

Institut Agronomique et Vétérinaire Hassan II

GENETIQUE DES POPULATIONS

Cours de 1^{ère} année Ingénieur

Formation en Agronomie

Prof. Ouafae Benlhabib

GENETIQUE DES POPULATIONS

La *génétique* est une science fondamentale de la Biologie dont les principes ont été découverts par Mendel en 1865. Elle étudie l'hérédité des caractères transmissibles. La *génétiques des populations* étudie le comportement des gènes et le dynamisme que subissent les populations. Elle a pour but de caractériser la structure génétique au sein d'un groupe d'individus appartenant à une même population et d'étudier l'évolution de celui-ci au cours des générations par des modèles élaborés.

La génétique des populations s'est développée suite à la volonté de concilier entre la théorie darwinienne de l'évolution et les données acquises transmission du matériel héréditaire au début du 20^{ème} siècle. Cette conciliation a été réalisée d'abord par le mathématicien Fisher R.A. 1929 (*The genetical theory of natural selection*), et les biologistes Wright S. 1931 (*Evolution in the mendelian populations*) et Haldane 1932 (*The causes of evolution*).

Les théories en **génétique des populations** ont précédé les constatations expérimentales. La population est considérée comme un groupe d'individus ou d'organismes diploïdes, vivant dans un espace limité. Les individus du groupe sont susceptibles de s'unir entre eux et d'engendrer une descendance. Le raisonnement les considère à leur naissance ou lors de leur reproduction. Chaque individu de la population a un certain potentiel de survie, un potentiel à engendrer une descendance et aussi une certaine probabilité de déplacement une certaine distance et vers une certaine direction. L'individu dans une population a aussi une influence, positive ou négative sur ses voisins. La population peut être également plus ou moins dispersée dans l'espace. L'ensemble de ces éléments démontre la complexité d'un raisonnement, la difficulté que peut rencontrer le généticien s'il considère chaque individu à part et l'impossibilité d'une description parfaite de la population.

Le raisonnement mathématique en **génétique des populations** est heureusement simplifié puisque chaque facteur génétique est considéré en terme de taux ou de

fréquence. Par exemple, le taux de mortalité, le taux de naissance, etc. Et pour une simplification extrême, chaque caractère génétique est pris à part. La population est aussi considérée comme un réservoir de gènes ou d'allèles ou chaque type d'allèle est présent en un grand nombre d'exemplaires avec une certaine fréquence.

I. STRUCTURE GENETIQUE DES POPULATIONS :

La **structure génétique** qui caractérise la population à un locus donné est déterminée par différentes fréquences, les fréquences génotypiques et les fréquences alléliques. Ces paramètres permettent de décrire la structure génétique d'une population et de la différencier des autres.

Les fréquences (%) des groupes sanguins *MN* se distribuent de la manière suivante chez deux populations humaines.

Fréquence	<i>MM</i>	<i>MN</i>	<i>NN</i>
Greenland	83.5	15.6	0.9
Ireland	31.2	51.5	17.3

La connaissance de la structure allélique est importante car c'est le changement des fréquences des gènes qui détermine l'évolution de la population.

1. Cas de diallélisme à codominance:

Soit les effectifs N_1 , N_2 et N_3 des génotypes respectifs *MM*, *MN* et *NN* les groupes sanguins et soit N l'effectif total de la population.

$$N = N_1 + N_2 + N_3$$

Les fréquences génotypiques sont notées et calculées comme suit:

$$\begin{aligned} \text{fréq (MM)} &= D = N_1/N \\ \text{fréq (MN)} &= H = N_2/N \\ \text{fréq (NN)} &= R = N_3/N \end{aligned}$$

La fréquence génotypique d'un génotype donné est égale au nombre total d'individus ayant ce génotype divisé par le nombre total d'individus de la population.

Parallèlement, la fréquence allélique d'un certain allèle est égale au nombre de ce type d'allèle divisé par le nombre total d'allèles dans la population.

$$\begin{aligned} \text{fréq } M = p &= (2N_1 + N_2) / 2N \\ &= N_1/N + N_2/2N \\ &= D + H/2 \\ \text{fréq } N = q &= (N_2 + 2N_3) / 2N \end{aligned}$$

$$\begin{aligned} &= N_2/2N + N_3/N \\ \text{fréq } N &= H/2 + R \end{aligned}$$

Les fréquences phénotypiques sont définies de la même manière.

2. Cas de diallélisme à dominance :

La détermination des fréquences alléliques dans le cas de relation de dominance et de récessivité entre les allèles nécessite une autre approche différente de la précédente. Le phénotype dominant A correspond à deux génotypes **AA** et **Aa**. Il est impossible de distinguer morphologiquement entre les deux génotypes. Le seul phénotype dont le génotype est connu est le phénotype récessif **aa**.

La méthode permettant pour déterminer les fréquences alléliques dans ces conditions suppose que la population est en équilibre. Ainsi, on peut déterminer les valeurs de **p** et **q** puisque la fréquence du phénotype récessif est égale dans ce cas au carré de la fréquence allélique de l'allèle **a**.

$$\text{fréq } (a) = \text{fréq } (aa) = q^2 \quad q = \sqrt{\text{fréq } (aa)}$$

La valeur de **q** est ainsi calculée. Ensuite, **p** est déduite à partir de la relation somme des fréquences est égale à l'unité.

$$p + q = 1 \quad p = 1 - q$$

De là sont déterminées les fréquences des deux génotypes dominants AA et Aa.

3. Cas de gène influencé par le sexe :

L'expression de la relation dominance-récessivité dans ce cas est variable d'un sexe à l'autre et est modifiée dans certains contextes environnementaux (influence hormonale). Le génotype hétérozygote produit des phénotypes différents chez les deux sexes. La relation dominance-récessivité est inversée. La détermination des fréquences alléliques est indirecte. Chez un sexe, la fréquence est déterminée en prenant la racine carré du génotype ou phénotype récessif ($q = \sqrt{\text{fréq } aa}$). De même la fréquence de l'autre allèle **p** est déduite de la racine carrée du phénotype récessif chez l'autre sexe.

chez les mâles	$[A] = AA \text{ et } Aa$	$\text{fréq } a = \sqrt{q^2}$
	$[a] = aa$	
chez les femelles	$[A] = AA$	$\text{fréq } A = \sqrt{p^2}$
	$[a] = Aa \text{ et } aa$	

II. EQUILIBRE DE HARDY-WEINBERG:

Un premier modèle a été proposé sur la structure génétique d'une population par G. H. Hardy un mathématicien Anglais et W. Weinberg un médecin Allemand en 1908. Leur raisonnement impose des conditions pour permettre de simplifier les formules.

H1 : le choix du conjoint ou partenaire se fait au hasard = hypothèse de panmixie

H2 : la population est de taille infinie ou la fréquence est égale à la probabilité d'un caractère = loi des grand nombre.

H3 : la fréquence des gènes n'est pas modifiée au cours des générations, câd que les gènes ou allèles sont incapables de mutations, la population ne subit ni migration ni sélection et que chaque individu de la population à la même chance à la procréation.

Hardy et Weinberg ont constaté et démontré ensuite qu'en conditions **panmictiques**, les fréquences génotypiques et celles alléliques restaient constantes au cours des générations.

1. Diallélisme à dominance totale :

Soit un allèle à deux états alléliques **A** et **a**. Chez une population à grand effectif, isolée et où les croisements se font au hasard, les fréquences alléliques sont notées **p** pour l'allèle dominant **A** et **q** pour l'allèle récessif **a**. Les fréquences génotypiques sont déterminées comme suit :

Echiquier :

<i>Gamètes (fréq)</i>	A (p)	a (q)
A (p)	AA (p²)	Aa (pq)
a (q)	Aa (pq)	aa (q²)

La **structure de Hardy-Weinberg** est donc :

$$\begin{aligned} \text{fréq AA} &= D = p^2 \\ \text{fréq Aa} &= H = 2 pq \\ \text{fréq aa} &= R = q^2 \end{aligned}$$

Les nouvelles fréquences alléliques après un cycle de reproduction seront :

$$\begin{aligned} p_{n+1} &= [2 p^2 + 2 pq] / 2 (p^2 + 2 pq + q^2) \\ p_{n+1} &= [2 p(p + q)] / [2 (p + q)^2] = p \\ p_{n+1} &= p_n \quad \text{et} \quad p_{n+1} + q_{n+1} = 1 \rightarrow q_{n+1} = q_n \end{aligned}$$

Par conséquent, étant donné les hypothèses faites, les fréquences alléliques restent constantes et ne varient pas au cours des générations.

Après un cycle de reproductions, les fréquences génotypiques seront:

<u>Croisements</u>	<u>Fréquences</u>	<u>Descendances</u>		
		AA	Aa	aa
AA x AA	D ²	D ²		
AA x Aa	2 DH	DH	DH	
Aa x aa	2 DR		DR	DR
Aa x Aa	H ²	1/4H ²	1/2H ²	1/4H ²
Aa x aa	2HR		HR	HR
aa x aa	R ²			R ²

$$\begin{aligned} \text{fréq AA} &= D_{n+1} = D^2 + DH + H^2/4 = (D + H/2)^2 = p^2 \\ \text{fréq Aa} &= H_{n+1} = (DH + 2DR + H^2/2 + HR) = 2(D+H/2) \\ &\quad (R+H/2) = 2pq \\ \text{fréq aa} &= R_{n+1} = (H^2/4 + HR + R^2) = (R + H/2)^2 = q^2 \end{aligned}$$

$$D_{n+1} = p^2 = D_n \quad H_{n+1} = 2pq = H_n \quad R_{n+1} = q^2 = R_n$$

Dans les conditions d'équilibre définies par *Hardy-Weinberg*, les fréquences génotypiques comme les fréquences alléliques restent constantes d'une génération à l'autre.

Enoncé de la loi de Hardy-Weinberg :

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

Si les accouplements sont panmictiques, les fréquences génotypiques se déduisent directement des fréquences alléliques selon la relation de la structure génétique de Hardy-Weinberg ; elles restent donc aussi constantes.

2. Polyallélisme :

Plus un gène est long, plus il offre la possibilité de mutation. Par conséquent, plus à celui-ci correspondront plusieurs allèles, plusieurs génotypes et une structure génétique bien définie dans les conditions de Hardy-Weinberg.

soit: gène \rightarrow n allèles

$\mathbf{a} \rightarrow a_1, a_2, \dots, a_i, \dots, a_n$

fréq (\mathbf{a}) $\rightarrow p_1, p_2, \dots, p_i, \dots, p_n$ $\sum_{i=1}^n p_i = 1$

Le nombre de génotypes possibles NG .

$a_i a_j \quad i=j$ homozygotes n génotypes

$i \neq j$ hétérozygotes $\binom{n}{2} = \frac{n!}{2! (n-2)!} = \frac{n(n-1)}{2}$

$$NG = n + \frac{n(n-1)}{2} = \frac{n(n+1)}{2}$$

Les fréquences génotypiques sont reportées dans l'échiquier suivant :

	<u>a_1</u>	<u>a_2</u>	<u>a_i</u>	<u>a_n</u>
<u>a_1</u>	$a_1 a_1$	$a_1 a_2$	$a_1 a_i$	$a_1 a_n$
<u>a_2</u>	$a_1 a_2$	$a_2 a_2$	$a_2 a_i$	$a_2 a_n$
<u>a_i</u>	$a_1 a_i$	$a_2 a_i$	$a_i a_i$	$a_i a_n$
<u>a_n</u>	$a_1 a_n$	$a_2 a_n$	$a_i a_n$	$a_n a_n$

$fréq(a_i a_j) = 2 p_i p_j \quad i \neq j$

$fréq(a_i a_i) = p_i^2 \quad i = j$

$\sum fréq a_i a_j = (p_1 + p_2 + \dots + p_i + \dots + p_n) = 1$

Loi Hardy-Weinberg:

Pour une population en équilibre de **Hardy-Weinberg** où il n'y a ni de sélection, ni de migration, à grand effectif et où les unions se font au hasard, et pour un gène neutre, c'est-à-dire non sollicité par la mutation, ni par la sélection, les fréquences génotypiques sont égales aux produits des fréquences alléliques des allèles qui les constituent.

Cas de triallélisme:

Considérons les groupes sanguins **A B O** chez l'homme. Deux des trois allèles I_A et I_B sont codominants et le troisième allèle i est récessif. C'est un cas de triallélisme.

Allèles	Fréquences	Phénotypes	Génotypes	réquences
I_A	$\text{fréq}(I_A) = p$	A	$I_A I_A$ $I_A i$	$p^2 + 2pr$
I_B	$\text{fréq}(I_B) = q$	B	$I_B I_B$ $I_B i$	$q^2 + 2qr$
i	$\text{fréq}(i) = r$	AB	$I_A I_B$	$2pq$
		O	ii	r^2

$$p + p + r = 1$$

Le polymorphisme des groupes sanguins (Langsteiner 1908)

	[O]	[A]	[B]	[AB]
Berlin	36,5	42,0	14,0	6,5
Espagne	41,5	46,5	9,2	2,2
Berbère Alger	39,0	37,6	18,6	4,6
Dakota (USA)	91,0	7,0	2,0	0,0

$$\text{fréq}(I_A) = \text{fréq } I_A I_A + \text{fréq } I_A i = p^2 + 2pr$$

$$\text{fréq}(I_B) = \text{fréq } I_B I_B + \text{fréq } I_B i = q^2 + 2qr$$

La méthode de calcul des fréquences alléliques à partir des fréquences phénotypiques est la suivante dans le cas de la population berbère d'Alger.

$$\text{fréq}[A] = 0,376 \quad \text{fréq}[B] = 0,186$$

$$\text{fréq}[AB] = 0,046 \quad \text{fréq}[O] = 0,390$$

$$\text{fréq}[O] = r^2 = 0,3908 \rightarrow r = \sqrt{0,3908} = 0,6251$$

$$\text{fréq}[A] + \text{fréq}[O] = p^2 + 2pr + r^2 = (p + r)^2$$

$$p + r = \sqrt{0,3768} \rightarrow$$

$$p = 0,7676 - r = 0,2510$$

$$p + q + r = 1$$

$$q = 1 - p - r = 1 - 0,6251 - 0,2510 = 0,1239$$

Les fréquences alléliques pour le reste des populations calculées de la même manière sont les suivantes :

	p	q	r
Berlin	0,28	0,12	0,60

<i>Berbère Alger</i>	0,25	0,13	0,62
<i>Espagne</i>	0,29	0,07	0,64
<i>Dakota</i>	0,35	0,01	0,955

Vérification de l'équilibre panmictique :

Une paire d'allèles codominants gouvernent la couleur de la robe chez les bovins, CrCr est rouge, Cr Cw est Brique et CwCw est blanche. Un échantillon est constitué de 180 vaches rouges, 244 briques et 80 blanches.

1. Quelles sont les fréquences alléliques ?
2. Est ce que la population est en équilibre panmictique ?

$\text{fréq Cr} = (2 \times 180 + 244) / 2 (180 + 244 + 80)$ $\text{fréq Cw} = (244 + 2 \times 80) / 2 (180 + 244 + 80)$ $p = \qquad q =$

Si nous supposons que la population est panmictique, les effectifs calculés seront les suivants:

$\text{eff [Cr]} = p^2 \times N =$ $\text{eff [CrCw]} = 2 q \times N = ?$ $\text{eff [Cw]} = q^2 \times N = ?$
--

Test X² :

$X_{ob}^2 = \sum_{i=1}^n (\text{eff ob.} - \text{eff cal.})^2 / \text{eff cal}$ $X_{ob}^2 = (? - 180)^2 / ? + (? - 244)^2 / ? + (? - 80)^2 / ?$ $X_{ob}^2 = ? \qquad X_{théo}^2 = ? (99\%) \text{ et } ? (95\%)$ $ddl = (3 - 1) - (2 - 1) = 1$
--

3. Cas de gène lié au sexe: (hétérosomal)

Chez la drosophile le nombre de chromosomes est de 2n = 8, 3 paires d'autosomes A et une paire de chromosomes sexuels, X et Y (=hétérosomes). Les mâles hétérogamétiques produisent deux types de gamètes, un à 3A + X (50%) et un deuxième à 3A + Y (50%). Les femelles homogamétiques forment 100% de gamètes à 3A + X. Le même principe de

déterminisme du sexe est retrouvé chez l'homme et plusieurs autres espèces animales. Chez les oiseaux et les lépidoptères, c'est les femelles qui sont hétérogamétiques (XY). Considérons chez la drosophile un caractère lié au sexe (porté par le chromosome X). Les allèles A et a ayant les fréquences p et q respectivement. Considérons aussi que la population est à l'état l'équilibre. Les gamètes produits vont s'unir par deux pour former des zygotes:

	<u>Mâles</u>		<u>Femelles</u>		
<i>Génotypes</i>	$X_A Y$	$X_a Y$	$X_A X_A$	$X_A X_a$	$X_a X_a$
<i>freq</i>	p	q	p^2	$2 pq$	q^2

Considérons le cas où les fréquences alléliques sont différentes au départ chez les deux sexes :

<i>Mâles</i>	$freq_m A = p_m$	$freq_m a = q_m$
<i>Femelles</i>	$freq_f A = p_f$	$freq_f a = q_f$

A chaque génération les chromosomes X des mâles proviennent des femelles de la génération d'avant, et les chromosomes X des femelles proviennent pour moitié des femelles et pour moitié des mâles de la génération qui a précédé.

G_{n-1}	$X_{m(n-1)} Y$	$X_{f(n-1)} X_{f(n-1)}$
G_n	$X_{f(n-1)} Y$	$X_{f(n-1)} X_{m(n-1)}$
$p_{m(n)} = p_{f(n-1)}$		
$p_{f(n)} = 1/2 p_{m(n-1)} + 1/2 p_{f(n-1)}$		

Evolution des fréquences alléliques chez les deux sexes :

$\Delta p_n = p_{f(n)} - p_{m(n)} = 1/2 p_{f(n-1)} + 1/2 p_{m(n-1)} - p_{f(n-1)}$
$\Delta p_n = - 1/2 (p_{f(n-1)} - p_{m(n-1)})$ donc $\Delta p_n = - 1/2 \Delta p_{n-1}$

A chaque génération, l'écart entre les fréquences des deux sexes est réduit de moitié.

$\Delta_n = (-1/2)^n \Delta_0$
$\forall \Delta_0 \quad 0 < \Delta_0 < 1 \quad \text{quant } n \rightarrow \infty \quad \Delta_n \rightarrow 0 \quad \rightarrow p_m = p_f$

La population évolue vers un état d'équilibre. La fréquence de l'allèle A dans la population est la fréquence moyenne entre les deux sexes. Les femelles contribuent pour les $2/3$ car elles apportent 2 chromosomes X , et les mâles pour $1/3$.

$$p = (p_m + 2p_f) / 3$$

$$p_n = 1/3 (p_{m(n)} + 2 p_{f(n)})$$

$$p_n = 1/3 (p_{m(n-1)} + p_{f(n-1)} + p_{m(n-1)})$$

$$p_n = 1/3 (2p_{f(n-1)} + p_{m(n-1)})$$

III. INFLUENCE DU REGIME DE REPRODUCTION

Le régime de reproduction décrit la manière dont les gamètes s'unissent pour former la génération suivante. Il influence fortement la structure génotypique. On distingue trois grandes catégories de plantes à fleurs: les autogames, les allogames et celles à multiplication végétative.

Le régime de reproduction est défini pour un locus et peut varier d'un locus à un autre chez une même espèce. Chez l'homme par exemple, on trouve une hétérogamie pour le sexe, une certaine homogamie pour les gènes de la couleur de la peau et la panmixie pour les groupes sanguins.

1. La panmixie : Il s'agit d'une union au hasard des gamètes. Ce régime confère à la population une structure génotypique comprenant $2pq$ hétérozygotes. Cette valeur constituera la référence pour classer les régimes de reproduction, en régime fermé où le taux d'hétérozygotie est inférieur à cette valeur, et en régime ouvert quand la proportion des hétérozygotes est supérieure à cette valeur.

2. Régimes fermés : parmi les régimes fermés l'autogamie, la consanguinité et homogamie.

a. Autogamie ou autofécondation : L'autogamie est un cas extrême de la consanguinité.

G_n	AA	Aa	aa
G_{n+1}	AA	AA	aa
	1	$1/4$	$1/4$

Les fréquences génotypiques à G_{n+1} en fonction des fréquences génotypiques initiales en G_n .

$$D_{n+1} = D_n + 1/4 H_n$$

$$H_{n+1} = 1/2 H_n$$

$$R_{n+1} = R_n + 1/4 H_n$$

Sous la forme matricielle, ces fréquences s'écrivent comme suit:

$$\begin{array}{l} D_{n+1} \\ H_{n+1} \\ R_{n+1} \end{array} = \begin{array}{c} \left[\begin{array}{ccc} 1 & 1/4 & 0 \\ 0 & 1/2 & 0 \\ 0 & 1/4 & 1 \end{array} \right] \begin{array}{l} D_n \\ H_n \\ R_n \end{array} \end{array}$$

donc

$$X_{n+1} = A \times X_n$$

Le développement par récurrence donne la formule suivante:

$$\begin{array}{l} X_{n+1} = A_n X_o \\ H_{n+1} = 1/2 H_n \quad H_{n+1} = (1/2)^2 H_{n-1} \dots H_{n+1} = (1/2)^n H_o \\ D_{n+1} = D_n + 1/4 H_n \\ D_{n+1} = D_{n-1} + 1/4 H_{n-1} + 1/4 H_n \\ D_{n+1} = D_o + 1/4 (H_o + H_1 + \dots H_n) \\ D_{n+1} = D_o + 1/4 H_o [1 + 1/2 + (1/2)^2 \dots + (1/2)^n] \\ D_{n+1} = D_o + 1/4 H_o [1 - (1/2)^n / 1/2] \\ D_{n+1} = D_o + 1/2 H_o [1 - (1/2)^n] \\ D_{n+1} = D_o + 1/2 H_o - H_o (1/2)^{n+1} \\ \text{quand } n \rightarrow \infty \quad D_n \rightarrow D_o + 1/2 H_o \\ R_n \rightarrow R_o + 1/2 H_o \quad \text{et} \quad H_n \rightarrow 0 \end{array}$$

b. Consanguinité : c'est un type de reproduction où les unions se font entre individus apparentés, c'est à dire ayant des ancêtres communs. Les unions consanguins entraînent une réduction des hétérozygotes dans la population au profit des homozygotes. Cette diminution concerne l'ensemble du génome comme l'autogamie. Tous les régimes consanguins autres que l'autogamie conduisent aussi, bien que plus lentement, à la disparition des hétérozygotes. Les croisements entre frères et soeurs augmentent également les génotypes homozygotes mais après un nombre de générations trois fois plus élevé.

Définitions :

Allèles identiques : On appelle allèles identiques, deux allèles provenant d'un même allèle ancestral sans mutation ni recombinaison.

Coefficient de consanguinité : (F défini pour individu, f défini pour une population).

Le coefficient de consanguinité F est la probabilité que les deux allèles sur deux loci homologues soient identiques. Pour une population, f est la probabilité que deux allèles sur deux loci homologues d'un individu quelconque de la population soient identiques.

Coefficient de parenté : ϕ est défini entre deux individus. C'est la probabilité qu'un locus d'un individu A porte l'allèle identique porté par le même locus chez un individu B .

Par définition le coefficient consanguinité d'un individu I est égal au coefficient de parenté de ses parents PM .

$$F(I) = \phi_{PM}$$

La formule qui permet de calculer ces coefficients :

$$F(I) = \phi_{PM} = \sum_{A=I}^n (1/2)^{n+n'+1} (1 + F_A)$$

A = Ancêtre commun aux deux parents

n = Nombre de générations entre le parent mâle et l'ancêtre A

n' = Nombre de générations entre le parent femelle et l'ancêtre A

F_A = Coefficient de consanguinité de l'ancêtre commun A

Le coefficient de consanguinité d'un individu est égale à la probabilité que deux allèles homologues du même locus soient identiques ou qu'ils dérivent du même ancêtre commun.

$$F(I) = \phi_{PM} = 1/2^n + 1/2^{n'} = (1/2)^{n+n'}$$

La probabilité que l'ancêtre soit homozygote et donc que ses deux allèles soient identiques est de $F(A)$ et la probabilité que l'ancêtre soit hétérozygote, c'est à dire que ses deux allèles soient différents est de $1 - F(A)$

La probabilité pour que I soit homozygote quand A est hétérozygote est de $1/2$.

$$F(I) = (1/2)^{n+n'} F(A) + 1/2 [1 - F(A)]$$

$$F(I) = (1/2)^{n+n'+1} [1 + F(A)]$$

Le même raisonnement est fait pour l'ensemble des ancêtres communs. Si A est inconnu on considère que $F(A) = 0$

$$F(I) = \sum_{A=I}^x 1/2^{n+n'+1} [1 + F(A)]$$

Cas particuliers:

a. Croisement entre **frère-soeur**. Ils ont leur père et leur mère en commun.

$$\begin{aligned}
 F(I) &= \varphi_{PM} A + \varphi_{PM} B \\
 &= \frac{1}{2}^{1+1+1} [1 + F(A)] + \frac{1}{2}^{1+1+1} [1 + F(B)] \\
 &= \frac{1}{2}^3 + \frac{1}{2}^3 = \frac{1}{2}^2 = \frac{1}{4}
 \end{aligned}$$

b. Croisement entre **cousins-germain**s.

$$\begin{aligned}
 F(I) &= \frac{1}{2}^{2+2+1} [1 + F(A)] + \frac{1}{2}^{2+2+1} [1 + F(B)] \\
 &= \frac{1}{2}^5 + \frac{1}{2}^5 = \frac{1}{2}^4 = 1/16
 \end{aligned}$$

c. Cas de l'**autofécondation**.

$$F(I) = \frac{1}{2}^{0+0+1} [1 + F(A)] = \frac{1}{2}$$

d. Cas d'une **population**:

Dans une population, il est rare d'avoir des unions complètement aléatoires. Une proportion de la population est consanguine et cette proportion est caractérisée par le coefficient de consanguinité f . La proportion panmictique est de $1 - f$.

Si on considère un couple allélique (A,a) avec les fréquences respectives p et q , les fréquences génotypiques se calculent comme suit:

$$\begin{aligned}
 \text{fréq } AA &= p^2 (1-f) + p f \\
 \text{fréq } Aa &= 2 p q (1-f) \\
 \text{fréq } aa &= q^2 (1 - f) + q f
 \end{aligned}$$

Entre deux générations, la fréquence du génotype dominant est calculée comme suit:

$$\begin{aligned}
 \text{fréq}_{n+1} AA &= p_n^2 (1-f) + p_n f = p_n^2 + f p_n (1-p_n) = p_n^2 + f p_n q_n \\
 \text{fréq}_{n+1} Aa &= 2 p_n q_n - 2 f p_n q_n \\
 \text{fréq}_{n+1} aa &= q_n^2 + f p_n q_n
 \end{aligned}$$

La consanguinité augmente donc la fréquence des génotypes homozygotes et réduit celle des individus hétérozygotes. Vu l'importance de la consanguinité, plusieurs méthodes permettent de calculer des paramètres spécifiques à la population.

L'indice de fixation F est un paramètre qui caractérise la structure de la population par l'écart à la condition d'équilibre.

$$H = (1 - F) \times 2 p q$$

$$F = 1 - H / 2 p q$$

La fréquence génotypique est déduite comme ci-après :

$$p = D + \frac{1}{2} H \quad D = p - \frac{1}{2} H = p - p q (1 - F)$$

$$D = p (1 - q) + p q F$$

$$D = p^2 + p q F$$

$$D = p^2 (1 - F) + p F$$

La structure de la population de **Wright** est écrite comme suit:

$$D = p^2 (1 - F) + p F$$

$$H = 2 p q (1 - F)$$

$$R = q^2 (1 - F) + q F$$

Connaissant D , H et R , F et p peuvent être calculés. F et p suffisent pour spécifier entièrement la structure génétique de la population.

p et q décrivent la structure génotypique

F permet de déterminer la façon dont les allèles sont associés.

Il détermine l'écart à la structure d'équilibre de Hardy-Weinberg.

F représente également la corrélation entre les gamètes qui s'unissent. F a été défini au départ comme coefficient de consanguinité par Wright. Seulement ce terme est très abusé car il n'y a pas que la consanguinité qui agit (voir ci-après).

La structure de Wright est trouvée autrement. Elle tient compte de l'influence de facteurs liés à l'union non aléatoire résultant d'un effectif limité ou de rencontre entre individus consanguins.

Soit un individu tiré au hasard dans une population. Deux de ses allèles sont identiques avec une probabilité f donc $fréq AA = p$ et $fréq aa = q$ et deux de ses allèles sont différents avec une probabilité $1 - f$ donc

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

La structure de Wright est donc dans ce raisonnement:

$$fréq AA = p^2 (1 - f) + p f$$

$$\text{fréq } aa = q^2 (1 - f) + qf$$

$$\text{fréq } Aa = 2pq (1 - f)$$

Il peut y avoir consanguinité dans une population d'effectif illimité si les croisements ne se font pas au hasard, mais préférentiellement entre individus apparentés. Dans une population d'effectif limité, même si les croisements se font au hasard, la consanguinité est inéluctable.

c. Homogamie : c'est un régime de reproduction où les unions se font entre individus phénotypiquement semblables. Comme la consanguinité, l'homogamie diminue le taux d'hétérozygotie à chaque génération.

Dans le cas de codominance, les unions qui ont lieu sont équivalents à ceux de l'autogamie puisque les croisements se font entre génotypes semblables. Les génotypes hétérozygotes diminuent de moitié à chaque génération, seulement dans le cas de l'homogamie, seul un locus est concerné.

Dans le cas de dominance, les phénotypes identiques s'unissent entre eux. Donc les individus récessifs ne forment que des récessifs, mais les dominants forment des dominants et des récessifs. Les hétérozygotes dans le groupe des dominants diminuent plus progressivement que dans le cas précédent et les croisements entre hétérozygotes deviennent de plus en plus rares.

L'autogamie, la consanguinité et l'homogamie conduisent à l'homozygotie totale quand elles sont absolues. La panmixie, quant à elle conduit à un taux d'hétérozygotie de $2pq$ en une seule génération. Dans les régimes partiels, la panmixie maintient toujours un taux d'hétérozygotie dans la population.

3. Régimes ouverts :

L'hétérogamie est l'union entre individus phénotypiquement dissemblables. Elle entraîne un taux d'hétérozygotie supérieur à $2pq$.

Les gènes sexuels (ou parasexuels) suivent un régime hétérogamique. Chez les mammifères, le sexe mâle est hétérogamétique XY et le sexe femelle homogamétique XX. En considérant des chromosomes X et Y comme des gènes, X a une fréquence de $\frac{3}{4}$ et Y de $\frac{1}{4}$. Le taux d'hétérozygotie dans un régime panmictique serait $2pq = 2 \cdot \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{8}$. En réalité, celui-ci est de $\frac{1}{2}$, il est équivalent à la proportion des mâles dans la population. Chez les plantes, les systèmes d'autoincompatibilité (gamétophytique ou sporophytique) empêchent les fécondations entre porteurs de mêmes allèles au locus d'incompatibilité et donc conduit à une hétérozygotie élevée pour le locus.

L'hétérogamie, à la différence des autres régimes précédents, modifie non seulement la structure génotypique, mais aussi la fréquence des allèles dans la population. Si les individus d'un sexe sont moins fréquents que ceux de l'autre sexe à une génération, ils se reproduisent plus. L'hétérogamie entraîne donc une forme de sélection qui favorise les gènes des plus rares ce qui les maintient à une fréquence d'équilibre précise.

IV. ETUDE DES FACTEURS EVOLUTIFS

Nous avons considéré jusqu'à présent des populations idéales à grands effectifs dont le polymorphisme génétique est contrôlé par des gènes inaptes à muter et tel qu'il ne se produit aucune sélection ni migration. A présent, nous allons considérer le cas où ces facteurs évolutifs agissent sur les populations et influencent la structure génétique de la population.

Influence d'un nombre fini d'individus:

Soit une population à N individus dont chacun produit un nombre de gamètes. A la génération suivante, le pool gamétique sera constitué de $2N$ types de gamètes.

Si on considère qu'à la génération G_0 les $2N$ gamètes sont différents $f_0 = 0$. A une génération n deux allèles ont une probabilité de $1 / 2N$ d'être différents et $(1 - 1 / 2N) f_{n-1}$ d'être identiques.

$f_n = 1/2N + [1 - (1 / 2N)] f_{n-1}$ $1 - f_n = (1 - 1 / 2N) (1 - f_{n-1})$ $1 - f_n = (1 - 1 / 2N)^n (1 - f_0)$

Le coefficient de consanguinité augmente autant plus vite que la taille de la population est petite. A terme $f_n = 1$.

La plupart des populations sont finies. La loi des grands nombres ne peut être appliquée. Les erreurs d'échantillonnage liées à l'effectif fini et cumulées sur de nombreuses générations peuvent devenir considérables. Elles introduisent une variation aléatoire de la fréquence des allèles au fil des générations. l'accumulation de ces fluctuations est appelée **dérive génétique**.

Soit une population à N individus ou toutes les conditions de Hardy-Weinberg sont satisfaites. Soit p_0 la fréquence de l'allèle A à la génération initiale G_0 et X_1 le nombre d'allèle A en G_1 une variable aléatoire dont la loi est binomiale de paramètres $2N$ et p_0 .

$$\begin{aligned} \text{Esp}(X_1) &= 2N p_0 \\ \text{Var}(X_1) &= 2N p_0 (1 - p_0) \end{aligned}$$

La fréquence de l'allèle A en G_1

$$\begin{aligned} p_1 &= X_1 / 2N \\ \text{Esp}(p_1) &= p_0 \\ \text{Var}(p_1) &= p_0 (1 - p_0) / 2N \end{aligned}$$

Si l'effectif N reste constant au cours des générations, la fréquence de l'allèle A en G_n

$$\begin{aligned} \text{Esp}(p_n) &= p_0 \\ \text{Var}(p_n) &= p_0 (1 - p_0) [1 - (1 - 1/2N)^n] \end{aligned}$$

D'autre part, la variance est le carré des écarts

$$\text{Var}(p_n) = \text{Esp}(p_n - p_0)^2$$

L'évolution des fréquences génotypiques dans le cas du génotype hétérozygote est de:

$$\begin{aligned} \text{Esp}(H) &= \text{Esp}[2 p_n (1 - p_n)] = 2 [\text{Esp}(p_n) - \text{Esp}(p_n^2)] \\ \text{or} \\ \text{Var}(p_n) &= \text{Esp}(p_n^2) - [\text{Esp}(p_n)]^2 \\ \text{Esp}(p_n) &= p_0 \\ \text{Var}(p_n) &= p_0 (1 - p_0) [1 - (1 - 1/2N)^n] \\ \text{d'ou} \\ \text{Esp}(p_n^2) &= \text{Var}(p_n) + [\text{Esp}(p_n)]^2 \\ \text{Esp}(p_n^2) &= p_0 (1 - p_0) [1 - (1 - 1/2N)^n] + p_0^2 \\ \text{donc} \\ \text{Esp}(H) &= 2 [p_0 - p_0 (1 - p_0) [1 - (1 - 1/2N)^n] - p_0^2] \\ \text{Esp}(H) &= 2 p_0 (1 - p_0) (1 - 1/2N)^n \\ n \rightarrow \infty \quad \text{Esp}(H) &\rightarrow 0 \end{aligned}$$

Les hétérozygotes tendent à disparaître de la population. Une homogénéisation de la structure est généralement atteinte avec p tendant vers 0 et q vers 1, ou inversement.

V INFLUENCE DE LA MIGRATION

Une population est sujette à une migration quand un groupe d'individus migre au sein de la population pour reconstituer une nouvelle population homogène.

Soit m la proportion des individus émigrants dans la nouvelle population P_1 .

Et donc $1-m$ la proportion des natifs de la population P_0 dans la nouvelle population P_1 .

Soit q_m la fréquence de l'allèle récessif chez le groupe des émigrants.

Et q_0 la fréquence de l'allèle récessif chez les natifs.

La fréquence de cet allèle dans la nouvelle population reconstituée est de

$$q_1 = q_0 (1 - m) + q_m \times m$$

Le changement de la fréquence allélique après une migration est de

$$\Delta q = q_1 - q_0 = q_0 (1 - m) + q_m m - q_0$$

$$\Delta q = (q_m - q_0) m$$

La variation de la fréquence allélique après une migration est proportionnelle au taux de migration m et à l'amplitude de la différence entre les fréquences alléliques des deux structures de départ, q_0 et q_m .

$$\Delta q = 0 \quad \text{quand} \quad q_m = q_0$$

VI. INFLUENCE DE LA MUTATION

C'est une modification au niveau des séquences de bases nucléiques sur les chromosomes. On définit le taux de mutation comme la probabilité qu'un gène A soit transformé en son allèle a .

Pour une espèce à reproduction asexuée, le taux de mutation est défini par cycle de division cellulaire. Et pour une espèce à reproduction sexuée, le taux de mutation est défini par cycle de reproduction.

Dans les conditions naturelles, les mutations inverses jouent en contre sens aux mutations directes. Les taux de mutations prennent des valeurs de l'ordre 10^{-4} et 10^{-8} . Les mutations inverses sont souvent moins fréquentes de l'ordre d' $1/10$. Vu la fréquence des mutations, l'effet de ce facteur par génération ne peut être que léger. Mais à long terme, celui-ci peut jouer un rôle essentiel dans les mécanismes de l'évolution des espèces.

$$A \xrightleftharpoons[u]{v} a$$

G_n	p_n	q_n
G_{n+1}	p_{n+1}	q_{n+1}

$$q_{n+1} = q_n (1-v) + p_n u$$

$$\Delta q = q_{n+1} - q_n$$

$$\Delta q = -q_n v + p_n u$$

A l'équilibre.

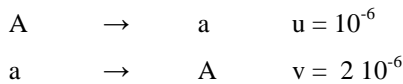
$$\Delta q = 0$$

$$v q = u p \rightarrow p/q = v/u$$

$$p_{eq} = v/(u+v) \quad q_{eq} = u/(u+v)$$

Exercice :

Soit une population d'une espèce unicellulaire haploïde, on considère un couple d'allèles (A,a). Les divisions cellulaires sont supposées synchronisées, le nombre de cellules est très grand et il n'y a pas de sélection et soient les taux de mutation



$$p_n = p_{n-1} (1-u) + q_{n-1} v$$

$$p_n = p_{n-1} (1-u) + (1-p_{n-1}) v$$

$$p_n = p_{n-1} (1-u-v) + v$$

$$p_n - v/(u+v) = p_{n-1} (1-u-v) + v - v/(u+v)$$

$$p_n - v/(u+v) = p_{n-1} (1-u-v) + [v(u+v-1)]/(u+v)$$

$$p_n - v/(u+v) = (1-u-v) [p_{n-1} - v/(u+v)]$$

$$p_n = v/(u+v) + (1-u-v)^n [p_0 - v/(u+v)]$$

$$n \rightarrow \infty \quad 0 < (1-u-v) < 1 \quad (1-u-v)^n \rightarrow 0$$

$$p_n \rightarrow v/(u+v) \quad \text{quand} \quad n \rightarrow \infty$$

Le nombre de générations nécessaires pour réduire la fréquence de l'allèle récessif de 50%

$$\Delta q = p_n - p = 1/2 (p_0 - p)$$

$$p_n - v/(u+v) = 1/2 (p_0 - p)$$

$$p_n - v/(u+v) = 1/2 [p_0 - v/(u+v)]$$

$$\rightarrow 1/2 = (1-u-v)^n \log 1/2 = n \log (1-u-v)$$

$$\rightarrow n = -\log 2 / \log (1-u-v) \rightarrow n = 230\,000 \text{ générations}$$

VII. INFLUENCE DE LA SELECTION

Dans la mesure où la fécondité est corrélée au fonctionnement du matériel héréditaire, certains allèles sont transmis plus fréquemment que d'autres aux descendants. Il y a

sélection contre un génotype lorsque sa contribution moyenne à la génération suivante est plus faible que celle d'un autre génotype. La sélection peut avoir lieu à deux niveaux du cycle.

1. La **sélection gamétique** a lieu au stade gamète quand un gamète porte un allèle déficient et que celui-ci a moins de chance de participer à la fécondation que les gamètes portant l'allèle dominant.

2. La **sélection zygotique** a lieu au stade zygote ou individus, quand la mortalité touche plus les individus ayant un certain génotype particulier que les autres génotypes.

Le **coefficient de sélection** s évalue la réduction relative d'un certain génotype par rapport au génotype le plus favorisé. Il traduit la fécondité relative des différents génotypes.

La **valeur sélective** ou Fitness σ d'un génotype est égale à sa contribution relative moyenne dans la formation d'une nouvelle la génération. La contribution du génotype le plus performant est égale à 1.

par définition $\sigma = 1 - s$

Les deux facteurs qui déterminent le coefficient de sélection ou la valeur sélective sont le taux de viabilité et le taux de fertilité.

soit (A,a)	AA	Aa	aa
G_n	p_n^2	$2 p_n q_n$	q_n^2
G_{n+1}	$\sigma_1 p_n^2$	$\sigma_2 2 p_n q_n$	$\sigma_3 q_n^2$

La valeur sélective moyenne de la population

$$C = \sigma_1 p_n^2 + \sigma_2 2 p_n q_n + \sigma_3 q_n^2$$

Les nouvelles fréquences génotypiques après un cycle de sélection:

$$\begin{aligned} \text{fréq AA} &= \sigma_1 p_n^2 / C \\ \text{fréq aa} &= \sigma_3 q_n^2 / C \\ \text{fréq Aa} &= \sigma_2 2 p_n q_n / C \end{aligned}$$

Les fréquences alléliques deviennent égales à:

$$\begin{aligned} p_{n+1} &= [2 \times \sigma_1 p_n^2 + \sigma_2 2 p_n q_n] / 2C \\ p_{n+1} &= [\sigma_1 p_n^2 + \sigma_2 p_n q_n] / C \end{aligned}$$

$$q_{n+1} = [\sigma_2 p_n q_n + \sigma_3 q_n^2] / C$$

La variation de la fréquence allélique:

$$\begin{aligned} \Delta p &= [\sigma_1 p_n^2 + \sigma_2 p_n q_n - p_n (\sigma_1 p_n^2 + \sigma_2 2 p_n q_n + \sigma_3 q_n^2)] / C \\ \Delta p &= [\sigma_1 p_n^2 (1 - p_n) + \sigma_2 p_n q_n (1 - 2 p_n) - \sigma_3 p_n q_n^2] / C \\ \Delta p &= p_n q_n [\sigma_1 p_n + \sigma_2 (q_n - p_n) - \sigma_3 q_n] / C \\ \Delta p &= p_n q_n [p_n (\sigma_1 - \sigma_2) + q_n (\sigma_2 - \sigma_3)] / C \\ \Delta p &= pq [p (\sigma_1 - \sigma_2) + q (\sigma_2 - \sigma_3)] / (\sigma_1 p^2 + \sigma_2 2pq + \sigma_3 q^2) \end{aligned}$$

$$\Delta q = pq [q (\sigma_3 - \sigma_2) + p (\sigma_2 - \sigma_1)] / (\sigma_1 p^2 + \sigma_2 2pq + \sigma_3 q^2)$$

Le changement de la fréquence allélique dans le cas d'une sélection est proportionnel aux fréquences alléliques p et q . L'amplitude de cette variation atteint son maximum quand $p = q = 0,5$.

Δq est également inversement proportionnelle à la valeur sélective moyenne C . Exprimée en fonction des coefficients de sélection.

$$\Delta q = pq [(s_1 - s_2) p + (s_2 - s_3) q] / [(1-s_1) p^2 + (1-s_2) 2pq + (1-s_3) q^2]$$

Dans le cas particulier de la dominance totale: $s_1 = 0$ $s_2 = 0$ $s_3 \neq 0$

$$\begin{aligned} \Delta q &= pq (-s q^3) / (p^2 + 2 pq + q^2 - sq^2) \\ \Delta q &= -spq^2 / (1 - sq^2) \end{aligned}$$

Et dans le cas où le génotype récessif est complètement éliminé: $s_3 = 1$

$$\begin{aligned} \Delta q &= -pq^2 / (1 - q^2) \\ p = 1 - q \rightarrow \Delta q &= -q^2 (1 - q) / [(1 - q) (1 + q)] \\ \Delta q &= -q^2 / (1 + q) \end{aligned}$$

Dans le cas particulier d'équilibre entre l'effet des mutations et celui de la sélection:

Mutation \rightarrow fréquence du gène récessif $\Delta q = up - vq$

Sélection \rightarrow gène complètement dominant $\Delta q = -spq^2 / (1 - sq^2)$

$$up - vq = -spq^2 / (1-sq^2)$$

Vu la valeur de la fréquence des mutations, on considère que le carré des taux de mutations inverses est négligeable devant l'effet des mutations directes.

$vq \ll up$ et aussi $sq^2 \ll 1$

$$\rightarrow up = -spq^2 \quad u = -sq^2$$

A l'équilibre, la fréquence de l'allèle récessif est réduite à l'équation:

$$q_e = \sqrt{u/s}$$