



HAL
open science

Theoretical aspects of applying sib–pair linkage tests to livestock species

Ku Götz, L Ollivier

► **To cite this version:**

Ku Götz, L Ollivier. Theoretical aspects of applying sib–pair linkage tests to livestock species. *Genetics Selection Evolution*, 1992, 24 (1), pp.29-42. hal-00893942

HAL Id: hal-00893942

<https://hal.science/hal-00893942>

Submitted on 11 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Theoretical aspects of applying sib-pair linkage tests to livestock species

KU Götz¹, L Ollivier²

¹ *Institut für Tierzucht und Haustiergenetik der Universität Göttingen, Albrecht-Thaer Weg 1, D-3400 Göttingen, Germany;*

² *Institut National de la Recherche Agronomique, Station de Génétique Quantitative et Appliquée, 78352 Jouy-en-Josas Cedex, France*

(Received 26 March 1991; accepted 12 December 1991)

Summary – The Haseman–Elston (HE) sib-pair linkage test in its original form is computationally simple but suffers from low power. With the advent of highly polymorphic markers, the exclusive use of fully informative matings (*ie* matings where the number of genes identical by descent for any sib pair can be inferred without error) for the HE test becomes feasible. This article examines the influence of highly polymorphic marker systems (5 alleles), large family sizes (6 full-sibs) and hierarchical breeding structures (mating ratio of 25) on the power of the HE test by means of simulation studies. Simulations are performed under the assumption that the costs of marker genotyping are a limiting factor for marker–QTL linkage studies. Consequently, the total number of individuals (parents and offspring) typed is fixed at 5 000 in each of the situations compared. The results show that the power of the HE test is considerably increased when both highly polymorphic markers and large full-sib families are available. For example, for a locus explaining 8% of the phenotypic variance the power of the test increases from 14 to 74% if the locus has 5 alleles instead of 2 and sibship size is 6 instead of 2. Hierarchical breeding structures tend to further increase the power of the test, for the example given from 74 to 79%.

linkage / marker gene / quantitative trait locus / Haseman–Elston test / power

Résumé – **Aspects théoriques de l'application du test de liaison génétique par les couples de germains aux espèces animales domestiques.** *Dans sa forme originelle, le test de liaison génétique de Haseman–Elston (HE), basé sur les couples de germains, est simple à calculer, mais statistiquement peu puissant. Avec des marqueurs hautement polymorphes, l'utilisation exclusive d'accouplements totalement informatifs (ie des accouplements permettant d'établir avec certitude le nombre de gènes d'origine identique pour n'importe quel couple de germains) peut être envisagée. Cet article examine, à l'aide de simulations, l'effet d'un système génétique hautement polymorphe (5 allèles également fréquents), d'une grande taille de fratrie (6 germains) et d'une structure d'élevage polygynique (25 femelles accouplées à chaque mâle) sur la puissance du test HE. Les simulations sont faites en supposant que le coût des typages génétiques est le facteur limitant des études de liaisons*

entre gènes marqueurs et locus de caractères quantitatifs. En conséquence, le nombre total d'individus typés (parents et descendants) est fixé à 5 000 dans chacune des situations comparées. Les résultats montrent que la puissance du test HE est considérablement accrue quand on dispose à la fois de marqueurs hautement polymorphes et de grandes fratries. Ainsi, pour un locus expliquant 8% de la variance phénotypique du caractère, la puissance du test au seuil de 5% est de 0,74 au lieu de 0,14 quand on passe de 2 à 5 allèles au locus marqueur et de 2 à 6 frères par fratrie. Les structures d'élevage polygyniques tendent à accroître encore la puissance du test, qui, dans l'exemple ci-dessus, passe de 0,74 à 0,79 avec une structure de 25 fratries issues du même père, par rapport à des couples de parents indépendants.

liaison génétique / gène marqueur / locus quantitatif / test de Haseman-Elston / puissance

INTRODUCTION

The sib-pair linkage method of Haseman and Elston (1972) is a tool for the detection of linkage between markers and quantitative trait loci (QTLs). The major advantage of the Haseman-Elston method is its computational ease, allowing fast screening of a large number of marker loci and traits. Further, the method is robust for a large variety of continuous distributions of the quantitative trait (Blackwelder and Elston, 1982). However, in its original form the method suffers from low power (Robertson, 1973), except when the effect of the QTL is very high and linkage between marker and QTL is tight (Blackwelder and Elston, 1982).

In typical animal breeding situations, the possibilities for the estimation of effects are often more advantageous than in human populations. Usually one has larger families, complete pedigree information and markers are available for parents and offspring. The advent of new, highly polymorphic markers such as minisatellites (Jeffreys *et al*, 1985) and microsatellites (Weber and May, 1988; Soller and Beckmann, 1990) increases the probability of informative matings and because of that also the probability for the detection of given linkage relationships.

The objective of this paper is to examine the power of the Haseman-Elston test in animal breeding situations under the assumption of the availability of highly polymorphic marker systems.

THEORY

The Haseman-Elston test

The linkage test of Haseman and Elston (1972) is based on the idea that the greater the number of alleles a pair of full-sibs shares identical by descent (*ibd*) at a marker locus which is linked to a QTL, the smaller the difference in the values of the quantitative trait which is affected by the QTL. A generalized description of the method is given by Elston (1990). The number of genes *ibd* at the trait locus for 2 full-sibs can either be 0, 1 or 2 but inference on this number depends on the parents'

and sibs' genotypes at the marker locus. The proportion of genes *ibd* at the marker locus for sib pair j (π_{jm}) is estimated from the parents' and offspring's genotypes and the regression of the squared difference of the sibs' phenotypic values on π_{jm} is calculated. If linkage between the marker and a QTL exists, this regression is expected to be negative.

Haseman and Elston (1972) assume random mating with respect to the marker locus, linkage equilibrium and no effect of the marker on the trait locus. The phenotypic value of the i th sib of the j th sib pair is assumed to be of the form:

$$x_{ij} = \mu + g_{ij} + e_{ij} \quad [1]$$

where μ is the overall mean, g_{ij} is the genotypic value at the trait locus and e_{ij} is the residual effect including genetic effects due to all loci other than the linked QTL. It is assumed that $e_j = e_{1j} - e_{2j}$ is a random variable with zero mean and variance σ_e^2 .

In a random mating population with a trait locus showing a given additive genetic variance (σ_a^2) and no dominance, the expectation of the squared difference of the 2 sibs' phenotypic values ($Y_j = [x_{1j} - x_{2j}]^2$) given the proportion of genes *ibd* at the trait locus (π_{jt}) is shown by Elston (1990) to be:

$$E(Y_j|\pi_{jt}) = 2\sigma_a^2 + \sigma_e^2 - 2\sigma_a^2\pi_{jt} \quad [2]$$

In practice π_{jt} is not known but has to be estimated by the number of genes *ibd* at the marker locus (π_{jm}). Elston (1990) shows the expectation of Y_j to be:

$$\begin{aligned} E(Y_j|\pi_{jm}) &= \sum_{\pi_{jt}} E(Y_j|\pi_{jt})P(\pi_{jt}|\pi_{jm}) \\ &= 2[1 - 2\theta(1 - \theta)]\sigma_a^2 + \sigma_e^2 - 2(1 - 2\theta)^2\sigma_a^2\pi_{jm} \end{aligned} \quad [3]$$

if there are only 2 alleles at the trait locus, where θ is the recombination frequency between the marker locus and the QTL. This expectation can be written in the form:

$$E(Y|\pi_{jm}) = \alpha + \beta\pi_{jm} \quad [4]$$

Blackwelder (1977; cited in Blackwelder and Elston, 1982) has shown that despite the non-normal distribution of Y , the distribution of the estimated regression coefficient (β) is asymptotically normal so that it is possible to use standard normal theory to test the hypothesis $H_0: \beta = 0$ against the one-sided alternative $H_1: \beta < 0$. Amos *et al* (1989) showed that the estimator of β is unbiased even if there is dominance at the trait locus, provided that information on the marker genotype of the parents is available.

Effect of marker polymorphism

Elston (1990) gives formulae for the estimation of π_{jm} from the parents' and sibs' genotypes at the marker locus. Since this study deals mainly with highly polymorphic markers, it is assumed here that a large number of informative matings

is available. In animal populations there are usually large full-sib families which makes it worthwhile to type the parents first and, according to these results, only the offspring of fully informative matings.

A fully informative mating in this sense is a mating where both parents are heterozygous and have at least 3 different alleles at the codominant marker locus which corresponds to mating types VI and VII in table II of Haseman and Elston (1972). Thus, the number of genes *ibd* for a sib pair can be inferred without error. The frequency of these matings in a given population depends on the number of alleles at the marker locus and their frequencies. In the general case of n alleles and unequal gene frequencies (p_i), the expected proportion of fully informative matings under random mating (*PFIM*) can be written as:

$$\begin{aligned}
 PFIM &= P(A_i A_j \times A_k A_l) + P(A_i A_j \times A_i A_k) + P(A_i A_j \times A_k A_j) \\
 &= \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2p_i p_j \left[\left(\sum_{k=1}^{n-1} \sum_{l=k+1}^n 2p_k p_l \right) - 2p_i p_j \right] \quad [5]
 \end{aligned}$$

Taking into account the number of animals usually tested in mapping experiments, it is unlikely for a system to be declared as highly polymorphic if one or two of the alleles show extreme frequencies. Therefore, the case of equal allelic frequencies ($p = 1/n$) is considered here and the proportion of informative matings is then given by:

$$PFIM = \left(\frac{n-1}{n} \right) \left(\frac{n-1}{n} - \frac{2}{n^2} \right) = \frac{(n-1)(n-2)(n+1)}{n^3} \quad [6]$$

Table I gives the expected *PFIMs* for various values of n . This proportion increases with increasing number of alleles. For loci with 9 alleles or more, less than 25% of the matings have to be rejected. For more than 15 alleles, the further increases in the proportion of fully informative matings are only small.

Table I. Expected proportion of fully informative matings (*PFIM*) in a random mating population with n alleles of equal frequencies.

n	<i>PFIM</i>	n	<i>PFIM</i>
3	0.296	9	0.768
4	0.469	10	0.792
5	0.576	15	0.863
6	0.648	20	0.898
7	0.699	30	0.932
8	0.738	40	0.949

Effect of breeding structure

Most animal populations have a very different structure from the human populations for whom this test was originally designed. In general the breeding structure for poultry, pigs and fish is favorable for effective testing because large full-sib families are available. Sheep and goat populations have an intermediate value for the application of the test whereas cattle populations show a very unfavorable structure.

Blackwelder and Elston (1982) show that, under the null hypothesis of no linkage, the $s(s - 1)/2$ sib pairs from a sibship of size s can be treated as independent without affecting the type I error rate, so that treating sib pairs as independent should provide a test with good power, and a correct nominal value. In addition to increased power, the study of large full-sib families requires typing of fewer parents resulting in a reduction of overall costs. Here, comparisons will be made for a given overall cost of typing, assuming that there is no limitation of the number of individuals measured for the trait of interest. In that case, if a proportion *PFIM* can be selected among the families available, the number of sibships (f) of size s which can be measured given a total number N of typed animals is:

$$f = \frac{N}{\left(s + \frac{2}{PFIM}\right)} \quad [7]$$

Table II gives the numbers of parents and offspring for the 3 variants which will be considered in the simulations. For the first variant ("standard"), for which all types of matings are used, 1 250 sib pairs from 1 250 families are generated, giving a total of 5 000 typed animals. From these 5 000 animals 2 500 have to be measured for the quantitative trait. The second variant has a *PFIM* of 0.576 (see table I for $n = 5$). From 1 587 typed couples $f = 914$ show a fully informative mating type. These 914 couples have a total of 1 828 progeny, resulting in a total of 5 002 ($2 * (1 587 + 914)$) typed animals. In the case of families with 6 sibs, 3 168 offspring from $f = 528$ families can be measured, an increase of 75% in the number of offspring as compared to the second variant.

Table II. Number of animals in the different categories with constant number of animals typed ($N = 5\ 000$).

<i>Variant</i>	<i>Sibs/ family</i>	<i>Parents typed</i>	<i>Families selected</i>	<i>Off- spring</i>	<i>Total typed</i>
Standard	2	2 500	1 250	2 500	5 000
Polymorphic	2	3 174	914	1 828	5 002
Polymorphic	6	1 834	528	3 168	5 002

Up to now, the families were assumed to be unrelated. However, in animal breeding populations one male is usually mated to several females. This gives rise to genetic covariances between families for the polygenic part of the genotype. Since

the variable Y_j is derived from the difference of the phenotypic values of 2 full-sibs, it is clear that there are no covariances between the Y_j 's of half-sib families that have one parent in common. Thus, the power of the Haseman-Elston test is not directly influenced by the genetic structure of animal populations.

However, if the cost of marker genotyping is a limiting factor and the families to be genotyped are selected in a 2-stage procedure, the number of measured offspring given a fixed number of assays can be considerably increased. We consider again the case of a population from which fully informative matings are selected for genotyping of the offspring with a fixed overall number of typed animals (N). In a first step only sires are typed and the heterozygotes are selected for genotyping of their mates. In a second step, heterozygous dams with at least one allele different from their mate's are selected and have their offspring genotyped. Then, the final number of families (f) measured for quantitative traits depends on: sibship size (s); mating ratio ($r =$ number of dams per sire); selection rate of sires (m_1); selection rate of dams given heterozygous sires (m_2).

The number of families is then:

$$f = \frac{N}{s + \frac{1}{m_2} \left(1 + \frac{1}{rm_1}\right)} \quad [8]$$

With n equally frequent alleles [6] can be written as $PFIM = m_1m_2$ with $m_1 = (n - 1)/n$ and $m_2 = [(n - 1)/n] - 2/n^2$. Note that [8] does not reduce to [7] when $r = 1$, because of the 2-step selection implied in [8]. Table III gives an example of the values of f for different mating ratios (r). For low values of r , the number of measured families increases rapidly with increasing mating ratio. The largest effect of this strategy can be observed for the case of a polymorphic marker in 2 sib-families. Beyond a ratio of 10 the value of f converges rapidly towards the limit for r equal to infinity, which is appropriate if only one male is used.

Table III. Number of families selected, given a total of 5 000 animals typed for different marker polymorphisms and mating ratios (r).

<i>Sibship size</i>	<i>No of marker alleles</i>	<i>Matings considered</i>	$r = 1$	$r = 10$	$r = 25$	$r = \infty$
2	2	All	1 250	1 613	1 644	1 666
2	5	Informative	976	1 403	1 445	1 475
6	2	All	625	704	710	714
6	5	Informative	606	661	670	678

Simulation studies

Simulations were performed to examine the impact of 4 factors on the power of the Haseman-Elston test: i) fully informative matings; ii) family size of 6 full-sibs; iii) within-family environmental correlation (c^2); and iv) a typical pig breeding

structure. Other factors varied were: 1) variance due to the linked QTL (relative to phenotypic variance) and 2) recombination frequency between the marker and the QTL.

METHODS

Data were simulated according to the following model in which e_{ij} of model (1) is developed in 5 terms:

$$x_{ij} = \mu + g_{ij} + \frac{1}{2}(bv_{sj} + bv_{dj}) + \phi_{ij} + ce_j + \varepsilon_{ij} \quad [9]$$

where

- x_{ij} = phenotypic value of the i th sib in the j th sibship
- μ = overall mean
- g_{ij} = effect of the QTL-genotype of sib i
- bv_{sj} = polygenic breeding value of the sire
- bv_{dj} = polygenic breeding value of the dam
- ϕ_{ij} = effect of Mendelian sampling on polygenic value
- ce_j = effect of common environment for the j th sibship
- ε_{ij} = random error

Table IV gives the range of variation for the different parameters. Each simulation was replicated 500 times. The sizes of the examined sibships were 2 and 6, respectively. For the larger sibship size the test was based on all possible sib pairs within the sibship, as proposed by Blackwelder and Elston (1982), resulting in 15 comparisons per sibship. The gene frequencies at the trait locus were $p = q = 0.5$ and additive gene action was assumed. At the marker locus the gene frequencies were 0.5 for the standard method (assuming 2 alleles) and 0.2 (corresponding to 5 alleles) for the case of fully informative matings. In the polyallelic case only the offspring of fully informative matings as defined above were considered as being typed and thus included in the analyses. All simulations were calculated for a total of 5 000 assays.

Table IV. Simulation parameters.

<i>Parameter</i>	<i>From</i>	<i>To</i>	<i>Step</i>
θ^1	0.0	0.1	0.05
QTL ²	0.02	0.16	*2
c^2 ³	0.0	0.25	0.25

¹ Recombination frequency; ² variance of quantitative trait locus effect relative to phenotypic variance; ³ variance of common family environment relative to phenotypic variance; h^2 : heritability (excluding QTL effect) assumed to be 0.25.

To check the significance level, simulations were performed under the null hypothesis (no effect of QTL or recombination frequency = 0.5) for all of the variants presented here. The empirical significance level was determined as the percentage of replications that, under the null hypothesis, exceeded the critical value of a 1-sided *t*-test with type I error of 5%. In none of the cases was this percentage significantly different from 5%, in a test based on the binomial distribution. The SAS univariate procedure (SAS, 1989) indicated no significant departure of the regression coefficients from normality.

RESULTS

In this section the *impact of fully informative matings* arising from highly polymorphic markers and of sibships of size 6 on the power of the Haseman–Elston test is examined. Columns 3 and 4 of table V show the effect of using only fully informative matings on the power of the Haseman–Elston test. The column for the standard version of the test shows the poor power of this method for the QTL effects considered here. When the QTL contributes 16% to the phenotypic variance, which is equivalent to 1.1 phenotypic standard deviations between the 2 homozygotes, the power is only 33%. The use of fully informative matings hardly increases power, except for higher QTL effects. However, even for the largest QTL effect the power of the test is below 50%.

Table V. Effect of standard *vs* fully informative matings and effect of family size on the power (in %) of the Haseman–Elston test, assuming a constant number of typed animals ($N = 5\,000$, $h^2 = 0.25$, $\alpha = 0.05$, 500 replications).

QTL ¹	θ^1	Fully informative		
		Standard 2 sibs 1 250 families	2 sibs 914 families	6 sibs 528 families
0.02	0.0	7	6	13
0.02	0.05	6	7	13
0.02	0.1	6	5	10
0.04	0.0	8	8	30
0.04	0.05	7	12	23
0.04	0.1	8	6	17
0.08	0.0	14	17	74
0.08	0.05	12	13	57
0.08	0.1	8	14	38
0.16	0.0	33	47	99
0.16	0.05	27	39	99
0.16	0.1	18	27	90
SE ²		1–2	1–2	0.4–2

¹ QTL, θ , h^2 , see footnotes to table IV; ² for an estimated power p (in %), $SE = (p(100 - p)/500)^{.5}$.

For lower QTL effects the power shows more or less erratic fluctuations with increasing recombination rate, while for a QTL effect of 16% the power is reduced

between 15 and 20 percentage points when the recombination rate increases from 0 to 0.1.

The *use of larger families* leads to a major increase in the power of the Haseman-Elston test. QTL effects of 8% can be detected with more than 50% power if the recombination rate is 0 or 0.05 (column 5, table V).

The effect of a *within-family environmental component* on the power of the test is given in table VI. This component reduces the within-family variance and thus increases the power of the test. The average increase is 59% of power for the 2-sib and 40% for the 6-sib families. In the latter situation, a QTL effect of 4% can be detected with nearly 50% power if a common environmental component is present and there is no recombination between the marker and the QTL.

Table VI. Effect of common environment on the power (in %) of the Haseman-Elston test (fully informative matings, $N = 5\ 000$, $h^2 = 0.25$, $\alpha = 0.05$, 500 replications).

QTL	θ	2-sib families ($N = 914$)		6-sib families ($N = 528$)	
		$c^2 = 0$	$c^2 = 0.25$	$c^2 = 0$	$c^2 = 0.25$
0.02	0.0	6	9	13	22
0.02	0.05	7	6	13	17
0.02	0.1	5	10	10	12
0.04	0.0	8	15	30	48
0.04	0.05	12	8	23	38
0.04	0.1	6	11	17	29
0.08	0.0	17	30	74	94
0.08	0.05	13	23	57	84
0.08	0.1	14	18	38	68
0.16	0.0	47	79	99	100
0.16	0.05	39	68	99	100
0.16	0.1	27	48	90	100
SE		1 - 2	1 - 2	0.4 - 2	0.4 - 2

QTL, θ , h^2 , c^2 : see footnotes to table IV.

In the simulations of *hierarchical breeding structures* the numbers of families for $r = 25$ were used. The results are given in table VII. It can be seen that in this situation the test based on 2-sib families is still not competitive. In the case of 6-sib families one should be able to detect a QTL effect of 8% with power between 48 and 79%. For smaller QTL effects power is not sufficient unless there is additional "support" from common environmental effects.

Soller and Genizi (1978) using the method of Jayakar (1970) presented calculations for half- and full-sib designs. The method of Soller and Genizi (1978) for a QTL contributing 4% of the phenotypic variance of the population has been compared to our simulation results, as summarized in table VIII. The base of the comparisons is an equal number of preselected matings for both tests (fully informative for Haseman-Elston, intercross for Soller and Genizi). The test of Soller and Genizi (1978) always has less power than the Haseman-Elston test for the two heritabilities that have been tested.

Table VII. Comparison of the powers (in %) of 3 different variants of the Haseman–Elston test for a total of 5 000 typed animals and a mating ratio of 25 ($N = 5\,000$, $h^2 = 0.25$, $c^2 = 0$, $\alpha = 0.05$, 500 replications).

<i>QTL</i>	θ	<i>Standard</i>	<i>Fully informative</i>	<i>Fully informative</i>
		<i>2 sibs</i> <i>1 644 families</i>	<i>2 sibs</i> <i>1 445 families</i>	<i>6 sibs</i> <i>670 families</i>
0.02	0.0	6	7	16
0.02	0.05	6	7	12
0.02	0.1	6	7	12
0.04	0.0	9	12	37
0.04	0.05	8	10	27
0.04	0.1	7	11	21
0.08	0.0	13	29	79
0.08	0.05	13	19	66
0.08	0.1	13	17	48
0.16	0.0	40	64	100
0.16	0.05	26	53	100
0.16	0.1	22	37	100
SE		1 – 2	1 – 2	0.4 – 2

QTL, θ , h^2 , c^2 : see footnotes to table IV.

Table VIII. Comparison of simulated powers of the Haseman–Elston test (fully informative matings) with calculated powers for the intercross design of Soller and Genizi (1978) (standard errors in brackets, 500 replications), assuming $QTL = 0.04$, $\theta = 0$ and $c^2 = 0$.

<i>No families</i> <i>(6 offspring)</i>	<i>Haseman–Elston</i> <i>(fully informative matings)</i>		<i>Soller and</i> <i>Genizi</i> <i>(intercross</i> <i>design)</i>
	$h^2 = 0.2$	$h^2 = 0.4$	
125	12 (1.4)	18 (1.7)	10
250	21 (1.8)	23 (1.9)	12
500	28 (2.0)	37 (2.1)	17
1 000	40 (2.2)	63 (2.2)	24

QTL, θ , h^2 , c^2 : see footnotes to table IV.

DISCUSSION

The present study confirms the findings of Robertson (1973) that the Haseman–Elston sib-pair linkage method in its original form has very low power. This is especially true if the variance explained by the QTL is small as compared to the residual variance, since the variance of Y_j is proportional to the fourth power of σ_e (Robertson, 1973). As a consequence, measures to increase the power of the test

should aim at a reduction of the residual variance. On the other hand, systematic environmental and sex effects that affect the difference between full-sibs would have to be eliminated. This, as well as the increase in power due to common environmental effects, leads to the recommendation that full-sib families should be reared together, as long as no competition effects occur.

Preselection of fully informative matings leads to an increase of the power of the test, especially for higher QTL effects. Furthermore, the power estimates found by simulation agree well with the values calculated according to the asymptotic formulae given by Blackwelder and Elston (1982) and Amos and Elston (1989). Thus, for a given mating structure and marker polymorphism, the necessary size of experiments for the detection of marker-QTL linkage can be calculated in advance for a given QTL effect. Power could also be increased by selective genotyping. Instead of selecting informative matings one would rather select sibships with at least one extreme individual for the trait investigated (see the 2-sib case recently analysed by Carey and Williamson (1991)).

Blackwelder and Elston (1982) concluded that a significance level of 5% as chosen in this study, would not be strong evidence for linkage in view of the low prior odds in favor of 2 genes being on the same chromosome. However, since this method is a preliminary instrument for the scanning of segregating populations, a relatively high type I error could be accepted in order to reduce the risk of rejecting possible QTLs.

A crucial question is whether power is influenced by using non-independent comparisons in large families. Results of Hodge (1984) suggest that the information contained in a sibship of size s approaches $2s - 3$ for $s > 4$. However, these findings could not be confirmed in simulation studies of Blackwelder and Elston (1982) and Amos *et al* (1989), who treated sibships of sizes 3 and 5 as in the present study. Their results indicate that neither type I error nor power are affected by treating all the comparisons in a sibship as independent. The same was true for the present study.

Furthermore, family sizes would not be equal in field studies. This is advantageous with respect to power since the expected number of comparisons per sibship increases with the variance of sibship size. A random variation of family size (s) in pig populations may be approximated by a Poisson distribution with parameter $\bar{s} = E(s) = V(s)$. Thus, as $E[s(s-1)/2] = [\bar{s}(\bar{s}-1) + V(s)]/2 = \bar{s}^2/2$, the average number of comparisons per family is always greater than with constant family size and twice as much for $\bar{s} = 2$. For pig populations, an average family size of 6 is a rather pessimistic assumption; in practice it would rather be between 8 and 9 and one such family would on average be equivalent to 32-40 families of 2 sibs.

The extension to hierarchical breeding structures increases the power if the families to be typed are selected in a 2-stage procedure. Relationships between families do not induce covariances between the Y_j and therefore need not be considered.

In a typical pig breeding situation the method can be used to detect QTLs of effects between 4 and 8% of the phenotypic variance with about 50% power for a total of 5 000 assays. This should include all economically interesting QTLs because in the supposed situation (linkage equilibrium) the phases must be determined for each sire in each generation if marker assisted selection is applied. Since the

determination of linkage phases causes additional cost, only QTLs with large effects are of interest.

Another question is whether the preselection of fully informative matings may give rise to false linkage. Preselection is made only with regard to the marker genotype. Therefore, false linkage should not occur as long as there is linkage equilibrium between the marker and the QTL. However, this cannot be proven until the QTL genotype can be determined directly. As suggested by an anonymous referee, inferences suggesting evidence in favor of linkage in the case of linkage disequilibrium should be correct. Furthermore, falsely positive evidence should not occur for unlinked loci which are in linkage disequilibrium when data for the parents are available. However, this has never been studied for sib-pair methods. The main reason for linkage disequilibrium in animal populations is hybridization between subpopulations that have been kept separate for several generations (Lande and Thompson, 1990). If this should be the case in the population considered, one would rather use the standard methodology of multiple regression which exploits the existing linkage disequilibria.

A comparison of the present results with those from other workers is difficult since only few investigations deal with the problem of detecting linkage within segregating populations. The comparison with the method of Soller and Genizi (1978) showed that the power of this method is inferior to the Haseman-Elston test. The reason is that the Soller and Genizi method favours half-sib families in the order of 1 500 animals and is thus better suited to dairy cattle populations.

Weller *et al* (1990) introduced the granddaughter design which leads to a considerable increase of power for a given number of assays compared with the Soller and Genizi (1978) design. It also depends on highly polymorphic markers and its range of application is limited to dairy cattle populations because this method requires large numbers of sons per sire and granddaughters per son to be effective.

Beckmann and Soller (1988) considered crosses between segregating populations. Their method is based on F_2 crosses and thus requires additional testing of F_1 individuals, with at least 2 generations and a special experimental design. It also implies 2 parental populations close to fixation ($p > 0.8$) for alternative QTL-alleles and a difference between the 2 homozygotes of at least 0.4 standard deviations. The authors suggest that this is more likely for traits like disease resistance than for the traditional performance traits.

CONCLUSIONS

With the advent of multiallelic markers, the Haseman-Elston sib-pair linkage method becomes more powerful for the detection of linkage between markers and QTLs. However, the marker should at least have 5 or more alleles at intermediate frequencies in order to reduce the number of parent animals to be typed. The method allows the detection of linkage within any segregating population. It can serve to indicate whether more sophisticated methods, that estimate recombination frequency and allele effects but require special mating plans, are appropriate. The method can also make use of multivariate data as shown by Amos *et al* (1990).

The power of the method increases considerably with increasing family size and with hierarchical breeding structures. Thus, it is especially useful for species with high reproductive rates such as pigs and poultry. Amos and Elston (1989) extended the test to any type of non-inbred relative pair. Thus, it could also be applied to dairy cattle where large half-sib families are available. However, on average twice as many half-sib pairs are needed as compared to full-sibs and at present there exists no possibility of combining different types of relatives in one analysis.

ACKNOWLEDGMENTS

The leave of KU Götz at the Station de Génétique Quantitative et Appliquée at INRA, Jouy-en-Josas, was supported by a grant from the Deutsche Forschungsgemeinschaft, which is gratefully acknowledged. The authors wish to thank two anonymous referees for valuable comments on the manuscript.

REFERENCES

- Amos CI, Elston RC (1989) Robust methods for the detection of genetic linkage for quantitative data from pedigrees. *Genet Epidemiol* 6, 349-360
- Amos CI, Elston RC, Wilson AF, Bailey-Wilson JE (1989) A more powerful robust sib-pair test of linkage for quantitative traits. *Genet Epidemiol* 6, 435-449
- Amos CI, Elston RC, Bonney GE, Keats BJB, Berenson GS (1990) A multivariate method for detecting genetic linkage, with application to a pedigree with an adverse lipoprotein phenotype. *Am J Hum Genet* 47, 247-254
- Beckmann JS, Soller M (1988) Detection of linkage between marker loci and loci affecting quantitative traits in crosses between segregating populations. *Theor Appl Genet* 76, 228-236
- Blackwelder WC (1977) Statistical methods for detecting genetic linkage from sibship data. Institute of Statistics, Mimeo Ser No 1114. University of North Carolina at Chapel Hill (cited in Blackwelder and Elston, 1982)
- Blackwelder WC, Elston RC (1982) Power and robustness of sib-pair linkage tests and extension to larger sibships. *Commun Statist-Theor Methods* 11 (5), 449-484
- Carey J, Williamson J (1991) Linkage analysis of quantitative traits: increased power by using selected samples. *Am J Hum Genet* 49, 786-796
- Elston RC (1990) A general linkage method for the detection of major genes. In: *Advances in Statistical Methods Applied to Livestock Production* (Hammond K, Gianola D, eds) Springer, Berlin, 495-506
- Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet* 1, 3-19
- Hodge SE (1984) The information contained in multiple sibling pairs. *Genet Epidemiol* 1, 109-122
- Jayakar SD (1970) On the detection and estimation of linkage between a locus influencing a quantitative character and a marker locus. *Biometrics* 26, 451-464
- Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable "minisatellite" regions in human DNA. *Nature (Lond)* 314, 67-73

- Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124, 743-756
- Robertson A (1973) Linkage between marker loci and those affecting a quantitative trait. *Behav Genet* 3, 389-391
- SAS Institute Inc (1989) *SAS/STAT User's Guide, Version 6*. SAS Institute Inc, Cary, NC, 4th edition, vol 1, 846 pp
- Soller M, Beckmann JS (1990) Molecular mapping of quantitative genes. In: *4th World Congr Genet Appl Livest Prod, Edinburgh, 23-27 July 1990* (Hill WG, Thompson R, Wooliams JA, eds) vol 13, 93-96
- Soller M, Genizi A (1978) The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. *Biometrics* 34, 47-55
- Weber JL, May PE (1988) An abundant new class of DNA polymorphism. *Am J Hum Genet* (suppl) 43, A161
- Weller JI, Kashi Y, Soller M (1990) Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *J Dairy Sci* 73, 2525-2537