

Université de Djelfa
Faculté des Sciences de la nature et de la vie

Département de Biologie

Domaine : Sciences de la nature et de la vie

Filière : Sciences Biologiques

Spécialité : Biologie Moléculaire (Licence)

Année universitaire 2019 /2020

Module : Génétique des Populations

III. Le modèle de Hardy-Weinberg

Objectifs : Ce chapitre vise à apprendre aux étudiants :

- la mise en évidence du modèle de Hardy-Weinberg
- la démonstration de l'équilibre de ce modèle
- et à connaître les principales hypothèses de son application

IV. Le modèle de Hardy-Weinberg

Le modèle de Hardy-Weinberg, encore appelé équilibre panmictique, a été mis en évidence au début du XX^{ème} siècle par plusieurs chercheurs, en particulier Hardy qui était mathématicien et Weinberg, médecin.

L'équilibre de Hardy-Weinberg reste le modèle théorique central de la génétique des populations. La notion d'équilibre dans le modèle de Hardy-Weinberg est assujettie à différentes hypothèses.

Ces hypothèses sont les suivantes :

- La population sur laquelle on étudie cette notion d'équilibre est panmictique. Les couples se forment au hasard (panmixie), et de même leurs gamètes se rencontrent au hasard (pangamie).
- La population est très grande en effectif, ceci pour diminuer très fortement les variations d'échantillonnage.
- Il ne doit y avoir dans la population, ni sélection, ni mutation, ni migration.
- Les générations successives sont discrètes (pas de superposition de générations dans les croisements).
- Les différents génotypes sont viables et féconds.

Variables d'état

Pour 1 locus autosomal d'un gène diallélique co-dominants autosomique (Allèles A_1 et A_2), on a :

Génotypes	A1A1	A1A2	A2A2	avec
Fréquences génotypiques F(G)	D	H	R	D+H+R=1
Effectifs	DN	HN	RN	(effectifs totaux N)

Fréquences alléliques F(A) :

de $A_1 = p = D + \frac{1}{2} H$

de $A_2 = q = R + \frac{1}{2} H$

avec $p + q = 1$

NB/ Les fréquences génotypiques F(G) permettent, dans tous les cas, de calculer les fréquences alléliques.

A. Système di allélique (Rappels)

Soit un gène autosomique A dans une population sous deux formes alléliques A1 et A2 de fréquences identiques dans les deux sexes. Comme il y a codominance, la distinction des 3 génotypes est possible.

- Si on désigne par p la fréquence de l'allèle A1 et q la fréquence de l'allèle A2 avec $p + q = 1$.
- Si on suppose que le nombre d'individus dans la population étudiée à la génération n, est très grand et que les croisements s'effectuent au hasard ainsi que leurs gamètes (panmixie et pangamie).
- Si on admet également qu'il n'y a pas de sélection, de mutation, de migration dans cette population.
- Si il n'y a pas de superposition de générations dans les croisements.
- Si les différents génotypes sont viables et féconds.

Les individus de la génération $n+1$ seront considérés comme les descendants de l'union au hasard d'un gamète mâle et d'un gamète femelle.

Par conséquent, si à la génération n la probabilité de tirer un allèle A1 est p, celle de produire après la fécondation un zygote A1A1 est : $p \times p = p^2$

de même pour A2, celle de produire un zygote A2A2 est : $q \times q = q^2$

La probabilité de produire un hétérozygote est : $pq + pq = 2pq$.

Enoncé de la loi de Hardy-Weinberg (HW)

Dans une population dont l'effectif est infini (très grand), panmictique (mariages au hasard), en l'absence de mutation et de sélection, la fréquence des génotypes sera le développement de $(p + q)^2$, p et q étant les fréquences alléliques.

Les fréquences alléliques permettent, seulement sous HW, de calculer les fréquences génotypiques,

avec $p^2 + 2pq + q^2 = (p + q)^2 = 1$

→ Les fréquences alléliques restent invariantes d'une génération à l'autre

→ Les fréquences génotypiques restent invariantes d'une génération à l'autre

1 locus diallélique autosomique

Les fréquences génotypiques seront données par le développement de $(p + q)^2$, p et q étant les fréquences des deux allèles dans cette population.

Génotypes	$A_1 A_1$	$A_1 A_2$	$A_2 A_2$
Génération n	p^2	$2pq$	q^2
Génération $n+1$	$F(A_1) = p^2 + \frac{1}{2}(2pq) = p^2 + pq = p(p+q) = p$ $F(A_2) = q^2 + \frac{1}{2}(2pq) = q^2 + pq = q(p+q) = q$		

On n'a donc pas de modification des fréquences alléliques :

à la génération n , on a p et q ;

à la génération $n+1$, on a p et q .

Les fréquences génotypiques à la génération $n+1$,
nouvelle fréquence de $(A_1 A_1)$ à la génération $n+1$,

Male : femelle	$p^2 (A_1 A_1)$	$2pq (A_1 A_2)$	$q^2 (A_2 A_2)$
$p^2 (A_1 A_1)$	$A_1 A_1$	$\frac{1}{2} (A_1 A_1)$	/////
$2pq (A_1 A_2)$	$\frac{1}{2} (A_1 A_1)$	$\frac{1}{4} (A_1 A_1)$	
$q^2 (A_2 A_2)$	/////		

nouvelle fréquence de $(A_1 A_1)$ à la génération $n+1$,

$$F(A_1 A_1) = p^2 \times p^2 + \frac{1}{2}(2pq) \times p^2 + \frac{1}{4}(2pq)^2$$

$$F(A_1 A_1) = p^4 + 2p^3q + p^2q^2$$

$$F(A_1 A_1) = p^2 (p^2 + 2pq + q^2)$$

$$F(A_1 A_1) = p^2$$

La fréquence du génotype $(A_1 A_1)$ ne change donc pas en passant de la génération n à la génération $n+1$.

Même démonstration pour les génotypes $(A_1 A_2)$ et $(A_2 A_2)$.

Ainsi, dès que la population est à l'équilibre de Hardy Weinberg, la structure génotypique ne varie plus.

Dans de très nombreux exemples les fréquences observées dans les populations naturelles sont conformes à celles attendues par la loi de Hardy-Weinberg.

B. Système multi alléliques

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

$$p_1^2 A_1 A_1 \quad p_2^2 A_2 A_2 \quad p_3^2 A_3 A_3 \quad p_k^2 A_k A_k$$

$$2p_1 p_2 A_1 A_2 \quad 2p_1 p_3 A_1 A_3 \quad 2p_1 p_k A_1 A_k \quad 2p_2 p_3 A_2 A_3 \quad 2p_2 p_k A_2 A_k \quad \text{etc}$$

Exemple: Les groupes sanguins du système ABO chez l'homme sont dus à l'existence de 3 allèles A, B et O dont les fréquences peuvent être appelées respectivement p, q, r . Une population à l'équilibre de Hardy-Weinberg aura alors les fréquences génotypiques suivantes :

$$p^2 AA \quad q^2 BB \quad r^2 OO \quad 2pq AB \quad 2pr AO \quad 2qr BO$$

Lorsqu'un gène présente plus de 2 états alléliques, la fréquence à l'équilibre des hétérozygotes H peut dépasser 50%, et elle est d'autant plus élevée que le nombre d'allèles est important. L'hétérozygotie maximale est atteinte lorsque tous les allèles ont même fréquence et sa valeur est

$$H_{\max} = 1 - 1/k \text{ où } k \text{ est le nombre d'allèles}$$

Explication

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.

C. Cas des gènes liés au sexe

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.

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)
XAXA = p^2	XAY = p
XAXa = $2pq$	XaY = q
XaXa = q^2	

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 autosomaux):

- 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 :

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

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 panmixtiques.

D. Application du modèle de Hardy-Weinberg (test de l'équilibre)

La loi de Hardy-Weinberg 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).

Dans les populations naturelles, le test de l'équilibre, pour des caractères codominants, (pour lesquels le calcul des fréquences alléliques est possible) permet de vérifier l'équilibre de Hardy-Weinberg. 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). Le test du Chi Deux nécessite le calcul de la distance X^2 permettant de tester l'hypothèse d'égalité entre la distribution observée et la distribution théorique (hypothèse H_0).

$$\chi^2 = \sum \frac{(\text{effectif théorique} - \text{effectif observé})^2}{\text{effectif théorique}}$$

La somme est effectuée sur tous les génotypes et la valeur X^2 est comparée à une valeur seuil, lue dans une table, en fonction de 2 paramètres : un risque α choisi par l'utilisateur qui est en général 5% et un nombre de degrés de liberté **ddl** égale à la différence entre le nombre de génotypes et le nombre d'allèles du système génétique étudié.

- si X^2 calculé est inférieur à X^2 seuil, H_0 est acceptée et on conclue que la population suit la loi de Hardy-Weinberg, donc équilibre
- si X^2 calculé est supérieur à X^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.. Explication

Exemple

Chez l'homme, le groupe sanguin MN est déterminé par un gène autosomal 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 un échantillon de population humaine (730 individus) 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:

$$MM = p^2 \times 730 = (0,178)^2 \times 730 = 23,1$$

$$MN = 2pq \times 730 = (2 \times 0,178 \times 0,822) \times 730 = 213,6$$

$$NN = q^2 \times 730 = (0,822)^2 \times 730 = 493,2$$

3- Test du Chi deux

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

la valeur seuil pour 3-2=1 degré de liberté et un risque de 5% est 3,84. La valeur de la statistique X^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 que cet échantillon de population est à l'équilibre de Hardy-Weinberg.

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 revanche, le fait qu'un échantillon soit trouvé non conforme à la loi implique que le fonctionnement génétique réel de la population est très éloigné du fonctionnement théorique, en particulier au niveau du système de croisement (homogamie, consanguinité) ou de la structure démographique de la population (fractionnement en sous-populations) qui modifient considérablement les fréquences relatives des homozygotes et des hétérozygotes. Cette situation conduit à rechercher les particularités démographiques, génétiques ou structurales de cette population.

Extrait de la Table de Khi-Deux (ddl=degrés de liberté : P = Seuil de signification (Probabilité))

Ddl	(p)	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,01	0,001
1		0,02	0,06	0,15	0,28	0,46	0,71	1,07	1,64	2,71	3,84	6,64	10,83
2		0,21	0,45	0,71	1,02	1,39	1,83	2,41	3,22	4,61	5,99	9,21	13,82
3		0,58	1,01	1,42	1,87	2,37	2,95	3,67	4,64	6,25	7,82	11,35	16,27
4		1,06	1,65	2,20	2,75	3,36	4,05	4,88	5,99	7,78	9,49	13,28	18,47
5		1,61	2,34	3,00	3,66	4,35	5,13	6,06	7,29	9,24	11,07	15,09	20,52
6		2,20	3,07	3,83	4,57	5,35	6,21	7,23	8,56	10,65	12,59	16,81	22,46
7		2,83	3,82	4,67	5,49	6,35	7,28	8,38	9,80	12,02	14,07	18,48	24,32
8		3,49	4,59	5,53	6,42	7,34	8,35	9,52	11,03	13,36	15,51	20,09	26,12
9		4,17	5,38	6,39	7,36	8,34	9,41	10,66	12,24	14,68	16,92	21,67	27,88
10		4,87	6,18	7,27	8,3	9,34	10,47	11,78	13,44	15,99	18,31	23,21	29,59
Non Significative											Significative		
Hypothèse nulle H0 Non rejetée (pour un risque > 5%)											H0 rejetée (risque < 5%)		