METHODS FOR THE DETECTION OF LINKAGE BETWEEN A MARKER GENE AND A LINKED POLYGENE AFFECTING A QUANTITATIVE TRAIT

Métodos para la detección del ligamento entre un gene marcador y un poligene ligado afectando a un carácter cuantitativo

Méthodes pour détecter la liaison entre un gène marqueur et un polygène lié, affectant un caractère quantitatif

M. SOLLER *

Two approaches for the detection of linkage between a marker gene and a polygene affecting a quantitative trait have been described in the literature. The first (1) is appropriate when marker gene and polygene are both at intermediate frequencies in the population of interest. This approach involves an analysis of variance between full sibs or half-sibs grouped according to their marker genotypes. The second (2) is appropriate when both marker gene and polygene are at high frequencies in the population of interest. This method involves crossing to a tester strain that is close to fixation for the alternative alleles of both marker gene and polygene; followed by an analysis of variance between backcross or F_2 progeny grouped according to their marker genotype. The purpose of this contribution is to obtain some notion of the size of the experiments that would be needed in order to detect linkage under each of the above experimental designs.

Initially we will assume complete linkage (proportion of recombination, $r = 0$). The effect of recombination will be discussed later. In the case of designs involving crosses to a tester line we assume complete fixation of alternative alleles in the population of interest and in the tester. We also assume that the marker gene has no pleiotropic effect on the quantitative trait; that all three marker genotypes, *MM, Mm* and *mm* can be distinguished phenotypically (as would be the case, e. g., for biochemical or blood group markers); and that the three quantitative genotypes, *AA, Aa* and *aa* have phenotypic values of $+D$, 0, and $-D$, respectively (i. e., no dominance for the polygene).

---

* Department of Genetics, The Hebrew University, Jerusalem, Israel.

1. *Population segregating for marker gene and polygene, halb-sib case*

This situation has been previously discusse din some detail by NEIMANN-SORENSEN and ROBERTSON (1961). In a population at linkage equilibrium for a segregating marker gene and a closely linked polygene, individuals heterozygous for the marker gene will be of four types with respect to the polygene, namely: *MA/ma*, *Ma/mA*, *MA/mA* and *Ma/ma*. Consider the double heterozygote, *MA/ma*. Assuming complete linkage, offspring of such an individual that can be identified as receiving the marker, *M* (i. e., *MM* offspring) will also receive the linked polygene, *A*; offspring that can be identified as receiving the marker, *m* (i. e., *mm* offspring) will also receive the linked polygene, *a*. Thus, in this case, we expect a difference, *d*, between the two groups of offspring out of the same parent. For the other double heterozygote, *Ma/mA*, we expect a difference of $-d$. For such parents, therefore, in an analysis of variance, the expected mean square (EMS) estimated from offspring grouped according to their marker genotype (*MM* or *mm*, *n* offspring per group), would be $\sigma^2_w + 1/2\,nd^2$. However, if p and q are the frequencies of the alleles *A* and *a* respectively, only $2pq$ of heterozygous marker parents will also be heterozygous for the linked polygene. In $p^2 + q^2$ of the cases, the heterozygous marker parent will be homozygous (*AA* or *aa*) for the linked polygene. In these cases the EMS estimated from offspring grouped according to their marker genotype would σ^2_w. It can be shown that in a mixed population of N parents (N fairly large), the pooled EMS between marker genotypes is $\sigma^2_w + 1/2\,(2pqd^2)$. Now $2pqd^2$ is the contribution of an additive *locus* to the total variance of a quantitative trait. The fraction of the total variance of the quantitative trait attributable to the *locus* in question would then be $2pdq^2/\sigma^2_w$. We will designate this fraction by D. [D is analygous to the heritability (h^2), except that heritability refers to that fraction of the total variance due to all additive gene effects, while D is the fraction of total variance due to the additive effects of a particular *locus*]. Thus if N parents are tested, and n offspring are examined in each of the two interesting marker classes (*MM* and *mm*) for each parent, the analysis of variance in this case takes the following form:

Source	df	EMS/σ^2_w
Between marker genotypes	N	$1 + 1/2\,nD$
Within marker genotypes	$2N\,(n\text{-}1)$	1

2. *Backcross to a tester strain*

Assuming that the initial genotypes of the population of interest and of the tester were *MA/MA* and *ma/ma*, respectively, then the backross offspring would be of two sorts, say, *MA/ma* and *ma/ma* (this would depend on the direction of the backcross). Again there would be a difference, *d*, between offspring grouped according to their marker genotype. However, in this case all F_1 parents will be heterozygous for both marker gene and polygene, and the sign of the difference

between offspring grouped according to their marker genotype will be the same for all F_1 parents. Hence, in the analysis of variance, all offspring having the same marker genotype can be grouped together irrespective of their parent. In this case the difference, d, between marker genotypes will contribute $1/2\,Nnd^2$ to the pooled EMS between marker genotypes. We can express this in terms of D, by defining D in this case as the fraction of the total variance that the *locus* would contribute if it were segregating in a population with $p = 0.5$. Then $1/2\,Nnd^2 = 1/2\,Nn\,(2pqd^2)/2pq = Nnd$ is the contribution of the linked quantitative *locus* to the EMS between marker genotypes. The entire analysis then takes the following form:

Source	*df*	EMS/σ^2_w
Between marker genotypes	1	$1 + Nnd$
Within marker genotypes	$2N\,(n\text{-}1)$	1

3. F_2 *out of a cross to a tester strain*

On pour usual assumptions, F_2 offspring will be *MA/MA*, *MA/ma* and *ma/ma*. The analysis of variance is based on the two homozygous classes, *MA/MA* and *ma/ma*. The difference between them will be $2d$, and it can readily be shown that this difference will contribute $4NnD$ to the EMS between the two marker genotypes. Thus, the analysis of variance will be the same as in the backcross design, except that the EMS between marker genotypes is $1 + 4NnD$.

SIZE OF THE EXPERIMENTS

For analyses of variance having the above form, simple manipulations of standard textbook formulas (3) yield

$$n = \frac{\Phi^2_N\,(N + 1)}{NkD}$$

where,

- n is the number offspring required per marker genotype, per parent.
- N is the number of parents.
- Φ_N is obtained from tables of powers of the F-test (3); and
- k is a function of the experimental design, $= 1/2$, for the segregating population; 1, for the backcross design, and 4, for the F_2 design.

The total number of offspring required will, of course, be much greater than n. There will be N parents, and two offspring groups are examined for each parent. In addition, in both the segregating design, and the F_2 tester design, heterozygous offspring are not included in the analyses, and these will comprise about half the total number of offspring in each case. Thus, the total number of offspring raised will be $2nN$ for the backcross design, and $4nN$ for the segregating and F-2 tester designs.

TABLE 1

N	$\Phi^2_N\ (N+1)/N$
1	10.60
5	3.31
10	2.10
15	1.58
20	1.32
40	0.83

Table I gives values of $\Phi^2_N (N+1)/N$, for selected values of N, Φ chosen so as to give Type I error, $\alpha = 0.05$, Type II error, $\beta = 0.10$. For the backross and F_2 designs, since offspring of all parents are pooled in the analysis of variance, Φ_1 should be used. These values can be converted to actual experimental sizes by substituting k and D according to the design and the trait, and by multiplying the solution for n by $2N$ or $4N$ as discussed above.

D is obviously the crucial unknown. D will depend on the total proportion of the variance due to additive genetic *loci* (i.e., on the heritability, h^2). But it will also depend on the number of *loci* (denoted by, b), contributing to h^2, specifically, $D = h^2/b$. The great question, therefore, on which the success or failure of experiments of the sort described above will depend, is the magnitude of b. If most of the additive genetic variance in the population comes from 5-10 segregating *loci*, then D for individual *loci* will often be in the range 0.01-0.10, and experiments of achievable magnitude may be able to detect closely linked polygenes. But if 50-100 *loci* contribute to the additive genetic variance in the population, then for individual *loci*, D will usually be in the range 0.001-0.01, and experiments of reasonable magnitude will generally be powerless to detect linkage. Be that as it may, Table 2 shows the numbers of offspring required for various values of D, according to the three designs. When heritabilities are in the

TABLE 2

Design	Number of offspring required							
	$D = 0.01$		$D = 0.02$		$D = 0.05$		$D = 0.10$	
	Per parent	total	Per parent	total	Per parent	total	Per parent	total
Segregating population:								
$N = 10$	1680	16800	840	8400	336	3360	168	1680
$N = 20$	1056	21120	528	10560	211	4224	106	2112
$N = 40$	644	25760	322	12880	129	5152	64	2576
Backcross *		2120		1060		424		212
F-2 *		1060		530		212		106

* In these designs only total number of offspring is given, since it is immaterial how many offspring each particular parent has.

range of 0.2 = 0.4, and b is in the range 5-10, then experiments based on 5,000-10,000 offspring in a segregating population, or 1000-2000 offspring in the back-cross of F_2 tester design should have good chances of detecting a closely linked polygene.

Effect of recombination

The effect of recombination in all three designs is to reduce the expected difference between the two marker classes by $(1\text{-}2r)$. Hence it will reduce the contribution of the linked polygene to the EMS between marker genotypes by $(1\text{-}2r)^2$. It is clear that once r is greater than 0.15-0.20, chances of detecting a linked polygene drop precipitously. However, even so, depending on the average number of chiasmata, each marker will cover about 20 % of its chromosome fairly effectively. Thus, if only b is not large, experiments involving a number of markers and a number of traits simultaneously, have good chances of being productive.

SUMMARY

Two designs for the detection of linkage between a marker gene and a polygene affecting a quantitative trait are evaluated. The first, appropriate when marker gene and polygene are both at intermediate frequencies, involves an analysis of variance between half-sib progeny groups. The second, appropriate when marker gene and polygene are at high frequencies in the population of interest, involves crossing to a tester strain, followed by a backcross of F_2 generation. If differences within the segregating population, or between the population of interest and the tester are determined by 5-10 *loci*, then experiments based on 5,000-10,000 offspring for the half-sib design, or 1000-2000 offspring for the backross of F_2 tester design, sohuld have good chances of detecting a polygene located within + 20 centimorgans of the marker.

RESUMEN

Se han valorado dos esquemas para la detección del ligamiento entre un gene marcador y un poligene que afecta a un carácter cuantitativo. El primero, adecuado cuando el gene marcador y el poligene son ambos de frecuencias intermedias, supone un análisis de la varianza entre el grupo de progenie de medio hermanos. El segundo, apropiado cuando el gene marcador y el poligene son de altas frecuencias en la población interesada, se efectúa mediante cruzamiento con una estirpe testigo, seguido por un cruce retrógado en la generación F_2. Si se determinan diferencias entre la población en segregación, o entre la población interesada y el testigo, en 5-10 *loci*, entonces los experimentos basados sobre 5.000-10.000 descendientes del esquema de medio hermanos, o sobre 1.000-2.000 descendientes para el cruce retrógrado de F_2, deberían dar una buena oportunidad para detectar un poligene situado dentro de + 20 centimorgans del marcador.

RESUME

On a évalué deux schémas pour détecter la liaison entre un gène marqueur et un polygène affectant un caractère quantitatif. La premier, adéquat quand le gène marqué et le polygène sont tous les deux de fréquences intermédiaires, suppose un analyse de la variance entre le group de lignée de demi-fratries. Le deuxième, adéquat quand le gène marqueur et le polygène ont de hautes fréquences dans la population intéressée, est réalisé au moyen de croisement avec une souche témoine, suivi d'un croisement rétrograde dans la génération F_2. Si l'on détermine des différences entre la population en ségrégation, ou entre la population intéressée et la témoine, dans les *loci* 5-10, alors, les expériences basées sur 5.000-10.000 descendants du schéma de demi-fratries, ou sur 1.000-2.000 descendants pour le croisement rétrograde de F_2, devraient fournir une excellente opportunité pour détecter un poligène placé dans $+ 20$ centimorgans du marqueur.

REFERENCES

1. NEIMAN-SÖRENSEN, A., and ROBERTSON, A. (1961): The association between blood groups and several production characteristics in three Danish cattle breeds. *Acta Agric. Scand.*, 11:163-196. HASEMAN, J. K., and ELSTON, R. C. (1972): The investigation of linkage between a quantitative trait and a marker *locus*. *Behavior Genetics*, 2:3-19. JAYAKAR, S. D. (1970): On the detection and estimation of linkage between a *locus* influencing a quantitative character, and a marker *locus*. *Biometrics*, 26:451-464.
2. THODAY, J. M. (1961): Location of polygenes. *Nature*, 191:368-370.
3. SCHEFFE, H. (1959): *The analysis of variance*. John Wiley and Sons, Inc., N. Y.