurs gamètes femelles. La quantité de pollen utilisé pour l'autofécondation est négligeable.

1) Combien de copies de leurs gènes les deux types d'individus laissent-ils dans leurs propres graines?

2) On appellera P le nombre de copies de leurs gènes laissées par les plantes de type (1) *via* leur pollen fécondant d'autres plantes. Les graines issues d'allofécondation germent h fois mieux que les graines issues d'autofécondation (pourquoi h ?). Combien de copies de leurs gènes les deux types d'individus laissent-ils à la génération suivante au stade plante? Comparer ces deux valeurs et en déduire l'évolution de la population.

Corrigé

1) Appelons g le nombre de graines produites par les individus; dans leurs propres graines, les individus s'autofécondent pour une fraction s (type (1)) laissent g copies de leurs gènes *via* leurs gamètes femelles mais aussi gs *via* le pollen qui a fécondé leurs propres ovules. Au total, ces individus laissent $g(1+s)$ copies de leurs gènes dans leurs propres graines. De même, les autres (type (2)) laissent $g(1+s+x)$ copies de leurs gènes dans leurs graines.

2) Puisque l'autofécondation ne consomme pas de pollen, les plantes de type (1) et de type (2) fécondent en espérance autant de plantes les unes que les autres. Par conséquent, les deux types de plantes auront laissé, dans les graines produites, des copies de leurs gènes en nombre : $g(1+s) + P$ pour les plantes de type (1) et $g(1+s+x) + P$ pour les plantes de type (2). Les graines issues d'allofécondation représentent respectivement $g(1-s)$ et $g(1-s-x)$ pour les deux types de plantes; le reste (soit $2sg$ et $2(s+x)g$) correspond aux gènes se trouvant dans des graines issues d'autofécondation. Remarquons que le fait que ces graines proviennent d'autofécondation aboutit au fait que ces graines portent deux génomes de leur plante mère (père) chacune. Après germination, l'hétérosis (qui justifie le paramètre h) agit en favorisant les individus issus d'allofécondation. Le nombre de copies du génome produites sera :

$$w_1 = gh(1-s) + 2sg + Ph \quad \text{pour le type (1)}$$

et $w_2 = gh(1-s-x) + 2(s+x)g + Ph \quad \text{pour le type (2)}$

La différence est donc de :

$$\begin{aligned} w_2 - w_1 &= gh(1-s-x) + 2(s+x)g + Ph - gh(1-s) - 2sg - Ph \\ &= g[h(1-s-x) + 2(s+x) - h(1-s) - 2s] \\ &= g[2x - hx] \\ &= gx(2-h) \end{aligned}$$

On voit que selon le signe de $2 - h$, la population va évoluer dans des directions opposées. Si h est inférieur à 2 (si les graines allofécondées sont moins de deux fois meilleures que les graines issues d'autofécondation), alors $w_2 - w_1$ est du signe de x . C'est-à-dire que c'est le génotype qui s'autoféconde le plus qui transmet le mieux ses gènes. L'espèce va devenir autogame.

Au contraire, si h est supérieur à 2 (si les graines allofécondées sont plus de deux fois meilleures que les graines issues d'autofécondation), alors $w_2 - w_1$ est de signe opposé à x . C'est-à-dire que c'est le génotype qui s'autoféconde le moins qui transmet le mieux ses gènes. L'espèce va devenir allogame.

2

Tétraploïdes

Dans une population, un locus présente deux allèles A et a .

1) Quels sont les génotypes en présence et leurs proportions en panmixie (en fonction des proportions respectives p et q de A et a) :

a) si l'espèce est diploïde?

b) s'il s'agit d'un autotétraploïde?

2) Mêmes questions en autogamie stricte et prolongée.

3) Mêmes questions (génotypes + fréquences en panmixie et autogamie) pour un allotétraploïde issu de la fusion de deux espèces possédant respectivement A en proportions p_1 et p_2 et a en proportions q_1 et q_2 .

Corrigé

1) Si l'espèce est diploïde, on trouve les célèbres proportions de Hardy-Weinberg :

$$p^2 AA, \quad 2pq Aa \quad \text{et} \quad q^2 aa$$

2) Si l'espèce est autotétraploïde, on trouve :

$$p^4 AAAA, \quad 4p^3q AAAa, \quad 6p^2q^2 Aaaa, \quad 4pq^3 Aaaa \quad \text{et} \quad q^4 aaaa$$

L'autogamie fait disparaître les hétérozygotes, on obtient donc :

- chez un diploïde : $p AA$ et $q aa$
- chez un autotétraploïde : $p AAAA$ et $q aaaa$

3) En panmixie :

$$\begin{aligned} &AAAA \quad p_1^2 p_2^2 \\ &AAAa \quad 2p_1^2 p_2 q_2 + 2p_1 q_1 p_2^2 \\ &AAaa \quad p_1^2 q_2^2 + q_1^2 p_2^2 + 4p_1 q_1 p_2 q_2 \\ &Aaaa \quad 2p_1 q_1 q_2^2 + 2q_1^2 p_2 q_2 \\ &aaaa \quad q_1^2 q_2^2 \end{aligned}$$

En autogamie :

$$\begin{aligned} &AAAA \quad p_1 p_2 \\ &AAAa \quad p_1 q_2 + q_1 p_2 \\ &aaaa \quad q_1 q_2 \end{aligned}$$

Les chémotypes de thym

Le thym (*Thymus vulgaris* L.) tire son odeur de monoterpènes. On a pu montrer que différents individus possédaient une odeur différente due à la production de molécules différentes. Ces molécules sont distinguables en chromatographie en phase gazeuse. Dans une population naturelle, on a prélevé des feuilles sur 200 individus et on les a analysées :

- 38 individus produisaient du Linalol, molécule non cyclique (phénotype [L]) ;
- 90 individus produisaient du Carvacrol, molécule phénolique (phénotype [C]) ;
- 72 individus produisaient du Thymol, molécule phénolique (phénotype [T]).

1) Quelques individus ont été transplantés en pépinière et plusieurs générations de croisements ont été réalisées. Les parents et descendants des différents croisements ont été analysés. Les différents croisements ont tous pu être classés dans une des disjonctions données dans le tableau ci-après (les phénotypes sont toujours indiqués entre crochets).

Quelle hypothèse génétique a-t-elle été utilisée? (On pourra, pour répondre, commencer par considérer le cas des croisements n'incluant pas [L], puis reprendre l'ensemble en regroupant les phénotypes [C] et [T] en un seul phénotype [P] et enfin conclure en reprenant un ou quelques croisements particuliers.)

Dans cette hypothèse, à quels génotypes les différents cas du tableau correspondent-ils?

Parents	Descendants		
	[L]	[C]	[T]
$[T] \times [T]$	0	0	1
$[C] \times [C]$	0	1	0
$[C] \times [C]$	0	$\frac{3}{4}$	$\frac{1}{4}$
$[C] \times [T]$	0	1	0
$[C] \times [T]$	0	$\frac{1}{2}$	$\frac{1}{2}$
$[L] \times [L]$	1	0	0
$[L] \times [L]$	$\frac{3}{4}$	$\frac{1}{4}$	0
$[L] \times [L]$	$\frac{3}{4}$	0	$\frac{1}{4}$
$[L] \times [L]$	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
$[L] \times [L]$	$\frac{3}{4}$	$\frac{3}{16}$	$\frac{1}{16}$
$[L] \times [C]$	1	0	0
$[L] \times [C]$	$\frac{1}{2}$	$\frac{1}{2}$	0
$[L] \times [C]$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$
$[L] \times [T]$	$\frac{1}{2}$	$\frac{3}{8}$	$\frac{1}{8}$
$[L] \times [T]$	1	0	0
$[L] \times [T]$	$\frac{1}{2}$	$\frac{1}{2}$	0
$[L] \times [T]$	$\frac{1}{2}$	0	$\frac{1}{2}$
$[L] \times [T]$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$

2) Toujours dans l'hypothèse ci-dessus, quels seraient les génotypes en présence dans la population d'origine :

- a)** Si le thym était autogame à 100 %?
- b)** Si le thym était allogame et se reproduisait en panmixie pour le caractère en question?
- c)** Si le thym était allogame mais que les abeilles se spécialisaient sur un type d'odeur (pollen transporté uniquement entre individus de même odeur)?

3) Dans chacun de ces trois cas, quelles seraient les fréquences des différents génotypes en présence en fonction des fréquences alléliques? En déduire les fréquences alléliques dans la population de départ pour les trois cas considérés. De quelle(s) hypothèse(s) supplémentaire(s) avez-vous besoin pour répondre à ces questions?

Corrigé

1) Les croisements entre les formes $[T]$ et $[C]$ sont typiques de la disjonction d'un locus à deux allèles. Les croisements $[T] \times [T]$ ne donnent que des formes $[T]$ tandis que les croisements $[C] \times [C]$ donnent $[C]$ et $[T]$. Ceci signifie que l'information génétique codant la forme $[T]$ peut être « cachée » dans des individus $[C]$, la réciproque n'étant pas vraie. Il s'ensuit que l'allèle déterminant la forme $[T]$ (nous l'appellerons c) est récessif et que celui déterminant la forme $[C]$ (nous l'appellerons C) est dominant. Les croisements $[C] \times [C]$ qui produisent des $[T]$ les produisent bien en proportion $1/4$ comme attendu. De même, les croisements $[C] \times [T]$ qui disjoignent donnent les proportions attendues $1/2 : 1/2$.

Si on regroupe $[C]$ et $[T]$ dans une même classe (soit $[P]$), on retrouve exactement le même schéma vis-à-vis de $[L]$; la forme $[P]$ est récessive face à $[L]$ qui est déterminé par un allèle dominant (qu'on appellera L).

2) On doit alors se demander si les allèles L , C et c se trouvent au même locus ou sur des loci différents. La dernière ligne du tableau fournit la réponse à cette question. En effet, le parent $[T]$ est nécessairement de génotype cc . Le parent $[L]$ est nécessairement de génotype L^* (* désigne soit L soit C soit c) : il ne peut pas être LL sinon tous les descendants seraient $[L]$, il ne peut pas non plus être LC sinon les descendants seraient $1/2[L] + 1/2[C]$; il ne peut pas non plus être Lc sinon les descendants seraient $1/2[L] + 1/2[T]$. Il y a donc deux loci. L'un porte deux allèles L et l , l'autre porte deux allèles C et c . Le premier est épistatique sur le second en ce que la présence de l'allèle L suffit à déterminer le phénotype $[L]$. Seuls les individus présentant le génotype ll à ce locus exprimeront le second locus déterminant les phénotypes $[C]$ et $[T]$. Les génotypes des parents des croisements sont donnés dans le tableau suivant.

a) Si le thym était autogame à 100 %, il n'y aurait que des homozygotes dans la population, soit les 4 génotypes : ***LLCC***, ***LLcc***, ***llCC*** et ***llcc***.

			Descendants		
Parents			[L]	[C]	[T]
[T] × [T]	<i>ll cc</i>	<i>ll cc</i>	0	0	1
[C] × [C]	<i>ll CC</i>	<i>ll C*</i>	0	1	0
[C] × [C]	<i>ll Cc</i>	<i>ll Cc</i>	0	3/4	1/4
[C] × [T]	<i>ll CC</i>	<i>ll cc</i>	0	1	0
[C] × [T]	<i>ll Cc</i>	<i>ll cc</i>	0	1/2	1/2
[L] × [L]	<i>LL **</i>	<i>L* **</i>	1	0	0
[L] × [L]	<i>Ll CC</i>	<i>Ll C*</i>	3/4	1/4	0
[L] × [L]	<i>Ll cc</i>	<i>Ll cc</i>	3/4	0	1/4
[L] × [L]	<i>Ll Cc</i>	<i>Ll cc</i>	3/4	1/8	1/8
[L] × [L]	<i>Ll Cc</i>	<i>Ll Cc</i>	3/4	3/16	1/16
[L] × [C]	<i>LL **</i>	<i>ll C*</i>	1	0	0
[L] × [C]	<i>Ll **</i>	<i>ll CC (ou Ll CC et ll Cc)</i>	1/2	1/2	0
[L] × [C]	<i>Ll cc</i>	<i>ll Cc</i>	1/2	1/4	1/4
[L] × [C]	<i>Ll Cc</i>	<i>ll Cc</i>	1/2	3/8	1/8
[L] × [T]	<i>LL **</i>	<i>ll cc</i>	1	0	0
[L] × [T]	<i>Ll CC</i>	<i>ll cc</i>	1/2	1/2	0
[L] × [T]	<i>Ll cc</i>	<i>ll cc</i>	1/2	0	1/2
[L] × [T]	<i>Ll Cc</i>	<i>ll cc</i>	1/2	1/4	1/4

N.B. Les étoiles signifient que l'allèle présent est indéterminé. D'autre part, lorsque les deux parents sont équivalents (comme à la deuxième ligne) on a choisi un ordre mais il pourrait bien sûr être inversé.

b) En panmixie, on trouverait 9 génotypes : ***LLCC***, ***LLCc***, ***LLcc***, ***LlCC***, ***LlCc***, ***Llcc***, ***llCC***, ***llCc*** et ***llcc***.

c) L'homogamie, en faisant se reproduire ensemble les individus [L], aboutirait à la disparition des hétérozygotes ***Ll*** ce qui laisse, pour les individus de type [L], trois génotypes : ***LLCC***, ***LLCc*** et ***LLcc***. Du fait que les [C] et les [T] se reproduisent également entre eux, les hétérozygotes ***Cc***

disparaissent quand ils sont associés au génotype **II**. Ce qui laisse deux génotypes : **IIcc** pour les [C] et **IIcc** pour les [7]. On peut remarquer que l'homogamie ne fait pas disparaître le génotype **Cc** quand il est associé à **LL** puisque parmi les plantes **LL**, il n'y a pas de différence d'odeur entre les **CC**, les **Cc** et les **cc**. À l'intérieur du groupe de reproduction des plantes **LL** (qui est un groupe fermé) il y a panmixie pour les allèles **C/c**.

3) Soient p_1 et $q_1 (=1 - p_1)$ les proportions respectives de **L** et **l** et p_2 et $q_2 (=1 - p_2)$ les proportions respectives de **C** et **c**.

En autogamie, les génotypes seraient en proportions :

$$p_1 p_2 \text{ LLCC}, p_1 q_2 \text{ LLc}, q_1 p_2 \text{ IIcC} \text{ et } q_1 q_2 \text{ IIcc}$$

On en déduit que $p_1 = 38/200 = 19\%$ (d'où $q_1 = 81\%$) et $p_2 = 90/162 = 10/18$ (d'où $q_2 = 8/18$).

En panmixie, on attendrait $p_1^2 p_2^2 \text{ LLCC}$, $2p_1^2 p_2 q_2 \text{ LLCc}$, $p_1^2 q_2^2 \text{ LLcc}$, $2p_1 q_1 p_2^2 \text{ LICC}$, $4p_1 q_1 p_2 q_2 \text{ LICc}$, $2p_1 q_1 q_2^2 \text{ LIcc}$, $q_1^2 p_2^2 \text{ IIcC}$, $2q_1^2 p_2 q_2 \text{ IIcC}$ et $q_1^2 q_2^2 \text{ IIcc}$.

On en déduit que $q_1^2 = 162/200 = 0,81$ d'où $q_1 = 90\%$ (et $p_1 = 10\%$) et $q_2^2 = 72/162 = 4/9$ d'où $q_2 = 2/3$ (et $p_2 = 1/3$).

En homogamie, à condition de supposer l'absence de déséquilibre de liaison entre les deux loci (ce qui dans ce cas n'est pas très raisonnable; en effet, les groupes **LL** et **II** sont totalement isolés et il est probable qu'ils ne restent pas semblables pour leurs proportions de **C**, etc.), on peut calculer les proportions des 5 génotypes : $p_1 p_2^2 \text{ LLCC}$, $2p_1 p_2 q_2 \text{ LLCc}$, $p_1 q_2^2 \text{ LLcc}$, $q_1 p_2 \text{ IIcC}$ et $q_1 q_2 \text{ IIcc}$.

On en déduit (de même que pour l'autogamie étant donné qu'on ne « voit » que les génotypes homozygotes) que $p_1 = 38/200 = 19\%$ (d'où $q_1 = 81\%$) et $p_2 = 90/162 = 10/18$ (d'où $q_2 = 8/18$).

4

Deux populations : l'effet Wahlund

Deux grandes populations (V et W) **isolées** et **panmictiques** de plantes hermaphrodites montrent un polymorphisme. Dans chacune, on trouve des plantes à fleurs blanches $[B]$ et des plantes à fleurs rouges $[R]$. Ce caractère ne modifie ni la fertilité ni la viabilité ni les conditions de fécondation des individus.

I À l'intérieur de chacune de ces deux populations, on a effectué des autofécondations et des croisements.

- Toutes les plantes $[B]$ autofécondées ou croisées entre elles ont toujours donné des descendants $[B]$.
- Certaines plantes $[R]$ autofécondées ou croisées avec n'importe quelle autre plante ont donné des descendants, tous $[R]$.
- Les autres plantes $[R]$ ont donné en autofécondation environ trois descendants $[R]$ pour un $[B]$.

Quel serait le résultat d'un croisement de ce dernier type de plantes avec une plante $[B]$? Et avec une autre plante $[R]$?

II Les fréquences des allèles responsables respectivement des types $[R]$ et $[B]$ sont p_v et q_v dans la population V et p_w et q_w dans la population W . Tous les résultats seront donnés en fonction de ces fréquences.

1) Quelles sont les proportions des génotypes et des phénotypes dans ces deux populations ?

2) On prélève un grand nombre de graines sur ces deux populations séparément :

- d'une part on effectue des croisements entre plantes de ces deux populations,
- d'autre part, on constitue une nouvelle population, U (qu'on supposera de taille infinie), issue d'un mélange d'un nombre égal de graines issues des deux populations V et W . Ces graines sont semées et les plantes produites constituent la génération 0 de la population U . Ces plantes se croisent entre elles en panmixie. Les graines ainsi produites, les plantes qui en sont issues correspondent à la génération 1 de la population U .

Nous allons envisager deux possibilités.

a) Les très nombreux croisements effectués entre plantes à fleurs blanches des deux populations donnent toujours des descendants à fleurs blanches. Qu'en déduisez-vous?

Quelles sont les proportions des allèles, des génotypes et des phénotypes chez les plantes de la génération 0 de la population U ? Ces proportions correspondent-elles aux proportions de Hardy-Weinberg? Pourquoi?

Quelles sont les proportions des différents génotypes dans les gamètes issus de la génération 0 de la population U ?

Quelles sont les proportions des allèles, des génotypes et des phénotypes chez les plantes de la génération 1 de la population U ?

Que deviendront ces fréquences dans les générations à venir? et à l'infini?

b) Les très nombreux croisements effectués entre plantes à fleurs blanches des deux populations donnent toujours des descendants à fleurs rouges. Qu'en déduisez-vous?

Quelles sont les proportions des allèles, des génotypes et des phénotypes chez les plantes de la génération 0 de la population U ? La loi de Hardy-Weinberg s'applique-t-elle? Pourquoi?

Quelles sont les proportions des différents génotypes dans les gamètes issus de la génération 0 de la population U ?

Quelles sont les proportions des allèles, des génotypes et des phénotypes chez les plantes de la génération 1 de la population U ?

Que deviendront ces fréquences dans les générations à venir? et à l'infini?

Corrigé

I Les autofécondations de plantes [R] peuvent donner des plantes [B]. On peut donc soupçonner que [R] est dominant sur [B]. Ceci est confirmé par les proportions 3/4, 1/4 obtenues en autofécondation. Un locus R/r déterminerait alors la couleur, avec RR et Rr déterminant le phénotype [R] et rr la couleur [B]. Ceci est confirmé par le fait que les croisements entre plantes [B] donnent toujours [B]. Le croisement de plantes Rr par une plante rr doit donner 1/2 [R] et 1/2 [B]. En croisement avec une autre plante [R], cette plante donnera soit uniquement des descendants [R] (si l'autre plante est RR) soit 3/4 [R] et 1/4 [B] (si l'autre plante est Rr).

II 1) Ces deux populations sont panmictiques, on va donc trouver les proportions suivantes :

Population V :	$p_v^2 RR, 2 p_v q_v Rr \text{ et } q_v^2 rr$
soit	$p_v^2 + 2 p_v q_v [R] \text{ et } q_v^2 [B]$
Population W :	$p_w^2 RR, 2 p_w q_w Rr \text{ et } q_w^2 rr$
soit	$p_w^2 + 2 p_w q_w [R] \text{ et } q_w^2 [B]$

2) a) Le croisement entre deux génotypes homozygotes récessifs (test de complémentation) permet de savoir, dans ce cas, que les gènes déterminant la couleur de la fleur dans les deux populations sont situés au même locus.

Le mélange en proportions égales des deux populations produit des proportions alléliques qui vont simplement être les moyennes des proportions précédentes :

$$p_u = \frac{1}{2}(p_v + p_w) \quad \text{et} \quad q_u = \frac{1}{2}(q_v + q_w)$$

Puisque ce sont des graines qui sont mélangées, les plantes qui en sont issues vont refléter les proportions de ces graines et les proportions des génotypes seront les moyennes de ces proportions. On trouvera donc, dans U , les génotypes dans les proportions :

$$\frac{1}{2}(p_v^2 + p_w^2)RR, Ho = (p_v q_v + p_w q_w)Rr \quad \text{et} \quad \frac{1}{2}(q_v^2 + q_w^2)rr$$

soit, pour les phénotypes :

$$\frac{1}{2}(p_v^2 + 2p_v q_v + p_w^2 + 2p_w q_w)[R] \quad \text{et} \quad \frac{1}{2}(q_v^2 + q_w^2)[B]$$

Les proportions de Hardy-Weinberg sont :

$$p_u^2 RR, 2p_u q_u Rr \quad \text{et} \quad q_u^2 rr$$

soit une proportion d'hétérozygotes de :

$$\begin{aligned} He &= 2p_u q_u = 2 \frac{1}{2}(p_v + p_w) \frac{1}{2}(q_v + q_w) \\ &= \frac{1}{2}(p_v q_v + p_w q_w + p_v q_w + p_w q_v) \end{aligned}$$

La différence entre la proportion d'hétérozygotes trouvée dans U (Ho) et celle qui serait attendue selon Hardy-Weinberg (He) est :

$$\begin{aligned} He - Ho &= \frac{1}{2}(p_v q_v + p_w q_w + p_v q_w + p_w q_v) - p_v q_v - p_w q_w \\ &= \frac{1}{2}(-p_v q_v - p_w q_w + p_v q_w + p_w q_v) \\ &= \frac{1}{2}[p_v(-q_v + q_w) + p_w(q_v - q_w)] = \frac{1}{2}(p_v - p_w)(q_w - q_v) \end{aligned}$$

Remarquons que $q_w - q_v = (1 - p_w) - (1 - p_v) = (p_v - p_w)$

d'où
$$He - Ho = \frac{1}{2}(p_v - p_w)^2$$

Cette valeur est positive tant que $p_v - p_w$ n'est pas nul, elle est nulle si $p_v - p_w = 0$. En d'autres termes, si deux populations panmixtiques sont différenciées, le mélange d'individus provenant de ces deux populations ne se comporte pas de façon panmixtique. Cet effet est appelé **effet Wahlund**. La non-panmixie sur l'ensemble vient du fait que les croisements ne se font pas au hasard sur l'ensemble mais seulement au sein de chaque groupe. Si les deux groupes sont différents, ces croisements endogames ont pour conséquence que les individus qui se rencontrent se ressemblent plus que si les croisements se faisaient au hasard et donc que la proportion d'hétérozygotes est toujours inférieure à celle qu'on attendrait en panmixie.

Il n'y a pas de pression évolutive, de ce fait les proportions dans U restent inchangées pour la suite des générations :

$$p_u = \frac{1}{2}(p_v + p_w) \quad \text{et} \quad q_u = \frac{1}{2}(q_v + q_w)$$

Ceci est vrai dans les gamètes et dans les plantes qui en seront issues.

La population U suit toutes les hypothèses de Hardy-Weinberg. Les proportions seront donc, dès la première génération, chez les graines et les individus qui en seront issus :

$$p_u^2 RR, 2p_u q_u Rr \text{ et } q_u^2 rr$$

b) Le test de complémentation est positif : les croisements entre génotypes homozygotes récessifs produisent le phénotype dominant. On peut donc en déduire que deux loci sont impliqués. On peut également en déduire que chacune des populations possède, à l'état fixé, l'allèle dominant correspondant à la couleur dans l'autre population. Les plantes de type [B] de la population V ont donc pour génotype $r_1 r_1 R_2 R_2$ tandis que celles de la population W ont pour génotype $R_1 R_1 r_2 r_2$. L'allèle R_1 a donc la proportion p_v dans V et 1 dans W , réciproquement, l'allèle R_2 est en proportion 1 dans V et p_w dans W .

La génération 0 de la population U aura alors les proportions suivantes :

$$\text{Pour les allèles : } \frac{1}{2}(1 + p_v)R_1 \text{ et } \frac{1}{2}q_v r_1,$$

$$\text{de même, } \frac{1}{2}(1 - p_w)R_2 \text{ et } \frac{1}{2}q_w r_2.$$

Pour les génotypes, la moitié des graines de la population U vient de la population V , soit :

$$R_1 R_1 R_2 R_2 : \frac{1}{2}p_v^2$$

$$R_1 r_1 R_2 R_2 : p_v q_v$$

$$r_1 r_1 R_2 R_2 : \frac{1}{2}q_v^2$$

L'autre moitié vient de la population W , soit :

$$R_1 R_1 R_2 R_2 : \frac{1}{2}p_w^2$$

$$R_1 R_1 R_2 r_2 : p_w q_w$$

$$R_1 R_1 r_2 r_2 : \frac{1}{2}q_w^2$$

Au total, ceci donne :

$$R_1 R_1 R_2 R_2 : \frac{1}{2}(p_v^2 + p_w^2)$$

$$R_1 r_1 R_2 R_2 : p_v q_v$$

$$r_1 r_1 R_2 R_2 : \frac{1}{2}q_v^2$$

$$R_1 R_1 R_2 r_2 : p_w q_w$$

$$R_1 R_1 r_2 r_2 : \frac{1}{2} q_w^2$$

soit, pour les phénotypes :

$$\frac{1}{2}(p_v^2 + p_w^2 + 2p_v q_v + 2p_w q_w) [R] \text{ et } \frac{1}{2}(q_v^2 + q_w^2) [B]$$

ce qui, remarquons-le, est similaire à ce qu'on avait trouvé en a).

Ici encore, la loi de Hardy-Weinberg ne s'applique pas puisque, pour chacun des deux loci, on a mélangé des génotypes provenant de populations différenciées (sauf dans le cas trivial où $p_v = p_w = 1$).

La génération 0 produira 3 types de gamètes dont les génotypes et leurs proportions sont donnés ci-dessous.

$$R_1 R_2 : \frac{1}{2}(p_v^2 + p_w^2 + p_v q_v + p_w q_w) = \frac{1}{2}(p_v + p_w),$$

$$R_1 r_2 : \frac{1}{2}(q_w^2 + p_w q_w) = \frac{1}{2} q_w$$

$$r_1 R_2 : \frac{1}{2}(q_v^2 + p_v q_v) = \frac{1}{2} q_v$$

L'association de ces gamètes ne changera pas les proportions alléliques (pas de pression évolutive) qui vont rester égales à $\frac{1}{2}(1 + p_v)R_1$ et $\frac{1}{2}q_v r_1$. De même, pour le locus R_2 , on aura $\frac{1}{2}(1 + p_w)R_2$ et $\frac{1}{2}q_w r_2$.

À la génération 1, les génotypes seront constitués par l'association au hasard des gamètes ci-dessus, soit (il peut être utile de construire le tableau pour ne pas oublier de compter deux fois les croisements qui se font dans les deux sens) :

$$R_1 R_1 R_2 R_2 : \frac{1}{4}(p_v + p_w)^2$$

$$R_1 R_1 R_2 r_2 : \frac{1}{2} q_w (p_v + p_w)$$

$$R_1 r_1 R_2 R_2 : \frac{1}{2} q_v (p_v + p_w)$$

$$R_1 r_1 R_2 r_2 : \frac{1}{2} q_v q_w$$

$$R_1 R_1 r_2 r_2 : \frac{1}{4} q_w^2$$

$$r_1 r_1 R_2 R_2 : \frac{1}{4} q_v^2$$

Seuls les deux derniers génotypes ci-dessus donnent le phénotype [B] et les proportions des phénotypes sont donc :

$$\frac{1}{4} [(p_v + p_w)^2 + 2(q_v + q_w)(p_v + p_w) + 2q_v q_w] \quad \text{de type [R]}$$

$$\frac{1}{4} (q_v^2 + q_w^2) \quad \text{de type [B].}$$

Dans les générations à venir, si les loci portant les allèles R_1/r_1 et R_2/r_2 ne sont pas totalement liés, le déséquilibre de liaison va disparaître sous l'action de la recombinaison et de la panmixie et on tendra, à l'infini, vers une situation d'indépendance statistique des deux loci avec les proportions suivantes dans les gamètes :

$$R_1 R_2 : \frac{1}{4} (1 + p_v)(1 + p_w)$$

$$R_1 r_2 : \frac{1}{4} q_w (1 + p_v)$$

$$r_1 R_2 : \frac{1}{4} q_v (1 + p_w)$$

$$r_1 r_2 : \frac{1}{4} q_v q_w.$$

Date de floraison

I Chez une espèce végétale allogame annuelle, l'étude de la date de floraison a montré qu'il existait deux types d'individus ne fleurissant pas au même moment; ces deux types de plantes sont rigoureusement identiques pour toutes leurs autres caractéristiques. Cette espèce peut pousser dans deux milieux : au soleil ou à l'ombre, ce qui modifie également la date de floraison. Les deux types de plantes seront notés [1] et [2] respectivement. Les dates de floraison sont données dans le tableau suivant.

En 1989, on a croisé une plante de type [1] et une plante de type [2]. En 1990, les graines issues de ce croisement ont été semées dans chacun des deux milieux et ont montré les caractéristiques données dans le tableau (ligne [1] x [2]). Dans chacun des deux milieux, on a laissé les plantes issues de ce croisement se féconder librement entre elles (*chacune des deux populations étant totalement isolée*) pour produire la floraison F2. Les générations suivantes seront également constituées par le semis des graines dans le milieu où se trouvait leur mère. Chaque milieu contient 4000 plantes. En F2, on a dénombré les plantes fleurissant à chaque date. Tous ces résultats sont donnés dans le tableau.

1) Quel déterminisme pouvez-vous proposer pour la date de floraison dans cette espèce?

2) Quels sont les effectifs espérés des différents génotypes parmi les plantes fleurissant en 1991 aux différentes dates dans les deux populations?

3) Les fréquences génotypiques vont-elles changer? Sous l'action de quel type de processus? Mêmes questions pour les fréquences alléliques.

4) Nous allons nous intéresser à la population se trouvant au soleil.

a) Quelles sont les proportions des allèles (p_1 et $q_1 = 1 - p_1$) dans le pollen au 1^{er} juin 1991?

b) Quelles sont, en fonction de p_1 et q_1 , les proportions des différents génotypes dans les graines produites par les plantes fleurissant au 1^{er} juin 1991?

c) Quelles seront, en fonction de p_1 et q_1 , les proportions p_2 et q_2 des allèles parmi les plantes fleurissant au 1^{er} juin 1992?

d) Comment va évoluer la population? *Une réponse qualitative est suffisante. On peut aussi calculer q_n en fonction de q_{n-1} puis de q_1 en utilisant la suite des inverses ($1/q_n$). À l'équilibre, quels seraient les effectifs de plantes fleurissant aux différentes dates dans la population se trouvant au soleil?*

5) Que va-t-il se passer dans la population se trouvant à l'ombre?

II Si les deux populations (soleil et ombre), au lieu d'être isolées, étaient proches (en lisière de forêt par exemple) et pouvaient librement échanger du pollen, que se passerait-il? *Pour répondre à cette question, aucun calcul n'est nécessaire. Il est utile de montrer que, même si les flux polliniques sont symétriques entre les deux populations, ces flux ne le sont pas du point de vue des allèles et d'en déduire que les allèles peuvent être piégés dans certains milieux. Quel serait l'état d'équilibre des dates de floraison dans les deux milieux? Les deux populations finales sont-elles différenciées? isolées? Ce processus vous semble-t-il de nature « adaptative »?*

Type	Milieu	
	Soleil	Ombre
[1]	1 ^{er} juin	15 juin
[2]	15 juin	30 juin
[1] x [2]	1 ^{er} juin	15 juin
(1990)	3000 : 1 ^{er} juin	3000 : 15 juin
F2 (1991)	1000 : 15 juin	1000 : 30 juin

Dates de floraison en fonction du type de plantes, ou du croisement. En deuxième génération (F2), les deux types de plantes sont en mélange, les chiffres donnent les effectifs de ces deux types sur un total de 4000 dans chaque environnement.

Corrigé

I 1) La date de floraison est déterminée conjointement par le génotype et le milieu. Dans chaque milieu, les résultats montrent que la différence entre les deux types de plantes peut être due à l'action d'un locus possédant deux allèles. L'allèle A , dominant, détermine la forme précoce [1], l'allèle récessif a , détermine la forme tardive [2].

2)

Soleil, 1 ^{er} juin :	1000 AA , 2000 Aa
Soleil, 15 juin :	1000 aa
Ombre, 15 juin :	1000 AA , 2000 Aa
Ombre, 30 juin :	1000 aa

3) Il y a homogamie, le nombre de descendants des différents génotypes n'est pas différent. Les fréquences génotypiques vont donc changer sous l'action de ce régime de reproduction, les fréquences alléliques ne doivent pas changer.

4) a) allèle A : $p_1 = \frac{1}{3} + \frac{1}{2} \times \frac{2}{3} = \frac{2}{3}$

allèle a : $q_1 = \frac{1}{2} \times \frac{2}{3} = \frac{1}{3}$

b) Les graines sont produites en panmixie parmi les plantes se reproduisant en même temps dans chaque population. Les proportions sont donc :

$$AA : p_1^2, \quad Aa : 2p_1q_1, \quad aa : q_1^2$$

c) Au 1^{er} juin, fleurissent les génotypes suivants :

$$AA : p_1^2 / (p_1^2 + 2p_1q_1) = p_1^2 / (1 - q_1^2)$$

$$Aa : 2p_1q_1 / (p_1^2 + 2p_1q_1) = 2p_1q_1 / (1 - q_1^2)$$

Les proportions des allèles seront donc :

$$A : p_2 = (p_1^2 + p_1q_1)/(1 - q_1^2) = p_1/(1 - q_1^2)$$

$$a : q_2 = p_1q_1/(1 - q_1^2)$$

d) Les allèles *a* vont disparaître du groupe de plantes fleurissant le 1^{er} juin. La population sera alors entièrement composée de plantes *AA* fleurissant le 1^{er} juin et de plantes *aa* fleurissant le 15. Les proportions des allèles ne changeant pas, il y aura 2000 plantes de chaque type.

De même qu'on a calculé q_2 , on peut calculer q_n :

$$q_n = p_{n-1} q_{n-1} / (1 - q_{n-1}^2)$$

ce qui se simplifie en remarquant que :

$$(1 - q_{n-1}^2) = (1 - q_{n-1})(1 + q_{n-1}) \text{ et que } (1 - q_{n-1}) = p_{n-1}$$

soit

$$q_n = q_{n-1} / (1 + q_{n-1})$$

ce qui donne $1/q_n = 1 + 1/q_{n-1} = (n-1) + 1/q_1$

On voit que quand n croît, $1/q_n$ augmente, ce qui signifie que q_n tend vers zéro. L'allèle *a* disparaît donc du groupe des plantes fleurissant au 1^{er} juin. Bien sûr, il ne disparaît pas de la population, il se retrouve dans le groupe des plantes fleurissant au 15 juin qui augmente en effectif.

5) Dans la population se trouvant à l'ombre, le même processus va entraîner la séparation des deux génotypes *AA* et *aa* (2 000 plantes chacun à l'équilibre) fleurissant respectivement au 15 et au 30 juin.

II Le résultat des échanges de gènes est schématisé ci-après. Les deux populations échangent des gènes le 15 juin mais seuls les gènes *a* migrent du milieu ensoleillé au milieu ombré tandis que l'allèle *A* ne migre que du milieu ombré au milieu ensoleillé. Sur l'ensemble, il n'y a aucune sélection. Les proportions globales des allèles *A* et *a* ne changent pas à l'échelle de l'ensemble des deux populations. En revanche, leur distribution change. Les allèles *A* sont « piégés » dans le milieu ensoleillé tandis que les allèles *a* sont « piégés », quand ils sont à l'état homozygote (au fur et à mesure que les gènes *A* partent, ceci est de plus en plus fréquent), dans le milieu ombré. Le résultat final est une divergence totale des deux populations, la population au soleil ayant fixé l'allèle *A* et la population à l'ombre ayant fixé l'allèle *a*. Les deux populations sont alors différentes pour des raisons environnementales et génétiques et fleurissent à des périodes radicalement différentes, ce qui crée un isolement total. Cet isolement est une conséquence de la « mécanique » génétique des populations et, quelles qu'en soient les conséquences, est issu d'un processus qui n'a rien d'adaptatif (l'histoire ne dit pas s'il est « bon » de fleurir plus ou moins tard dans chacun des milieux).

	Soleil		Ombre
1 ^{er} juin	AA + Aa ↑ A ↓ aa		a
15 juin		← A a →	AA + Aa a ↓ aa
30 juin			

Dates de floraison et flux de gènes entre les deux populations.

6

Trois allèles

Dans une population infinie et isolée de plantes hermaphrodites annuelles coexistent, en un locus A , trois allèles A_1 , A_2 et A_3 en fréquences respectives p , q et r . On supposera que ces trois allèles sont neutres.

1) Supposons que la population est panmictique.

a) Quelles proportions des différents génotypes peut-on attendre parmi les individus de cette population?

b) Toujours si la population est panmictique, quels génotypes peut-on attendre, et en quelles proportions, dans la descendance d'une plante de génotype A_1A_1 ? et d'une plante de génotype A_2A_3 ?

2) Supposons maintenant que chaque plante de la population produit une fraction t de ses graines en panmixie et une fraction $(1 - t)$ en autogamie.

a) Dans ce cas, quels génotypes peut-on attendre, et en quelles proportions, dans la descendance d'une plante de génotype A_1A_1 ? et d'une plante de génotype A_2A_3 ?

b) Soit H_n la proportion du génotype A_1A_2 à la génération n dans la population. Combien vaut H_n en fonction de p , q , t et H_{n-1} ? Quelle est la limite de H_n quand n croît? (On pourra utilement exprimer le résultat en utilisant le paramètre F .)

c) En déduire les proportions d'équilibre des trois génotypes hétérozygotes, puis celles des trois génotypes homozygotes (on se souviendra que les proportions des allèles A_1 , A_2 et A_3 restent respectivement p , q et r).

3) Ce locus détermine la couleur de la fleur : A_1 donne une couleur violette, A_2 une couleur jaune et A_3 une couleur blanche. Les relations de dominance sont les suivantes : $A_1 > A_2 > A_3$.

a) Quelles couleurs de fleurs peut-on obtenir dans la descendance par autofécondation d'une plante à fleurs violettes? à fleurs jaunes? à fleurs blanches?

b) Quelles sont, en fonction de p , q et r , les proportions des trois phénotypes si la population est panmictique?

c) Quel serait l'effet d'une augmentation de la proportion d'autogamie sur les proportions de ces phénotypes?

N.B. Les questions 2 et 3 sont largement indépendantes.

Corrigé

	A_1A_1	A_1A_2	A_1A_3	A_2A_2	A_2A_3	A_3A_3
1) a)	p^2	$2pq$	$2pr$	q^2	$2qr$	r^2
b)	p	q	r	0	0	0
	0	$p/2$	$p/2$	$q/2$	$(q+r)/2$	$r/2$
2) a)	$1-t+tp$	tq	tr	0	0	0
	0	$tp/2$	$tp/2$	$tq/2 +$ $(1-t)/4$	$(1-tp)/2$	$tr/2 +$ $(1-t)/4$

b) $H_n = t \, 2pq + (1-t) H_{n-1}/2.$

Quand n croît, H_n tend vers H tel que $H = t \, 2pq + (1-t) H/2$ soit $H = 2pq \, 2t/(1+t)$, c'est-à-dire $H = 2pq(1-F)$ avec $F = (1-t)/(1+t)$.

	A_1A_1	A_1A_2	A_1A_3	A_2A_2	A_2A_3	A_3A_3
c)	p^2	$2pq(1-F)$	$2pr(1-F)$	q^2	$2qr(1-F)$	r^2
	$+ Fp(q+r)$			$+ Fq(p+r)$		$+ Fr(p+q)$

En effet, cherchons la proportion de A_1A_1 à l'équilibre; soit X cette valeur, nous savons que $p = X + pq(1-F) + pr(1-F)$, proportion de l'allèle A_1 . On en déduit que $X = p^2 + Fp(q+r)$.

3) a) Violettes : les trois. Jaunes : jaunes ou blanches. Blanches : blanches uniquement.

- b)** Violettes : $p^2 + 2pq + 2pr$
 Jaunes : $q^2 + 2qr$
 Blanches : r^2

c) La proportion du phénotype violet, exprimé par tout individu possédant un allèle A_1 , va diminuer avec la diminution d'hétérozygotie.

Celle du phénotype blanc, double récessif, ne peut qu'augmenter avec l'homozygotie.

La proportion du phénotype jaune évoluera dans une direction ou l'autre en fonction des proportions des différents allèles. En effet, elle augmente quand un allèle A_2 (jaune) n'est plus masqué par un A_1 (violet dominant), ce qui arrive souvent quand p est grand mais elle diminue quand un A_2A_3 (jaune du fait de la dominance de A_2 sur A_3) est détruit, ce qui arrive souvent quand r est grand. Au total, la proportion du phénotype jaune doit augmenter s'il y a beaucoup de violet et diminuer s'il y a beaucoup de blanc.

Démonstration formelle

Le phénotype violet a pour proportion :

$$p^2 + Fp(q + r) + 2pq(1 - F) + 2pr(1 - F) = p^2 + 2pq + 2pr - F(pq + pr)$$

valeur qui décroît quand F croît.

Le phénotype blanc a pour proportion :

$$r^2 + Fr(p + q)$$

valeur qui augmente quand F croît.

Le phénotype jaune a pour proportion :

$$q^2 + Fq(p + r) + 2qr(1 - F) = q^2 + 2qr + Fq(p - r)$$

valeur qui augmente quand F croît si $p > r$ et qui diminue quand F croît si $r > p$.

La cyanogénèse du trèfle

Chez le trèfle blanc (*Trifolium repens*), qui est une espèce allogame, on a repéré l'existence de deux formes de plantes. Certaines produisent du cyanure quand on écrase les feuilles (forme cyanogénique), d'autres non (forme acyanogénique). Le cyanure (HCN) est produit par la dégradation d'un glucoside (Ac) par une enzyme (Li). Les deux produits Ac et Li restent dans des compartiments différents de telle façon que la réaction ne se produit que lorsque la feuille est endommagée.

On a constaté qu'on pouvait différencier quatre types de plantes. Les plantes de type T_4 produisent du cyanure. Les plantes de type T_1 , T_2 et T_3 ne produisent pas de cyanure mais lorsqu'on écrase ensemble des feuilles T_2 et T_3 , on obtient un dégagement de cyanure (alors que les mélanges $T_1 + T_2$ et $T_1 + T_3$ ne donnent pas de cyanure).

Les croisements (F1) effectués entre des plantes prélevées dans la nature donnent les résultats suivants.

$$T_1 \times T_1 \rightarrow T_1$$

$$T_1 \times T_2 \rightarrow T_2 \text{ ou } \frac{1}{2}T_1 + \frac{1}{2}T_2$$

$$T_1 \times T_3 \rightarrow T_3 \text{ ou } \frac{1}{2}T_1 + \frac{1}{2}T_3,$$

$$T_2 \times T_2 \rightarrow T_2 \text{ ou } \frac{1}{4}T_1 + \frac{3}{4}T_2$$

$$T3 \times T3 \rightarrow T3 \text{ ou } \frac{1}{4}T1 + \frac{3}{4}T3,$$

$$T4 \times T4 \rightarrow T4 \text{ ou } T4 + T3 \text{ ou } T4 + T2 \text{ ou } T4 + T3 + T2 + T1$$

$$T2 \times T4 \rightarrow T4 \text{ ou } T4 + T3 \text{ ou } T4 + T2 \text{ ou } T4 + T3 + T2 + T1$$

$$T3 \times T4 \rightarrow T4 \text{ ou } T4 + T3 \text{ ou } T4 + T2 \text{ ou } T4 + T3 + T2 + T1$$

$$T1 \times T4 \rightarrow T4 \text{ ou } T4 + T3 \text{ ou } T4 + T2 \text{ ou } T4 + T3 + T2 + T1$$

$$T2 \times T3 \rightarrow T4 \text{ ou } T4 + T3 \text{ ou } T4 + T2 \text{ ou } T4 + T3 + T2 + T1$$

I Montrer que ces résultats peuvent s'expliquer en faisant intervenir deux loci comportant chacun deux allèles. On appellera ces couples d'allèles respectivement A/a et L/l et on supposera que ces deux loci se comportent indépendamment l'un de l'autre. En déduire les proportions attendues dans les deux derniers croisements ci-dessus.

Comment les résultats auraient-ils été modifiés si le trèfle avait été une espèce strictement autogame dans la nature (comme les pois de Mendel)?

II Soient, pour une population donnée, supposée panmictique pour ce caractère, p_1 et $q_1 (=1 - p_1)$ les proportions respectives de A et a et p_2 et $q_2 (=1 - p_2)$ les proportions respectives de L et l .

1) Admettons qu'il n'y a pas de déséquilibre de liaison entre ces deux loci.

a) Quels types de plantes, et en quelles proportions, attendriez-vous dans des graines prélevées dans cette population sur des plantes mères de type $T1$?

b) Même question si, avant de prélever les graines, on autoféconde une fraction s des fleurs de la plante concernée.

c) Quelles sont, en fonction de p_1 , q_1 , p_2 et q_2 , les proportions respectives des formes $T1$, $T2$, $T3$ et $T4$ dans la population.

d) Dans une population, on a trouvé les effectifs suivants :

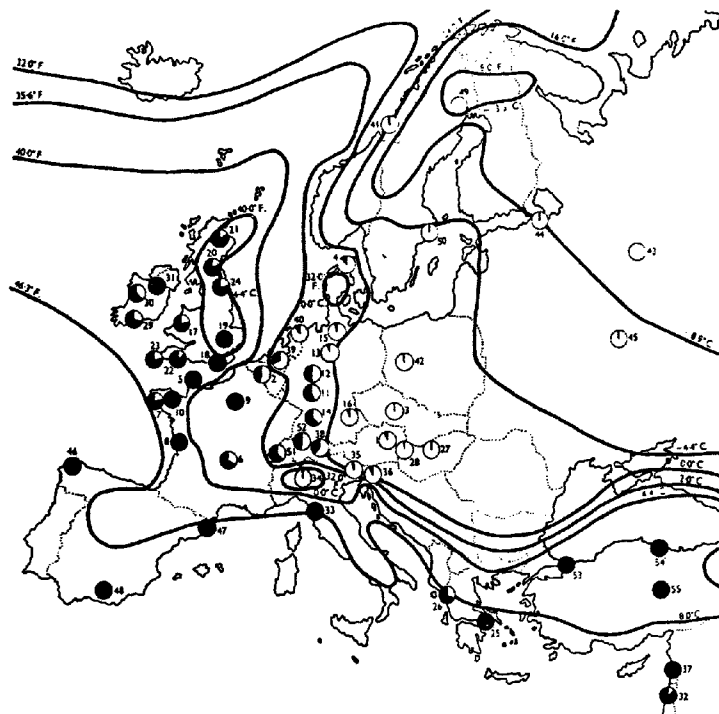
$$T1 : 44 \quad T2 : 46 \quad T3 : 116 \quad T4 : 154$$

Pouvez-vous, en supposant la panmixie, estimer les valeurs de p_1 et p_2 ? L'hypothèse d'indépendance statistique (absence de déséquilibre de liaison) des deux loci vous paraît-elle acceptable (un χ^2 à 1 degré de liberté a 95 % de chances d'être compris entre 0 et 3,84)?

2) Admettons maintenant qu'il y a un déséquilibre de liaison de valeur D . Quelle serait, en fonction de p_1 , q_1 , p_2 , q_2 et D , la proportion des formes $T1$, $T2$ et $T3$ dans la population?

III Le cyanure est un poison au goût assez fort (odeur d'amande amère). Pouvez-vous proposer une hypothèse concernant les raisons évolutives conduisant les plantes à produire les produits Ac et Li ? Une étude à l'échelle de l'Europe (publiée en 1941) a montré la répartition suivante des plantes cyanogéniques et acyanogéniques. L'auteur a superposé les courbes des

isothermes de janvier. Pouvez-vous, en utilisant entre autres vos connaissances sur les facteurs qui peuvent expliquer l'existence d'un cline, proposer une (ou des) hypothèse(s) concernant la distribution de ce polymorphisme?



Répartition des formes cyanogéniques (en noir) et acyanogéniques (en blanc). Les isothermes de janvier ont été ajoutées sur la carte. (D'après DADAY, H., 1954. Gene frequencies in wild populations of *Trifolium repens*. I. Distribution by latitude. *Heredity*, **8**, p. 65-66.)

Corrigé

I Les croisements incluant $T1$ et $T2$ peuvent s'expliquer avec un locus à deux allèles A et a avec A , dominant sur a , déterminant le type $T2$. De même pour ceux incluant $T1$ et $T3$, on peut supposer un locus avec deux allèles, L et l où L , dominant, détermine $T3$. Le fait qu'en écrasant ensemble $T2$ et $T3$, on obtient du cyanure laisse penser que l'un (disons $T2$) produit le glucoside (et pas l'enzyme) et l'autre ($T3$) produit l'enzyme (et pas le glucoside). L'apparition de $T4$ dans la descendance de $T2 \times T3$ est un phénomène de complémentation. $T1$ correspond au génotype $aa ll$, $T2$ au génotype $A* ll$, $T3$ au génotype $aa L*$ et $T4$ au génotype $A* L*$ (* est mis quand l'allèle présent est indéterminé).

$T1 \times T4 \rightarrow T4$ s'obtient si le parent $T4$ a pour génotype $AA LL$

$T1 \times T4 \rightarrow \frac{1}{2}T4 + \frac{1}{2}T3$ " $Aa LL$

$T1 \times T4 \rightarrow \frac{1}{2}T4 + \frac{1}{2}T2$ " $AA Ll$

$T1 \times T4 \rightarrow \frac{1}{4}T4 + \frac{1}{4}T3 + \frac{1}{4}T2 + \frac{1}{4}T1$ " $Aa Ll$

$T2 \times T3 \rightarrow T4$ parent $T2$ $AA ll$ et parent $T3$ $aa LL$

$T2 \times T3 \rightarrow \frac{1}{2}T4 + \frac{1}{2}T3$ " $Aa ll$ " $aa LL$

$T2 \times T3 \rightarrow \frac{1}{2}T4 + \frac{1}{2}T2$ " $AA ll$ " $aa Ll$

$T2 \times T3 \rightarrow \frac{1}{4}T4 + \frac{1}{4}T3 + \frac{1}{4}T2 + \frac{1}{4}T1$ " $Aa ll$ " $aa Ll$

Si le trèfle était autogame, il n'y aurait pas d'hétérozygotes. Les croisements auraient donné des résultats toujours homogènes :

$$\begin{aligned} T1 \times T1 &\rightarrow T1, T1 \times T2 \rightarrow T2, T1 \times T3 \rightarrow T3, T2 \times T2 \rightarrow T2, \\ T3 \times T3 &\rightarrow T3, T4 \times T4 \rightarrow T4, T2 \times T4 \rightarrow T4, T3 \times T4 \rightarrow T4, \\ T1 \times T4 &\rightarrow T4, T2 \times T3 \rightarrow T4 \end{aligned}$$

II Soient, pour une population donnée, supposée panmictique pour ce caractère, p_1 et $q_1 (=1 - p_1)$ les proportions respectives de A et a et p_2 et $q_2 (=1 - p_2)$ les proportions respectives de L et l .

1) a) Dans des graines prélevées sur des individus de type $T1$, on devrait trouver :

- des descendants aa ll ($T1$) en proportion $(1 - p_1)(1 - p_2)$,
- des plantes Aa ll ($T2$) en proportion $p_1(1 - p_2)$,
- des plantes aa Ll ($T3$) en proportion $(1 - p_1)p_2$,
- et des plantes Aa Ll ($T4$) en proportion p_1p_2 .

b) L'autofécondation produit uniquement des génotypes aa ll . De ce fait, la plante produira des graines se répartissant de la façon suivante :

- aa ll ($T1$) en proportion $s + (1 - s)(1 - p_1)(1 - p_2)$,
- Aa ll ($T2$) en proportion $(1 - s)p_1(1 - p_2)$,
- aa Ll ($T3$) en proportion $(1 - s)(1 - p_1)p_2$,
- Aa Ll ($T4$) en proportion $(1 - s)p_1p_2$.

c) Les proportions respectives des formes $T1$, $T2$, $T3$ et $T4$ dans la population sont données par le tableau suivant.

	$LL + Ll \ p_2^2 + 2 \ p_2q_2$	q_2^2
$AA + A \ p_1^2 + 2 \ p_1q_1$	$T4 \ \begin{pmatrix} p_1^2 + 2p_1q_1 \\ p_2^2 + 2p_2q_2 \end{pmatrix}$	$T2 \ (p_1^2 + 2p_1q_1)q_2^2$
$aa \ q_1^2$	$T3 \ q_1^2(p_2^2 + 2p_2q_2)$	$T1 \ q_1^2q_2^2$

d) Dans une population, on a trouvé les effectifs suivants :

$$T1 : 44 \quad T2 : 46 \quad T3 : 116 \quad T4 : 154$$

$$q_1^2 = (116 + 44)/360 = 4/9 \Rightarrow p_1 = 1/3, q_1 = 2/3$$

$$q_2^2 = (44 + 46)/360 = 1/4 \Rightarrow p_2 = q_2 = 1/2$$

Si les deux loci sont indépendants, les proportions espérées dans chaque case du tableau sont données par le produit de la probabilité de se trouver dans la ligne considérée par la probabilité de se trouver dans la colonne considérée. L'effectif espéré s'obtient en multipliant la probabilité par l'effectif total.

$$T1 : 40 \quad T2 : 50 \quad T3 : 120 \quad T4 : 150$$

d'où un χ^2 de 4^2 ($1/40 + 1/50 + 1/120 + 1/150$) = 0,96 ce qui est loin en dessous de la valeur 3,84. L'hypothèse d'indépendance statistique (absence de déséquilibre de liaison) peut donc être conservée.

2) Les proportions des génotypes sont alors données par le tableau suivant.

	$AL p_1 p_2 + D$	$Al p_1 q_2 - D$	$aL q_1 p_2 - D$	$al q_1 q_2 + D$
$AL p_1 p_2 + D$	$T4$	$T4$	$T4$	$T4$
$Al p_1 q_2 - D$	$T4$	$T2$	$T4$	$T2$
$aL q_1 p_2 - D$	$T4$	$T4$	$T3$	$T3$
$al q_1 q_2 + D$	$T4$	$T2$	$T3$	$T1$

On voit que les proportions des différents types sont :

$$T1 : (q_1 q_2 + D)^2$$

$$T2 : (p_1 q_2 - D)^2 + 2 (p_1 q_2 - D) (q_1 q_2 + D)$$

$$T3 : (q_1 p_2 - D)^2 + 2 (q_1 p_2 - D) (q_1 q_2 + D)$$

III Le cyanure est un poison au goût assez fort qui n'est pas produit tant que la feuille n'est pas endommagée. Il est donc probable que sa production serve à diminuer la prédation par les herbivores. On sait, en particulier, que les gastéropodes (escargots, limaces) peuvent être repoussés par certains composés produits par les plantes. Cependant, le cyanure pourrait, en fait, empoisonner la plante elle-même, ce qui l'oblige à garder le glucoside et l'enzyme dans des compartiments différents.

La répartition observée à l'échelle de l'Europe montre un cline très net. Ce cline est difficile à expliquer. En effet, il semble que le froid hivernal joue un rôle dans la détermination des proportions observées. Ce rôle peut être de rendre les herbivores moins nombreux, ou moins efficaces ou, dans un autre ordre d'idées, de tuer plus facilement les plantes cyanogéniques (on pourrait penser qu'elles s'empoisonnent du fait d'une moins bonne étanchéité des compartiments). Cependant, ce type de mécanisme devrait *a priori* aboutir à de larges zones homogènes séparées par un cline beaucoup plus restreint dû à la migration. Il semble, en effet, inenvisageable que le trèfle migre à chaque génération sur des distances de l'ordre du millier de kilomètres. Il faut alors imaginer qu'en chaque point, la sélection détermine une proportion d'équilibre fonction de la température, sans que la migration joue un rôle important dans ce cline.

Corrigés des exercices

Chapitre A

Exercice A.1.1

- 1) $f(AA) = p$, $f(aa) = q$
- 2) $f(Aa) = 2q$, $f(AA) = 1 - 2q$
- 3) $f(Aa) = 2p$, $f(aa) = 1 - 2p$

Exercice A.1.2

$$f(A_1) = 0,88, f(A_2) = 0,12, f(B_1) = 0,84, f(B_2) = 0,02, f(B_3) = 0,14$$

Exercice A.1.3

1) Un seul allèle est présent chez les mâles, leur génotype est X_oY (phénotype roux) ou X_oY (« non roux »). Les femelles sont de génotypes X_oX_o (rousses), X_oX_o (« écaille de tortue »), ou X_oX_o (« non rousses »).

$$2) f_f(O) = \frac{2F_o + F_e}{2F}, \quad f_m(O) = \frac{M_o}{M}, \quad f_t(O) = \frac{2F_o + F_e + M_o}{2F + M}$$

Exercice A.2.1

1) Un locus biallélique contrôle la couleur, R , allèle dominant (rouge), b , allèle récessif (blanc).

- a) $1/2$ de plantes à fleurs blanches, $1/2$ à fleurs rouges.
- b) Si la plante à fleurs rouges est homozygote, 100 % de plantes à fleurs rouges, si elle est hétérozygote, $3/4$ de plantes à fleurs rouges, $1/4$ de plantes à fleurs blanches.

$$2) f(b) = 0,4, f(R) = 0,6. 192 \text{ hétérozygotes attendus.}$$

3) Descendants d'une plante à fleurs blanches : 40 % d'homozygotes (phénotype fleurs blanches), 60 % d'hétérozygotes (rouges). Descendants d'une plante à fleurs rouges : 50 % d'hétérozygotes, 20 % d'homozygotes récessifs, 30 % d'homozygotes dominants, soit 80 % de plantes à fleurs rouges et 20 % de plantes à fleurs blanches.

Exercice A.2.2

Daphnies : $f(A_1) = 0,437$ et $f(A_2) = 0,563$, $\chi^2 = 0,19$, avec un degré de liberté. Le test est non significatif, on ne rejette pas l'hypothèse de panmixie.

Coléoptère : $f(B_1) = 0,741$ et $f(B_2) = 0,259$, $\chi^2 = 41,85$, avec un degré de liberté. Le test est très significatif, on rejette l'hypothèse de panmixie (on suppose qu'il s'agit d'un effet Wahlund).

Exercice A.2.3

En faisant l'hypothèse que la population se reproduit en panmixie, on a $q = \sqrt{\frac{36}{100}} = 0,6$ soit $p = 0,4$. On ne peut pas tester cette hypothèse, car elle a servi pour calculer les fréquences alléliques et on a ainsi utilisé toute l'information disponible dans l'échantillon. Si on cherchait à calculer les effectifs théoriques, on retomberait exactement sur les effectifs observés : il ne reste aucun degré de liberté.

Exercice A.2.4

$$1) f(T^a T^a) = p^2, f(TT) = q^2, f(t^b t^b) = r^2, f(T^a T) =$$

$$(= 2)pq, f(T^a t^b) = 2pr, f(Tt^b) = 2qr$$

$$2) f(\text{abyssin}) = f(T^a T^a) + f(T^a T) + f(T^a t^b) = p^2 + 2pq + 2pr$$

$$= p(p + 2q + 2r) = p(1 + q + r)$$

$$f(\text{tigre}) = f(TT) + f(Tt^b) = q^2 + 2qr = q(q + 2r)$$

$$f(\text{marbré}) = f(t^b t^b) = r^2$$

$$3) r = 0,23, q = 0,55, p = 0,22$$

Exercice A.2.5

1) Il faut faire l'hypothèse que la population est panmictique et qu'il n'y a pas de différence de viabilité des trois génotypes entre le stade zygote et la naissance. L'existence de mutation et de sélection après la naissance (gène subléta) n'invalide pas le calcul. Les personnes touchées par la maladie sont en fréquence $q^2 = 0,0004$. La fréquence de l'allèle muté est donc de 2 % environ. La fréquence attendue de personnes porteuses est de $2pq \approx 0,04 = 4 \%$ soit une personne sur 25.

$$2) \text{ Proportions d'allèles mutés chez les porteurs : } \frac{pq}{pq + q^2} = p = 98 \%,$$

$$\text{chez les personnes atteintes : } \frac{q^2}{pq + q^2} = q = 2 \%.$$

Exercice A.2.6

$$1) G_0 : p_f = 1, p_m = 0, p = 2/3$$

$$G_1 : p_f = 1/2, p_m = 1, p = 2/3$$

$$G_2 : p_f = 3/4, p_m = 1/2, p = 2/3$$

$$2) p_m(n) = p_f(n-1), p_f(n) = 1/2[p_f(n-1) + p_m(n-1)]$$

$$p(n) = 1/3[2p_f(n) + p_m(n)]$$

$$= 1/3p_f(n-1) + p_m(n-1) + p_f(n-1) = p(n-1)$$

$$E_n = -1/2[p_m(n-1) - p_f(n-1)] = (-1/2)E_{n-1} = (-1/2)^n E_0$$

$$3) \text{ Équilibre : } p_f = p_m = 2/3, p = 2/3$$

Exercice A.2.7

1) La fréquence de l'allèle est égale à la fréquence des garçons daltoniens soit $q = 8,01 \%$. En supposant que la population est panmictique, on attend une fréquence égale à q^2 de filles daltoniennes soit $0,642 \%$.

2) Sur 9072 filles on attend donc 58 filles daltoniennes. Elles sont moins nombreuses que prévu. Il semble en fait que plusieurs gènes interviennent dans l'anomalie.

Exercice A.3.1

$$2) N > 384$$

Exercice A.3.2

$$1) D_n = (D_0 + 1/2H_0)^2 \quad H_n = 2(D_0 + 1/2H_0)(R_0 + 1/2H_0)$$

$$R_n = (R_0 + 1/2H_0)^2$$

$$2) D_n = D_{n-1} + 1/4H_{n-1} \quad H_n = 1/2H_{n-1} \quad R_n = R_{n-1} + 1/4H_{n-1}$$

$$\text{À l'équilibre} \quad D_e = D_0 + 1/2H_0 = p, \quad H_e = 0, \quad R_e = q$$

$$3) H_n = 2tpq + \frac{1-t}{2}H_{n-1}, \text{ équilibre pour : } H_e = \frac{4tpq}{1+t} \quad F_e = \frac{1-t}{1+t}$$

Exercice A.3.3

$$s = 2y - 4x$$

$$q = \frac{4x - y}{1 - 2y + 4x}$$

Exercice A.3.4

1) Allogamie obligatoire, 100 % d'hétérozygotes; ce locus induit l'allogamie, ce qui favorise la panmixie pour les autres loci.

2) Trois allèles au minimum, trois génotypes 1, 2, 3.

$$f(1) = 1/2f(2) + 1/2f(3) = 1/2[1-f(1)], f_e = 1/3$$

3) Grand nombre d'allèles.

Exercice A.3.5

1) Un locus : allèle A, dominant, forme précoce (1), allèle a, récessif, forme tardive (2).

2) Croisement F2 : 1/4 AA, 1/2 Aa, 1/4 aa soit 3/4 de précoces et 1/4 de tardifs dans chaque milieu.

3) Absence de pressions évolutives : fréquences alléliques constantes.

Régime de reproduction (ici, homogamie) : modification des fréquences génotypiques.

$$4) p_1 = 1 - q_1 = 2/3$$

$$-f(AA) = p_1^2, f(Aa) = 2p_1q_1, f(aa) = q_1^2$$

$$-p_2 = \frac{p_1^2 + p_1q_1}{p_1^2 + 2p_1q_1} = \frac{1}{1 + q_1} \Rightarrow q_2 = \frac{q_1}{1 + q_1}$$

soit $\frac{1}{q_n} = \frac{1}{q_{n-1}} + 1 = n - 1 + \frac{1}{q_1}$

Si n est grand, q tend vers 0 : il ne restera que des génotypes précoces AA parmi les plantes fleurissant le 1^{er} juin et uniquement des aa parmi celles du 15 juin.

5) Même phénomène à l'ombre.

6) Allèles A « piégés » dans le milieu ensoleillé, allèles a « piégés » dans le milieu à l'ombre : à terme génotypes AA au soleil et aa à l'ombre exclusivement.

Exercice A.3.6

$$\Phi_{AB} = \frac{1}{8} \quad f_C = \frac{1}{4} \quad f_D = \frac{1}{16} \quad f_E = \frac{1}{8}$$

Exercice A.3.7

$$f_K = \Phi_{JJ} = \frac{1}{2} + \frac{1}{2}f_J = \frac{1}{2} + \frac{1}{2}\left(\frac{1}{2} + \frac{1}{2}f_I\right) = \frac{3}{4} \text{ car } f_I = 0$$

Après n générations :

$$f_n = \frac{1}{2} + \frac{1}{2}f_{n-1} \Rightarrow 1 - f_n = \frac{1}{2}(1 - f_{n-1}) = \left(\frac{1}{2}\right)^n (1 - f_0)$$

$$\Rightarrow f_n = 1 - \left(\frac{1}{2}\right)^n \text{ si } f_0 = 0$$

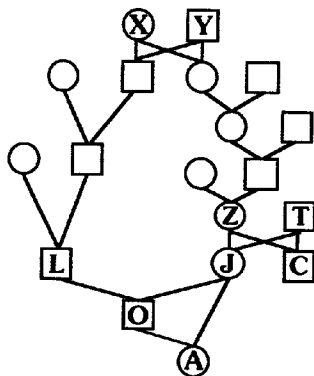
Exercice A.3.8

$$f_n = \frac{1}{4}(1 + f_{n-2}) + \frac{1}{2}f_{n-1}$$

Exercice A.3.9

Entre Œdipe et Jocaste, il y a trois chaînes de parenté : par X (neuf maillons), par Y (neuf maillons) et $O-J$ (un maillon). Le coefficient de consanguinité d'Antigone vaut donc 0,252.

Entre Antigone et Créon, il y a six chaînes de parentés. Une par X (dix maillons), une par Y (dix maillons), deux par Z ($A-O-J-Z-C$, quatre maillons, $A-J-Z-C$, trois maillons), et deux par T (trois et quatre maillons). Leur coefficient de parenté vaut donc 0,188.

**Exercice A.4.1**

Si X est le nombre d'allèles A , $P[X = k] = C_N^k p^k q^{N-k}$ avec $N = 4$ et $p = q = 1/2$, pour $k = 0, 1, 2, 3$ ou 4 , $P = 6,25\%$, 25% , $37,5\%$, 25% ou $6,25\%$. Le polymorphisme est conservé si $k = 1, 2$ ou 3 , soit une probabilité de $87,5\%$.

Exercice A.4.2

1) 206 heures, soit 8,6 jours.

2) 6570 ans...

3) a) La dérive va éliminer l'un ou l'autre.

b) La dérive peut éliminer A ou a mais, si a est éliminé, la mutation va le réintroduire. Seul A peut être perdu; on va donc vers la fixation de a .

Exercice A.4.3

1) Un allèle est soit perdu soit fixé.

$$2) \frac{1}{N}, \left(1 - \frac{1}{N}\right) \frac{1}{N}, \left(1 - \frac{1}{N}\right)^2 \cdot \frac{1}{N}.$$

$$\mathcal{P}(X = g) = 1/N(1 - 1/N)^{n-1}, E(\mathcal{P}(X)) = N$$

3) PO'11

```

DEFINT A-Z : REM Initialisation-----
DEFINT A-Z
DIM r(1000), d(1000), a(1000), ad(1000)
RANDOMIZE TIMER
SCREEN 9
gt = 40 : REM Nombre de générations
n = 20 : REM Nombre d'individus
lar = 580 : REM Largeur de l'écran
hau = 200 : REM Hauteur de l'écran
inter = lar / gt : REM Intervalle entre deux générations
ind = hau / n : REM Intervalle entre deux individus
FOR j = 1 TO n : a(j) = j : NEXT j : REM Numérotation des ancêtres

FOR i = 1 TO gt : REM Les générations-----

FOR j = 1 TO n : REM Tirage du nombre de descendants de chaque repli-
cateur
r = INT(n * RND + 1) : REM Le replicateur r a un descendant;
d(r) = d(r) + 1 : REM le nombre de descendants de r augmente donc d'1
NEXT j

FOR r = 1 TO n : REM Pour chaque replicateur :
IF d(r) = 0 THEN 10 : REM s'il n'a pas de descendant, on l'oublie;
FOR k = 1 TO d(r) : REM sinon, on prend chaque descendant,
de = de + 1 : REM, on le numérote,
ad(de) = a(r) : REM et on lui attribue le même ancêtre qu'à son parent.
c = 1 + a(r) - INT(15 * INT(a(r) / 15)): REM On attribue une couleur à chacun
LINE (i * inter, r * ind)-(i + 1) * inter, de * ind), c : REM On trace l'arbre
NEXT k
10 NEXT r

de = 0 : REM et on prépare la génération suivante.
FOR r = 1 TO n : d(r) = 0 : a(r) = ad(r) : NEXT r
NEXT i : REM fin-----
END

```

Exercice A.4.4

L'effectif génétique est de sept individus.

Exercice A.4.5

$$p' = \frac{w_A p}{w_A p + w_a q} \Rightarrow \Delta p = p q \frac{w_A - w_a}{w_A p + w_a q}$$

Il n'y a pas d'équilibre polymorphe possible.

Exercice A.4.6

$$1) p' = \frac{3/4p + q}{3/4p + 2q} \Rightarrow p' - p = \frac{q(1 - 5/4p)}{3/4p + 2q}$$

$$2) \text{ Oui : } p_e = 4/5$$

Exercice A.4.7

1) Un locus « couleur des pattes » : allèle blanc (B) dominant, allèle jaune récessif (j), un locus « couleur du plumage » : allèle blanc (Bl) dominant, allèle roux (r) récessif. Souche A : au locus pattes génotype jj , au locus plumage génotype Bl^* (* = allèle quelconque), avec l'hypothèse de panmixie : $f(Bl) = 1 - f(r) = 1 - \sqrt{1} \% = 0,9$. Souche B : locus pattes, génotype B^* , $f(B) = 1 - f(i) = 1 - \sqrt{3} \% = 0,827$, locus plumage, génotype rr .

$$2) s = 74,4 \%$$

$$3) s' = 77,5 \%, \text{ soit } 3 \% \text{ de gain.}$$

Exercice A.4.8

$$1) q_0 = 0,15, 30 \% \text{ d'hétérozygotes, } 70 \% \text{ d'homozygotes.}$$

$$2) \frac{1}{q_n} = \frac{1}{q_{n-1}} + 1 = \frac{1}{q_0} + n$$

avec $q_0 = 0,15$ et $q_n = 0,06$, $n = 10$ (≈ 250 ans).

Exercice A.4.9

$$1) P_i(n) = (1 - m_i)P_i(n-1) + m_i P_j(n-1)$$

$$P_j(n) = (1 - m_j)P_j(n-1) + m_j P_i(n-1)$$

$$2) \text{ Équilibre quand } \Delta P_i = 0 \text{ (ou } \Delta P_j = 0) \text{ soit } P_i(n) = P_j(n).$$

$$3) E_n = (1 - m_i - m_j)^n E_0, \text{ il faudra environ } 23 \text{ générations.}$$

Exercice A.4.10

1) $f(A_1) = 0,6$, $f(A_2) = 0,4$, $\chi^2 = 5,55$ à comparer dans les tables à un χ^2 à un degré de liberté soit 3,84 (à 5 %) ou 6,64 (à 1 %). L'hypothèse de la conformité à une structure de Hardy-Weinberg est rejetée.

2) Lieu 1 : $f(A_1) = 0,785$, $f(A_2) = 0,215$, $\chi^2 = 0,02$

Lieu 2 : $f(A_1) = 0,415$, $f(A_2) = 0,585$, $\chi^2 = 0,1$.

Hypothèses non rejetées. Le déficit correspond à un effet Wahlund.

Exercice A.5.1

Trois paramètres indépendants permettent de décrire tous les déséquilibres génétiques :

$$D_1 = g_{A_1B_1} - p_{A_1}p_{B_1} = 0,1044$$

$$D_2 = g_{A_2B_1} - p_{A_2}p_{B_1} = -0,1596$$

$$D_3 = g_{A_3B_1} - p_{A_3}p_{B_1} = 0,0552$$

Exercice A.5.2

génotypes : $\begin{array}{cc} A & B \\ \hline A & B \end{array} \begin{array}{cc} a & b \\ \hline a & b \end{array}$ soit :

Génération G_0 :

gamètes : $\underline{A} \underline{B} \underline{a} \underline{b}$

fréquence : $p \ q$

	$A(p)$	$a(q)$
$B(p)$	p	0
$b(q)$	0	q

On a donc $D_0 = pq$.

Génération G_1 :

génotypes : $\begin{array}{cc} AB & AB \\ \hline AB & ab \end{array} \begin{array}{cc} ab \\ \hline ab \end{array}$

fréquences : $p^2 \ 2pq \ q^2$ soit :

gamètes : $\underline{A} \underline{B} \underline{A} \underline{b} \underline{a} \underline{B} \underline{a} \underline{b}$

fréquences : $p \cdot rpq \ rpq \ rpq \ q \cdot rpq$

	$A(p)$	$a(q)$
$B(p)$	$p \cdot rpq$	rpq
$b(q)$	rpq	$q \cdot rpq$

On a donc $D_1 = (1 - r)pq$.

Génération G_n issue de l'union au hasard des gamètes :

$\underline{A} \underline{B} \quad \underline{A} \underline{b} \quad \underline{a} \underline{B} \quad \underline{a} \underline{b}$

en fréquences : $p^2 + D_{n-1}$ $pq - D_{n-1}$ $pq - D_{n-1}$ $q^2 + D_{n-1}$

Soit :

	$A(p)$	$a(q)$
$B(p)$	$p^2 + D_{n-1}$	$pq - D_{n-1}$
$B(q)$	$pq - D_{n-1}$	$p^2 + D_{n-1}$

Où : $D_{n-1} = (1-r)^{n-1} D_0 = (1-r)^{n-1} pq$.

Application numérique : $p = q = 1/2$, $r = 1/2$.

Déséquilibre génétique	Génération	$f\left(\frac{a\bar{b}}{a\bar{b}}\right) = (q^2 + D)^2$	$f\left(\frac{aB}{a\bar{B}}\right) = (pq - D)^2$
$D_0 = 1/4$	$G_1 :$	0,25	0
$D_1 = 1/8$	$G_2 :$	0,141	0,016
$D_2 = 1/16$	$G_3 :$	0,098	0,035
$D_3 = 1/32$	$G_4 :$	0,079	0,048
$D_4 = 1/64$	$G_5 :$	0,071	0,055
$D_\infty = 0$	$G_\infty :$	0,0625	0,0625

Chapitre B

Exercice B.1

1) Si les parents possèdent au même gène le même allèle inactif, les descendants seront galactosémiques. Sinon ils seront normaux.

2) $q = 10^{-3}$.

3) Effet de la dérive, donc fixation de l'allèle de galactosémie.

Exercice B.2

1) Allèle récessif : $q = 2 \times 10^{-3}$,

allèle dominant $2pq + q^2 (\approx 2q) = 4 \times 10^{-6}$, $q = 2 \times 10^{-6}$

2) Allèle dominant : $u = 10^{-6}$, récessif : $u = 4 \times 10^{-6}$

3) Risque = $f q + (1-f) q^2$, $f = 1/16$, $q = 2 \times 10^{-3}$,

risque = $1,3 \times 10^{-4}$

$$P(\text{enfant atteint/parents cousins}) = P(\text{parents cousins/enfant atteint}) \times \\ P(\text{enfant atteint}) / P(\text{parents cousins}) = 0,15 \times 4 \times 10^{-6} / 5 \times 10^{-3} = 1,2 \times 10^{-4}$$

d'où $q = 1,9 \times 10^{-3}$.

Exercice B.3

1) $q = \sqrt{u}$, fréquences hétérozygotes $\approx 2pq = 2\sqrt{u(1-\sqrt{u})} \approx 2\sqrt{u}$

2) Nombre moyen d'allèles létaux par individu = $2n\sqrt{u}$

3) Fardeau pour un locus $L = u$

Fardeau pour N locus $L = 1 - (1 - u)^N \approx 1 - e^{-Nu}$

4) 63 allèles létaux : la population humaine a fonctionné en petits effectifs (d'où consanguinité), hypothèse de population infinie non respectée.

Chapitre C

Exercice C.1

Si on appelle q la proportion de l'allèle létal, on sait que cette proportion dans la population dépend du régime de reproduction.

- Chez une espèce autogame, q va être de l'ordre de $2u$, la proportion d'hétérozygotes est alors de $4u$ et, en autofécondation, on attend $\frac{1}{4}4u = u$ plantes présentant le génotype homozygote létal. Chez le millet, la proportion des plantules albinos est de 1/10 000 ou moins, ce qui est compatible avec la valeur du taux de mutation produisant l'allèle létal.

- Chez une espèce allogame, la proportion attendue de l'allèle létal est de $q = \sqrt{u}$. La proportion d'hétérozygotes est alors de l'ordre de $2\sqrt{u}$ et, en autofécondation, on attend une proportion de $\frac{1}{4}2\sqrt{u} = \frac{1}{2}\sqrt{u}$ du génotype homozygote létal. Cette situation ressemble à celle du mil où on a trouvé 12/10 000 plantules albinos.

- Chez une espèce reproduite asexuellement (par bouturage par exemple), on s'attend que l'allèle létal puisse venir se fixer à l'état hétérozygote dans l'ensemble d'un clone. Si le clone est entièrement hétérozygote, son autofécondation fournira en espérance $\frac{1}{4}$ d'homozygotes létaux. C'est manifestement ce qui s'est passé dans la variété de vigne considérée où on a trouvé 2522/10 000 plantules albinos.

On voit que le millet étant autogame, on pourra commercialiser des lignées pures de cette espèce. Le mil est allogame, l'hétérosis interdit donc de commercialiser des variétés consanguines, seules des populations ou des hybrides entre lignées pourront être exploités par les agriculteurs. Quant à la vigne, ce sont bien entendu des clones qui constituent les cépages qui sont cultivés.

Exercice C.2

On suppose la population infinie afin de ne pas prendre en compte les effets d'échantillonnage se produisant à chaque génération et aboutissant à une dérive génétique. L'application numérique correspond à $p = (10 + 20)/90 = \frac{1}{3}$ d'où $q = \frac{2}{3}$.

1) Si les deux allèles n'ont pas d'effet sur la valeur sélective, leurs proportions restent constantes au cours des générations et les proportions des génotypes produits en panmixie sont :

$$p^2 A_1 A_1, \quad 2pq A_1 A_2 \quad \text{et} \quad q^2 A_2 A_2$$

L'application numérique donne $\frac{1}{9} A_1 A_1$, $\frac{4}{9} A_1 A_2$ et $\frac{4}{9} A_2 A_2$, ce qui correspond aux proportions initiales dans ce cas.

2) Les valeurs sélectives sont telles que l'hétérozygote est deux fois meilleur que les homozygotes qui sont équivalents entre eux. Il s'agit d'un cas de superdominance. L'équilibre stable des proportions est obtenu pour $p = q = \frac{1}{2}$.

À la génération $g + 1$, on aura :

$$p' = (p^2 + 2pq)/\bar{W} \quad \text{avec} \quad \bar{W} = p^2 + 4pq + q^2 = 1 + 2pq.$$

Les proportions des génotypes seront $p'^2 A_1 A_1$, $2p'q' A_1 A_2$ et $q'^2 A_2 A_2$.

L'application numérique donne $p' = \frac{5}{13}$ (ce qui est supérieur à $\frac{1}{3} = \frac{5}{15}$, ce qui ne nous étonne pas puisque l'équilibre est à $\frac{1}{2}$) et $q' = \frac{8}{13}$. Les proportions des génotypes dans la population sont alors :

$$\frac{25}{169} A_1 A_1, \quad \frac{80}{169} A_1 A_2 \quad \text{et} \quad \frac{64}{169} A_2 A_2$$

Au bout d'un temps infini, les proportions des allèles seront $p = q = \frac{1}{2}$ et celles des génotypes seront $\frac{1}{4} A_1 A_1$, $\frac{1}{2} A_1 A_2$ et $\frac{1}{4} A_2 A_2$.

3) Si la sélection portait sur la viabilité, globalement, les résultats seraient les mêmes pour les zygotes (avec une génération de décalage). Cependant, les adultes auraient subi la sélection et leurs proportions intégreraient ce

phénomène. Les zygotes de la génération suivante seraient constitués par panmixie à partir des gamètes de la génération g . Les différences de viabilité agiraient ensuite. Les proportions des génotypes dans la population d'individus adultes de la génération $g + 1$ seraient :

$$\frac{p'}{W}A_1A_1 \quad 2 \times \frac{2pq}{W}A_1A_2 \quad \text{et} \quad \frac{q^2}{W}A_2A_2$$

avec toujours $W = p^2 + 4pq + q^2 = 1 + 2pq$.

À l'équilibre, pour les mêmes raisons que précédemment, on arrivera à

$$p = q = \frac{1}{2}.$$

L'application numérique donne :

$$\frac{1}{13} = \frac{13}{169}A_1A_1, \quad \frac{8}{13} = \frac{104}{169}A_1A_2 \quad \text{et} \quad \frac{4}{13} = \frac{53}{169}A_2A_2.$$

Remarquons que les gamètes produits par ces plantes porteront A_1 en fréquence $p' = (13 + 104/2)/169 = 65/169 = 5/13$. On retrouve bien les mêmes valeurs que dans la question précédente et la génération suivante de zygotes sera la même que celle de la question 2. Remarquons aussi que la population adulte ne correspond pas aux proportions de Hardy-Weinberg puisque la sélection s'est exercée après le stade zygote. À l'infini, les proportions des allèles seront bien sûr de $p = q = \frac{1}{2}$ comme on l'a vu mais les proportions des génotypes adultes seront :

$$\frac{1}{6}A_1A_1, \quad \frac{2}{3}A_1A_2 \quad \text{et} \quad \frac{1}{6}A_2A_2.$$

Lexique

Allèles (Bateson et Saunders 1902) : gènes homologues, qui assurent la même fonction mais de façon différente (exemples : allèle « pois jaune », allèle « pois vert »). Un allèle est une classe d'isoaction, c'est-à-dire un ensemble de gènes déterminant le même phénotype.

Allogamie : régime de reproduction dans lequel les individus se croisent les uns avec les autres (par opposition à **autogamie**).

Apparentement : situation dans laquelle des individus (apparentés) ont des ancêtres communs.

Autofécondation : fécondation entre deux gamètes (mâle et femelle) provenant du même individu.

Autogamie : régime de reproduction dans lequel les individus se reproduisent en autofécondation.

Coalescence des gènes dans une population (Kingman 1982) : fait que tous les gènes homologues d'une population d'effectif limité descendent d'un seul gène ancêtre, trouvé parmi les gènes présents un certain nombre de générations auparavant.

Consanguinité : régime de reproduction dans lequel les croisements se font entre individus apparentés. Résultat de tels croisements.

Dérive génétique (Wright 1921) : dans une population d'effectif limité, fluctuation aléatoire de la fréquence des gènes au fil des générations liée à un effet d'échantillonnage.

Déséquilibre de liaison (Kimura 1956) : dans une population, association non aléatoire d'allèles appartenant à des locus différents.

Dominance (Mendel 1866) : à l'origine, masquage de l'effet phénotypique d'un allèle (**récessif**) par un autre (**dominant**). Par extension, interactions entre gènes homologues.

Épistasie (Bateson 1907) : à l'origine, masquage de l'effet phénotypique d'un gène par un autre gène situé à un locus différent. Par extension, interaction entre gènes non homologues.

Fardeau génétique (Muller 1950) : diminution de la valeur sélective d'une population due au maintien d'individus ayant des valeurs sélectives inférieures à celle du meilleur individu de la population.

Fitness : (de l'anglais) autre appellation de la valeur sélective.

Gène (Johannsen 1909, à partir de pangène [De Vries 1903] issu de pangensis [Darwin 1868]) : unité de fonction du matériel héréditaire, séquence d'ADN transcrite.

Gènes homologues : gènes normalement appariés à la méiose. Des gènes homologues ont nécessairement la même fonction.

Gènes identiques par descendance : gènes issus par copie sans mutation d'un même gène ancêtre.

Génome (Wrinkler 1920) : ensemble de l'information génétique portée par un individu.

Génotype (Johannsen 1909) : Constitution génétique d'un individu. composition allélique du ou des locus étudiés chez un individu.

Hétérogamie : régime de reproduction dans lequel les unions se font entre individus phénotypiquement dissemblables.

Hétérozygote (Bateson et Saunders 1902) : individu possédant en un locus des gènes appartenant à des allèles différents.

Homogamie : régime de reproduction dans lequel les unions se font entre individus phénotypiquement semblables.

Homozygote (Bateson et Saunders 1902) : individu possédant en un locus des gènes du même allèle.

Linkage (Morgan 1911) : liaison génétique entre deux locus.

Locus (Morgan *et al.* 1915) : à l'origine, emplacement du gène sur le chromosome. Le locus est défini comme une classe d'homologie. Un locus est désigné par le nom de la fonction des gènes homologues qui le constituent (exemple : le locus de la couleur du pois).

Migration : passage de gènes d'une population dans une autre, sous forme d'individus, de gamètes.

Mutation (de Vries 1901) : erreur dans la reproduction conforme du message génétique.

Neutralisme : théorie selon laquelle une grande part du polymorphisme observé dans les populations naturelles est neutre, sans signification adaptative.

Panmixie (Weismann 1895) : régime de reproduction dans lequel les gamètes s'associent au hasard vis-à-vis du ou des locus considérés.

Phénotype (Johannsen 1909) : expression du génotype dans un milieu donné. Résultat de l'observation d'un caractère par un expérimentateur.

Pléiotropie (Plate 1910) : situation dans laquelle un gène agit simultanément sur plusieurs caractères.

Polymorphisme (Ford 1940) : coexistence dans une population de deux ou plusieurs formes discontinues dans des proportions telles que la plus rare ne peut être maintenue par la mutation récurrente. Le polymorphisme génétique est la coexistence dans une population de plusieurs allèles au locus considéré.

Population (Johannsen 1903) : ensemble d'individus se reproduisant entre eux par la voie sexuée.

Pression évolutive : action exercée sur une population qui a pour effet de modifier la structure allélique (hasard, sélection, mutation, migration).

Récessif (Mendel 1866) : voir **dominance**.

Recombinaison (Bridges et Morgan 1923) : processus permettant à de nouvelles combinaisons des caractères parentaux d'apparaître dans la descendance. L'association allélique de deux locus liés, observée chez les parents, est modifiée et aboutit à des associations appelées *recombinées*.

Régime de reproduction : description de la façon dont les gamètes s'associent, vis-à-vis du gène ou des gènes considérés, pour former la génération suivante.

Ségrégation (Bateson et Saunders 1902) : processus se traduisant par la séparation de formes alléliques du même locus, initialement associées.

Sélection (Darwin 1858) : processus entraînant une reproduction supérieure ou inférieure chez certains individus du fait de leur génotype. La sélection peut être naturelle ou exercée par l'Homme pour l'amélioration des espèces animales ou végétales.

Structure allélique (ou génique) : dans une population, fréquences des différents allèles présents au locus considéré.

Structure génétique : description de la structure allélique et de la structure génotypique d'une population.

Structure génotypique : dans une population, fréquences des différents génotypes au locus considéré.

Superdominance (Hull 1946) : relation de dominance dans laquelle les hétérozygotes ont une valeur supérieure aux deux homozygotes correspondants pour le caractère considéré.

Valeur sélective d'un génotype : succès reproducteur des individus du génotype sous l'effet de la sélection. Elle se mesure par le nombre moyen de descendants laissés par ce génotype à la génération suivante.

Bibliographie

- P.H. Gouyon, J.P. Henry et J. Arnould – *Les Avatars du gène*. Pour la science, Belin, 1997.
- D.L.Hartl – *Génétique des populations*, traduit par N. Borot. Médecine-Sciences, Flammarion, 1995.
- M. Kimura – *La Théorie neutraliste de l'évolution moléculaire*, traduit par C. Montgelard. Nouvelle bibliothèque scientifique, Flammarion, 1994.
- J. Maynard Smith – *Evolutionary genetics*. Oxford University Press, 1989.
- M. Solignac, G. Périquet, D. Anxolabéhère et C. Petit – *Génétique évolutive*. Hermann, 1995.
- Tome I. *La Variation, les gènes dans les populations*.
- Tome II. *L'Espèce, l'évolution moléculaire*.

Index

A

Abeilles 64
 ADN X
 Allèle 3
 Allèle nul 95
 Allogame 20, 83
 Allozyme 95
 Anémie falciforme 69, 103
 Apparentés 14
 Autogame 20, 83
 Autogamie 13
 Auto-incompatibilité 61
 Auto-stop 50
 Avantage du rare 105
 Avoine 107

B

Bactéries 65
 Biodiversité 108

C

Chats 55
 Coalescence 33
 Coefficient de consanguinité 18
 Coefficient de parenté 18
 Coléoptères 58
 Consanguin 14
 Consanguinité 13, 14
 Coquelicot 70
 Cryptopolymorphisme 77
 Cyanure 149

D

Daltonisme 59
 Daphnies (*Daphnia cucullata*) 57
 Dates de floraison 61
 Déficit d'hétérozygotes 41
 Dérive 26
 Déséquilibre de liaison 46
 Déséquilibre gamétique 47
 Diploïdes 3, 64
 Dominance 5
 Drépanocytose 103
 Drosophile 59

E

Effectif génétique 28
 Effet de fondation 30
 Effet délétère 85
 Effet Hill-Robertson 50
 Effet Wahlund 40
 Électrophorèse X, 95
 Évolution X, 3

F

Fardeau 84
 Fardeau génétique 109
 Fertilité 34
 Fixation 27
 Fixées 28
 Forme de la Terre 1
 Fréquence IX, X, 3

G

Galactosémie 89
 Gènes identiques 14
 Génotype IX
 Goulot démographique 28
 Groupes sanguins 94

H

Haploïde 3, 64
 Hardy-Weinberg 9
 Hémoglobine 103
 Hermaphrodites 60
 Hétérogamie 19
 Hétérosis 103
 Hétérozygotes 4
 Hêtres (*Fagus sylvatica* L.) 55
 Homogamie 13
 Homozygotes 4
 Horloge moléculaire 112

I

Indice de fixation 20
 Isozymes 95

K

Kimura 112

L

Létal 76
 Locus 3

M

Maïs 68
 Maladies génétiques 75
 Malaria 103
 Marqueurs X
 Migration 37
 Modèle dit « en île » 38
 Monomorphe 93
 Mucoviscidose 58
 Mutation IX, 36

N

Neutre 94

P

Panmixie 7, 11
Phénotype IX
Polymorphisme 93
Polymorphisme protégé 106
Populations naturelles X
« pool » de gènes 3
Poules 68
Pressions 37
Protéines 97

R

Récessif 20
Régime de reproduction 11
Reproduction asexuée 20
RFLP 97

S

Sélection IX
Sélection apostatique 105

Sélection fréquence-dépendante 105
Sexe 19
Sites de restriction 97
Speonomus zophosinus 58
Superdominance 102

T

Taille 26
Tay-Sachs 89
Trèfle 61
Trifolium repens 149

V

Vaches 68
Valeur sélective 33
Viabilité 33
Vigueur hybride 103

X

X2 60