

Recherche d'effets quantitatifs associés aux génotypes AA et Aa à un locus marqueur avec dominance complète : optimisation de l'épreuve de descendance pour l'identification des génotypes

P Mérat, F Minvielle

Institut national de la recherche agronomique, centre de Jouy-en-Josas, laboratoire de génétique factorielle, 78352 Jouy-en-Josas Cedex, France

(Reçu le 18 juin 1991 ; accepté le 23 avril 1992)

Résumé – Lorsqu'on identifie par épreuve de descendance les individus de génotype AA et Aa à un locus marqueur avec dominance complète, en vue de rechercher des effets quantitatifs associés, on montre que, pour une précision maximale de la comparaison des moyennes de ces génotypes, il existe des valeurs optimales pour le nombre d'individus soumis à l'épreuve de descendance et celui des descendants par parent dans l'hypothèse d'une limitation matérielle du nombre total des descendants. Ces valeurs optimales dépendent des fréquences alléliques et de la différence moyenne entre génotypes.

gène majeur / dominance / identification des génotypes / épreuve de descendance / taille de famille

Summary – Research of quantitative effects associated with the AA and Aa genotypes at a marker locus with complete dominance : optimization of progeny-test for the identification of genotypes. When individuals of AA or Aa genotype at a marker locus with complete dominance are identified so as to detect quantitative associated effects, it is shown that for maximum accuracy of the comparison between mean values of these genotypes, there are optimal values for the number of progeny-tested individuals and that of progeny per parent under the hypothesis of a limit to the total number of progeny. This optimal value depends on the allelic frequencies and on the average difference between genotypes.

major gene / dominance / genotype identification / progeny-test / family size

INTRODUCTION

Il existe beaucoup de travaux relatifs à des méthodes de détection de gènes majeurs (Elston et Stewart, 1971 ; Hanset, 1982 ; Le Roy *et al.*, 1989), à l'effet de tels gènes sur la stratégie d'amélioration génétique (Smith et Webb, 1981), ou encore à l'existence de gènes majeurs influant sur des caractères quantitatifs dans une espèce (chez la poule : Mérat, 1990).

Concernant le problème particulier de la détection d'effets quantitatifs associés à un gène décelé par un effet visible, cette détection implique la comparaison pour le caractère quantitatif étudié de différents génotypes au locus incriminé. Souvent, le gène marqueur présente une dominance et l'on se contente de comparer des individus des phénotypes récessif et dominant de même origine parentale, par exemple de génotype aa et Aa , obtenus par un croisement de type $Aa \times aa$.

S'il apparaît une différence quantitative entre ces 2 génotypes, il peut être indiqué de chercher si les génotypes AA et Aa diffèrent eux aussi de ce point de vue. Dans le cas de dominance complète, assez fréquente par exemple chez la poule (Smyth, 1990 ; Somes, 1990), cette recherche est plus laborieuse que la première, car elle nécessite d'identifier les génotypes par un test de descendance : chaque individu de phénotype dominant ne peut *a priori* avoir son génotype identifié que par un croisement avec un « testeur » récessif aa ou hétérozygote Aa . Pour ce testage, on dispose en général d'une capacité limitée en ce qui concerne en particulier le nombre total des descendants que l'on peut faire naître ou élever à un moment donné. Ceci étant, il est possible de tester un plus grand nombre de parents avec peu de descendants pour chacun, ou moins de parents avec davantage de descendants pour chacun, et il est souhaitable de savoir s'il existe une solution optimale. N'ayant pas connaissance de travaux antérieurs sur ce point, nous avons voulu en faire ici une première étude théorique.

MATÉRIEL ET MÉTHODES

Hypothèses et notations

Nous considérerons le cas d'une dominance complète de l'allèle A , de sorte que, supposant d'abord l'absence d'informations sur la structure familiale de la population, nous ne pourrions identifier le génotype (AA ou Aa) parmi les individus de phénotype $[A]$ que par test de descendance (*progeny-test*). Nous supposerons qu'il y a un nombre maximal N de descendants pouvant être obtenus par *progeny-test*, à partir de n individus testés (par croisement avec un récessif aa , ou à défaut avec un hétérozygote) et x descendants par parent ($N = nx$).

Soient p la fréquence de l'allèle récessif a dans la population supposée en équilibre de Hardy-Weinberg dans un modèle déterministe, $q = 1 - p$, et d la différence moyenne, supposée non nulle, entre les génotypes AA et Aa , pour un caractère quantitatif.

Objectif recherché et formulation

Le but est de savoir s'il y a dans ces conditions un nombre optimal de descendants par parent testé – et corrélativement un nombre optimal d'individus testés – selon

les valeurs de p et de d . L'optimum choisi est la valeur pour laquelle la variable Z qui compare statistiquement, pour le caractère étudié, la moyenne des hétérozygotes identifiés à celle des individus non reconnus hétérozygotes a une valeur maximale : Z tendra à être abaissée par une plus grande proportion d'individus réellement hétérozygotes mais non identifiés comme tels, et en conséquence confondus avec les homozygotes dominants, si le nombre de descendants par parent est limité ; par ailleurs, Z tendra à croître avec le nombre d'individus testés dans les 2 groupes dont la moyenne est comparée.

Supposant un nombre de descendants égal quel que soit le génotype du parent, une proportion attendue $1 - (1/2)^x$ d'individus hétérozygotes testés sera identifiée en croisement avec un récessif aa (probabilité qu'il y ait au moins un descendant aa parmi x enfants). Nous négligerons les variations possibles de la proportion réellement obtenue autour de la valeur attendue.

La fréquence des hétérozygotes identifiés parmi les individus testés (de phénotype dominant) dans la population sera donc :

$$\frac{2pq}{q^2 + 2pq} [1 - (1/2)^x]$$

De même, la fréquence des individus testés non identifiés comme hétérozygotes (comprenant les individus AA et les Aa non reconnus) sera

$$\frac{q^2 + 2pq(1/2)^x}{q^2 + 2pq}$$

Si l'on attribue la valeur moyenne 0 au génotype Aa et d à AA (ignorant ici les effets génétiques autres que ceux associés au locus A sur le caractère considéré), la différence moyenne entre les 2 groupes obtenus sera

$$m_2 - m_1 = \frac{d q^2}{q^2 + 2pq(1/2)^x}$$

Appelant respectivement n_1 et n_2 le nombre d'individus testés classés dans les 2 groupes définis ci-dessus, on a :

$$\begin{aligned} n_1 + n_2 &= \frac{N}{x} \\ n_1 &= \frac{N}{x} \frac{q^2 + 2pq(1/2)^x}{q^2 + 2pq} \text{ et} \\ n_2 &= \frac{N}{x} \frac{2pq[1 - (1/2)^x]}{q^2 + 2pq} \end{aligned}$$

Supposant que la variance intragénotype du caractère considéré (AA ou Aa) est égale à 1 (données exprimées en variable normale réduite), appelons respectivement S_1^2 et S_2^2 la variance des individus testés reconnus hétérozygotes et celle des individus non reconnus comme tels. On a $S_1^2 = 1$ et S_2^2 représente la variance d'un mélange

de 2 distributions, chacune de variance unité, celle des individus AA et celle des Aa non identifiés. Reprenant la notation m_2 pour la valeur moyenne de ce groupe et rappelant que la moyenne du génotype AA a été prise pour d et celle du génotype Aa pour 0, on peut écrire :

$$S_2^2 = 1 + f_{Aa}(-m_2)^2 + f_{AA}(d - m_2)^2$$

en appelant respectivement f_{Aa} et f_{AA} les fréquences des génotypes Aa et AA dans le groupe «mélange». On peut montrer que :

$$f_{Aa} = \frac{2pq(1/2)^x}{q^2 + 2pq(1/2)^x}$$

$$f_{AA} = \frac{q^2}{q^2 + 2pq(1/2)^x}$$

En reprenant la valeur obtenue précédemment pour m_2 , on voit que l'expression complète de S_2^2 est

$$S_2^2 = 1 + \frac{2pq(1/2)^x}{q^2 + 2pq(1/2)^x} \cdot \left[\frac{-dq^2}{q^2 + 2pq(1/2)^x} \right]^2$$

$$+ \frac{q^2}{q^2 + 2pq(1/2)^x} \cdot \left[d - \frac{dq^2}{q^2 + 2pq(1/2)^x} \right]^2$$

Finalement Z est défini de la manière habituelle :

$$Z = \frac{m_2 - m_1}{\sqrt{\frac{s_1^2}{m_1} + \frac{s_2^2}{m_2}}}$$

Z peut être considéré en première approximation comme une variable normale réduite dont la valeur, pour N et d données, dépend de x et de p .

Pour p donné, l'expression de la dérivée de Z par rapport à x n'est pas simple et son annulation donne une équation qui ne paraît pas soluble algébriquement, mais un calcul par points indique l'existence de valeurs optimales de x et précise ces valeurs pour différentes fréquences du récessif et différentes valeurs de d .

Un calcul analogue peut être fait si l'on suppose que le testeur est de génotype Aa et non aa. Dans ce cas, la probabilité de n'obtenir aucun descendant récessif dans la descendance d'un individu devient $(3/4)^x$ au lieu de $(1/2)^x$.

RÉSULTATS ET DISCUSSION

Le tableau I présente, en fonction des considérations qui précèdent, et pour diverses valeurs de p et d , le nombre optimal x_M de descendants à obtenir par parent testé en croisement avec un récessif aa , pour maximiser la variable Z dans le test statistique de comparaison des moyennes.

Tableau I. Identification des individus AA et Aa par épreuve de descendance avec homozygote récessif. Nombre optimal de descendants par parent en fonction de la fréquence de l'allèle a (p) et de la différence entre AA et Aa (d).

p	d							
	0,1	0,3	0,5	0,7	0,9	1,1	1,3	1,5
0,1	1	1	1	1	1	1	1	1
0,2	1	1	1	1	1	1	1	1
0,3	1	1	1	1	1	1	1	1
0,4	1	1	1	1	1	2	2	2
0,5	2	2	2	2	2	2	2	3
0,6	2	2	2	3	3	3	3	3
0,7	3	3	3	3	4	4	4	4
0,8	4	4	4	4	5	5	5	5
0,9	6	6	6	6	6	6	6	6

Le tableau II montre les valeurs optimales lorsque le testeur est de génotype Aa.

Tableau II. Identification des individus AA et Aa par épreuve de descendance avec hétérozygote. Nombre optimal de descendants par parent en fonction de la fréquence de l'allèle a (p) et de la différence entre AA et Aa (d).

p	d							
	0,1	0,3	0,5	0,7	0,9	1,1	1,3	1,5
0,1	1	1	1	1	1	1	1	1
0,2	1	1	1	1	1	1	1	1
0,3	1	1	1	1	1	2	3	3
0,4	2	2	2	2	3	4	4	5
0,5	4	4	4	4	5	5	6	6
0,6	6	6	6	6	7	7	7	8
0,7	8	8	8	8	9	9	9	10
0,8	10	10	10	10	10	10	10	10
0,9	10	10	10	10	10	10	10	10

En réalité, la variable Z n'est pas rigoureusement une variable réduite, puisque l'un des groupes dont les moyennes sont comparées est un mélange de distributions. Cependant, dans les cas usuels où la proportion d'individus Aa non reconnus est limitée ainsi que la valeur de d , et compte tenu de la robustesse du test de comparaison des moyennes, il est vraisemblable que ce fait puisse, en général, être négligé et n'affecte pas de façon appréciable les conclusions qui suivent.

En testage avec un récessif aa, on constate que pour une fréquence $p < 0,5$, le test optimal est le plus souvent obtenu avec un seul descendant par parent, et corrélativement avec le nombre maximal de parents testés.

Pour $p \geq 0,5$, l'optimum se déplace vers des valeurs plus grandes de x . Cependant, l'optimum ne dépasse 4 que pour $p = 0,8$ et 6 pour $p = 0,9$. Dans le cas d'un testage avec un conjoint Aa, la valeur optimale de x n'est égale à l'unité que pour $p \leq 0,3$ et atteint 10 à partir de $p = 0,8$. On note d'autre part que les variations de d ont très peu d'influence sur le résultat.

Le gain de précision obtenu en choisissant pour x une valeur optimale peut être relativement important. À d et N égaux, le tableau III compare la variable Z obtenue d'une part avec x optimal et d'autre part avec x égal respectivement à 1 et à 6 pour diverses valeurs de p et quelques valeurs de d . On voit que des valeurs optimales augmentent la valeur Z dans une proportion pouvant dépasser 50% dans certains cas.

Tableau III. Identification des individus AA et Aa par épreuve de descendance avec homozygote récessif. Rapport de la variable Z obtenue avec un nombre de descendants par individu (x) égal à 1 ou 6 à la valeur obtenue avec x optimal, en fonction de la fréquence de l'allèle a (p) et de la différence entre AA et Aa (d).

p	d							
	0,1		0,5		1,0		1,5	
	Comparaison avec $x = 1$ $x = 6$							
0,1	1	0,60	1	0,61	1	0,63	1	0,65
0,2	1	0,64	1	0,65	1	0,68	1	0,72
0,3	1	0,68	1	0,69	1	0,73	1	0,79
0,4	1	0,73	1	0,75	1	0,79	0,98	0,83
0,5	1	0,80	0,99	0,81	0,98	0,83	0,95	0,88
0,6	0,97	0,86	0,96	0,87	0,93	0,88	0,90	0,91
0,7	0,90	0,91	0,90	0,92	0,88	0,93	0,84	0,95
0,8	0,80	0,97	0,79	0,97	0,78	0,97	0,75	0,98
0,9	0,62	1	0,62	1	0,62	1	0,60	1

On pourrait enfin considérer le cas où une information complémentaire est disponible, à savoir la connaissance des familles où un individu récessif au moins est apparu à partir du croisement de parents de phénotypes dominants, lors de la reproduction de la génération à tester. Ces familles correspondent au croisement de type Aa \times Aa. Parmi leurs descendants de phénotype [A] à soumettre individuellement au *progeny-test* pour identifier leur génotype, la fréquence du récessif dans ces familles est égale à $1/2$.

Pour p prenant une telle valeur, on a vu précédemment que la valeur optimale x_M est 2 ou 3 avec un testeur aa. Dans les autres familles, la fréquence p' du récessif sera différente de p caractérisant l'ensemble de la population si $p \neq 1/2$. L'examen de la fréquence respective des génotypes AA et Aa dans ces familles suggère que les valeurs optimales de x peuvent alors être modifiées, de façon relativement faible cependant.

RÉFÉRENCES

- Elston RC, Stewart J (1971) A general model for the genetic analysis of pedigree data. *Hum Hered* 21, 523-542
- Hanset R (1982) Major genes in animal production, examples and perspectives : cattle and pigs. In : *2nd World Congr Genet Appl Livest Prod*. Editorial Garsi, Madrid, 4-8 oct 1982, 5, 439-453
- Le Roy P, Elsen JM, Knott S (1989) Comparison of four statistical methods for detection of a major gene in a progeny-test design. *Genet Sel Evol* 21, 341-357
- Mérat P (1990) Pleiotropic and associated effects of major genes. In : *Poultry Breeding and Genetics* (Crawford RD, ed) Elsevier, Amsterdam, 429-467
- Smith C, Webb AJ (1981) Effects of major genes on animal breeding strategies. *Z Tierz Züchtungsbiol* 98, 161-169
- Smyth JR Jr (1990) Genetics of plumage, skin and eye pigmentation in chickens. In : *Poultry Breeding and Genetics* (Crawford RD, ed) Elsevier, Amsterdam, 109-167
- Somes RG Jr (1990) Mutations and major variants of plumage and skin in chickens. In : *Poultry Breeding and Genetics* (Crawford RD, ed) Elsevier, Amsterdam, 169-208