

## Chapitre III. Structure génétique de population idéale.

### Population théorique idéale :

#### Définition :

Le devenir de la variabilité génétique d'une population au cours des générations (la transmission des différents allèles et leurs fréquences) est au premier abord très difficile à prévoir. Outre la difficulté à identifier une population, c'est-à-dire les limites du groupe d'individus sur lequel calculer les fréquences alléliques, de très nombreux facteurs peuvent modifier la fréquence de ces allèles (mutations, migrations, différence de survie ou fécondité entre individus). De plus, il faut considérer la transmission simultanée de très nombreux gènes polymorphes qui peuvent interagir entre eux et ne sont donc pas indépendants.

Une première étape pour contourner ces difficultés est d'aborder la transmission des caractères dans un cas simple, appelé population théorique idéale, qui se définit par les caractéristiques suivantes:

- 1) Population d'organismes diploïdes à reproduction sexuée et à générations non chevauchantes (aucun croisement entre individus de générations différentes).
- 2) Population d'effectif infini où les croisements sont entièrement aléatoires.
- 3) Population close génétiquement (absence de flux migratoires).
- 4) Tous les individus, quel que soit leur génotype, ont la même capacité à se reproduire et à engendrer une descendance viable (absence de sélection).
- 5) Absence de mutation et de distorsion de ségrégation méiotique (un individu Aa produira toujours 50% de gamètes A et 50% de gamètes a).

Parmi toutes ces caractéristiques, le croisement au hasard des individus, appelé système de reproduction panmictique, est l'hypothèse la plus importante.

Cette hypothèse suppose que les individus ne choisissent pas leur partenaire sexuel ni en fonction de leur génotype, ni en fonction de leur phénotype = panmixie et que la rencontre des gamètes se fait au hasard = pangamie.

## 2. L'équilibre de Hardy-Weinberg

Dans une population théorique idéale, les fréquences des allèles et des génotypes au cours des générations suivent une loi simple appelée loi de Hardy-Weinberg qui constitue le modèle de référence en génétique des populations. Cette loi doit son nom à Hardy, mathématicien anglais et Weinberg, médecin allemand, qui l'ont établie indépendamment en 1908. La loi de Hardy-Weinberg stipule que les fréquences alléliques et les fréquences génotypiques (c'est-à-dire la structure génétique de la population) reste stable de génération en génération. On dit alors que la population est à l'équilibre et il existe une relation simple entre les fréquences alléliques et les fréquences génotypiques (Masson et Hove, 2011, 2014).

## 3. Transmission d'un gène à 2 allèles

Supposons qu'une population soit polymorphe pour un caractère gouverné par un locus à deux allèles A et a et que les fréquences des génotypes AA, Aa et aa soient les mêmes dans les deux sexes, respectivement D, H et R (avec  $D+H+R = 1$ ).

Les fréquences alléliques à la génération  $G_0$  seront :

Pour l'allèle A  $p_0 = (D_0 + H_0)/2$

Pour l'allèle a  $q_0 = (R_0 + H_0)/2$  avec  $p_0 + q_0 = 1$

Dans une population théorique idéale, ces fréquences seront également les fréquences des différentes catégories de gamètes (identiques dans les deux sexes) soit  $p_0$  pour les gamètes qui portent l'allèle A et  $q_0$  pour les gamètes qui portent l'allèle a (absence de mutation, de sélection et de distorsion méiotique).

La formation d'un nouvel individu de la génération suivante  $G_1$  est alors le résultat de deux tirages au sort indépendants, l'un parmi les gamètes mâles, l'autre parmi les gamètes femelles (croisement au hasard = panmixie).

Les fréquences des différents génotypes de la génération suivante  $G_1$  résultent alors de la répétition de ce simple tirage au sort qui donnera les fréquences génotypiques suivantes :

$AA = p_0^2$   $Aa = 2 p_0 q_0$   $aa = q_0^2$

Dans une population théorique idéale, ces fréquences seront également celles des adultes reproducteurs de la génération  $G_1$  (absence de sélection), pour lesquels les fréquences

alléliques seront pour A  $p_1 = p_0^2 + p_0q_0 = p_0(p_0+q_0) = p_0 = p_0$  pour a  $q_1 = q_0^2 + p_0q_0 = q_0(p_0+q_0) = q_0 = q_0$

Les fréquences alléliques n'ont donc pas changé, ce qui donnera à la génération suivante  $G_2$  les mêmes fréquences génotypiques qu'à la génération précédente soit  $p^2$ ,  $AA$ ,  $2pq$   $Aa$ , et  $q^2$   $aa$ .

Le système est donc stable aussi bien en ce qui concerne les fréquences alléliques que les fréquences génotypiques.

On dit qu'on est à l'équilibre de Hardy-Weinberg dont la loi peut se s'énoncer de la façon suivante : Dans une population théorique idéale, les fréquences alléliques et les fréquences génotypiques restent stables de génération en génération. Les fréquences génotypiques sont déterminées à partir des fréquences alléliques par une relation simple qui correspond au développement du binôme  $(p+q)^2$  dans le cas d'un locus à deux allèles A de fréquence p et a de fréquence q, soit les fréquences génotypiques  $p^2$   $AA$ ,  $2pq$   $Aa$  et  $q^2$   $aa$ .

## Fréquences génotypiques constantes

### Fréquences génotypiques

| Gamètes mâles ou femelles | A (p)            | a (q)                |
|---------------------------|------------------|----------------------|
| A (p) a (q)               | AA $p^2$ Aa (pq) | Aa (pq) aa ( $q^2$ ) |

### Fréquences géniques:

|   |   |
|---|---|
| <p>pour l'allèle A</p> $p_1 = p^2 + pq$ $= p(p+q)$ $= p \times 1 \quad p_1 = p$ | <p>Pour l'allèle a</p> $q_1 = q^2 + pq$ $= q(p+q)$ $= q \times 1$ $q_1 = q$ |
|---|---|

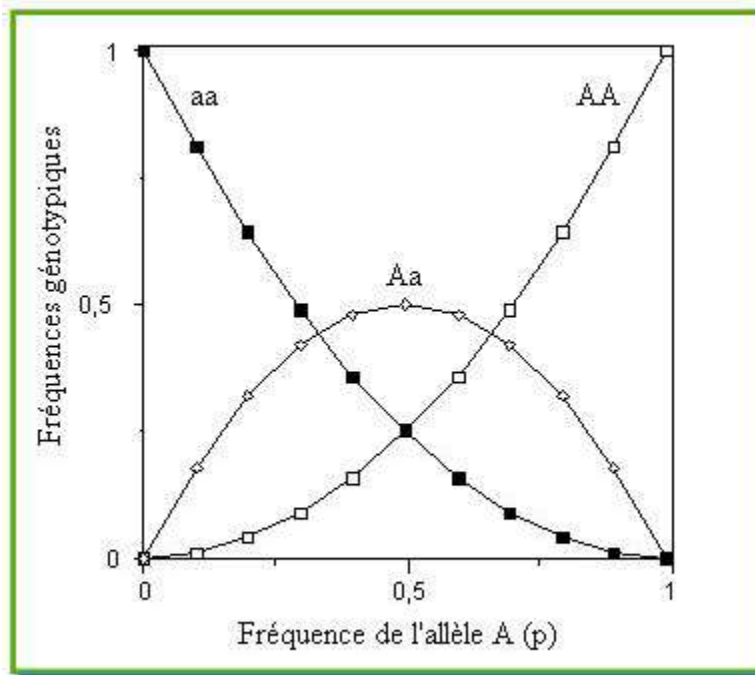
Conclusion:

D'une génération à la suivante: Fréquences alléliques constantes

Fréquences génotypiques constantes

$p^2$  AA,  $2pq$  Aa,  $q^2$  aa.

Cette relation entre fréquences alléliques et fréquences génotypiques peut être visualisée par la figure suivante:



On peut remarquer que les proportions mendéliennes 1/4, 1/2, 1/4 que l'on trouve lorsque l'on croise deux hétérozygotes est un cas particulier de la loi de Hardy-Weinberg lorsque  $p = q = 0,5$ . Chaque hétérozygote Aa a en effet comme fréquence allélique  $f(A)=1/2$  et  $f(a)=1/2$ .

#### 4. Systèmes multialléliques

La loi de Hardy-Weinberg s'applique également à des gènes qui existent sous plus de 2 états alléliques.

L'équilibre correspond alors à l'association aléatoire des différents allèles pour former les génotypes dont la fréquence reste stable de génération en génération.

Pour un locus à  $k$  allèles  $A_1, A_2, A_3, \dots, A_k$ , il y aura en théorie  $(k(k+1))/2$  génotypes différents dans la population.

Si les fréquences de ces différents allèles sont respectivement  $p_1, p_2, p_3, \dots, p_k$ , les fréquences des différents génotypes seront données par le développement de  $(p_1 + p_2 + p_3 + \dots + p_k)^2$

soit  $p_1^2 A_1 A_1$      $p_2^2 A_2 A_2$      $p_3^2 A_3 A_3$      $p_k^2 A_k A_k$

|           |           |           |           |           |     |
|-----------|-----------|-----------|-----------|-----------|-----|
| $2p_1p_2$ | $2p_1p_3$ | $2p_1p_k$ | $2p_2p_3$ | $2p_2p_k$ | etc |
| $A_1A_2$  | $A_1A_3$  | $A_1A_k$  | $A_2A_3$  | $A_2A_k$  |     |

### Explication

Si les  $k$  allèles ont même fréquence ( $p_1=p_2=p_3=...=p_k$ ), chaque allèle a une fréquence de  $1/k$  dans la population.

Il y a donc :  $k \times (1/k)^2 = 1/k$  homozygotes -  $1/k$  hétérozygotes

Par exemple, une population à l'équilibre peut comporter de plus de 90% d'hétérozygotes à un locus donné lorsqu'il existe plus de 10 allèles ayant les mêmes fréquences.

## 5. Application et utilisation du modèle de Hardy-Weinberg

### 5.1. Test de l'équilibre

Une question centrale est de savoir si la loi de Hardy-Weinberg établie pour une population théorique idéale s'applique également aux populations naturelles.

Cette loi s'appuie en effet sur un raisonnement probabiliste, ne s'applique en théorie qu'à des populations d'effectif infini, et suppose remplies toute une série de conditions qui ne sont jamais respectées dans la nature (absence de mutation, migration, sélection).

L'application de la loi de Hardy-Weinberg, peut être vérifiée pour des caractères codominants, pour lesquels le calcul des fréquences alléliques est possible. C'est le test de l'équilibre

Le principe du test est simple et peut être résumé en 3 étapes:

1- échantillonnage d'une population, dénombrement des effectifs génotypiques réels (possible grâce à la codominance) et calcul des fréquences alléliques réelles parmi les  $N$  individus échantillonnés soit  $p = f(A)$  et  $q = f(a)$

2- calcul des effectifs génotypiques attendus dans une population théorique idéale qui aurait le même effectif et les mêmes fréquences alléliques que la population étudiée soit

$AA = p^2 \times N$   $Aa = 2pq \times N$   $aa = q^2 \times N$

3- comparaison des effectifs observés et des effectifs attendus (comparaison des deux distributions) par un test statistique du Chi Deux (ou d'autres tests).

$$\chi^2 = \sum (\text{effectifs observés} - \text{effectifs théorique})^2 / \text{effectifs théoriques}$$

La somme est effectuée sur tous les génotypes et la valeur  $\chi^2$  est comparée à une valeur seuil, lue dans une table  $\chi^2$ , en fonction de 2 paramètres :

- un risque  $\alpha$  choisi par l'utilisateur qui est en général 5% et un nombre de degrés de liberté d.d.l. égale à la différence entre le nombre de génotypes et le nombre d'allèles du système génétique étudié.

- Si  $\chi^2$  calculé est inférieur à  $\chi^2$  seuil,  $H_0$  est acceptée et on conclut que la population suit la loi de Hardy-Weinberg donc est à l'équilibre.

- si  $\chi^2$  calculé est supérieur à  $\chi^2$  seuil,  $H_0$  est rejetée et on conclut que la population ne suit pas la loi de Hardy-Weinberg avec un risque  $\alpha = 5\%$  de se tromper

Table de  $\chi^2$ 

La table donne la probabilité  $\alpha$  pour que  $\chi^2$  égale ou dépasse une valeur donnée, en fonction du degré de liberté (d.d.l.).

| d.d.l. \ $\alpha$ | 0,90   | 0,50   | 0,30   | 0,20   | 0,10   | 0,05   | 0,02   | 0,01   | 0,001  |
|-------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1                 | 0,0158 | 0,455  | 1,074  | 1,642  | 2,706  | 3,841  | 5,412  | 6,635  | 10,827 |
| 2                 | 0,211  | 1,386  | 2,408  | 3,219  | 4,605  | 5,991  | 7,824  | 9,210  | 13,815 |
| 3                 | 0,584  | 2,366  | 3,665  | 4,642  | 6,251  | 7,815  | 9,837  | 11,345 | 16,266 |
| 4                 | 1,064  | 3,357  | 4,878  | 5,989  | 7,779  | 9,488  | 11,668 | 13,277 | 18,467 |
| 5                 | 1,610  | 4,351  | 6,064  | 7,289  | 9,236  | 11,070 | 13,388 | 15,086 | 20,515 |
| 6                 | 2,204  | 5,348  | 7,231  | 8,558  | 10,645 | 12,592 | 15,033 | 16,812 | 22,457 |
| 7                 | 2,833  | 6,346  | 8,383  | 9,803  | 12,017 | 14,067 | 16,622 | 18,475 | 24,322 |
| 8                 | 3,490  | 7,344  | 9,524  | 11,030 | 13,362 | 15,507 | 18,168 | 20,090 | 26,125 |
| 9                 | 4,168  | 8,343  | 10,656 | 12,242 | 14,684 | 16,919 | 19,679 | 21,666 | 27,877 |
| 10                | 4,865  | 9,342  | 11,781 | 13,442 | 15,987 | 18,307 | 21,161 | 23,209 | 29,588 |
| 11                | 5,578  | 10,341 | 12,899 | 14,631 | 17,275 | 19,675 | 22,618 | 24,725 | 31,264 |
| 12                | 6,304  | 11,340 | 14,011 | 15,812 | 18,549 | 21,026 | 24,054 | 26,217 | 32,909 |
| 13                | 7,042  | 12,340 | 15,119 | 16,985 | 19,812 | 22,362 | 25,472 | 27,688 | 34,528 |
| 14                | 7,790  | 13,339 | 16,222 | 18,151 | 21,064 | 23,685 | 26,873 | 29,141 | 36,123 |
| 15                | 8,547  | 14,339 | 17,322 | 19,311 | 22,307 | 24,996 | 28,259 | 30,578 | 37,697 |
| 16                | 9,312  | 15,338 | 18,418 | 20,465 | 23,542 | 26,296 | 29,633 | 32,000 | 39,252 |
| 17                | 10,085 | 16,338 | 19,511 | 21,615 | 24,769 | 27,587 | 30,995 | 33,409 | 40,790 |
| 18                | 10,865 | 17,338 | 20,601 | 22,760 | 25,989 | 28,869 | 32,346 | 34,805 | 42,312 |
| 19                | 11,651 | 18,338 | 21,689 | 23,900 | 27,204 | 30,144 | 33,687 | 36,191 | 43,820 |
| 20                | 12,443 | 19,337 | 22,775 | 25,038 | 28,412 | 31,410 | 35,020 | 37,566 | 45,315 |
| 21                | 13,240 | 20,337 | 23,858 | 26,171 | 29,615 | 32,671 | 36,343 | 38,932 | 46,797 |
| 22                | 14,041 | 21,337 | 24,939 | 27,301 | 30,813 | 33,924 | 37,659 | 40,289 | 48,268 |
| 23                | 14,848 | 22,337 | 26,018 | 28,429 | 32,007 | 35,172 | 38,968 | 41,638 | 49,728 |
| 24                | 15,659 | 23,337 | 27,096 | 29,553 | 33,196 | 36,415 | 40,270 | 42,980 | 51,179 |
| 25                | 16,473 | 24,337 | 28,172 | 30,675 | 34,382 | 37,652 | 41,566 | 44,314 | 52,620 |
| 26                | 17,292 | 25,336 | 29,246 | 31,795 | 35,563 | 38,885 | 42,856 | 45,642 | 54,052 |
| 27                | 18,114 | 26,336 | 30,319 | 32,912 | 36,741 | 40,113 | 44,140 | 46,963 | 55,476 |
| 28                | 18,939 | 27,336 | 31,391 | 34,027 | 37,916 | 41,337 | 45,419 | 48,278 | 56,893 |
| 29                | 19,768 | 28,336 | 32,461 | 35,139 | 39,087 | 42,557 | 46,693 | 49,588 | 58,302 |
| 30                | 20,599 | 29,336 | 33,530 | 36,250 | 40,256 | 43,773 | 47,962 | 50,892 | 59,703 |

Exemple : avec d.d.l. = 3, pour  $\chi^2 = 0,584$  la probabilité est  $\alpha = 0,90$ .

**Exemple :**

Chez l'homme, le groupe sanguin MN est déterminé par un gène à deux allèles codominants M et N, ce qui permet d'attribuer un génotype à chaque individu échantillonné, puis d'estimer les fréquences alléliques dans la population.

Une étude portant sur 730 aborigènes australiens a donné les résultats suivants :

22 MM, 216 MN, 492 NN

1- Calcul des fréquences p et q des allèles M et N :

$$p = (22 + 1/2 \times 216) / 730 = 0.178 \text{ pour l'allèle M.}$$

$$q = (492 + 1/2 \times 216) / 730 = 0.822 \text{ pour l'allèle N.}$$

2- Calcul des effectifs théoriques attendus des différentes catégories génotypiques :

$$MM = p^2 \times 730 = (0.178)^2 \times 730 = 23,1$$

$$MN = 2pq \times 730 = (2 \times 0.178 \times 0.822) \times 730 = 213,6$$

$$NN = q^2 \times 730 = (0.822)^2 \times 730 = 493,22$$

**3- Test du Chi deux**

$$\chi^2 = (22-23,1)^2/23,1 + (216-213,6)^2/213,6 + (492-493,2)^2/493 = 0,083$$

La valeur seuil pour  $3-2 = 1$  degré de liberté et un risque de 5% est 3.84. La valeur de la statistique  $\chi^2$  étant très inférieure à la valeur seuil, on conclut qu'il n'y a pas de différence significative entre la distribution observée et la distribution théorique.

On admet donc que la population d'aborigènes australiens est à l'équilibre de HardyWeinberg. Le fait qu'une population soit considérée à l'équilibre de Hardy-Weinberg après un test statistique n'implique pas que toutes les conditions d'application de cette loi soient respectées (effectif infini, absence de mutation, absence de sélection, etc).

L'hypothèse la plus importante qui doit être respectée est la panmixie. Un équilibre génétique instantané apparent peut donc être observé même si la population est soumise à une forte sélection. On en conclut que le fonctionnement génétique global est "proche" du fonctionnement théorique et ce n'est qu'en prenant en compte la dimension temporelle que l'on peut apprécier l'état génétique d'une population et prévoir son évolution.



En conclusion, dans la plupart des cas, le modèle de Hardy-Weinberg constitue un bon descripteur de la structure génétique des populations naturelles car l'hypothèse de panmixie est souvent respectée alors que les effets des mutations, migration, sélection ne sont pas assez forts pour faire diverger les fréquences Génotypiques des proportions du modèle de Hardy-Weinberg. Cette loi peut alors être utilisée pour faire des prévisions notamment dans le domaine médical. Cet équilibre de Hardy-Weinberg ne s'applique pas obligatoirement à tous les gènes d'une même population et peut ou non être rejeté en fonction du système génétique considéré.

### Exercice d'application

L'analyse du polymorphisme de l'enzyme estérase 1 dans un échantillon de 300 personnes a révélé l'existence de 3 niveaux de migration (E1, E2 et E3) à l'origine de 6 génotypes dont les effectifs sont les suivants :

$$E1E1 = 72 \quad E2E2 = 24 \quad E3E3 = 15 \quad E1E2 = 99 \quad E1E3 = 57 \quad E2E3 = 33$$

Solution

$$\text{Fréquences alléliques : } E1 = 0,5 = p ; E2 = 0,3 = q ; E3 = 0,2 = r$$

Test de l'équilibre par un test Chi 2 ( $\chi^2$ ) : Hypothèse testée : "la population étudiée suit la loi de Hardy Weinberg

### Effectifs observés Effectifs théoriques

$$15 \quad r^2 \times N = 12$$

$$72 \quad p^2 \times N = 75$$

$$24 \quad q^2 \times N = 27$$

$$99 \quad 2pq \times N = 90$$

$$57 \quad 2pr \times N = 60$$

$$33 \quad 2qr \times N = 36$$

$$\chi^2 = \sum \frac{(\text{effectifs observés} - \text{effectifs théoriques})^2}{\text{effectifs théoriques}}$$

d'où  $X^2 = 2,5$   $X^2$  seuil (5% ; 3ddl) = 7,81

3 ddl car 6 classes génotypiques - 3 (2 fréquences alléliques et l'effectif total)

Donc on accepte  $H_0$  : la population suit la loi de Hardy Weinberg.

## 5.2. Estimation des fréquences alléliques

Lorsque la variabilité d'un caractère est due à un gène à 2 allèles avec une forme allélique totalement dominante sur l'autre (A dominant sur a), seuls 2 phénotypes peuvent être distingués dans la population :

- le phénotype [A] correspondant aux génotypes AA et Aa
- le phénotype [a] correspondant au génotype aa

C'est le cas par exemple de nombreuses maladies génétiques chez l'homme qui sont dues à un allèle récessif (mucoviscidose, phénylcétonurie).

Contrairement à un système codominant, il n'est pas possible de calculer les fréquences alléliques dans la population car les proportions respectives des génotypes AA et Aa ne sont pas connues

Le modèle de Hardy-Weinberg va permettre de donner une estimation de ces fréquences à partir de la fréquence du phénotype homozygote récessif qui est égale à  $q^2$  si la population est conforme au modèle

La résolution de cette équation à une inconnue permet d'estimer la fréquence  $q$  de l'allèle récessif dans la population en calculant la racine carrée de la fréquence des homozygotes récessifs.

La fréquence  $p$  de l'allèle dominant est obtenue par différence à 1.

A partir de cette estimation des fréquences alléliques, et toujours sous l'hypothèse de conformité au modèle de Hardy-Weinberg, on obtient une estimation de la fréquence des homozygotes AA et des hétérozygotes Aa parmi les individus de phénotype [A] c'est-à-dire la probabilité qu'un individu de phénotype [A] soit homo- ou hétérozygote :

-fréquence des homozygotes parmi les individus

$$[A] = p^2/(p^2+2pq) \text{ ou } p^2/(1-q^2)$$

-fréquence des hétérozygotes parmi les individus

$$[A] = 2pq/(1-q^2).$$

**Exemple :**

Chez l'homme, une étude portant sur le système Rhésus a recensé 14% d'individus Rhésus négatif. Sachant que l'allèle Rh<sup>+</sup> est dominant sur l'allèle Rh<sup>-</sup>, l'estimation de la fréquence de l'allèle Rh<sup>-</sup> est  $q = 0.37$  (racine carrée de 0.14) sous l'hypothèse que la population suit la loi de Hardy-Weinberg. On peut en déduire la fréquence des individus Rh<sup>+</sup>Rh<sup>+</sup> et Rh<sup>+</sup>Rh<sup>-</sup> parmi les individus Rhésus positif, respectivement  $p^2/(p^2+2pq) = 0.45$  et  $2pq/(p^2+2pq) = 0.55$

**5.3. Diagnostic et conseil génétique**

La loi de Hardy-Weinberg permet de faire des prévisions sur le génotype d'un individu lorsque l'on connaît la population dont il est issu. Ce calcul est utilisé en génétique humaine pour calculer la probabilité qu'un individu soit atteint d'une anomalie génétique.

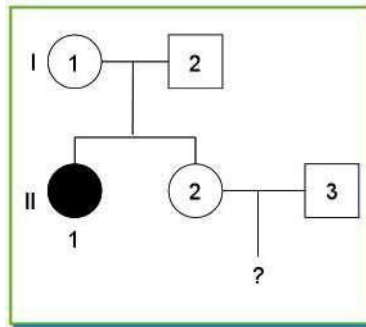
C'est le conseil génétique.

Le calcul du risque d'apparition d'une anomalie génétique chez un individu donné dépend de plusieurs paramètres :

- du déterminisme du caractère et des relations de dominance
- entre les allèles
- de la fréquence du gène responsable de la maladie dans la population
- de la généalogie de l'individu notamment des phénotypes
- des ascendants, descendants et collatéraux

Pour une maladie autosomique récessive déterminée par un allèle  $a$  de fréquence  $q$ , la probabilité qu'un individu dont on ne connaît ni la généalogie ni le phénotype soit atteint par cette maladie correspond à la fréquence de ce phénotype dans la population soit  $q^2$ .

Le diagnostic s'affine considérablement lorsque l'on dispose de plus d'informations par exemple dans la généalogie suivante où il s'agit de calculer la probabilité que le couple formé des individus sains II<sub>2</sub> et II<sub>3</sub> donne naissance à un enfant atteint de l'anomalie ce qui nécessite que les deux parents, tous les deux sains, soient hétérozygotes



Pour l'individu II3, aucune information n'est disponible, excepté son propre phénotype.

La probabilité que cet individu soit porteur de l'allèle a est  $2pq/(p^2+2pq)$  c'est à dire la fréquence des individus hétérozygotes parmi les sains dans la population.

L'individu II2 ayant une soeur atteinte, leurs parents sont obligatoirement tous les deux hétérozygotes et la probabilité est alors de  $2/3$  pour que II2 soit hétérozygote sachant qu'il est lui-même non atteint. ( $2/3$  et non  $1/2$  car le phénotype de l'individu II2 est connu). La probabilité pour que le couple II2 x II3 donne naissance à un enfant atteint de l'anomalie est alors la suivante:

$$2pq/(p^2 + 2pq) \times 2/3 \times 1/4$$

Soit : proba (père Aa) x proba (mère Aa) x proba (enfant aa sachant que les parents Aa).

## 6. Transmission des gènes liés au sexe

L'étude des gènes liés au sexe n'est pas anecdotique car chez certains organismes, ils constituent une grande partie de l'ADN codant. C'est par exemple le cas chez la drosophile dont plus d'un tiers des gènes sont porté par le chromosome X.

Pour les caractères portés par les chromosomes sexuels, les deux sexes ont des constitutions génétiques différentes et il faut distinguer : - Le sexe homogamétique qui porte les deux mêmes chromosomes sexuels (femme XX chez les mammifères, certains insectes dont la drosophile ; mâle ZZ chez certains crustacés et papillons). Ce sexe est donc diploïde pour ce chromosome. Le sexe hétérogamétique, qui porte deux chromosomes sexuels différents (ou un seul) donc haploïde ou hémizygote (mâles XY chez les mammifères, femelles WZ chez les crustacés et papillons). Ce sexe est haploïde (hémizygoïde) pour ce chromosome. Les

deux sexes ont donc des contributions génétiques différentes et s'ils sont en fréquence égale (sex-ratio équilibrée), le sexe homogamétique détient pour les gènes concernés les 2/3 du pool génétique de la population, et le sexe hétérogamétique 1/3 seulement.

a/Lorsque les fréquences alléliques sont les mêmes chez les mâles et les femelles,  $p_m = p_f = p$  et  $q_m = q_f = q$ , la loi de Hardy-Weinberg s'applique au sexe homogamétique et les fréquences génotypiques dans le sexe hétérogamétique sont directement déduites des fréquences alléliques. Chez l'homme, pour un gène porté par X et présentant 2 allèles A et a de fréquences p et q, les fréquences génotypiques dans chacun des deux sexes pour une population à l'équilibre seront :

|                                       | Femme (XX)      | Homme (XY)  |
|---------------------------------------|-----------------|-------------|
| Fréquences génotypiques à l'équilibre | $X^A X^A = p^2$ | $X^A Y = p$ |
|                                       | $X^A X^a = 2pq$ | $X^a Y = q$ |
|                                       | $X^a X^a = q^2$ |             |

Cette différence de structure génétique des sous-populations mâle et femelle a un effet spectaculaire dans le cas où un allèle est récessif. Le phénotype récessif est beaucoup plus fréquent dans le sexe hétérogamétique que dans le sexe homogamétique, où il peut être exceptionnel si la fréquence de l'allèle est faible.

C'est l'exemple classique du daltonisme chez l'homme qui est dû à une mutation récessive sur un gène porté par X.

**Maladie récessive liée au chromosome X**

|                  | Garçon |        | Filles         |                |
|------------------|--------|--------|----------------|----------------|
| <b>Phénotype</b> | Sain   | Malade | Saine          | Saine (Malade) |
| <b>Génotype</b>  | A      | A (I)  | AA             | Aa(H)<br>aa    |
| <b>Fréquence</b> | P      | q      | P <sup>2</sup> | 2Pq            |

On connaît I, l'incidence de la maladie dans la population masculine = q

La fréquence des filles hétérozygotes, H est donc égale à :

$$2 \times q \times (1 - q)$$

Si q est très petit :  $H = 2I$  (il y a deux fois plus de femmes hétérozygotes que de garçon atteints).

**En Europe:**

- la fréquence de l'allèle récessif est de l'ordre de  $q = 0,04$ .
  - l'anomalie est fréquente chez les hommes ( $q = 4\%$ ) mais très rare chez les femmes ( $q^2 = 0,16\%$ ).
  - la fréquence des hétérozygotes est élevée chez les femmes (de l'ordre de  $2pq = 8\%$ ).
- Rappelons que ces femmes ont une chance sur deux de transmettre l'anomalie à chacun de leurs descendants mâles.

b/ Lorsque les fréquences alléliques sont différentes chez les mâles et les femelles,  $p_m \neq p_f$  et  $q_m \neq q_f$ ,

la contribution différentielle des deux sexes à la descendance maintient cette différence pendant plusieurs générations et l'égalité des fréquences alléliques n'est obtenue que progressivement (contrairement aux caractères autosomiaux):

Exemple: Daltonisme chez l'homme qui est dû à une mutation récessive sur un gène porté par X.

- chaque mâle XY reçoit un chromosome X de sa mère, donc les fréquences alléliques chez les mâles à la génération t correspondent aux fréquences alléliques chez les femelles de la génération précédente t-1:

$$f(A) : p_{mt} = p_{ft-1}$$

$$f(a) : q_{mt} = q_{ft-1}$$

- chaque femelle XX reçoit un chromosome X de sa mère et un chromosome X de son père donc les fréquences alléliques chez les femelles à la génération t correspondent à la moyenne des fréquences alléliques des deux sexes de la génération précédente t-1 :

$$f(A): p_{ft} = (p_{mt-1} + p_{ft-1})/2$$

$$f(a): q_{ft} = (q_{mt-1} + q_{ft-1})/2$$

Dans l'ensemble de la population, les fréquences alléliques globales des allèles A et a sont les moyennes de la fréquence de ces allèles dans les deux sexes pondérées par leurs contributions relatives soit les coefficients 1/3 et 2/3 lorsqu'il y a autant de mâles que de femelles :

**population :**

$$f(A): p = 2/3 p_f + 1/3 p_m.$$

$$f(a): q = 2/3 q_f + 1/3 q_m.$$

Si la proportion des sexes est inégale, la pondération tient compte des effectifs  $N_m$  des mâles et  $N_f$  des femelles :

$$\text{population : } f(A): p = (p_m.N_m + 2.p_f.N_f) / (N_m + 2N_f)$$

Au cours des générations, l'évolution des fréquences alléliques dans chacun des deux sexes fluctue autour de cette valeur d'équilibre avec une différence qui s'inverse et diminue de moitié à chaque génération. Cette fluctuation conduit à l'égalité des fréquences alléliques dans les deux sexes après plusieurs générations de croisements panmictiques. Fluctuation

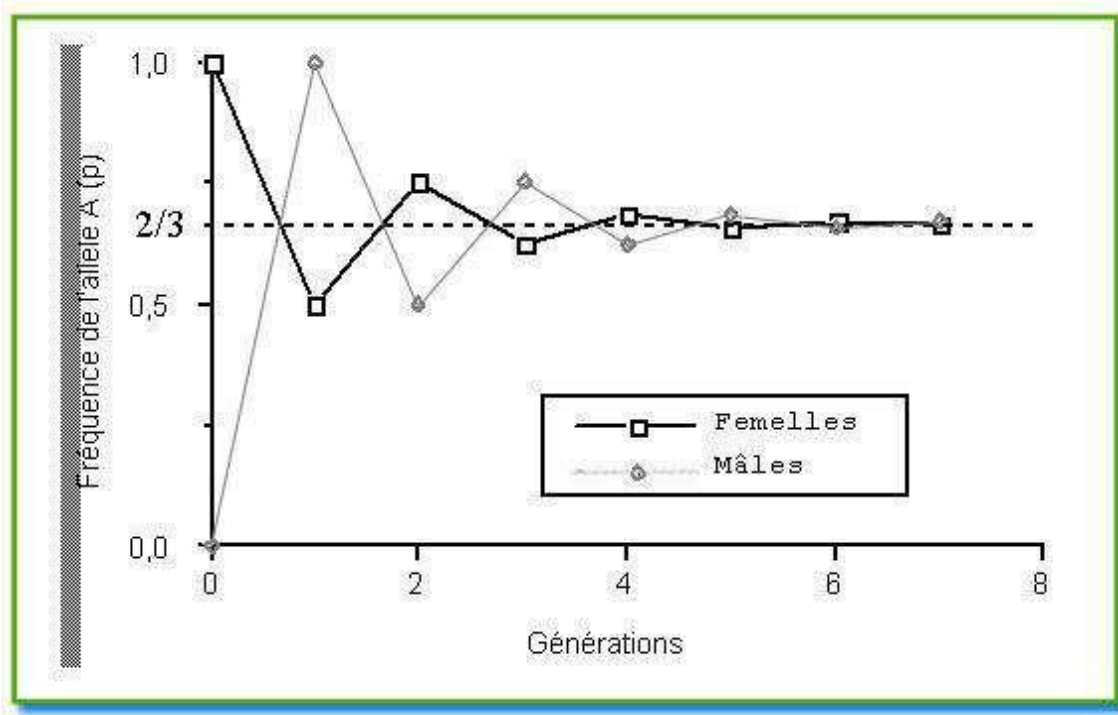
Si à la génération  $G_0$  tous les mâles sont  $X_aY$  ( $p_{m0} = 0$ ) et toutes les femelles sont homozygotes  $XAXA$  ( $p_{f0} = 1$ ), la fréquence de l'allèle A dans la population est  $p_0 = 2/3$ .

A la première génération, la fréquence allélique chez les mâles est  $p_{m1} = p_{f0} = 1$  et chez les femelles :  $p_{f1} = (p_{m0} + p_{f0})/2 = 0,5$ .

A la génération suivante, on a  $p_{m2} = 0,5$  et  $p_{f2} = 0,75$ , etc

Il se produit donc des oscillations des fréquences alléliques dans les deux sexes, en opposition de phase, qui s'amortissent progressivement comme le montre la figure ci- dessous.

Les fréquences alléliques dans l'ensemble de la population restent invariables (ici  $2/3$  pour  $A$ ) et c'est vers cette valeur d'équilibre que convergent les fréquences alléliques des deux sexes. A l'équilibre,  $q_m = q_f = 2/3$ . La structure génétique est alors la suivante : - chez les mâles :  $p$  individus  $XAY$  ( $2/3$ ) et  $q$  individus  $XaY$  (ici  $1/3$ ); - chez les femelles,  $p^2$  individus  $XAXA$  ( $4/9$ ),  $2pq$  individus  $XAXa$  ( $4/9$ ).



Go tous les mâles sont  $XaY$  ( $p_{m0} = 0$ )

Go toutes les femelles sont homozygotes  $XAXA$  ( $p_{f0} = 1$ )

La fréquence de l'allèle  $A$  dans la population est  $p_0 = 2/3$ .

G1 La fréquence allélique chez les mâles est  $p_{m1} = p_{f0} = 1$

G1 les femelles :  $p_{f1} = (p_{m0} + p_{f0})/2 = 0,5$ .

G2 A la génération suivante, on a  $p_{m2} = 0,5$  et  $p_{f2} = 0,75$ , etc



## 7. Transmission de plusieurs gènes et déséquilibre gamétique

Dans les populations naturelles, un très grand nombre de caractères sont génétiquement variables. Il est donc nécessaire d'étudier la transmission simultanée de plusieurs gènes polymorphes. C'est en effet l'ensemble du pool génétique qui se transmet d'une génération à l'autre, c'est à dire pour une population de  $N$  individus diploïdes l'ensemble des  $2N$  exemplaires de chacun des gènes dont certains sont présents sous plusieurs formes alléliques. L'étude de la transmission simultanée des caractères polymorphes se complique rapidement car elle concerne un grand nombre de gènes pouvant ou non être portés par les mêmes chromosomes (gènes physiquement liés) ou interagissant entre eux par des relations d'épistasie. Seule l'étude du cas de deux loci alléliques peut être abordée de façon simple



Chez l'escargot *Cepaea nemoralis*, la variabilité de l'ornementation de la coquille est déterminée par plusieurs gènes dont les allèles ne sont pas toujours associés au hasard. L'allèle responsable de la couleur jaune de la coquille (gène C) est par exemple souvent associé à celui responsable de la présence de bandes (gène B)

L'analyse de la composition génétique d'une population à plusieurs loci permet de mieux comprendre son fonctionnement à la fois au niveau des effectifs, des flux migratoires et des pressions de sélections qui s'exercent sur les caractères concernés.

D'un point de vu appliqué, l'étude dès l'association alléliques à des loci étroitement liés

permet de cartographier les gènes et de diagnostiquer des maladies génétiques à partir de marqueurs moléculaires non impliqués dans la maladie.

### 7.1. Equilibre gamétique à 2 loci

Dans une population théorique idéale, les allèles des différents loci sont associés au hasard donc sont statistiquement indépendants.

Cette indépendance statistique résulte du processus de recombinaison qui est maximum pour des gènes indépendants et d'autant plus faible que les gènes sont proches les uns des autres sur le même chromosome.

La notion d'équilibre intègre donc non seulement l'association au hasard des allèles d'un même gène pour former les génotypes (loi de Hardy-Weinberg) mais également l'association au hasard des allèles de différents gènes.

Cet équilibre à plusieurs loci peut être facilement formalisé dans le cas de 2 gènes à 2 allèles avec :

- Au locus A les allèles A et a de fréquences p et q.
- Au locus B les allèles B et b de fréquences r et s.

### 7.2. L'équilibre gamétique

Correspond à l'association au hasard de tous les allèles au niveau des gamètes, c'est-à-dire des haplotypes. Les fréquences des 4 catégories de gamètes correspondent alors au produit des fréquences des allèles qui forment ces gamètes soit :

Fréquence des gamètes AB = pr

Fréquence des gamètes Ab = ps

Fréquence des gamètes aB = qr

Fréquence des gamètes ab = qs On distingue souvent les associations de type couplage (ou cis), correspondant aux associations AB et ab, des associations de type répulsion (ou trans) Ab et aB. Cette distinction n'est cependant possible que lorsque des relations de dominance existent entre les allèles d'un même locus.

$$f(AB) \times f(ab) = f(Ab) \times f(aB)$$

$$f(aB) \times p_r \times q_s = p_s \times q_r$$

$$p_q r_s = p_q r_s$$

Au niveau des génotypes, cet équilibre se traduit à la fois par une association au hasard des allèles à un même locus et une association au hasard des génotypes des différents loci. Les fréquences génotypiques sont alors :

$$\text{Au locus A } p^2 \text{ AA } 2pq \text{ Aa } q^2$$

$$\text{aa Au locus B } r^2 \text{ BB } 2rs \text{ Bb}$$

$$s^2 \text{ bb}$$

Les fréquences des 9 génotypes possibles correspondent alors aux produits des fréquences génotypiques à chaque locus (indépendance statistique entre génotypes des deux loci) :  $AABB = p^2 r^2$   $AaBB = 2pq r^2$   $aaBB = q^2 r^2$

$$AABb = 2 p^2 r s$$

$$AaBb = 4pqrs$$

$$aaBb = 2 q^2 rs$$

$$AAbb = p^2 s^2$$

$$Aabb = 2pq s^2$$

$$aabb = q^2 s^2$$

**Ces fréquences correspondent également à la fusion au hasard des gamètes qui forment ces génotypes soit par exemple la fusion au hasard de 2 gamètes AB chacun de fréquence  $pr$  pour le génotype donc  $(pr)^2 = p^2 r^2$**