

# éléments de génétique quantitative

Louis  
OLLIVIER

2<sup>ème</sup> édition revue et augmentée



MIEUX COMPRENDRE

 **INRA**  
EDITIONS



# éléments de génétique quantitative

2<sup>ème</sup> édition revue et augmentée

© INRA, 2002

ISBN : 2-7380-1029-6

ISSN : 1144-7605

La première édition de cet ouvrage a été publiée par les éditions Masson.

© INRA et Masson, Paris 1981

ISBN : 2-225-68684-X

Le code de la propriété intellectuelle du 1<sup>er</sup> juillet 1992 interdit la photocopie à usage collectif sans autorisation des ayants droit. Le non respect de cette disposition met en danger l'édition, notamment scientifique. Toute reproduction, partielle ou totale, du présent ouvrage est interdite sans autorisation de l'éditeur ou du Centre français d'exploitation du droit de copie (CFC), 20 rue des Grands-Augustins, 75006 Paris.

# éléments de génétique quantitative

2<sup>ème</sup> édition revue et augmentée

L. OLLIVIER

## MIEUX COMPRENDRE

*Ouvrages parus dans la même collection :*

### **Génie génétique**

Une histoire, un défi

E. HEBERLE-BORS

trad. M.L. SPIRE, R. JUDOR

2001, 304 p.

### **L'eau dans l'espace rural**

Vie et milieux aquatiques

A. NEVEU, C. RIOU, R. BONHOMME,

P. CHASSIN, F. PAPY, éd.

2001, 300 p.

### **Principes de virologie végétale**

Génome, pouvoir pathogène, écologie des virus

S. ASTIER, J. ALBOUY, Y. MAURY, H. LECOQ

2001, 488 p.

### **Le grain de blé**

Composition et utilisation

P. FEILLET

2000, 310 p.

### **Biology of lactation**

J. MARTINET, L.-M. HOUEBINE, H.H. HEAD

1999, 686 p.

### **Principes des techniques de biologie moléculaire**

Denis TAGU

1999, 136 p.

### **Sol : interface fragile**

Pierre STENGEL et Sandrine GELIN

1998, 222 p.

### **Les marqueurs moléculaires en génétique et biotechnologies végétales**

Dominique DE VIENNE

1998, 200 p.

### **Assimilation de l'azote chez les plantes**

Aspects physiologique, biochimique et moléculaire

Jean-François MOROT-GAUDRY (éd.)

1997, 422 p.

### **L'eau dans l'espace rural**

Production végétale et qualité de l'eau

C. RIOU, R. BONHOMME, P. CHASSIN,

A. NEVEU, F. PAPY (éd.)

1997, 414 p.

### **La pomme de terre**

P. ROUSSELLE, Y. ROBERT

et J.-C. CROSNIER (éd.)

1996, 640 p.

### **Vie microbienne du sol et production végétale**

Pierre DAVET

1996, 380 p.

### **Nutrition des ruminants domestiques**

R. JARRIGE, Y. RUCKEBUSH,

C. DEMARQUILLY, M.-H. FARCE,

et M. JOURNET (éd.)

1995, 921 p.

### **Biologie de la lactation**

Jack MARTINET,

et Louis-Marie HOUEBINE

1993, 587 p.

### **Amélioration des espèces végétales cultivées**

Objectifs et critères de sélection

André GALLAIS et Hubert BANNEROT

1992, 768 p.

### **La régression non linéaire : méthodes et applications en biologie**

Sylvie HUET, Emmanuel JOLIVET,

et Antoine MESSÉAN

1992, 250 p.

### **L'épidémiologie en pathologie végétale : mycoses aériennes**

Frantz RAPILLY

1991, 318 p.

### **Les oligo-éléments en agriculture et en élevage**

Yves COÏC et Marcel COPPENET

1989, 114 p.

# Avant-propos de la 2<sup>e</sup> édition

La première édition de ce manuel de génétique quantitative remonte à 1981. Le livre faisait partie de la collection des *Actualités scientifiques et agronomiques de l'INRA* (Institut National de la Recherche Agronomique), publiée aux éditions Masson (Paris). Cette édition étant aujourd'hui épuisée, et la génétique quantitative ayant connu depuis vingt ans de nombreux développements, la préparation d'une nouvelle édition des *Éléments de génétique quantitative* paraissait souhaitable. Cette nouvelle édition conserve la même structure générale en cinq chapitres, consacrés successivement, à la situation de base des populations qui correspond à l'équilibre de Hardy-Weinberg, et aux quatre principaux facteurs d'évolution des populations que sont leur taille souvent limitée, la sélection, les mutations (avec la considération de gènes individuellement identifiés) et les migrations (avec la prise en considération des croisements).

Au cours de ces vingt dernières années, la génétique quantitative a continué à étendre son champ d'action. Un progrès majeur a sans aucun doute été l'application de la méthode du *meilleur prédicteur linéaire sans biais* de Henderson (méthode connue sous l'acronyme anglais BLUP) à l'évaluation génétique des reproducteurs. La théorie du BLUP fut proposée dès le début des années 1950, comme une extension de la théorie des indices de sélection, théorie elle-même établie dans ses grandes lignes au début des années 1940. Le BLUP est maintenant devenu un outil universel d'évaluation génétique dans les populations animales, grâce essentiellement aux progrès réalisés dans les moyens de calcul automatique. La méthode est succinctement décrite au chapitre 3.

La génétique moléculaire a constitué une avancée majeure de la biologie de la 2<sup>e</sup> moitié du XX<sup>e</sup> siècle. Ses perspectives d'application dans l'amélioration génétique avaient été analysées dès le début des années 1960, mais elles connurent un net regain de faveur à partir des années 1980, avec notamment les progrès dans la connaissance du génome des espèces animales et végétales utiles à l'Homme. Depuis lors, cette connaissance a continué de progresser et des cartes génétiques très détaillées sont maintenant disponibles dans la plupart de ces espèces. Le chapitre 4 présente les principes de base de la mise en évidence et de l'utilisation en sélection de gènes de nature diverse ou de simples segments d'ADN qui balisent le génome.

Quelques compléments de moindre importance ont par ailleurs été apportés à divers chapitres du livre. C'est le cas pour les méthodes d'estimation des paramètres génétiques au chapitre 1. Le chapitre 2 a été complété par l'étude des populations fractionnées à l'aide des indices de fixation de Sewall Wright, et par des notions de démographie utiles dans le calcul des effectifs génétiques. La théorie de la sélection à deux locus a été ajoutée dans le chapitre 3. Le chapitre 5 inclut une présentation des méthodes de sélection en croisement, qui vient en complément de la méthode classique de la sélection récurrente réciproque.

L'optique maintenue dans ce livre est de se situer à un niveau élémentaire, ce qui a obligé à restreindre le détail des calculs et parfois de schématiser des situations très complexes. Le traité *Genetics and Analysis of Quantitative Traits* de Lynch et Walsh (1998) est à recommander pour une étude plus approfondie des derniers développements de la discipline. Le 2<sup>e</sup> volume de ce traité, *Evolution and Selection of Quantitative Traits*, est en préparation et peut d'ores et déjà être consulté sur la Toile à l'adresse suivante : [http://nitro.biosci.arizona.edu/zbook/volume\\_2/vol2.html](http://nitro.biosci.arizona.edu/zbook/volume_2/vol2.html). D'autre part, la génétique des populations, qui est brièvement traitée au début de chaque chapitre, est une discipline également en pleine expansion, tout particulièrement avec les compléments apportés aux études de protéines par l'examen des séquences d'ADN. D'excellents manuels de génétique des populations ont récemment été publiés, parmi lesquels on peut retenir la 3<sup>e</sup> édition du classique *Principles of Population Genetics* de Hartl & Clark (1997). Rappelons enfin l'existence du manuel classique de Falconer, *Introduction to Quantitative Genetics*, dont le succès est attesté par ses quatre éditions depuis 1960, et qui demeure une référence de base.

La plupart des références bibliographiques de la précédente édition ont été gardées, et un nombre relativement limité de nouvelles références a été ajouté. L'exhaustivité n'a donc pas été recherchée dans ce domaine, sachant que les ouvrages qui viennent d'être mentionnés sont tous complétés par des listes des références récentes les plus significatives.

#### REMERCIEMENTS

Au cours de la préparation de cette réédition, j'ai bénéficié de nombreuses discussions avec des collègues français et étrangers. Je voudrais tout particulièrement remercier mes collègues du Département de Génétique Animale de l'INRA, Jean-Jacques Colleau, Jean-Louis Foulley, Hubert de Rochambeau, ainsi que le Professeur Étienne Verrier de l'Institut National Agronomique (Paris). Je voudrais aussi remercier le Professeur Jean Générumont (Université Paris-Sud, Orsay) de m'avoir communiqué ses notes de cours, dont j'ai tiré parti dans le chapitre premier. Des commentaires m'ont été aimablement fournis par les Professeurs Leo Dempfle (Munich-Weihenstephan, Allemagne) et John James (Sydney, Australie), auxquels je suis reconnaissant pour leur examen scrupuleux des parties nouvelles ajoutées à la première édition.

J'ai également bénéficié du soutien que m'ont apporté dans la préparation du manuscrit les équipes de la Station de Génétique Quantitative et Appliquée de l'INRA, et notamment Hervé Lagant et Sylvie Nugier, ainsi que le Secrétariat de la Station, en la personne de Marie-Laure Le Paih et de Françoise Bouchain.

Il m'est agréable aussi d'exprimer ma reconnaissance à Bernard Bibé, Chef du Département de Génétique Animale de l'INRA, pour le soutien qu'il m'a apporté.

Je voudrais tout particulièrement remercier le Service INRA Éditions d'avoir accepté la prise en charge de ce travail et pour tout le soin mis à le réaliser. Jeanine Hommel, Claudine Geynet, Dominique Bollo et Joëlle Veltz ont à ce titre droit à toute ma reconnaissance.

Je remercie aussi les éditions Masson d'avoir autorisé la réédition par l'INRA de ce livre.

A Jouy en Josas le 28 février 2002

Louis Ollivier



# Table des matières

<b>Avant-propos de la 1<sup>ère</sup> édition</b> .....	11
<b>1. Populations en équilibre de Hardy-Weinberg</b> .....	15
1.1 Description génétique des populations .....	15
1.11 Cas d'un locus (loi de Hardy-Weinberg) .....	15
1.12 Cas de deux locus considérés conjointement (loi de Robbins) .....	18
1.13 Mise en évidence d'équilibres de Hardy-Weinberg .....	20
1.2 Génétique des caractères quantitatifs dans les conditions de l'équilibre de Hardy-Weinberg .....	21
1.21 Cas d'un caractère quantitatif exclusivement déterminé par deux gènes allèles .....	21
1.211 <i>Relation entre le génotype et la moyenne.</i> <i>Notion de valeur génétique</i> .....	22
1.212 <i>Variance due au génotype. Notion de variance génétique</i> .....	25
1.22 Cas d'un caractère déterminé par plusieurs locus, additifs et indépendants .....	28
1.23 Cas d'un caractère déterminé par des locus non additifs (épistasie) .....	28
1.24 Cas d'un caractère influencé par le milieu .....	30
1.25 Cas de deux caractères quantitatifs considérés conjointement (les paramètres génétiques d'une population) .....	31
1.3 Parenté et consanguinité. Ressemblance entre individus apparentés .....	33
1.31 Définitions .....	33
1.32 Calcul des coefficients de parenté et de consanguinité .....	34
1.33 Ressemblance entre individus apparentés pour des caractères quantitatifs (covariances entre apparentés) .....	36
1.4 L'estimation des paramètres génétiques .....	39
1.41 Méthodes d'estimation basées sur la régression linéaire .....	40
1.42 Méthodes d'estimation basées sur l'analyse de la variance .....	43
1.43 Estimations combinées .....	49
1.44 Méthodes d'estimation basées sur le maximum de vraisemblance .....	49

<b>2. Populations d'effectif limité et consanguinité</b> .....	53
2.1 Notion de dérive génétique : consanguinité automatique .....	53
2.11 Les fréquences géniques dans les populations d'effectif limité .....	53
2.12 Les fréquences génotypiques dans le cas de locus bialléliques. Consanguinité et hétérozygotie .....	55
2.13 Généralisation aux populations subdivisées. Indices de fixation et diversité génique .....	56
2.14 Les fréquences gamétiques dans les populations d'effectif limité : évolution du déséquilibre de liaison .....	58
2.15 La consanguinité dans les populations d'effectif limité (effectif génétique) .....	59
2.16 Le cas des générations imbriquées. Notion d'intervalle de génération ..	62
2.2 Consanguinité systématique .....	64
2.21 Croisement frère $\times$ sœur .....	65
2.22 Croisement parent $\times$ descendant .....	65
2.23 Croisement demi-frère $\times$ demi-sœur .....	66
2.3 Effets de la consanguinité sur les caractères quantitatifs .....	67
2.31 Effets de la consanguinité sur la moyenne d'un caractère quantitatif ....	67
2.32 Effets de la consanguinité sur la variance d'un caractère quantitatif ....	69
<b>3. La sélection dans les populations d'effectif illimité</b> .....	73
3.1 La sélection sur les génotypes .....	73
3.11 La sélection à un locus .....	73
3.12 La sélection à deux locus .....	76
3.2 La sélection sur les caractères quantitatifs .....	77
3.21 La sélection dans le cas d'un seul caractère : prédictions théoriques ....	77
3.211 <i>Effet de la sélection sur la moyenne</i> .....	77
3.212 <i>Effet de la sélection sur la variance</i> .....	80
3.22 La sélection dans le cas de plusieurs caractères (les méthodes de sélection) .....	88
3.221 <i>Critère et objectif de sélection uniques</i> .....	88
3.222 <i>Critère et objectif de sélection multiples :</i> <i>les indices de sélection</i> .....	90
3.23 Généralisation des indices de sélection (la prédiction des valeurs génétiques par la méthode BLUP) .....	97
3.231 <i>Prise en compte d'une information inégale</i> <i>sur les candidats à la sélection</i> .....	97
3.232 <i>Prise en compte des facteurs non génétiques</i> <i>(le modèle mixte)</i> .....	98
3.233 <i>Un exemple de calcul de BLUP</i> .....	100

3.24	Les plans de sélection : aspects biologique et économique .....	101
3.241	<i>Recherche de la réponse maximale à la sélection (aspect biologique)</i> .....	102
3.242	<i>Aspects économiques de la sélection</i> .....	105
3.25	Mesure de la réponse à la sélection et du progrès génétique .....	109
3.251	<i>Expériences de sélection : paramètres génétiques réalisés</i> .....	109
3.252	<i>Mesure du progrès génétique dans les populations animales</i> ....	111
<b>4.</b>	<b>Les mutations et les gènes à effet quantitatif</b> .....	<b>115</b>
4.1	Effets des mutations sur les fréquences géniques .....	115
4.11	Les mutations dans les populations d'effectif illimité sans sélection ....	115
4.12	Les mutations dans les populations d'effectif illimité avec sélection ....	116
4.13	Les mutations dans les populations d'effectif limité .....	118
4.131	<i>Effets de la dérive dans le cas de deux allèles pouvant muter l'un vers l'autre (sans sélection)</i> .....	118
4.132	<i>Effets de la dérive sur l'équilibre mutation-sélection</i> .....	118
4.133	<i>Effets de la dérive quand le nombre d'allèles susceptibles d'apparaître par mutation est très grand (évolution moléculaire)</i> .....	119
4.2	Effets des mutations sur les caractères quantitatifs .....	120
4.21	Mutations et variabilité polygénique .....	121
4.22	Recherche de gènes à effet quantitatif .....	121
4.221	<i>Mise en évidence de gènes majeurs</i> .....	122
4.222	<i>Mise en évidence directe de l'effet d'un gène sur un caractère quantitatif</i> .....	123
4.223	<i>Mise en évidence de l'effet d'un gène sur un caractère quantitatif à l'aide de locus marqueurs</i> .....	124
4.23	Utilisation des gènes à effet quantitatif dans l'amélioration génétique : la sélection assistée par marqueur .....	127
4.231	<i>Le cas des gènes identifiés ou associés statistiquement à des gènes marqueurs (utilisation de locus connus ou de gènes candidats)</i> .....	127
4.232	<i>Le cas des gènes en équilibre de liaison avec des gènes marqueurs (absence de locus connus)</i> .....	128
<b>5.</b>	<b>Les migrations et les croisements</b> .....	<b>133</b>
5.1	Effets des migrations sur les fréquences géniques .....	133
5.11	Les migrations dans les populations d'effectif illimité sans sélection ....	133
5.12	Les migrations dans les populations d'effectif illimité avec sélection ...	134
5.13	Les migrations dans les populations d'effectif limité (sans sélection) ...	134
5.2	Effets des migrations sur les caractères quantitatifs (diffusion de l'amélioration génétique) .....	135

5.21 Migration sans sélection .....	135
5.22 Migration avec sélection (retard génétique) .....	135
5.23 Cas général .....	136
5.3 Effets du croisement sur les caractères quantitatifs .....	140
5.31 Le croisement simple : effet d'hétérosis .....	141
5.32 Les différents systèmes de croisement .....	143
5.33 Utilisation du croisement dans l'amélioration génétique .....	151
5.331 <i>Recherche des meilleurs croisements :</i> <i>sélection entre croisements</i> .....	151
5.332 <i>Sélection sur les résultats de croisement :</i> <i>sélection récurrente</i> .....	154
<b>Conclusion</b> .....	161
<b>Références bibliographiques</b> .....	165
<b>Annexe</b> .....	175
<b>Symboles utilisés</b> .....	179
<b>Index</b> .....	181

# Avant-propos de la 1<sup>e</sup> édition

Depuis les premiers temps de l'agriculture, l'Homme a utilisé des espèces animales et végétales pour répondre à ses besoins, principalement pour se nourrir et se vêtir. La recherche d'animaux et de plantes permettant de mieux satisfaire ces besoins a donc toujours été au premier rang des préoccupations humaines. Le développement de la génétique, par contre, est très récent si l'on considère la longueur de l'évolution qui a vu la transformation progressive des espèces sauvages en espèces domestiques ou cultivées. La génétique a élucidé les mécanismes de cette transformation et elle fournit en outre les moyens de mieux contrôler l'évolution des animaux et des plantes actuellement utilisés par l'Homme.

Les travaux de Mendel ont été le point de départ d'une attitude nouvelle face au matériel biologique : l'intérêt n'est plus porté à l'individu en lui-même mais à des ensembles d'individus très nombreux ou populations. Car, si les caractères d'un individu sont imprévisibles, ceux d'une population obéissent à des lois statistiques. La « génétique des populations » se constitue en tant que discipline scientifique dès le début de ce siècle, puisque son point de départ théorique, la loi de Hardy-Weinberg, remonte à 1908. Les développements ultérieurs sont dus principalement aux travaux de Fisher, Wright et Haldane. Comme le hasard intervient aux deux étapes essentielles du cycle de l'être vivant, la méiose et la fécondation, il n'est pas surprenant que cette branche de la génétique repose sur le calcul des probabilités. L'aspect probabiliste de la génétique a été mis en évidence d'une manière particulièrement féconde par Malécot dans son livre « Les mathématiques de l'hérédité ». Ce livre, paru en 1948 et longtemps méconnu même en France, constitue maintenant l'un des ouvrages classiques dans ce domaine. La génétique « quantitative » est une discipline qui repose sur les mêmes fondements, mendéliens et probabilistes, que la génétique des populations. Elle se distingue cependant de celle-ci par son objet qui est l'étude des relations existant entre les caractères à variation continue et le patrimoine héréditaire. En fait, la génétique quantitative permet de concilier la discontinuité essentielle du matériel héréditaire et la variation continue des caractères mesurables. Elle est le fondement théorique de l'amélioration génétique animale et végétale. Les applications aux animaux sont dues essentiellement à l'école de Lush, dont le livre *Animal Breeding Plans* (1937) eut un énorme retentissement. De plus, Lush a formé toute une génération de généticiens qui ont essaimé à travers le monde et contribué à diffuser son enseignement.

Plus récemment, ces aspects de la génétique ont été présentés dans d'excellents ouvrages en langue anglaise, principalement ceux de Kempthorne (1957), de Falconer (1960), de Crow et Kimura (1970) et, d'une manière très approfondie, dans le traité de génétique des populations en 4 volumes publié par Wright entre 1968 et 1978. Ces deux derniers ouvrages, comme ceux en langue française de Malécot (1966) et de Jacquard (1970), sont

surtout axés sur la génétique des populations. Ce qui va suivre s'en inspirera largement. Cependant, dans la présentation, on tentera d'imbriquer étroitement les aspects, souvent traités séparément, que constituent la génétique des populations et la génétique quantitative. La première étudie les caractères qualitatifs associés à des gènes identifiables et l'évolution des populations naturelles alors que la seconde traite des caractères à variation continue déterminés par des gènes nombreux, pour la plupart non identifiables individuellement, et de l'évolution des populations soumises à l'action de l'Homme. En fait, les facteurs de l'évolution des populations naturelles et des populations « artificielles » ne sont pas de nature différente et on peut dire qu'en agriculture l'animal et la plante s'adaptent aux besoins de l'Homme selon des processus analogues à ceux par lesquels dans la nature ils s'adaptent à leur milieu. Pour Darwin, d'ailleurs, la domestication a été le modèle sur lequel il a basé sa théorie de la sélection naturelle.

C'est pourquoi, après avoir présenté le cas des populations satisfaisant à la loi de Hardy-Weinberg, nous aborderons successivement les principaux facteurs d'évolution des populations, qu'elles soient naturelles ou non. Ces facteurs sont, essentiellement :

- le fractionnement en sous-populations d'effectif limité,
- la sélection,
- les mutations,
- les migrations et le croisement.

L'ampleur du sujet nous a amené à restreindre les développements théoriques au minimum nécessaires à la compréhension des méthodes d'amélioration génétique couramment employées dans les différentes espèces, et l'évolution naturelle des populations est, pour la même raison, délibérément réduite au profit de l'étude de leur amélioration par l'Homme. Malgré tout, quelques développements analytiques, de portée générale ou qui peuvent servir d'exemples, sont conservés.

Précisons que seules sont considérées les populations d'individus diploïdes à sexes séparés. Ce qui suit concerne donc avant tout les espèces animales domestiques (mammifères et oiseaux), mais la plupart des notions présentées s'appliquent aussi aux végétaux à reproduction allogame. Pour un tableau complet des méthodes d'amélioration génétique végétale, on peut consulter l'ouvrage de Demarly (1977).

La présentation choisie est volontairement analytique. Ainsi, bien que les facteurs d'évolution qui viennent d'être mentionnés n'interviennent pas d'une manière indépendante les uns des autres et qu'ils agissent souvent ensemble, les effets combinés de deux ou plusieurs facteurs ne seront que partiellement traités. Dans le même ordre d'idées, il faut souligner que *la présentation repose très fortement sur l'hypothèse d'additivité des effets génétiques et de milieu*. En conséquence, les notions d'interaction génotype-milieu et d'épistasie, par exemple, seront définies succinctement mais les complications qui en résultent ne seront que brièvement évoquées. Le lecteur sera donc fréquemment renvoyé aux ouvrages qui ont été mentionnés plus haut pour complément d'information.

Il faut aussi souligner que l'accent est mis sur l'aspect théorique, qui consiste à déduire des conséquences, ou des prédictions, à partir de quelques hypothèses. Mais la démarche inverse, qui consiste à « remonter » des observations à la théorie, apparaît également tout au long du livre, à travers les problèmes d'« estimation ». Cependant, l'estimation statistique fait appel à des méthodes très variées et souvent complexes dont il n'est donné qu'un aperçu. En particulier, il n'a pas été possible d'inclure les récents développements de l'école d'Henderson dans ce domaine, qui font suite notamment aux méthodes proposées par Henderson (1953).

Nous supposons connus les fondements biologiques de l'hérédité ainsi que les principes de base du calcul des probabilités. Quelques notions de calcul matriciel sont également nécessaires pour l'étude des indices de sélection et des migrations. Un rappel de ces notions se trouve, par exemple, dans le livre de Jacquard (1974). Les variables aléatoires auxquelles il est fait appel le plus fréquemment sont les variables de Bernoulli, binomiale, de Poisson et normale. Leurs lois sont supposées connues ainsi que les opérations applicables aux variables aléatoires en général, soient l'espérance (ou moyenne), la variance et la covariance.





# Populations en équilibre de Hardy-Weinberg

## 1.1 Description génétique des populations

### 1.1.1 Cas d'un locus (loi de Hardy-Weinberg)

Une population se définit comme un ensemble d'individus se reproduisant entre eux. Pour décrire du point de vue génétique une population, on a recours aux notions de *fréquence génotypique* et de *fréquence génique*, qui définissent une *structure génétique*. Une fréquence génotypique est le rapport du nombre d'individus d'un génotype sur le nombre total d'individus de la population. Si on considère, par exemple, un seul locus autosomique<sup>1</sup>, avec deux allèles  $A_1$  et  $A_2$ , on a trois génotypes possibles :

Génotype	Nombre d'individus	Fréquence génotypique
$A_1A_1$	$n_1$	$P = n_1/N$
$A_1A_2$	$n_2$	$2Q = n_2/N$
$A_2A_2$	$n_3$	$R = n_3/N$
	$(n_1 + n_2 + n_3 = N)$	$(P + 2Q + R = 1)$

Une fréquence génique est le rapport du nombre de gènes d'une catégorie sur le nombre total de gènes présents au locus considéré. Dans l'exemple ci-dessus ce nombre est égal à  $2N$ . Le nombre de gènes  $A_1$  est  $2n_1 + n_2$ , puisque chaque individu  $A_1A_1$  possède deux exemplaires de ce gène et chaque individu  $A_1A_2$  en possède un seul.

1. Un locus est l'emplacement qu'occupe un gène sur un *chromosome*, qui lui-même peut être un *autosome* (chaque individu en possède deux exemplaires) ou un *hétérosome*, appelé aussi *chromosome sexuel* (l'individu en possède deux homologues, XX, ou deux différents, XY, selon le sexe). Nous adopterons la convention habituelle de considérer les noms latins comme invariables, soit un *locus*, des *locus*.

La fréquence du gène  $A_1$  est donc :

$$p = \frac{2n_1 + n_2}{2N} = P + Q$$

La fréquence du gène  $A_2$  est :

$$q = \frac{n_2 + 2n_3}{2N} = Q + R, \text{ et donc } p + q = 1$$

Dans l'exemple ci-dessus on voit que ces fréquences sont directement *mesurables*. Il n'en est pas toujours ainsi, soit que l'on ne dispose que d'un échantillon de la population, soit que l'on ne puisse distinguer tous les génotypes, par exemple quand les deux génotypes  $A_1A_1$  et  $A_1A_2$  ont le même phénotype. Il faut alors recourir à des méthodes *d'estimation*, qui sont variées et qu'il n'est pas possible de présenter ici (voir par exemple Kempthorne, 1957).

**Loi de Hardy-Weinberg.** — Cette loi, établie en 1908 indépendamment par l'anglais Hardy et l'allemand Weinberg, s'énonce comme suit :

*Dans une population fermée, d'effectif illimité, non soumise à la sélection ni aux mutations, et où les unions se font au hasard, les fréquences géniques restent constantes de génération en génération, de même que les fréquences génotypiques, celles-ci pouvant se déduire de la connaissance des fréquences géniques.*

Une population est dite en équilibre de Hardy-Weinberg lorsqu'elle satisfait à toutes les conditions de validité de cette loi. Examinons-les une à une :

1) *population fermée* : cette condition exclut les migrations et les croisements entre populations, qui peuvent modifier les fréquences géniques ;

2) *effectif illimité* : les fréquences géniques et génotypiques peuvent varier d'une génération à l'autre du seul fait du hasard, mais ces variations tendent à s'annuler dès que l'effectif de la population devient très grand ;

3) *absence de sélection* : c'est-à-dire qu'on suppose que le nombre de descendants d'un individu est indépendant de son génotype ;

4) *absence de mutations* : les mutations sont des événements aléatoires, de très faible probabilité, qui font apparaître des gènes nouveaux ou font disparaître des gènes existant dans la population et modifient donc les fréquences géniques ;

5) le régime de la *panmixie*, ou union au hasard, s'oppose aux systèmes suivants :

– homogamie, quand les unions se font préférentiellement entre les individus qui se ressemblent, quant à leur phénotype (homogamie phénotypique) ou à leur génotype (homogamie génotypique),

– hétérogamie, quand les unions se font préférentiellement entre des individus dissemblables,

– endogamie, quand les unions se font préférentiellement entre des individus apparentés (consanguinité),

– exogamie, quand les unions se font préférentiellement entre des individus non apparentés.

La panmixie est un mode d'union très répandu dans les populations animales. On peut du moins souvent considérer que la réalité s'écarte suffisamment peu de cette hypothèse pour que les erreurs faites en l'adoptant soient négligeables. Remarquons dès maintenant que,

contrairement aux conditions précédentes, les écarts à la panmixie ne modifient pas les fréquences géniques mais seulement les fréquences génotypiques (voir chap. 2).

Certains auteurs englobent sous le nom de panmixie l'ensemble des conditions de validité de la loi de Hardy-Weinberg. En fait, la panmixie, au sens strict, est compatible avec toutes les éventualités possibles par ailleurs, pour les conditions (1) et (4), et elle n'est donc qu'une des conditions nécessaires pour que cette loi s'applique.

**Démonstration de la loi Hardy-Weinberg.** — Considérons le cas d'un locus avec 2 allèles et une population infinie où les fréquences des trois génotypes ont des valeurs quelconques à la génération  $n$  :

$$\begin{array}{ll} A_1A_1 & P \\ A_1A_2 & 2Q \\ A_2A_2 & R \end{array}$$

telles que  $P + 2Q + R = 1$ .

La génération suivante sera obtenue, dans les conditions de la panmixie, en supposant les trois génotypes également fertiles (absence de sélection) et non susceptibles de muter, par combinaison au hasard des gamètes porteurs de  $A_1$  dont la fréquence est  $P + Q = p$ , et des gamètes porteurs de  $A_2$  dont la fréquence est  $Q + R = q$ .

		Gamète mâle	
		$A_1$	$A_2$
Gamète femelle	Fréquence	$p$	$q$
	$A_1$	$p$	$p^2 A_1A_1$
$A_2$	$q$	$pq A_1A_2$	$q^2 A_2A_2$

L'hypothèse d'un effectif illimité permet d'assimiler fréquence et probabilité. Les quatre combinaisons possibles, avec leurs probabilités, figurent ci-dessus. On voit que les fréquences des trois génotypes dans la génération  $n + 1$  sont :

$$p^2 A_1A_1, 2pq A_1A_2, q^2 A_2A_2.$$

La fréquence de  $A_1$  dans cette génération est  $p^2 + pq = p(p + q) = p$  et la fréquence de  $A_2$  est  $pq + q^2 = q(p + q) = q$ . Ces fréquences n'ayant pas varié, les fréquences génotypiques de la génération  $n + 2$  resteront également inchangées. L'équilibre est donc atteint au bout d'une seule génération à partir d'une population de structure génétique quelconque.

*Généralisation de la loi de Hardy-Weinberg au cas de plusieurs allèles.* — Si on considère deux quelconques des allèles possibles à un locus donné, dont les fréquences sont  $p_i$  et  $p_j$ , les fréquences génotypiques correspondantes seront :

$$p_i^2 A_iA_i, 2p_i p_j A_iA_j \text{ et } p_j^2 A_jA_j.$$

Ainsi, dans le cas de trois allèles  $A_1, A_2, A_3$ , avec des fréquences  $p, q, r$ , on a six génotypes dont les fréquences sont :

$$\begin{array}{lll} p^2 A_1A_1 & q^2 A_2A_2 & r^2 A_3A_3 \\ 2pq A_1A_2 & 2pr A_1A_3 & 2qr A_2A_3 \end{array}$$

*Cas des gènes liés au sexe.* — Considérons un locus situé sur le chromosome X, avec deux allèles possibles  $A_1$  et  $A_2$ . Dans le sexe homogamétique (femelle chez les mammifères, mâle chez les oiseaux) à 2 chromosomes X, supposons les fréquences génotypiques suivantes :

$$P A_1 A_1, 2Q A_1 A_2 \text{ et } R A_2 A_2 \text{ avec } P + 2Q + R = 1$$

et, dans le sexe hétérogamétique, à un seul chromosome X :

$$S A_1 \text{ et } T A_2, \text{ avec } S + T = 1 ;$$

$P, Q, R, S$  et  $T$  étant quelconques.

Les fréquences de  $A_1$  dans le sexe homogamétique XX (supposé femelle) et dans le sexe hétérogamétique XY (supposé mâle) sont respectivement :

$$p_f = P + Q \text{ et } p_m = S.$$

Pour calculer la fréquence génique globale  $p$  dans la population, il faut tenir compte du fait que les femelles contiennent deux fois plus de chromosomes X que les mâles, et on a :

$$p = (2p_f + p_m)/3 = [2(P + Q) + S]/3.$$

Les fréquences de  $A_1$  à la génération suivante seront :

$$p'_f = 1/2(p_m + p_f) \text{ et } p'_m = p_f$$

D'où

$$p'_f - p'_m = -1/2(p_f - p_m) \text{ et } p' = (2p'_f + p'_m)/3 = p.$$

La différence entre les fréquences dans les deux sexes est diminuée de moitié et changée de signe à chaque génération. On aboutit donc rapidement à l'égalité des fréquences géniques dans les deux sexes. Cette fréquence d'équilibre est égale à la fréquence génique initiale  $[2(P + Q) + S]/3$ .

## 1.12 Cas de deux locus considérés conjointement (loi de Robbins)

Si, au lieu de considérer un seul locus, nous en considérons deux, les fréquences des divers génotypes possibles ne peuvent plus se déduire des fréquences géniques, comme dans le cas précédent, mais elles résultent des *fréquences gamétiques*, que l'on définit comme le pourcentage de gamètes porteurs de deux (ou plusieurs) gènes appartenant à des locus différents. L'évolution des fréquences gamétiques au cours des générations ne peut généralement pas se déduire, comme nous allons le voir, de la seule connaissance des fréquences géniques.

Considérons deux gènes A et B, à *deux locus autosomiques*. Appelons  $P_{AB}$  la fréquence des gamètes porteurs à la fois de A et de B à la génération  $n$ , et  $P'_{AB}$  la fréquence du même type de gamète à la génération suivante. La relation entre  $P_{AB}$  et  $P'_{AB}$  peut être établie en faisant le raisonnement suivant, dû à Malécot. Considérons un individu pris au hasard et cherchons la probabilité  $P'_{AB}$  pour qu'il produise un gamète AB. Les deux gènes A et B réunis sur le même gamète peuvent provenir soit d'un même gamète de la génération précédente, soit chacun d'un des gamètes parentaux. Ces deux événements ont des probabilités respectives de  $1 - c$  et  $c$ ,  $c$  étant le taux de recombinaison entre les deux locus ( $c$  varie de 0 à 1/2 : la valeur 0 correspond à une *liaison génétique* très étroite et la valeur 1/2 à l'indépen-