



Loi et équilibre de Hardy-Weinberg

Les cours ou manuels d'initiation à la génétique des populations commencent presque toujours par un exposé de la « loi de Hardy-Weinberg ». Par ailleurs, de nombreux articles relatifs à la diversité intraspécifique font état de tests statistiques de l'« équilibre de Hardy-Weinberg ». Le non-spécialiste est légitimement en droit de se demander quelles sont exactement les relations entre « loi » et « équilibre » de Hardy-Weinberg, ce qui implique évidemment une compréhension des notions de base sous-jacentes. Le but de cet article est de présenter ces notions et, en insistant sur la logique des raisonnements qui conduisent à quelques résultats très simples, de dégager les limites de validité des résultats théoriques et de leurs éventuelles vérifications concrètes.

Définitions préliminaires

Population

Nous désignerons ainsi un ensemble d'individus appartenant à la même espèce, tel qu'un de ses membres a une probabilité très élevée, s'il se reproduit, d'y trouver ses partenaires et d'y laisser ses descendants, alors qu'il a une probabilité bien plus faible de trouver un partenaire et de laisser des descendants dans une autre population. L'espèce est supposée diploïde et à reproduction sexuée exclusive.

Structure génotypique (ou composition génotypique)

Sauf situation extrêmement particulière, une population est formée d'individus génétiquement différents. Nous nous intéresserons à la part de la variation qui est transmise

d'une génération à l'autre par la ségrégation mendélienne d'une seule série allélique autosomique. Soient $A_1, A_2, \dots, A_i, \dots, A_n$ les différents allèles. Les individus de la population peuvent être classés selon leur génotype et l'on peut dénombrer ceux qui présentent un génotype donné. A chaque classe génotypique correspond donc un effectif. Le rapport de cet effectif à l'effectif total de la population est la fréquence du génotype considéré. L'ensemble de toutes les fréquences génotypiques constitue la structure génotypique de la population.

Fréquences alléliques

En admettant que chaque individu de génotype $A_i A_i$ possède deux exemplaires de l'allèle A_i et que chaque hétérozygote de type $A_i A_j$ en possède un seul, il est aisé de dénombrer les allèles A_i présents dans la population et de calculer le rapport de ce nombre au nombre total de gènes A (celui-ci n'est autre que le double de l'effectif de la population, puisque nous ne considérons que le cas d'une série allélique autosomique). On obtient ainsi la fréquence de l'allèle A_i et on peut obtenir de même les fréquences des autres allèles de la série.

La structure génotypique de Hardy-Weinberg

Si l'on connaît la structure génotypique d'une population, on en déduit immédiatement les fréquences alléliques. En revanche, la connaissance des fréquences alléliques, bien qu'elle fournisse une certaine information sur la structure génotypique, ne permet pas de la

déterminer de façon exacte, sauf avec des hypothèses bien précises.

La panmixie des gamètes et ses conséquences

Supposons que la population étudiée appartienne à une espèce dans laquelle les gamètes des deux sexes sont émis de manière synchrone dans le milieu environnant au sein duquel ils s'unissent deux à deux ultérieurement (de telles espèces existent, par exemple des oursins et des étoiles de mer). La formation d'un zygote peut être considérée comme une épreuve aléatoire consistant en deux tirages, celui d'un gamète mâle et celui d'un gamète femelle. Chaque gamète peut porter l'un ou l'autre des allèles $A_1, A_2, \dots, A_i, \dots, A_n$. Désignons par p'_i et p''_i , quel que soit i , les probabilités respectives de porter l'allèle A_i pour un gamète mâle et pour un gamète femelle. Admettons que, quels que soient i et j , la probabilité pour un zygote d'être formé par l'union entre un gamète mâle portant A_i et un gamète femelle portant A_j soit égale au produit $p'_i * p''_j$. C'est dire que les états alléliques du gène A chez les gamètes mâle et femelle qui s'unissent pour former un zygote sont indépendants en probabilité. Cela définit une loi particulière d'union des gamètes deux à deux dite union au hasard ou panmixie, ou encore pangamie.

Sous ces hypothèses, la probabilité pour un zygote d'être homozygote $A_i A_i$ est $p'_i * p''_i$, tandis que celle d'être hétérozygote $A_i A_j$ est $p'_i * p''_j + p'_j * p''_i$. Ces expressions prennent des formes particulièrement simples si p'_i et p''_i sont égales quel que soit i . En désignant par p_i la valeur commune, les probabilités pour un zygote d'être

respectivement homozygote $A_i A_i$ et hétérozygote $A_i A_j$ sont respectivement p_i^2 et $2p_i p_j$. S'il se forme un très grand nombre de zygotes, l'application intuitive de la « loi des grands nombres » montre que la proportion effectivement observée de zygotes d'un génotype donné ne doit pas s'écarter beaucoup de la probabilité calculée. Parmi des zygotes en nombre infiniment grand, constituant une nouvelle génération ou une nouvelle cohorte de la population étudiée, les fréquences génotypiques seraient exactement égales aux probabilités calculées. On montre, en outre, aisément que la fréquence d'un allèle quelconque A_i dans la nouvelle génération ou cohorte est exactement égale à p_i . Une structure génotypique telle que la fréquence de tout génotype homozygote soit égale au carré de la fréquence allélique correspondante et celle de tout génotype hétérozygote égale au double du produit des deux fréquences alléliques correspondantes a été décrite indépendamment, dès 1908, par le mathématicien G.H. Hardy [1] et par le médecin W. Weinberg [2]. On peut donc légitimement parler de structure génotypique de Hardy-Weinberg. D'après ce qui précède, une condition suffisante à son obtention dans une population de zygotes est la vérification simultanée des hypothèses :

- effectif infiniment grand de cette population (1) ;
- même loi de probabilité de l'état allélique du gène A chez les gamètes mâles et femelles dont elle est issue (2) ;
- unions au hasard (panmixie) de ces gamètes (3).

Aucune hypothèse n'est, en revanche, nécessaire concernant la manière dont a été obtenue la « population de gamètes » qui engendrent les zygotes. Rien ne s'oppose en particulier à ce que certains de ces gamètes soient issus de reproducteurs ayant vécu sur un autre territoire (phénomène de migration), à ce que des gamètes soient porteurs d'allèles mutants, ni à ce que des phénomènes de sélection aient eu lieu à la génération précédente, pourvu que soit respectée *in fine*

l'hypothèse 2 de « symétrie » entre les deux sexes.

On remarque qu'une écriture symbolique de la structure génotypique de Hardy-Weinberg est obtenue en effectuant le développement de :

$$(p_1 A_1 + p_2 A_2 + \dots + p_i A_i + \dots + p_j A_j + \dots + p_n A_n)^2 = \sum p_i^2 A_i A_i + \sum 2p_i p_j A_i A_j$$

Rapprochement sexuel

et structure de Hardy-Weinberg

Chez la plupart des espèces animales, la reproduction implique une séquence comportementale plus complexe que la simple émission synchrone de gamètes dans le milieu extérieur. Nombreuses sont celles qui pratiquent un accouplement, comme les mammifères, les oiseaux ou les insectes, ou au moins un rapprochement sexuel, comme beaucoup de poissons. On parle, chez de telles espèces, de panmixie, non à propos de la loi d'union des gamètes, mais à propos de la loi d'union deux à deux des reproducteurs, ou loi de formation des couples. La panmixie consiste alors en l'indépendance en probabilité des génotypes des deux partenaires d'un même couple. On peut démontrer que la panmixie des reproducteurs a les mêmes conséquences que la panmixie des gamètes, à condition toutefois que soient aussi vérifiées deux hypothèses, celle de l'absence de sélection par fécondité différentielle (les couples des différentes catégories, définies par les génotypes des deux partenaires, engendrent en moyenne le même nombre de zygotes) et celle de la conformité aux probabilités mendéliennes parmi les zygotes issus de chaque catégorie de couples (pas de distorsion de ségrégation, pas de sélection gamétique, unions au hasard des gamètes formés par chacun des deux partenaires).

Une condition suffisante à la réalisation de la structure génotypique de Hardy-Weinberg dans une population de zygotes est donc la vérification simultanée des hypothèses suivantes :

- effectif infiniment grand (1) ;
- égalité de toutes les fréquences alléliques chez les parents mâles et femelles (2) ;

– unions au hasard de ces reproducteurs (3) ;

- pas de fécondité différentielle (4) ;
- stricte applicabilité des lois de Mendel (5).

Le résultat reste valable si la génération précédente a connu des phénomènes sélectifs, des migrations et des mutations, pourvu que l'hypothèse 2 reste vérifiée.

Remarques

- Nous n'avons considéré ici que deux types de cycles biologiques, le second étant particulièrement répandu parmi les métazoaires. Pour d'autres situations, par exemple celle des plantes à fleurs, des modèles similaires peuvent être construits, permettant de définir, dans chaque cas, un corps d'hypothèses constituant une condition suffisante à la réalisation de la structure génotypique de Hardy-Weinberg. Il comporte toujours une hypothèse d'« union au hasard » ou panmixie au sens large. Dans une acception plus restrictive, la panmixie désigne l'union au hasard des reproducteurs d'une espèce à rapprochement sexuel, le terme de pangamie étant réservé à l'union au hasard des gamètes.

- Cette structure génotypique n'a *a priori* aucune raison d'être une structure d'équilibre. En effet, par définition, une structure génotypique d'équilibre est une structure qui ne varie pas au cours des générations successives. Or les corps d'hypothèses énoncés ci-dessus laissent de toute évidence la porte ouverte à diverses causes de variations, par exemple l'introduction, par mutations ou par migrations, d'allèles antérieurement absents. Rien ne s'oppose à ce que des générations successives de la même population présentent toutes des structures de Hardy-Weinberg, mais que ces structures diffèrent les unes des autres par les fréquences alléliques.

- Certains écarts aux hypothèses ne causent que des déviations minimales par rapport à la structure de Hardy-Weinberg. C'est le cas d'une « dissymétrie » entre les deux sexes (hypothèse 2). Supposons par exemple que, parmi les gamètes qui vont s'unir pour former des zygotes, la

probabilité de porter un allèle donné soit $p - \varepsilon$ dans un des sexes et $p + \varepsilon$ dans l'autre. Si toutes les autres hypothèses sont vérifiées, la fréquence du génotype homozygote correspondant dans la population de zygotes sera $p^2 - \varepsilon^2$, peu différente de p^2 si ε n'est pas très grand. Ainsi, dans le cas simple d'une variation due à un seul couple d'allèles A_1-A_2 , de probabilités respectives 0,3 et 0,7 chez les gamètes mâles, 0,5 et 0,5 chez les gamètes femelles, les fréquences génotypiques attendues dans la population de zygotes sont 0,15 pour le génotype A_1A_1 , 0,50 pour A_1A_2 et 0,35 pour A_2A_2 , alors que si la population avait une structure de Hardy-Weinberg avec les mêmes fréquences alléliques (0,4 et 0,6), les fréquences des trois génotypes seraient respectivement 0,16, 0,48 et 0,36. De telles déviations n'ont des chances d'être détectées dans une population réelle qu'au prix d'un important effort d'échantillonnage! Une remarque similaire pourrait être faite à propos de la fécondité différentielle. Quant au cas où l'effectif est fini, si l'union des gamètes est panmictique, et si les hypothèses 2 et 3 sont réalisées, la déviation à l'échelle de la population de zygotes dans son ensemble est assimilable à une erreur d'échantillonnage, ce qui revient à peu près à dire qu'elle est statistiquement indétectable.

• Les raisonnements justifiant la structure génotypique de Hardy-Weinberg s'appliquent à une population ou à une cohorte de zygotes à l'instant même de leur formation. Or, l'étude d'une population réelle porte nécessairement, sauf cas tout à fait exceptionnels, sur des ensembles d'individus qui ont atteint un stade de développement plus ou moins avancé. L'histoire d'un tel ensemble est le plus souvent émaillée d'événements susceptibles d'introduire diverses déviations par rapport à la structure génotypique de la population de zygotes originelle. En conséquence, il convient d'être extrêmement prudent sur les conclusions biologiques qui peuvent être tirées des résultats des tests statistiques de conformité à la structure génotypique de Hardy-Weinberg, que l'on détecte ou non une dévia-

tion significative. Ces tests peuvent apporter des présomptions sur la réalisation ou la non-réalisation de telle ou telle des hypothèses, par exemple sur le « système de croisement » (panmixie ou autre), mais ces présomptions doivent être étayées par des arguments complémentaires, tirés notamment de la connaissance de l'« histoire naturelle » de l'espèce étudiée.

L'équilibre de Hardy-Weinberg

La situation d'équilibre idéale

Considérons le corps d'hypothèses suivant :

- effectif infiniment grand ;
- population isolée (pas de migrations) ;
- pas de mutations ;
- aucune sorte de sélection.

Il a pour conséquence l'invariance des fréquences alléliques au cours des générations successives.

Si, maintenant, on ajoute à ces hypothèses celle de la panmixie, on réalise *ipso facto* une condition suffisante à l'obtention de la structure génotypique de Hardy-Weinberg, qui est alors stable au cours des générations successives. Il est clair que de telles hypothèses ne peuvent guère être vérifiées dans une population réelle, mais elles permettent de définir une situation idéale à laquelle on peut se référer, notamment pour une étude théorique des effets d'un facteur susceptible d'introduire un écart aux hypothèses. C'est cette situation idéale qu'avaient étudiée Hardy et Weinberg, et il est licite de parler à son propos d'équilibre de Hardy-Weinberg.

Structure de Hardy-Weinberg et polymorphisme équilibré

Certaines formes de sélection (avantage de l'hétérozygote, par exemple), les mutations et certains régimes de migrations sont de nature à entretenir le polymorphisme d'une population. Cela ne constitue nullement un obstacle à la réalisation, à chaque génération, d'une structure de Hardy-Weinberg, pourvu que soient respectées les hypothèses présentées et discutées (voir page ??, la structure génotypique de Hardy-Weinberg), de telle sorte que la population peut parfaitement tendre vers un état

d'équilibre caractérisé par une structure de Hardy-Weinberg. Un tel équilibre n'a pas du tout la même signification que celui de la situation idéale qui vient d'être présentée et, du reste, ce n'est pas celui qu'avaient prédit les deux pionniers de la génétique des populations. L'application de l'expression « équilibre de Hardy-Weinberg » à cette situation est donc discutable.

Commentaires généraux

La présentation qui vient d'être faite est notablement plus complexe que celles qui sont ordinairement données de la « loi de Hardy-Weinberg » dans les ouvrages élémentaires. Voici, à titre d'exemples, deux énoncés extraits : (1) l'un d'un manuel à l'usage d'étudiants ; (2) l'autre d'un dictionnaire de génétique.

1. *Dans une population panmictique à l'équilibre, ne présentant ni sélection, ni mutations ni migrations et d'effectif élevé, la proportion des allèles et des génotypes est, d'une génération à l'autre, absolument constante.*

2. *Loi qui établit le principe selon lequel, dans une population panmictique d'effectif supposé illimité, et en l'absence de mutation, sélection et dérive génétique, pour un locus autosomique à deux allèles, l'un A de fréquence p et l'autre a de fréquence q identiques chez les mâles et les femelles, les diverses fréquences génotypiques possibles se stabilisent en une seule génération, soit p² pour AA, 2pq pour Aa et q² pour aa.*

L'énoncé 1, isolé de son contexte, est incomplet, puisqu'il ne mentionne pas l'expression des fréquences génotypiques en fonction des fréquences alléliques (cette expression n'est donnée qu'après de longs commentaires). Il fait par ailleurs l'hypothèse inutile d'un état d'équilibre dont il faut du reste bien réaliser qu'il s'agit d'un équilibre démographique (stabilité de l'effectif total de la population). L'énoncé 2 pêche quant à lui par omission d'une hypothèse (celle de l'absence de migrations) et par redondance (si l'effectif est illimité, il n'y a pas de dérive génétique). Enfin, ni l'un ni l'autre ne distingue, parmi les hypothèses retenues, celles qui sont responsables de la stabilité des fréquences alléliques de celles qui

sont responsables de la relation entre fréquences génotypiques et fréquences alléliques. Ce dernier défaut se retrouve dans de très nombreux ouvrages de tous niveaux, français ou étrangers. Il est à la source d'une ambiguïté de l'expression «équilibre de Hardy-Weinberg». Nombreuses sont en effet les publications relatives au polymorphisme dans lesquelles sont présentés les résultats de «tests de l'équilibre de Hardy-Weinberg» qui ne sont en réalité que des tests de la structure génotypique de Hardy-Weinberg, car les observations portent sur un seul échantillon, donc ne fournissent aucune information sur l'éventuelle stabilité de cette structure au cours du temps et *a fortiori* sur ses causes.

Il serait raisonnable de désigner sous le nom de «loi», «théorème», voire «principe» de Hardy-Weinberg une proposition telle que «un corps d'hypothèses du type de ceux qui sont présentés dans cet article. La structure génotypique de Hardy-Weinberg) implique la réalisation immédiate de la structure génotypique de Hardy-Weinberg». L'expression «équilibre de Hardy-Weinberg» devrait être en toute rigueur réservée à la situation extrêmement particulière dans laquelle une telle structure est invariante de génération en génération du fait de la réalisation permanente de toutes les hypothèses: panmixie, absence de migrations, absence de mutations, absence de sélection, effectif infiniment grand. C'est du reste la position prise dans un traité classique [3].

Exemple

Dans l'espèce humaine, un couple d'allèles autosomique, ici désigné, pour des raisons de simplicité par **M-N**, détermine le système de groupes sanguins M-N, avec trois phénotypes correspondant aux trois génotypes. Il est donc aisé d'apprécier, par un échantillonnage convenable, la structure génotypique d'une population. Li [4] présente ainsi un échantillon de 1 482 adultes d'une ville japonaise, se répartissant en 406 individus de génotype **MM**, 744 **MN** et 332 **NN**. La fréquence de l'allèle **M** dans cet échantillon est $(2 \cdot 406 + 744) / 2 \cdot 1482 = 1556 / 2 \cdot 1482$. On prévoit donc que si la population d'où est tiré cet échantillon a une structure génotypique de Hardy-Weinberg, la fréquence du génotype **MM** doit y être peu différente de $(1556 / 2 \cdot 1482)^2$; le nombre d'individus présentant ce génotype dans un échantillon de 1 482 individus doit donc être peu différent de $1482 \cdot (1556 / 2 \cdot 1482)^2 = 1556^2 / 4 \cdot 1482 = 408,42$. Des calculs similaires montrent que les effectifs des génotypes **MN** et **NN** doivent être proches respectivement de 739,16 et 334,42. Il n'est pas besoin d'effectuer un test statistique pour voir que les effectifs observés dans l'échantillon ne s'écartent pas de façon significative de ces effectifs attendus. L'hypothèse selon laquelle la population étudiée a une structure génotypique de Hardy-Weinberg n'est donc pas rejetée. On en déduit que la loi de formation des couples dans cette popu-

lation ne s'écarte probablement guère de la panmixie, du moins en ce qui concerne le système **M-N**. Ce n'est toutefois pas rigoureusement démontré, puisque l'échantillon examiné est un échantillon d'adultes et non de zygotes, et il serait absurde de conclure à la stabilité au cours du temps des fréquences des allèles **M** et **N** dans la population au vu de ces seuls résultats ■

Jean Générmont

Professeur des universités. Biologie évolutive et Dynamique des populations, Ura Cnrs 2154, Université Paris-Sud, bâtiment 446, 91405 Orsay Cedex, France.

RÉFÉRENCES

1. Hardy GH. Mendelian proportions in a mixed population. *Science* 1908; 28: 49-50.
2. Weinberg W. Über den Nachweis der Vererbung beim Menschen. *Jahreshefte Verein F Vaterl Naturk in Württemberg* 1908; 64: 368-82.
3. Crow JF, Kimura M. *An introduction to population genetics theory*. New York: Harper and Row, 1970.
4. Li CC. *Human genetics. Principles and methods*. New York: McGraw-Hill Book Company, 1961.

TIRÉS À PART

J. Générmont.