

Génétique Quantitative et Dynamique des Populations

I- Rappels sur la génétique classique et quantitative

1- Génome

- L'ensemble du matériel génétique d'un individu ou d'une espèce codé dans son ADN (Homo sapiens) ou dans son ARN (rétrovirus).

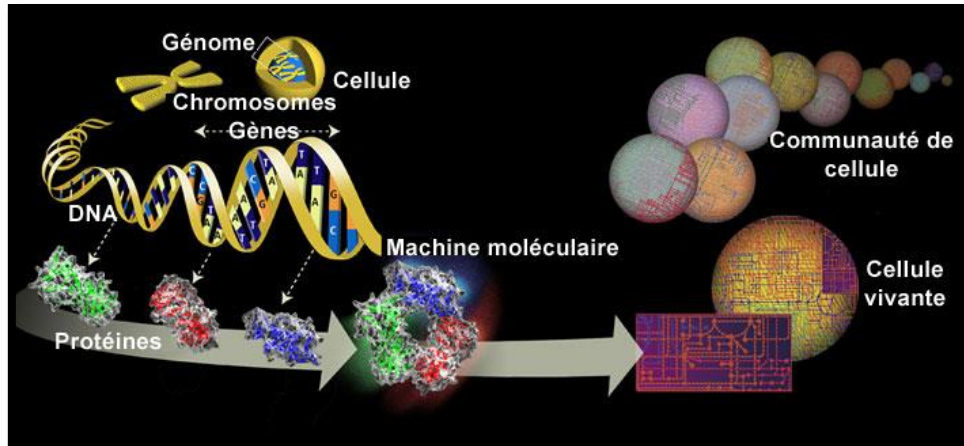


Figure 01. Du génome à la cellule

2- Génotype

- Le génotype est l'ensemble ou une partie donnée de la composition génétique (information génétique) d'un individu. Le génotype d'un individu est donc la composition allélique de tous les gènes d'un individu.
- Un gène est une partie du génome dont la séquence sert à coder une molécule qui intervient dans la vie de la cellule (enzyme, ARNt, ARNr...), il est caractérisé par la fonction qu'il assure.
- Le terme allèle désigne chacune des différentes formes ou versions possibles d'un même gène, relatives au même caractère.
- On peut avoir dans les populations naturelles plusieurs séquences différentes pour un même locus. On parle de différents allèles.
- Ce polymorphisme génétique est dû à l'apparition de mutations qui sont des variations de la séquence nucléotidique.

3- Phénotype

- L'état d'un caractère observable (caractère anatomique, morphologique, moléculaire, physiologique, ou éthologique) chez un organisme vivant.



Figure 02. Polymorphisme du phénotype

4- Relations génotype/phénotype

- Le concept de phénotype est défini par opposition au génotype, l'identité des allèles qui caractérise le génome d'un individu.
- Pour certains traits simples, la correspondance entre le génotype et le phénotype est directe (caractère qualitatif).
- La plupart des caractères (les caractères quantitatifs) dépendent de multiples gènes, et l'influence du milieu (l'environnement) peut être un facteur déterminant. Dans ce cas, le génotype ne permet pas de prévoir précisément le phénotype de l'individu, mais seulement d'estimer sa valeur moyenne.
- La composition allélique de chaque individu pour chaque gène va produire un phénotype particulier à chaque fois. Par exemple, la couleur de la peau peut varier suivant la composition allélique des individus pour les gènes impliqués dans la voie de biosynthèse de la mélanine.

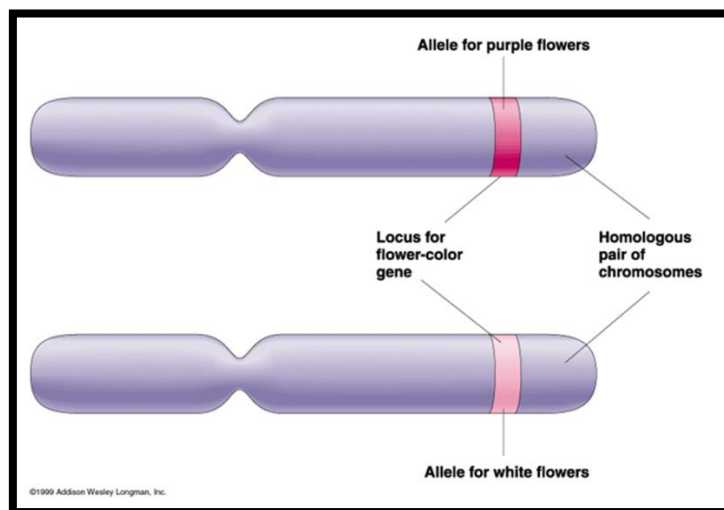


Figure 03. Les Allèles dans les chromosomes homologues

5- Monohybridisme

- Le monohybridisme est le produit de croisement entre deux individus différant par un seul caractère.

- Les différentes formes d'un caractère étant généralement contrôlées par différents allèles d'un même gène, on croise des individus de lignée pure par exemple des individus à fleurs jaunes avec des individus à fleurs bleues et on observe la couleur des fleurs obtenues. On pourra ainsi définir quel est l'allèle dominant et quel est le récessif.
- Lorsque le croisement concerne deux caractères différents, on parle de dihybridisme, et ainsi de suite.

6- Le dihybridisme

- Le dihybridisme est le croisement entre deux individus appartenant à deux lignées homozygotes (pures) qui diffèrent entre elles par deux caractères ou deux gènes, donc deux couples d'allèles localisés soit sur des autosomes (chromosomes non sexuels) soit sur des hétérosomes (chromosomes sexuels).

7- polyhybridisme

- Croisement entre deux parents qui diffèrent par de nombreux caractères héréditaires.
- Le terme polyhybride représente les individus hétérozygotes pour deux caractères (deux gènes) ou plus.
- Les lois fondamentales de la génétique concernent tous les individus qu'ils soient Monohybrides ou polyhybrides.

8- Lois de Mendel

- La génétique classique débute avec les travaux du moine autrichien, Gregor Mendel (1822-1884), qui expérimenta pendant 9 ans sur le Pois (*Pisum sativum*) pour confirmer ses théories de l'hérédité.
- G. Mendel a publié le résultat de ses études en 1866. À cette époque on ignorait tout de la méiose et des chromosomes mais, en proposant l'existence d'unités héréditaires (qui seront appelées gènes, en 1906 par le biologiste danois Wilhem Johannsen), Mendel fonda la Génétique. De cette hypothèse Mendel a tiré un certain nombre de principes connus sous le nom de lois de Mendel et applicables à tout eucaryote ayant une méiose normale.
- La redécouverte des lois de Mendel en 1900, puis leur combinaison avec la découverte des chromosomes, considérés comme le support physique de l'hérédité, est à l'origine de la fondation de la génétique formelle au début du XXe siècle.
- Cette loi résulte du fait qu'il y a ségrégation des chromosomes homologues lors de l'anaphase de la première division méiotique (anaphase I).

Première loi : uniformité des hybrides de première génération

- Les deux allèles d'un gène déterminant un caractère se disjoignent (ségrègent) lors de la formation des gamètes : une moitié des gamètes contient l'un des allèles et l'autre moitié contient l'autre.
- Si l'on croise deux individus d'une même espèce homozygotes relativement à un caractère, tous les descendants de la première génération, qui sont appelés hybrides F1, sont identiques relativement à ce caractère, c'est-à-dire tous hétérozygotes.

Deuxième loi : Loi de disjonction des allèles

- Si les descendants d'un croisement impliquant deux lignées pures différant par un seul caractère présentent tous le même phénotype, ce phénotype (caractère) est qualifié de dominant (l'autre caractère est dit récessif).
- Lorsqu'on croise entre eux, deux des individus de générations F1, on obtient une génération F2 dans laquelle on trouve les deux versions de la couleur des fleurs dans des proportions bien définies : trois descendants à fleurs rouges (1 RR homozygote + 2 Rr hétérozygotes) et un descendant à fleurs blanches (rr homozygote).
- Cette loi est dite « de ségrégation des caractères dans la génération F2 ».

Troisième loi : indépendance de la transmission des caractères

- La ségrégation d'un couple d'allèles est indépendante de celle d'un autre couple d'allèles.
- Cette règle ne s'applique que si les gènes responsables des caractéristiques se situent sur différents chromosomes ou s'ils sont éloignés sur le même chromosome. C'est le partage d'allèles dans des gamètes différents.

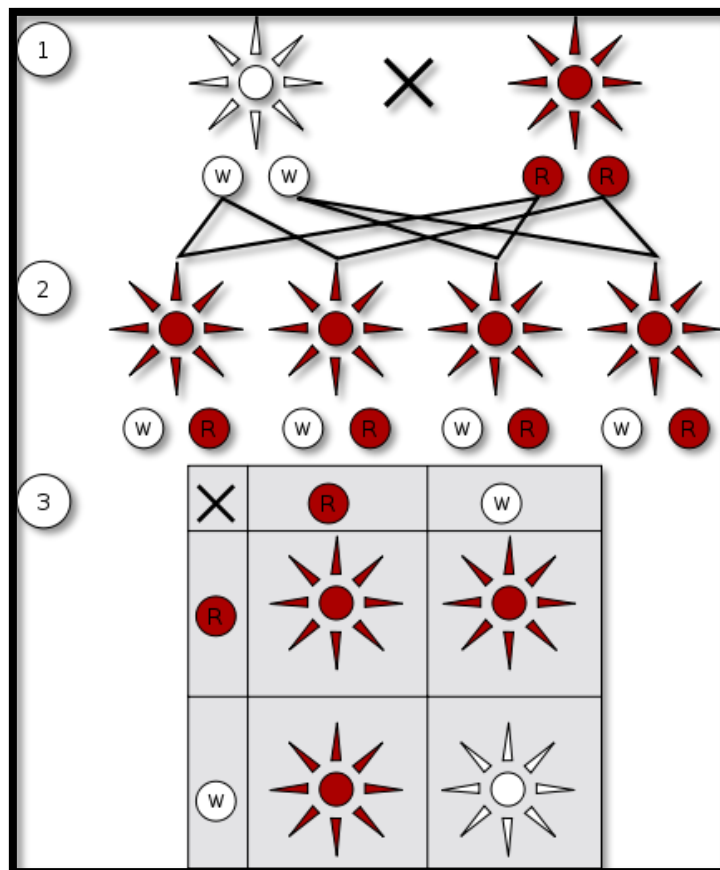


Figure 04. Première et deuxième lois

- 1 - Croisement de pois à fleurs rouges avec des pois à fleurs blanches (tous deux homozygotes pour ce trait, WW et RR).
- 2 - Génération F1 : tous les individus sont rouges car l'allèle rouge est dominant et le blanc est récessif). Alors que les parents étaient tous deux homozygotes (respectivement WW et RR), toute la F1 est hétérozygote (RW).
- 3 - Génération F2 : les formes rouges et blanches montrent un rapport de 3:1.

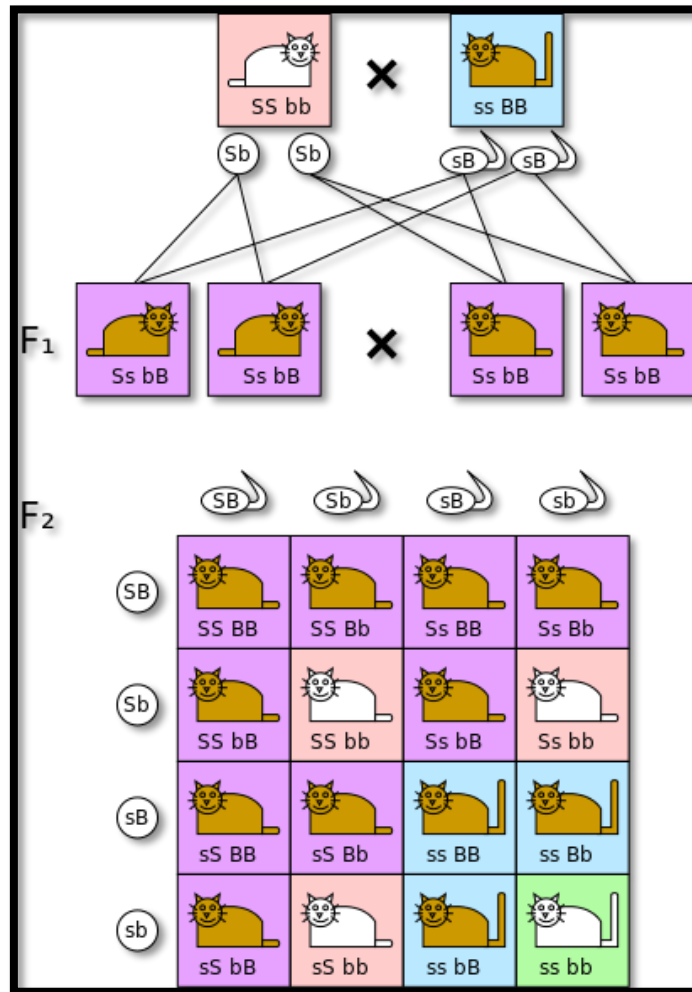


Figure 05. Troisième loi

(L'échiquier de Punnett des deux caractéristiques (poils blancs/bruns, queue courte/longue, où "brun" et "court" devraient être dominants) donne en génération F2 des phénotypes variés dans le rapport de 9:3:3:1. (S = court (short), s = long, B = brun, b = blanc)).

9- La génétique quantitative

- La génétique quantitative est la génétique des caractères qui peuvent donner lieu à des mesures, que ce soient des caractères à variation continue (tels que le poids ou la taille d'un organisme) ou discontinue (à déterminisme complexe), c'est-à-dire résultant de plusieurs facteurs génétiques ou non (On parle également de génétique multifactorielle).
- La génétique quantitative s'appuie sur la génétique des populations et les statistiques.
- Un caractère quantitatif est un caractère mesurable dont la variation est continue. Ce caractère peut donc prendre une infinité de valeurs possibles (par exemple le poids, la longueur...).
- Elle s'intéresse à des caractères à variation continue et à déterminisme complexe, c'est-à-dire gouvernés par plusieurs facteurs génétiques et plusieurs facteurs non génétiques.
- L'expression de ces caractères résulte généralement d'interactions complexes entre de nombreux facteurs de l'environnement et le programme génétique de l'individu, lui-même constitué d'un grand nombre de gènes s'exprimant successivement ou simultanément.

Exemples de caractères quantitatifs

- **Caractères biométriques** : Taille des individus, poids, croissance, Pression artérielle, taux de cholestérol, glycémie, Nombre de soies de l'abdomen de la drosophile.
- **Caractères agronomiques** : Taille de portée chez les animaux, production laitière Teneur en huile chez le Maïs, Nombre de grains par épi de Blé, Date de floraison chez le Blé.
- **Maladies multifactorielles / maladies "monogéniques"** : Diabète Prédilection à l'obésité.
- **Caractères impliqués dans l'adaptation** : Précocité floraison, fertilité, tolérance facteurs du milieu.

La description de la variabilité phénotypique

- La variation ou variabilité phénotypique au sein d'une population peut être décrite par une distribution à partir de la mesure d'un nombre élevé d'individus. Cette distribution doit être relativement voisine d'une distribution théorique normale c'est-à-dire la variation doit suivre une loi normale.

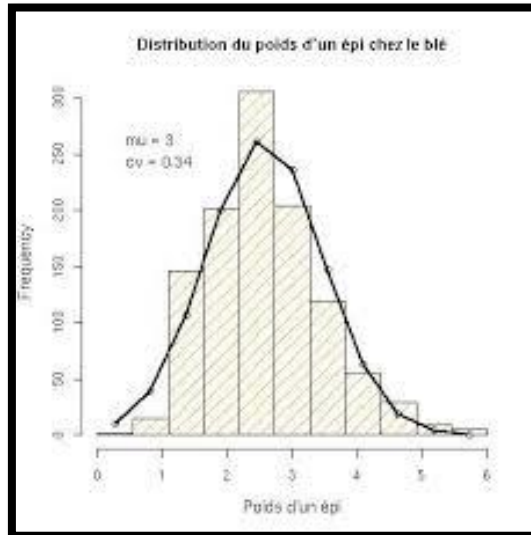


Figure 06. Exemple de caractère quantitatif où la variation suit une loi normale

Effets génétiques et environnementaux

- Le phénotype d'un individu résulte de la combinaison de facteurs génétiques et environnementaux. Pour une analyse génétique du caractère, il est nécessaire, au moins dans un premier temps, de supposer l'additivité de ces facteurs ; autrement dit, un milieu donné agit de la même façon sur des génotypes variés. Cependant l'utilisation de ces résultats dans des milieux différents implique nécessairement de valider ces mêmes résultats dans chaque milieu.
- Dans le modèle polygénique, la valeur phénotypique s'exprime de la façon suivante : $P = G + E$ où P représente la valeur phénotypique, G la valeur génétique et E (environnement) qui est l'effet de milieu contrôlé et/ou non contrôlé.
- La variation phénotypique s'exprime de la façon suivante : $V_P = V_G + V_E + e$ où V_P est la variation phénotypique, V_G la variation génétique et **e** représente l'**épistasie** qui désigne l'interaction existant entre deux ou plusieurs gènes. Il y a épistasie lorsqu'un ou plusieurs gènes

(dominants ou récessifs) masquent ou empêchent l'expression de facteurs situés à d'autres lieux génétiques.

II- Introduction à la génétique des populations

1-Introduction

- L'expérimentation génétique est essentiellement fondée sur des croisements entre souches pures. Au contraire, les populations naturelles sont composées d'organismes génétiquement différents pour un grand nombre de gènes. La génétique des populations se propose d'abord d'évaluer l'importance de cette diversité génétique.
- Au-delà de cette estimation, la génétique des populations se propose aussi de dégager des lois permettant de rendre compte du maintien ou de l'évolution dans l'espace et dans le temps de cette diversité génétique. Partant du postulat que l'évolution des espèces est, en dernière analyse, celle de leur patrimoine génétique, la génétique des populations apparaît alors comme le noyau dur obligatoire de toute théorie de l'évolution.
- Les différences génétiques entre individus, autrement dit le polymorphisme génétique des populations, sont alors conçues comme un patrimoine, un capital adaptatif, un gage de survie de l'espèce aux variations de l'environnement selon le principe darwinien de la « survie du plus apte » ; ce qui suppose évidemment l'existence préalable de telles différences génétiques entre individus. À sa naissance la génétique des populations réalise une sorte de synthèse entre la vision darwinienne de l'évolution et la vision mendélienne de l'hérédité, mais ses développements mathématiques ultérieurs ont débordé le cadre strictement darwinien.
- Les modèles théoriques de la génétique des populations s'attachent à l'étude des mécanismes gouvernant l'évolution de la diversité génétique au niveau d'un gène ou deux, rarement plus ; au-delà, la démarche analytique est pratiquement impossible mais peut être remplacée par les simulations informatiques. C'est pourquoi nous aborderons dans un premier temps la définition et la mesure du polymorphisme génétique au niveau d'un seul gène, et la mise en évidence de la structure génétique d'une population pour ce gène.
- La mesure du polymorphisme pour un grand nombre de gènes permet de fournir une estimation de la diversité génétique globale existante non seulement entre individus d'une même population mais aussi entre populations, notamment chez l'homme.
- L'analyse de la diversité génétique entre populations humaines permet d'aborder le sujet si controversé de la définition des races. Elle permet aussi d'entreprendre une analyse phylogénétique des populations humaines dont le but est la mise en évidence de l'âge et de l'origine de l'homme à travers une recherche, dans les gènes, de l'histoire du peuplement de la planète.

2- Les différents types de polymorphismes utiles en génétique des populations

- À tout moment, une mutation peut modifier le génome contenu dans le noyau d'une cellule. Si une mutation affecte le génome d'une cellule somatique, un sous-clone muté se développera chez l'individu affecté (par exemple une tumeur). Une mutation somatique n'est pas transmissible, au contraire d'une mutation survenant dans la lignée germinale peut être

transmise à la descendance par les quelques gamètes qui en sont porteurs. Dans ce cas le patrimoine génétique de la population se trouve « enrichi » puisque sa diversité est accrue.

2.1. Le polymorphisme génique

- Les mutations géniques affectent un gène dans l'une de ses séquences et peuvent affecter sa fonction et retentir sur le ou les phénotypes des caractères dans lequel ce gène est impliqué. Selon sa nature (changement, perte ou insertion d'une ou plusieurs paires de bases) et son site dans le gène, une mutation peut se révéler être :
 - a) **Sur le plan fonctionnel :**
 - ✓ **Une mutation de perte de fonction**, si elle entraîne la diminution ou l'absence de produit (mutation quantitative) ou la présence d'un produit moins actif ou inactif (mutations qualitatives) ;
 - ✓ **Une mutation de gain de fonction** si elle entraîne une surproduction (mutations quantitatives), ou la présence d'un produit plus actif, ou d'un produit doué d'une propriété nouvelle, absente du produit « sauvage » (mutations qualitatives), éventuellement toxique dans le cas de certaines maladies dominantes.
 - b) **Sur le plan phénotypique :**

L'effet d'une mutation peut se révéler :

 - ✓ « **Dominant** » vis-à-vis de l'effet de l'allèle sauvage si l'hétérozygote pour ces deux allèles est de phénotype muté ;
 - ✓ « **Récessif** » vis-à-vis de celui de l'allèle sauvage si l'hétérozygote pour ces deux allèles est de phénotype sauvage et que le phénotype muté n'est observable que chez les porteurs de deux allèles mutés du gène.

2.2. Le polymorphisme chromosomique

- Des mutations peuvent survenir à une autre échelle au sein du génome et affecter le nombre ou la structure des chromosomes.
- À petite échelle, celle des individus, ces mutations sont souvent associées à des pathologies et sont le plus souvent éliminées par la sélection.
- Mais à grande échelle, celle de l'évolution, on sait désormais qu'elles ont joué un rôle important en remplaçant les uns par rapport aux autres des blocs de gènes affectant le développement. Les cytogénéticiens ont depuis longtemps précisé le nombre et l'ampleur des remaniements chromosomiques existant entre le caryotype à 46 chromosomes de l'homme et celui à 48 du chimpanzé.
 - a) **Les mutations chromosomiques** sont des modifications de la structure des chromosomes sans que le nombre en soit changé; elles correspondent à des translocations de fragments chromosomiques, des inversions, des fusions centriques, des délétions ou des duplications. Il ne faut pas croire qu'elles sont rares.
 - b) **Les mutations génomiques** ne concernent plus un fragment de chromosome mais un ou plusieurs chromosomes dans leur totalité ; ce sont les changements de ploïdie par perte d'un lot haploïde de chromosomes (monoploïdie) ou par gain d'un ou plusieurs lots (tri-, tétraploïdie), et les aneuploïdies par perte ou gain d'un ou plusieurs chromosomes

(monosomie si un chromosome est absent ; trisomie quand il y a un chromosome surnuméraire).

- Ces mutations sont souvent pathologiques et létales car elles perturbent profondément le déroulement du programme génétique, notamment lors du développement embryonnaire. Mais leur rôle a peut-être été essentiel lors de certaines étapes de l'évolution (dans le cas de la macroévolution par exemple).

3- Les différents types et niveaux de perception du polymorphisme génétique

- Les premières études de génétique ont porté sur des caractères morphologiques (taille, forme, couleur des organismes ou de certaines de leurs parties, yeux, ailes, tige, fleur, graine...). Chez l'homme, on s'intéressa assez vite aux phénotypes cliniques (atteint ou non atteint) associés aux maladies héréditaires. Tous ces phénotypes, même quand ils sont gouvernés par les allèles d'un seul gène, ne sont qu'une conséquence indirecte, physiologique ou cellulaire, en aval de l'effet primaire de ces allèles, ou de ce qu'on nomme, pour les maladies génétiques, les mutations pathogènes.
- Le développement de la biochimie des protéines, de l'enzymologie et de l'immunologie, puis de la biologie moléculaire du gène, a permis d'avoir accès soit au produit du gène, la chaîne peptidique, soit à la séquence du gène lui-même.
- Aussi dans bien des cas, l'analyse biochimique des protéines ou moléculaire du gène donne accès à des polymorphismes sans traduction perceptible à l'échelle morphologique ou physiologique et permet de percevoir, à cette échelle, un polymorphisme génétique non perceptible à l'échelle de l'organisme.

REMARQUES

- Le développement considérable de la génétique et de ses applications dans nombre de domaines de la biologie ou la médecine, l'a parfois transformée, notamment chez le grand public et les médias, en sciences des « certitudes ». C'est par exemple le sens implicitement donné à des formules comme le « programme génétique » ou le fait qu'un caractère soit « génétiquement déterminé ». Or, c'est oublier que l'expression des gènes ne saurait être conçue comme indépendante du milieu, de l'environnement au sein duquel ces gènes s'expriment.
- Quelques exemples sont utiles pour bien rappeler et comprendre que la diversité entre individus d'une même espèce ou population, peut résulter de leur diversité génétique mais aussi de la diversité des milieux au sein desquels ils se sont construits (La nervation de l'aile de drosophile).

4- Mesure de la diversité génétique et composition génétique d'une population

- L'espèce est par définition un groupe génétiquement fermé au sein duquel les organismes sont susceptibles, par l'alternance méiose-fécondation de séparer ou de réunir les divers allèles de chacun des gènes et de concevoir des combinaisons génétiques nouvelles par la recombinaison génétique.
- Cependant tous les individus d'une même espèce, s'ils sont potentiellement susceptibles de réaliser ce brassage, peuvent en être pratiquement empêchés quand des barrières limitent les possibilités de croisements entre certains individus.

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

4.1. Mesure de la diversité génétique de phénotypes codominants (groupe sanguin MN)

- Le typage est réalisé, comme pour le groupe ABO ou rhésus, par un test d'hémagglutination, les individus appartenant aux groupes [M], [N] ou [MN] selon que leurs hématies sont respectivement reconnues par l'anticorps anti-M ou l'anticorps anti-N ou les deux, parce qu'elles sont porteuses d'une chaîne peptidique spécifiée par l'allèle L^M et/ou l'allèle L^N .

Groupe sanguin	[M]	[MN]	[N]
Effectifs observés (total : 1000)	350	500	150
Fréquences phénotypiques	$350/1000 = 0.35$	$500/1000 = 0.5$	$150/1000 = 0.15$
Génotypes	L^M/L^M	L^M/L^N	L^N/L^N
Fréquences génotypiques	D	H	R

Tableau 01. Les échantillons aléatoires d'une population (système MN)

4.2. Mesure de la diversité génétique de phénotypes dominants et récessifs (groupe sanguin ABO)

- Il est toujours possible d'estimer les fréquences phénotypiques (tableau) mais ici il est impossible d'en déduire les fréquences génotypiques, du moins pour les phénotypes dominants présentant plusieurs possibilités génotypiques ; dans ce cas la fréquence du phénotype, par exemple celle de [A], est égale à la somme des fréquences des deux génotypes, I^A/I^A et I^A/I^O , mais on ne peut connaître la valeur individuelle de la fréquence de chaque génotype, indispensable pour estimer les fréquences alléliques selon les formules vues précédemment.

Groupe sanguin	[A]	[B]	[AB]	[O]
Effectifs observés (total:500)	213	56	13	216
Fréquences phénotypiques	$213/500 = \mathbf{0.426}$	$56/500 = \mathbf{0.112}$	$13/500 = \mathbf{0.03}$	$216/500 = \mathbf{0.432}$
Génotypes	I^A/I^A ou I^A/I^O	I^B/I^B ou I^B/I^O	I^A/I^B	I^O/I^O

Tableau 02. Les échantillons aléatoires d'une population (système ABO)

5- Degré de polymorphisme et d'hétérozygotie

- On peut établir une mesure plus globale de la diversité génétique d'une population en intégrant sur un grand nombre de gènes, la diversité allélique établie pour chacun d'entre eux. Cette diversité globale peut être appréhendée par le « degré de polymorphisme » et le « degré d'hétérozygotie ».

- Connaître la constitution génétique d'une population pour un seul gène polymorphe ne permet pas de savoir si cette population est ou n'est pas très polymorphe. Elle l'est si un grand nombre de gènes sont polymorphes ; ce qui suppose, en pratique d'avoir étudié un grand nombre d'individus pour un grand nombre de gènes.
- Si on étudie un très grand nombre d'individus pour un gène, on trouvera toujours un ou quelques allèles très rares correspondant notamment au bruit de fond des mutations *de novo*, dont la plupart, disparaissent en quelques générations du pool génique. Sur cette base un gène polymorphe est un gène qui présente au moins deux formes alléliques de fréquence supérieure ou égale à 1 %, et le degré de polymorphisme d'une population est défini par son pourcentage de gènes polymorphes.
- Or si un gène polymorphe présente un allèle A1 très fréquent, à 99 % et un allèle rare A2, à 1 %, la très grande majorité des individus sera constituée d'homozygote A1/A1 et une minorité sera hétérozygote A1/A2 (les individus A2/A2 étant sans doute exceptionnels, comme on le verra plus tard grâce à la relation de Hardy-Weinberg). Par contre si un gène polymorphe présente deux allèles B1 et B2 de fréquences sensiblement égales (1/2), un grand nombre d'individus sera hétérozygote.
- Dans les deux cas précédents, on obtient pour l'hétérozygotie des valeurs respectives de 1,98 % et 50 %, ce qui montre bien que la population n'est pas génétiquement dans la même situation de diversité pour ces deux gènes di-alléliques.

III- Le modèle général d'Hardy-Weinberg

- Le modèle de Hardy-Weinberg repose sur plusieurs conditions qu'il sera utile d'introduire en considérant le cycle vital d'un organisme à sexes séparés, entre les stades adultes reproducteurs de deux générations successives (figure04). Les étapes critiques du cycle vital pour lesquelles chaque condition du modèle de Hardy-Weinberg sera introduite, ont été numérotées sur la figure suivante :

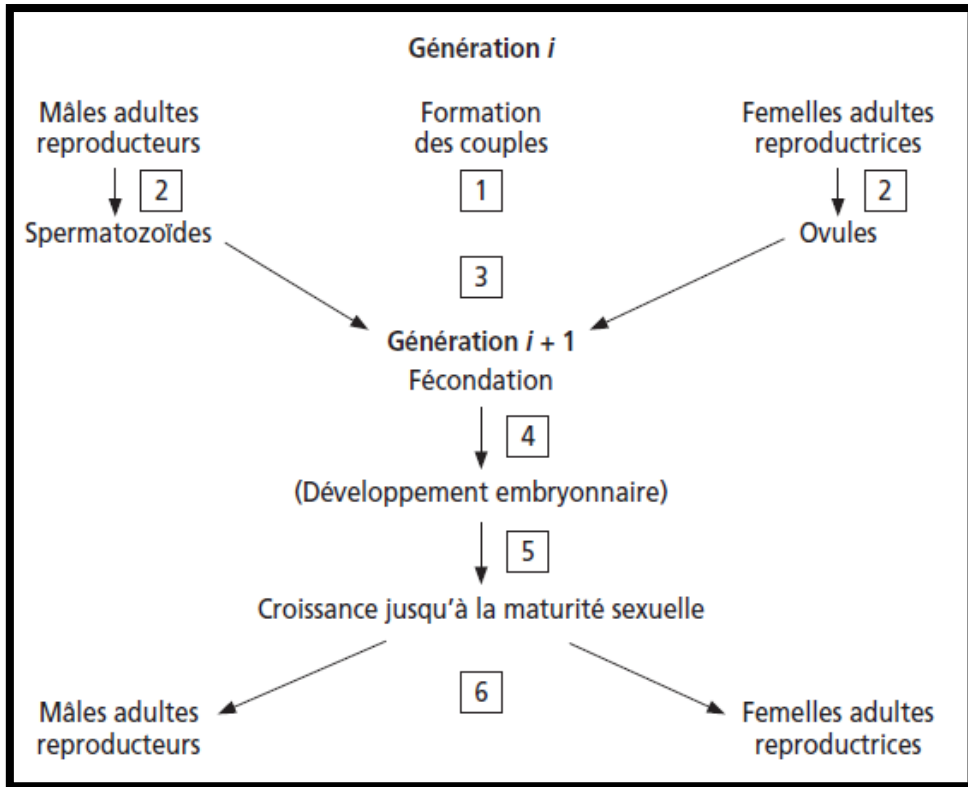


Figure 07. Étapes du cycle vital où sont introduites des conditions du modèle de Hardy-Weinberg.

1- Le modèle d'HARDY-WEINBERG

- Le modèle théorique général de Hardy-Weinberg sera établi dans le cas le plus simple, le plus général et le plus utile d'un gène autosomique di-allélique, pour une population d'organismes à sexes séparés, présentant des générations séparées.
- Considérons une population constituée des adultes reproducteurs de la génération i. Dans le cas d'un gène di-allélique (allèles A1 et A2), la composition génétique de la population est constituée des :
 - ✓ Trois génotypes possibles : A1/A1 A1/A2 A2/A2
 - ✓ De fréquences génotypiques : D H R
- Ces fréquences D, H et R sont quelconques et seront considérées, en première analyse, égales dans les deux sexes et on sait en déduire les fréquences des allèles A1 et A2, respectivement nommées p et q, soit :

$$p = D + H/2 \quad \text{et} \quad q = R + H/2$$

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

1-1 Établissement du modèle de Hardy-Weinberg par le cycle vital

A- Condition de panmixie

- La première étape du cycle vital (1 dans la figure) est la formation de couples reproducteurs. Des règles d'union peuvent exister : on fera l'hypothèse qu'ils se forment au hasard, les couples sont dits panmictiques (condition de panmixie). Dans ce cas, on peut générer six types possibles de couples (tableau).

B- Condition d'effectif infini de la population

- On sait que la fréquence d'un évènement est égale à sa probabilité si le nombre de tirages est très grand (loi des grands nombres) ; par exemple la fréquence des « piles » peut être égale à 0,7 sur dix tirages mais ne peut, sur 100 000 tirages, s'écarter notablement de sa probabilité égale à 0,5.
- Afin de pouvoir considérer que les fréquences des couples ou des génotypes sont égales à leurs probabilités respectives, nous considérerons que la population est de taille infinie (concrètement suffisamment grande pour y appliquer la loi des grands nombres aux évènements étudiés).
- Cette condition d'effectif infini s'ajoute à la condition de panmixie. Le tableau de formation des couples et de leurs descendants se présente alors ainsi :

Types de couples	Fréquences des couples	Fréquences des descendants A1/A1	Fréquences des descendants A1/A2	Fréquences des descendants A2/A2
A1/A1 x A1/A1	D ²	1	0	0
A1/A1 x A1/A2	2D H	1/2	1/2	0
A1/A1 x A2/A2	2D R	0	1	0
A1/A2 x A1/A2	H ²	1/4	1/2	1/4
A1/A2 x A2/A2	2H R	0	1/2	1/2
A2/A2 x A2/A2	R ²	0	0	1
Total	1	D ² +DH+H ² /4	DH+2DR+H ² /2+HR	H ² /4+HR+R ²

Tableau 03. Fréquences des couples et de leurs descendants.

C- Condition d'absence de mutations

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

D- Condition d'absence de sélection gamétique

- Les deux types de gamètes formés par les hétérozygotes sont théoriquement dans les proportions 50 :50, mais lors de la fécondation (4 dans le cycle vital), ces proportions peuvent avoir été modifiées si un mécanisme de sélection gamétique (3 dans le cycle vital) a joué en faveur de l'un et en défaveur de l'autre.

E- Condition d'absence de sélection zygotique

- Les proportions des génotypes chez les descendants, à la fécondation, ne sont maintenues, au stade final de reproducteur adulte, que si une nouvelle condition est réalisée : l'absence de sélection zygotique (pendant l'embryogenèse ou la croissance ; 5 dans le cycle vital).
- Autrement dit les trois génotypes sont supposés avoir la même espérance de vie et ne présentent, du moins pour le gène considéré, aucune mortalité différentielle.

F- condition d'absence de sélection et de migration

- On obtient ainsi, avec les conditions citées ci-dessus, les trois sommes figurant en bas du tableau, mais cela suppose encore deux dernières conditions pour qu'on puisse les considérer comme les fréquences génotypiques des adultes reproducteurs de la génération $i+1$:
 - ✓ Il faut d'abord considérer que les couples sont également fertiles (condition d'absence de sélection en matière de fertilité différentielle).
 - ✓ Il faut ensuite considérer que ces sommes, ces fréquences génotypiques, n'ont pas été modifiées dans le temps par l'adjonction à la population d'individus extérieurs (condition d'absence de migrations).
- Sous toutes ces hypothèses, les nouvelles fréquences génotypiques deviennent pour les trois génotypes :

	A1/A1	A1/A2	A2/A2
égales à :	p^2 (D)	$2pq$ (H)	q^2 (R)

1-2 Établissement du modèle de Hardy-Weinberg par le schéma de l'urne gamétique

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

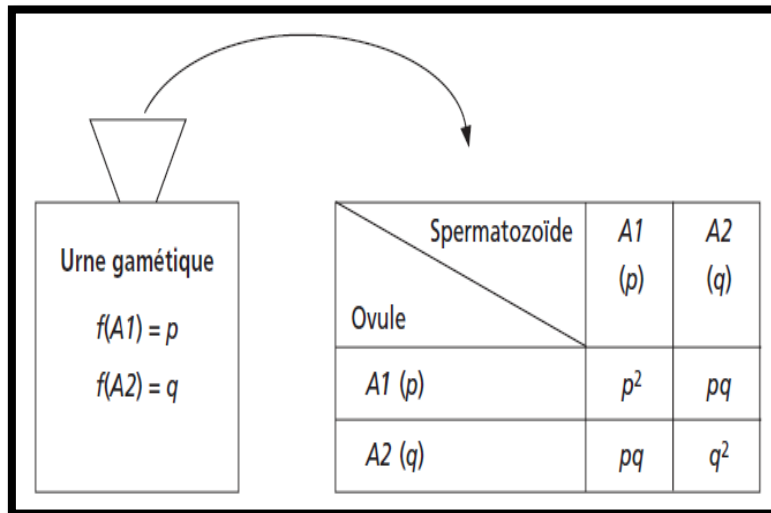


Figure 08. Schéma de l'urne gamétique et conséquences mathématiques sur la diversité génétique.

1-3 Bilan du modèle de Hardy-Weinberg

- Les nouvelles fréquences génotypiques correspondent soit au carré des fréquences alléliques pour les homozygotes, soit au double produit des fréquences alléliques pour l'hétérozygote.
 - ✓ Les trois génotypes sont : A1/A1 A1/A2 A2/A2
 - ✓ Leurs fréquences sont égales à : p^2 (D) $2pq$ (H) q^2 (R)
- La relation ainsi établie entre les fréquences alléliques et les fréquences génotypiques est appelée « relation de Hardy-Weinberg » ou « relation panmictique » car elle découle directement de l'hypothèse panmictique.

Remarque : Cette relation mathématique permet, en supposant que l'hypothèse panmictique soit valide, de remonter aux fréquences génotypiques (donc phénotypiques) quand on ne connaît que les fréquences alléliques.

Exercice : Calculez la composition génétique des différents groupes du système ABO.

- Les fréquences alléliques sont inchangées à la génération suivante. En effet, selon la formule de calcul des fréquences alléliques à partir des fréquences génotypiques, on a :
 - ✓ $f(A1) = D + H/2 = p^2 + 2pq/2 = p^2 + pq = p(p + q) = p$
 - ✓ $f(A2) = R + H/2 = q^2 + 2pq/2 = q^2 + pq = q(q + p) = q$
- Cette stabilité de la composition génétique de la population, au fil des générations, est appelée « équilibre de Hardy-Weinberg », avec les conditions déjà citées et qu'on regroupe comme suit :

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

- Le modèle de Hardy-Weinberg est le modèle central de la génétique des populations.
- Pourtant le nombre et l'importance des conditions sous-jacentes devraient le faire apparaître comme un modèle théorique, abstrait et irréaliste. On sait bien que les mutations existent et sont la source de la variation génétique ayant permis la sélection et l'évolution. Pourtant l'étude de la plupart des gènes dans les populations naturelles donne des résultats compatibles avec ce modèle. Comment expliquer ce paradoxe ?
- Tout simplement par le fait que certaines des conditions peuvent parfaitement être réalisées et que les autres, bien qu'illégitimes, n'ont d'effet perceptible que sur une longue échelle de temps. On peut donc les négliger sur l'espace de quelques générations.
- Le modèle de Hardy-Weinberg tel qu'il est défini permet de conclure qu'une population panmictique, de grande taille, sans mutations, ni sélection, ni migration, maintient son polymorphisme génétique en l'état.
- En effet les fréquences alléliques, p et q à la génération i , n'ont pas varié à la génération suivante. La diversité en termes de fréquences alléliques est donc stable. Par contre les fréquences génotypiques D , H et R , choisies quelconques à la génération i , ont varié pour prendre des valeurs particulières p^2 , $2pq$, q^2 qui représentent $(p + q)^2$ le développement du carré de la somme des fréquences alléliques.
- Cette relation entre les fréquences alléliques p et q d'une part et les fréquences génotypiques p^2 , $2pq$, q^2 d'autre part, est une conséquence directe de la panmixie qui revient à réaliser les fécondations comme deux tirages aléatoires indépendants dans une urne.
- Cette relation de Hardy-Weinberg est très utile car elle va permettre d'estimer les fréquences alléliques dans l'étude d'un gène présentant, selon les allèles, des phénotypes dominants ou récessifs.
- Dès que les fréquences génotypiques ont les valeurs p^2 , $2pq$, q^2 , à la génération $i + 1$, elles restent inchangées dans les générations ultérieures tant que les conditions de Hardy-Weinberg sont maintenues. En effet, l'urne gamétique $(p + q)$ restant elle-même inchangée, les fréquences génotypiques qui en sont issues, par panmixie, demeurent égales à p^2 , $2pq$, q^2 .
- Cette situation d'équilibre des fréquences alléliques et des fréquences génotypiques appelée « équilibre de Hardy-Weinberg » est atteinte en une génération pour un gène autosomique.

1-4 Établissement de l'équilibre quand les fréquences alléliques diffèrent entre sexes :

- On avait supposé, dans l'établissement du modèle, que les fréquences alléliques étaient les mêmes dans les deux sexes. Si tel n'est pas le cas, il est facile de voir, en reprenant le schéma de l'urne gamétique, qu'il faut une première génération de panmixie pour obtenir des fréquences alléliques égales dans les deux sexes, puis une deuxième évidemment, pour obtenir les fréquences génotypiques de l'équilibre de Hardy-Weinberg.
- En effet, supposons qu'à la génération g_0 , la composition génétique d'une population soit différente dans chacun des sexes, ce qui peut arriver lors d'une fusion de population, on aura:

Sexe	Femelle			Mâle		
Génotypes	A1/A1	A2/A1	A2/A2	A1/A1	A2/A1	A2/A2
Fréquences	D	H	R	d	h	r
Fréquences alléliques	$p = D + H/2$ $q = R + H/2$			$u = d + h/2$ $v = r + h/2$		

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

Génotypes (mâle ou femelle)	A1/A1	A1/A2	A2/A2
Fréquences	$p u$	$p v + q u$	$q v$

Fréquences alléliques $P = p u + (p v + q u)/2$

$Q = q v + (p v + q u)/2$

- Les fréquences alléliques ont pris une valeur différente de celles de la génération précédente, mais égales dans les deux sexes, qui ont désormais la même urne gamétique (P et Q).
- Par contre les fréquences génotypiques ne vérifient pas la relation de Hardy-Weinberg ($p.u$ n'est pas égal à P^2 , ni $q.v$ à Q^2). Mais dès la génération suivante, g_2 , les fréquences génotypiques vérifieront cette relation en étant égales à P^2 , $2PQ$ et Q^2 , si les conditions de Hardy-Weinberg ont été maintenues.

2- Estimation des fréquences alléliques d'un gène responsable de l'albinisme :

- Il existe plusieurs formes d'albinisme dont chacune est dépendante d'un gène. La forme la plus fréquente résulte d'une mutation touchant le gène de structure de la tyrosinase, la première enzyme de la chaîne de biosynthèse des mélanines. Les individus albinos sont déficients en tyrosinase et sont porteurs de deux copies mutées du gène ; ils sont homozygotes a/a . Ce phénotype est récessif car les hétérozygotes A/a , ayant une copie fonctionnelle A du gène, sont capables d'assurer la synthèse des mélanines.
- On observe, dans une population africaine, un phénotype albinos (par déficience de tyrosine) pour 10 000 habitants.

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

- Le tableau permet de voir que l'estimation directe des fréquences alléliques est impossible, car on ne peut estimer indépendamment les trois fréquences génotypiques, D, H et R (on ne peut estimer que D + H d'une part, et R d'autre part : 2ième ligne du tableau).

- Cependant, si on fait l'hypothèse que la population obéit aux conditions de l'équilibre de Hardy-Weinberg, il est possible, par la relation de Hardy-Weinberg, d'associer les fréquences des trois génotypes présents dans la population aux fréquences alléliques p , pour l'allèle fonctionnel A , et q , pour l'allèle muté a (3^{ème} et 4^{ème} lignes du tableau).
- Dans l'exemple, la fréquence des hétérozygotes, $2pq$, est égale à $198/10\ 000$, et proche de la valeur simplifiée $2q$ soit 1 habitant sur 50. Cette fréquence peut paraître considérablement élevée en regard de celle des albinos. Il faut bien comprendre, qu'en régime panmictique, la fréquence des hétérozygotes est souvent supérieure à celle des homozygotes. Quand un allèle est très rare (q proche de zéro), la plupart des exemplaires de cet allèle sont présents dans des individus hétérozygotes et très peu dans des homozygotes.
- Bien évidemment cette situation n'a pas la même conséquence selon que l'allèle rare est récessif ou dominant. Dans le premier cas, l'effet de l'allèle récessif n'est perceptible que chez les homozygotes, et le phénotype associé sera beaucoup plus rare que la fréquence des hétérozygotes. Dans le deuxième cas, au contraire, l'effet de l'allèle dominant est perceptible chez les hétérozygotes et les homozygotes ; comme ceux-ci sont très rares, le phénotype associé aura la fréquence des hétérozygotes.

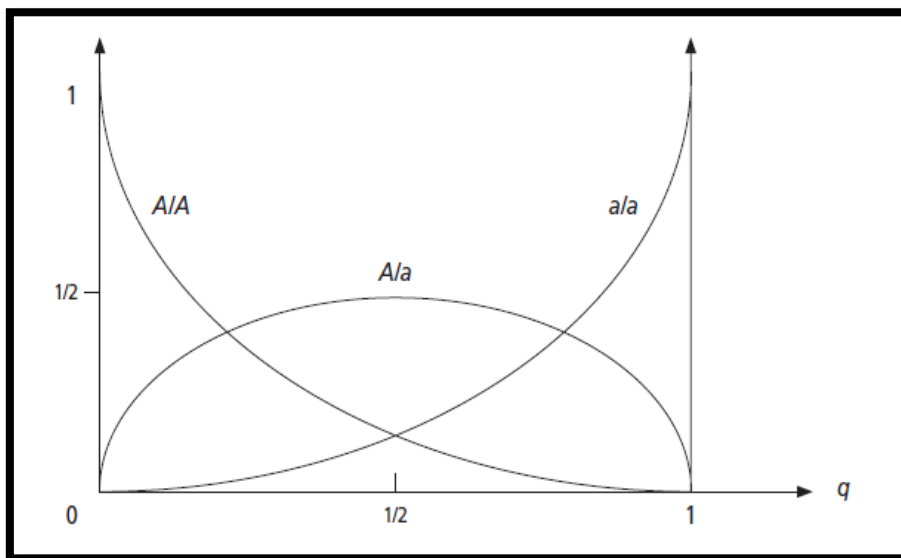


Figure 09. Fréquences des différents génotypes en fonction de la valeur de q [équations $p^2 = (1-q)^2$, $2pq = 2q(1-q)$ et q^2].

3- Tests statistiques de vérification de la conformité au modèle d'Hardy-Weinberg :

- Il est utile de savoir confronter les observations faites dans une population naturelle, en supposant qu'elles sont représentatives de sa composition génétique, aux valeurs théoriques attendues sous l'hypothèse que cette population est à l'équilibre de Hardy-Weinberg, pour le gène considéré. Cette démarche, prend la forme d'un test statistique d'hypothèse utilisant la distribution d'une variable de χ^2 .
- Reprenons l'exemple du groupe sanguin MN dans un échantillon d'une population européenne (voir le tableau) :

Groupes sanguins	[M]	[MN]	[N]
Effectifs observés (total : 1 000)	350	500	150
Fréquences phénotypiques	350/1000= 0.35	500/1000= 0.5	150/1000= 0.15

- Les phénotypes étant codominants, les fréquences génotypiques sont égales aux fréquences phénotypiques ; de ce fait les fréquences alléliques sont directement accessibles, donc :
- Si cette population est panmictique et à l'équilibre de Hardy-Weinberg, on doit s'attendre à ce que les fréquences génotypiques ne soient pas très différentes, aux variations d'échantillonnage près, des valeurs p^2 , $2pq$ et q^2 , soit $0,6^2$, $2 \times 0,6 \times 0,4$ et $0,4^2$.
- Sous l'hypothèse de Hardy-Weinberg, un échantillon aléatoire théorique de 1 000 individus aura une composition génétique où l'effectif de chacun des trois génotypes sera égal à sa fréquence respective (p^2 , $2pq$ et q^2) multipliée par 1 000, soit :
- La question que se propose de résoudre le test statistique, sera de savoir si on peut considérer que les différences entre effectifs observés et effectifs attendus (on dit aussi théoriques) peuvent ou ne peuvent pas être considérées comme résultant du seul hasard d'échantillonnage. Si la réponse du test à cette question est « oui », l'hypothèse théorique de Hardy-Weinberg sera acceptée ; dans le cas contraire, elle sera rejetée.
- Pour cela, dans chaque classe i , on calcule le carré de l'écart entre effectifs observé (o_i) et théorique (t_i), que l'on rapporte à l'effectif théorique t_i , soit : $(o_i - t_i)^2/t_i$

Donc, on obtient dans notre exemple :

Groupes sanguins	[M]	[MN]	[N]
Effectifs observés (total : 1 000)	350	500	150
Effectifs théoriques (total : 1 000)	$p^2 \times 1000 = 360$	$2pq \times 1000 = 480$	$q^2 \times 1000 = 160$
$(o_i - t_i)^2/t_i$	0.277	0.833	0.625
$\Sigma (o_i - t_i)^2/t_i$	1.736		

La variable définie comme la somme de ces écarts $\Sigma (o_i - t_i)^2/t_i$ suit une loi de χ^2 .

- La fonction χ^2 est une somme de carrés et peut prendre des valeurs quelconques dans le domaine de variation $[0 ; + \infty]$.
- Les mathématiciens ont étudié la « fonction de densité de probabilité du χ^2 », qui permet de calculer la probabilité p avec laquelle un χ^2 peut voir sa valeur observée, atteindre ou dépasser telle ou telle valeur désignée par $\chi^2_{(p)}$.

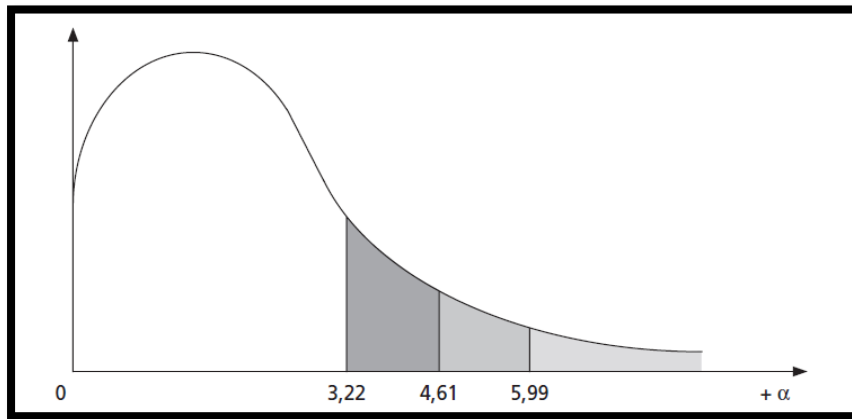


Figure 10. Graphe d'un χ^2 à deux degrés de liberté.

- Des tables permettent de déterminer pour un χ^2 , en fonction de son nombre de degré de liberté (voir plus loin), les valeurs $\chi^2_{(p)}$ de son domaine de variation $[0 ; +\infty]$, qui peuvent être atteintes ou dépassées, par le hasard d'échantillonnage, avec une probabilité p égale à 50 %, 40 %, 30 %, 20 %, 10 %, 5 %, 2.5 %, ou 1 %. Ces valeurs sont appelées seuil au risque p %.
- Le test consiste à prendre une décision qui est, soit d'accepter l'hypothèse en considérant que les écarts résultent du hasard d'échantillonnage, soit de rejeter l'hypothèse en considérant que les écarts sont trop importants (la valeur observée du χ^2 est trop élevée) pour qu'ils puissent être dus à un hasard d'échantillonnage.
- Dans l'exemple ci-dessus le degré de liberté est égal à 3 moins 1 (pour la somme des effectifs théoriques) moins 1 (pour l'estimation de p) soit 1 degré de liberté ($ddl = 1$).
- Dans ces conditions la valeur seuil du χ^2 qui n'est dépassée que 5 fois sur 100, est égale à 3,84 (pour un χ^2 à deux degrés de liberté, cette valeur seuil est égale à 5,99).
- Dans notre exemple la valeur observée du χ^2 est de 1,73. Comme elle est inférieure à la valeur seuil 3,84, pour laquelle le risque d'erreur est de 5 %, on accepte l'hypothèse : la population étudiée est panmictique et peut être considérée comme à l'équilibre de Hardy-Weinberg.