Genet. Sel. Evol. 32 (2000) 547–560
© INRA, EDP Sciences

Original article

Selective genotyping for QTL detection using sib pair analysis in outbred populations with hierarchical structures

Dimitrios G. CHATZIPLIS[a,b,*], Chris S. HALEY[a]

[a] Roslin Institute (Edinburgh), Roslin, Midlothian, EH25 9PS, United Kingdom
[b] Present address: Ross Breeders Ltd, Newbridge Midlothian, EH28 8SZ Scotland, United Kingdom

(Received 31 January 2000; accepted 13 July 2000)

Abstract – A simulation study illustrates the effects of the inclusion of half-sib pairs as well as the effects of selective genotyping on the power of detection and the parameter estimates in a sib pair analysis of data from an outbred population. The power of QTL detection obtained from samples of sib pairs selected according to their within family variance or according to the mean within family variance within half sib family was compared and contrasted with the power obtained when only full sib pair analysis was used. There was an increase in power (4–16%) and decrease in the bias of parameter estimates with the use of half-sib information. These improvements in power and parameter estimates depended on the number of the half sib pairs (half sib family size). Almost the same power as that obtained using all the available sib pairs could be achieved by selecting only 50–60% the animals. The most effective method was to select both full and half sib pairs on the basis of high within full sib family variance for the trait in question. The QTL position estimates were in general slightly biased towards the center of the chromosome and the QTL variance estimates were biased upwards, there being quite large differences in bias depending on the selection method.

selective genotyping / QTL detection / sib pair analysis / outbred populations

Résumé – Génotypage sélectif pour la détection des QTL par l'analyse de fratries dans des populations non consanguines à structures familiales. Cette étude par simulation a porté sur les effets de l'inclusion des paires de demi-frères et du génotypage sélectif sur la puissance de détection des QTL et l'estimation de leur position et de leur variance dans des populations non consanguines et à structure familiale de germains et demi-germains. On a calculé la puissance de détection des QTL dans l'analyse des paires de demi-frères sélectionnés en fonction de leur variance intra-famille, et on l'a comparée à celle obtenue quand seulement des paires de germains étaient utilisées. Quand l'information sur les demi-frères a été incluse, on a observé une augmentation de la puissance de détection (4–16 %) et une diminution

* Correspondence and reprints
E-mail: dimitrios_chatziplis@rossbr.com

du biais de l'estimation des paramètres. Ces améliorations dépendaient du nombre de paires de demi-frères (taille des familles). Ainsi, en incluant les demi-frères, on atteignait sensiblement le même pouvoir de détection avec 40–50% d'animaux en moins que pour l'analyse à partir des germains seulement. La méthode la plus efficace a consisté à analyser conjointement des frères et des demi-frères provenant de familles avec une grande variance pour le caractère étudié. Les estimations des positions de QTL étaient légèrement biaisées en direction du milieu du chromosome, alors que celles de la variance des QTL étaient surestimées. Ces biais différaient fortement selon la méthode de sélection des données utilisée.

génotypage sélectif / QTL / germains / population ouverte

1. INTRODUCTION

Sib pair analysis is a method for QTL detection, which utilises information from full sib families, and which has been described for single [12] and bracket markers [8]. The principle behind sib pair analysis is that for any locus a pair of full sibs will share 0 or 1 or 2 alleles Identical By Descent (IBD) from their parents. Sib pairs, which share alleles IBD for a gene controlling a trait, are likely to have similar phenotypes for the trait compared to sib pairs which do not share alleles. Hence, the association of differences between the phenotypes of sib pairs and the number of alleles shared IBD at a marker provides a means for identification of a QTL near that marker. A simple test to detect these associations is the regression of the squared phenotypic difference between a sib pair on the proportion of alleles the sib pair shares identical by descent (IBD) [12].

The power of the sib pair methods increases with increasing family size [2, 10] and is comparable with other QTL detection methods using analysis of variance when applied in animal populations (*e.g.* pigs, [10]). Moreover, it has been shown [1] that the same regression method for the detection of QTL [12] can be used for any type of outbred relatives (*e.g.* half sibs, first cousins etc.). Götz and Hamann [9] have shown that full sib and half-sib information can be used simultaneously in a combined analysis, with a resulting improvement in the power of detection and the parameter estimates.

In a previous paper [6] it was shown that by selection of full sib pairs from families with high within family variance, the power of QTL detection can be maintained with a reduced number of genotyped individuals. In this paper, selected samples containing both full and half sibs were analyzed jointly in a single analysis. This was applied in populations with two different hierarchical structures. Furthermore, the power and parameter estimates from samples selected on the basis of the within family variance and comprising both full and half sibs were compared with the power and parameter estimates from randomly selected samples and samples only of full sibs.

2. MATERIALS AND METHODS

It has been shown that the expectations of the squared phenotypic differences conditional on the proportion of alleles shared IBD at a marker locus between

full sib pairs, when there is no dominance variation present, are [12]:

$$E(Y_j|\hat{\pi}_{jm}) = [\sigma^2_{e_{FS}} + 2(1 - 2\theta + 2\theta^2)\sigma^2_g] - 2(1 - 2\theta)^2\sigma^2_g\hat{\pi}_{jm} \quad (1)$$

where,

$E(Y_j|\hat{\pi}_{jm})$ = expected mean of squared phenotypic differences conditional on the estimated proportion of alleles shared IBD at a marker locus;

$\hat{\pi}_{jm}$ = estimated proportion of alleles shared IBD at a marker locus;

$\sigma^2_{e_{FS}}$ = contains environmental variance and covariance of full sibs and any order effect;

σ^2_g = additive genetic variance of the QTL;

θ = recombination fraction between the marker and the QTL;

Y_j = squared phenotypic differences of sib pairs.

Amos and Elston [1] have shown that the same holds for half sib pairs, with the only differences being the regression constant and the "environmental variance" component, which is that of half sibs ($\sigma^2_{e_{HS}}$):

$$E(Y_j|\hat{\pi}_{jm}) = [\sigma^2_{e_{HS}} + 2(1 - \theta + \theta^2)\sigma^2_g] - 2(1 - 2\theta)^2\sigma^2_g\hat{\pi}_{jm} \quad (2)$$

Therefore, the regression coefficient of the squared phenotypic differences on the proportion of alleles shared IBD at a marker locus is identical for both full and half-sib pairs. However, since the regression constants are different, the two regression lines are parallel to each other. Moreover, the proportion of alleles that a half-sib pair can share IBD can only be 0 or 0.5, whereas in the case of full sibs it can be 0, 0.5 or 1.

Götz and Hamann [9] suggested that by correcting for the differences between the regression constants:

$$\Delta_\alpha = \sigma^2_{e_{HS}} - \sigma^2_{e_{FS}} + 2\theta(1 - \theta)\sigma^2_g \quad (3)$$

both full and half sibs can be used in one combined analysis. However, the parameters of equation (3) (*e.g.* θ, σ^2_g etc.) are not known in advance, therefore, Δ_α must be estimated. The use of the expectations of the squared phenotypic differences conditional on the proportion of alleles shared IBD from the two regression equations (1) and (2), obtained when full and half sibs are analyzed independently can provide such an estimate. However, the proportion of alleles shared IBD in half sib pairs can only be 0 or 0.5. Therefore, the correction factor (Δ_α) is estimated from the squared phenotypic differences of the two regression equations of full (1) and half (2) sibs both under the condition of $\hat{\pi}_{jm} = 0.25$:

$$\hat{\Delta}_\alpha = (\hat{Y}_{HS}|\hat{\pi}_{jm}) - (\hat{Y}_{FS}|\hat{\pi}_{jm}). \quad (4)$$

Correction of the squared phenotypic differences of full or half sibs for the correction factor Δ_α, can provide the data for a combined analysis of full and half sibs [9]. Nevertheless, Δ_α has to be estimated separately for each marker.

2.1. Selection methods

Two selection schemes based on the within family variance were used for the selection of both full and half-sib samples. A randomly selected sample of full and half sibs and a sample only of selected full sibs were used for comparison. For all methods of selection, the size of the sample for analysis was varied by altering the intensity of selection from a population of fixed size.

2.1.1. Within family variance for full sib families (WFV)

The within family phenotypic variance was calculated for each full-sib family. Families with the highest within family variance were selected for analysis regardless of the half sib family from which they came. All the possible full sib pairs from selected families were included in the analysis. In addition, where they occurred, half-sib relationships between the selected full-sib families were also included in the analysis. Therefore, all the possible half-sib pairs between selected full-sib families with high within family variance were used in the analysis.

2.1.2. Mean within family variance within half sib families (MWFV)

In this selection scheme the mean within full-sib family variance from the full-sib families included within a half-sib family was calculated. Half-sib families with the highest mean within full-sib family variance were selected. All possible half or full sib pairs from selected families were included in the analysis.

2.1.3. Full sibs only (FS)

Families were selected according to their within family variance as in the WFV approach. However, only the full sib pairs were used in the analysis.

2.1.4. Random selection (RS)

Entire full sib families were selected at random, without replacement, and all their sib pairs (half and full) were used in the analysis.

2.2. Analysis

Regression of the squared phenotypic difference (Y_j) on the proportion of alleles shared IBD at a marker locus [12] was used for the detection of a linked QTL. The correction factor (Δ_α) was calculated for each marker from the regression equation of full and half-sibs, independently, as in equation (4). Then the correction factor (Δ_α) was subtracted from the squared phenotypic differences of half sibs and all values were used in one single regression analysis.

Single marker [11] and interval mapping [8] analyses were used for this parameter estimation and the results of parameter estimates were compared.

2.3. Simulations

Simulated data were used to evaluate the power of the alternative selection schemes. Initially, a biallelic marker ($p = q = 0.5$) completely linked ($\theta = 0$) with a biallelic QTL accounting for $0.1\sigma_P^2$ (10% of the phenotypic variance) was simulated. For simplicity the phenotypic variance of the trait was considered to be one ($\sigma_P^2 = 1$). No selection was applied on the parents and loci were assumed to be in linkage equilibrium.

The scenario of a QTL completely linked to the marker was chosen because increasing the distance between the marker and the QTL has a similar negative effect on power of detection in all selection schemes [4]. A marker completely linked with a QTL should give a clearer comparison between the selection schemes.

Two different population structures were simulated; both of which had the same full sib family size (8) and the same total number of progeny (800). In the first structure (PS1) 50 sires were mated with 100 dams (2 dams/sire) resulting in 3 200 half-sib pairs. In the second structure (PS2) 10 sires were mated with 100 dams (10 dams/sire), producing 28 800 half-sib pairs. In both structures 2 800 full sib pairs were produced. These population structures were selected because in previous studies of two pig nucleus populations from a commercial company (unpublished data) the number of dams per sire varied between 2 and 10. Therefore, the above simulations should indicate the maximum and minimum power in two extreme scenarios.

In order to evaluate the effect of repeated correlated tests and obtain estimates of position and variance contributed by the QTL, the simulations were extended. Parents with a 100 cM chromosome were generated with markers spaced at 10 cM intervals. Each marker had eight alleles at equal frequency (0.125). The biallelic QTL was simulated at 25 cM in all cases, and symmetrically placed at 5 cM between two markers. The same family size (8) was used, but only the PS2 population structure was simulated. No selection was applied to the parents and loci were assumed to be in linkage equilibrium. A fixed level of the trait heritability (0.4) and common family environment variance (0.2) were used. The parents were mated at random to produce the offspring generation that provides the sib pairs used in the analysis.

In all cases the phenotypes of an additive quantitative trait were determined by: (a) a QTL with two alleles at equal frequency, (b) a polygenic effect created by 10 additional biallelic ($p = q = 0.5$) trait loci of equal effect, which were independent of each other and the QTL (unlinked), (c) an environmental component (normally distributed) and (d) a common environmental effect (normally distributed). The model of the simulated phenotype was:

$$x_{ijk} = \mu + q_{ijk} + p_{ijk} + c_{ij} + e_{ijk} \tag{5}$$

where, x_{ijk} = phenotypic value of animal k in family ij
μ = overall mean;
q_{ijk} = effect of animal ijk's QTL genotype;
p_{ijk} = effect of animal ijk's polygenic genotype;
c_{ij} = effect of common litter environment;
e_{ijk} = environmental effect.

The empirical thresholds, at the 5% level of significance, were obtained for every level of selection (sample selection intensity) and for the different selection methods by simulations under the null hypothesis ($\sigma^2_{QLT} = 0$) using 1 000 replicates. In the simulations under the null hypothesis, all the genetic variance were due to the polygenic effect. The same starting seed value was used for both null ($\sigma^2_{QLT} = 0$) and alternative hypotheses ($\sigma^2_{QLT} = 0.1\sigma^2_P$) in order to avoid any effect on Type I errors. For the power calculations, the appropriate threshold for each method and level of selection was used.

3. RESULTS

3.1. Power

The power of detection was the percentage of 1 000 replicates, under the alternative hypothesis ($\sigma^2_{QLT} = 0.1\sigma^2_P$), exceeding the empirical threshold simulated in each case. In Figure 1 the scaling of power of detection under different sample sizes of four selection schemes for a population with 2 dams/sire (PS1) is presented. Selecting the samples of both half and full sib pairs based on high within full sib family variance (WFV) was the most powerful selection method for all sample sizes (*i.e.* all selection intensities). The gain in power of detection from selecting according to the WFV rather than random selection of samples (RS) was rather substantial, except where most families are genotyped (at low selection intensities). The gain in power of detection from the inclusion of the half sibs in the analysis was rather marginal, equivalent to an absolute increase in the power of detection of about 4% on average.

When the alternative population structure was used (PS2), with 10 dams / sire and hence more potential half sib pairs, the use of the mean within family variance in a half sib family (MWFV) as the selection criterion was most powerful in small sample sizes (less than 280 individuals) (Fig. 2). However, this advantage disappeared in bigger sample sizes and the selection of full and half-sib pairs from families with high WFV became more powerful as the sample size increased. The gain in power of QTL detection from including the half sibs in the analysis was much larger in this population structure (PS2), being about 16% in average. The overall power of QTL detection was consequently increased from a maximum of about 66% in PS1 (Fig. 1) to a maximum of about 81% in PS2 (Fig. 2).

By appropriate selection of substantially fewer individuals, in both population structures, almost the same power of QTL detection could be achieved, as when all individuals were genotyped. However, the optimum sample size changed in the two different population structures. For example, in the PS1 structure with 50% of the individuals, almost the same power as when genotyping all individuals can be achieved; in the PS2 structure at least 60% of the individuals have to be typed in order to achieve almost the same power as with all individuals (Figs. 1 and 2).

3.2. Parameter estimates

Since the gain in power from the inclusion of half sibs in the PS1 structure (2 dams/sire) was rather marginal (Fig. 1), only the PS2 structure (10 dams/sire) was used for the parameter estimation.

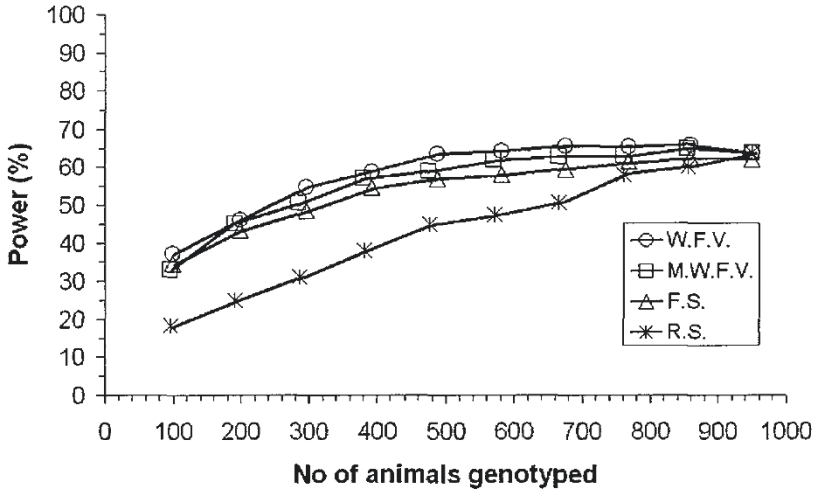

Figure 1. Power of alternative selective genotyping schemes at different selection intensities. Data were simulated with one QTL of medium effect ($\sigma^2_P = 1$, $h^2 = 0.4$, $\sigma^2_{QLT} = 0.25\sigma^2_G = 0.1\sigma^2_P$) completely linked to a biallelic marker. The population consisted of 50 sires with 2 dams/sire and 800 progeny (PS1) (maximum number of animals genotyped 950: 800 progeny and 150 parents). The power was the percentage of 1 000 replicates, under the alternative hypothesis ($\sigma^2_{QTL} = 0.1\sigma^2_P$), exceeding the empirical threshold simulated in each case.

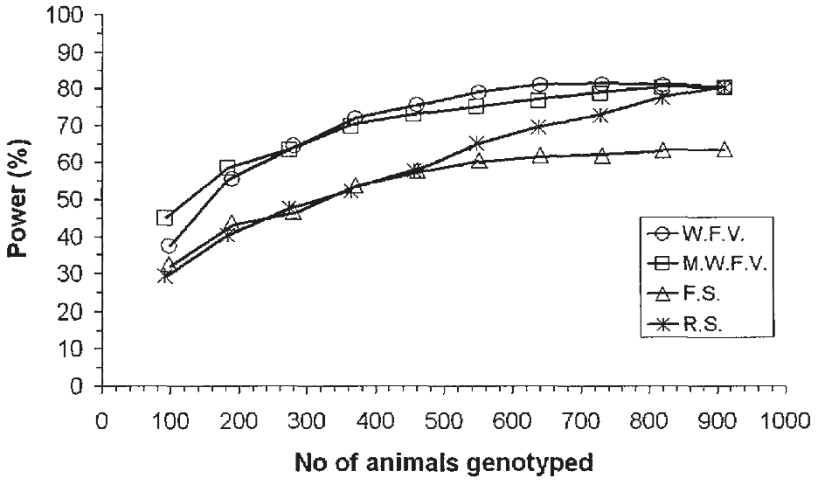

Figure 2. Power of alternative selective genotyping schemes at different selection intensities. Data were simulated with one QTL of medium effect ($\sigma^2_P = 1$, $h^2 = 0.4$, $\sigma^2_{QTL} = 0.25\sigma^2_G = 0.1\sigma^2_P$) completely linked to a biallelic marker. The population consisted of 10 sires with 10 dams/sire and 800 progeny (PS2) (maximum number of animals genotyped 910: 800 progeny and 110 parents). The power was the percentage of 1 000 replicates, under the alternative hypothesis ($\sigma^2_{QTL} = 0.1\sigma^2_P$), exceeding the empirical threshold simulated in each case.

As can been seen in Table I the QTL position estimates, averaged over all 1 000 replicates, tend to be biased towards the middle of the chromosome for all selection methods. However, the use of selected half sib information (WFV, MWFV) improved the position estimates substantially over the estimates obtained from the use of only selected full sib pairs (FS), where the QTL was positioned outside the correct interval in all sample sizes. Selection of the sib pairs from families with high WFV produced marginally less biased position estimates than the other methods (MWFV, RS). Furthermore, relatively unbiased position estimates, similar to those obtained using all the animals, can be obtained by using only 60% of the total number of animals (Tab. I).

The QTL variance estimates were biased upwards when the samples of both full and half sibs were selected on the basis of WFV (Tab. II). Similar biases were observed when only full sib pairs were used in the analysis. Selection of the samples on the basis of the MWFV within a half sib family or random selection seemed to produce less biased (although still inflated) QTL variance estimates with randomly selected samples yielding the best estimates.

The method of analysis did not seem to have an important influence on parameter estimation, with single marker and interval mapping analysis producing similar parameter estimates (both position and QTL variance) (Tabs. I and II). However, interval mapping had the tendency to overestimate the QTL variance slightly more than the single marker analysis (Tab. II).

4. DISCUSSION

Inclusion of half-sib information can improve the power of QTL detection when a sib pair analysis is used. The gain in power increases with increasing half-sib family size (Figs. 1 and 2, [9]). Selective genotyping can decrease the amount of genotyping without significantly reducing the power of QTL detection (Figs. 1 and 2). Selecting the animals to be genotyped based on the within full sib family variance (WFV) is generally the best selection method, although it yields substantially less total sib pairs when large half-sib families are used (Fig. 3).

When full sib pairs are selected from families with high WFV, the power increases and reaches a maximum at a certain sample size and remains almost stable as the sample size further increases [5, 6]. This is because sib pairs coming from families with low WFV are not segregating for the QTL and thus, are "uninformative" for the analysis. Sib pairs from such families are not adding any extra information in the analysis and consequently to the power of detection.

However, at high selection intensities (small sample size), selection of sib pairs on the basis of the MWFV within a half-sib family is a more powerful selection method (Fig. 2). This is because when selecting on the basis of MWFV at high selection intensities, the number of half-sib pairs is high with the majority of them being very informative (only one or two heterozygous sires with all their dams are selected). On the contrary, when selecting on the basis of WFV, although the number of full sib pairs is the same, the number of half sib pairs, though very informative, is much lower than when the selection is based on the MWFV (Fig. 3). Consequently, this results in lower power.

Table I. Mean position estimates in cM for single and interval mapping analysis over 1 000 replications with increasing sample size. Standard errors of the estimates are given in parentheses. The simulated position was at 25 cM with markers at 20 and 30 cM.

	Sample size									
	~ 100		~ 350		~ 550		~ 700		910	
Selection Method	Single marker	Interval mapping	Single marker	Interval mapping	Single marker	Interval mapping	Single marker	Interval mapping	Single marker	Interval mapping
WFV	36.0	36.2	30.8	31.4	29.2	29.3	28.9	29.4	28.4	28.4
	(0.9)	(0.9)	(0.7)	(0.7)	(0.6)	(0.6)	(0.7)	(0.6)	(0.6)	(0.5)
MWFV	37.9	38.0	32.0	32.3	30.3	30.5	28.7	29.1	28.4	28.4
	(0.9)	(0.9)	(0.7)	(0.7)	(0.7)	(0.6)	(0.6)	(0.6)	(0.6)	(0.5)
RS	40.8	42.4	32.0	32.3	30.0	31.0	29.3	29.7	28.4	28.4
	(1.0)	(0.9)	(0.8)	(0.7)	(0.7)	(0.7)	(0.6)	(0.6)	(0.6)	(0.5)
FS	36.9	37.3	32.7	33.0	32.3	32.8	31.6	32.2	31.2	31.7
	(1.0)	(0.9)	(0.7)	(0.7)	(0.7)	(0.7)	(0.7)	(0.7)	(0.7)	(0.7)

Table II. Mean QTL variance estimates as proportion of the total phenotypic variance, for single and interval mapping analysis, over 1 000 replications with increasing sample size. Standard errors of the estimates are given in parentheses. The simulated variance was $0.1\sigma^2_P$.

	Sample size									
	~ 100		~ 350		~ 550		~ 700		910	
Selection Method	Single marker	Interval mapping	Single marker	Interval mapping	Single marker	Interval mapping	Single marker	Interval mapping	Single marker	Interval mapping
WFV	0.566	0.623	0.246	0.267	0.182	0.199	0.143	0.155	0.111	0.121
	(0.010)	(0.010)	(0.004)	(0.004)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
MFWV	0.185	0.286	0.175	0.192	0.146	0.159	0.127	0.137	0.111	0.121
	(0.008)	(0.007)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
RS	0.195	0.195	0.135	0.148	0.124	0.134	0.115	0.126	0.111	0.121
	(0.006)	(0.005)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
FS	0.579	0.630	0.246	0.265	0.186	0.201	0.149	0.162	0.124	0.135
	(0.010)	(0.010)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)

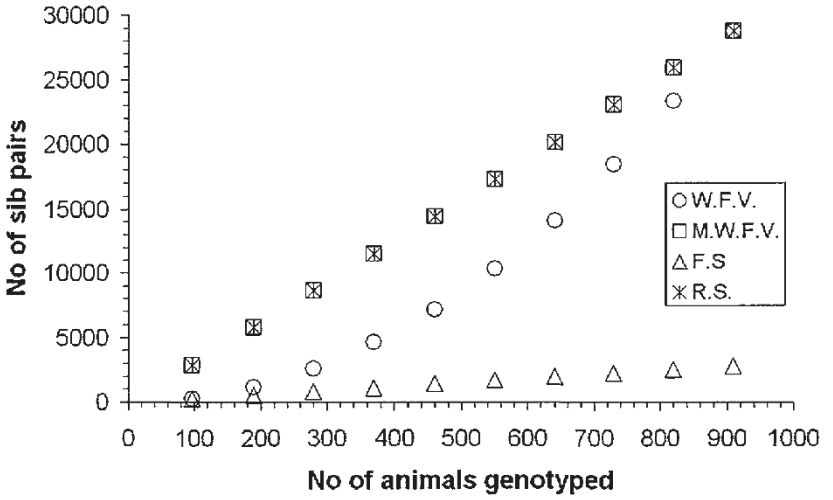

Figure 3. The number of sib-pairs generated in each selective genotyping scheme. The numbers shown represent the mean of 1 000 replicates. The population size simulated and hence the maximum number of animals genotyped was 910 (800 progeny and 110 parents) and the population structure was 10 dams/sire (PS2), resulting in a maximum of 2 800 full sib pairs and 28 800 half sib pairs.

When the selection pressure is relaxed (large sample size) and is based on the WFV the extra full or half-sib pairs included are still very informative. However, when the selection is based on the MWFV the number of "non-informative" half or full-sib pairs increases because less "informative matings" are entering the sample (matings with homozygous dams). Consequently, there is a change in relative power between the two methods (Fig. 2). The above case is more likely to appear in a population structure with large half sib families (*e.g.* PS2) where the differences in the number of half-sib pairs between the two selection methods (WFV and MWFV) are larger (Fig. 3). This is also the reason why this situation does not appear in population structures with small half sib families (*e.g.* PS1: Fig. 1).

This difference in the available half-sib pairs between different population structures (*e.g.* PS1, PS2) appears to be the reason for the difference in the sample size needed to reach the maximum in the two population structures ($\sim$ 50% for PS1 and $\sim$ 60% for PS2). For example, decreasing the selection intensity in population structures with large half-sib families (*e.g.* PS2) vastly increases the number of available half sib pairs for analysis (Fig. 3), with some of them being "informative". Therefore, when large half-sib families are used some additional power could be gained by genotyping an extra 10%.

The same scaling of power, between the selection schemes, was also observed when the QTL was 5 cM away from a polymorphic marker on a 100-cM chromosome (data not shown). However, in this paper only the case of an additive biallelic QTL with equal allelic frequencies and a QTL variance of $0.1\sigma_P^2$ is considered. Overall power will be less for smaller QTL-effects. However, QTL

detection will still be possible with larger sample sizes. When the population size is large enough to allow an increase of the sample size, selection according to the WFV seems to increase its advantage over the other selection methods [4]. Nevertheless, the mode of action of the QTL (additive, dominant, recessive etc.) and its allelic frequencies are parameters that are not known in advance. Others [3,7,14] have considered a wide variety of models (additive, dominant, and recessive) as well as different allelic frequencies and concluded that these parameters have an effect on the efficiency of different selective genotyping schemes. However, these authors considered selective genotyping schemes on small sized families (= 2) that are based on the distribution of the phenotypes of the sibs rather than the phenotypic variance of the trait. In these selection schemes one is looking for informative sib pairs *per se*. These informative sib pairs will change in alternative selection schemes according to the allelic frequency of the QTL and its mode of action. On the contrary, when selecting for the WFV of the trait, one is looking for informative matings (*e.g.* matings of heterozygous parents) which, consequently, would result in informative sib pairs. Selective genotyping on the basis of the WFV of the trait seems to be as effective even in cases of rare QTL alleles ($p = 0.2$) or dominant QTL [4,5]. For example in cases of a rare QTL allele, although the informative matings are reduced they could still be distinguished by a selection scheme based on the WFV of the trait in question. In cases of small family sizes (= 2), where the use of the WFV is not applicable, the selection of concordant and discordant sib pairs seems to be overall the most effective selection scheme [4,5]. Note that larger families selected on the basis of high variance in which a QTL is segregating will have a mixture of concordant and discordant sib-pairs. The value of a mix of concordant and discordant sib-pairs has been demonstrated by Eaves and Meyer [7], where a variety of QTL modes of action and allelic frequencies were used. However, we do not have enough evidence to determine how much such parameters (rare QTL allele, recessive) would affect the effectiveness of selection on the basis of WFV. Nevertheless, it is expected that the overall power of detection, in a population of a fixed size, would be reduced in such cases (rare QTL allele, recessive QTL) due to reduction in the number of informative matings.

A comparison of efficiency of selective genotyping schemes under different population structures and different models can only be suggestive of which analysis and selective genotyping methods are the most suitable in specific cases. It is the authors'belief that there is not an overall "best" analytical or selective genotyping method. The method of selection and analysis of the available information depends on many parameters, both known (population structure, capital investment etc.) and unknown (QTL mode of action and allelic frequency). The most objective way to determine the most powerful and cost effective design for different scenarios would be the use of a simulation study under a variety of analytical and selective genotyping methods, tailored to the population of interest. However, sib pair analysis seems to be at least comparable, if not more powerful, with other methods in hierarchical populations. In a direct comparison with an intercross design [15], sib pair analysis was more powerful when a family size of six was used [10]. Moreover, in a single marker, with a least squares analysis of 20 sires with 100 half-sib progeny each, and a

QTL, 5 cM away from the marker, accounting for $\sim 0.148\sigma_P^2$ [13], the power was 80% ($P > 0.01$). In our analysis, although the marker was completely linked with the QTL, 81.3% ($P > 0.05$) power could be achieved using a smaller QTL ($0.1\sigma_P^2$) and substantially less individuals ($< 25\%$). However, since the QTL size, population size, linkage etc. differ between the two studies a large-scale simulation study would be more appropriate in order to be able to make an objective comparison.

The parameter estimates (QTL position and variance), are not greatly affected by selection when the sample sizes are sufficiently large ($> 50\%$ of the total population) (medium selection intensities) and are slightly improved when half sibs are included (Tabs. I–II; [9]). However, when the selection intensity is very high (less than 35% of the total population), the position estimates of the QTL are outside the correct interval. Moreover, the QTL variance is inflated when the samples are selected and this inflation is higher with increasing selection intensity of the sample. The reasons for these biases are: (a) the tendency of the method to position the QTL in the middle of a chromosome when the information decreases (decreasing number of sib pairs with increasing selection intensity) and (b) inflation of the regression coefficient due to selection. Moreover, in simulations, there is an overall overestimation of the regression coefficient due to the fact that in every replicate the marker with the highest test statistics (most negative) is selected, something that has "inflating" effects on the regression coefficients. Consequently, since the test is one-sided, the most negative regression coefficients predominate, resulting in an overestimation of the parameters. In addition, in single marker analysis when the two markers used for the parameter estimation have different signs, estimates in the range of real values cannot be obtained. This has been reported [9] to produce biases in the estimation of QTL variance in full sib, half sib and combined sib pair analysis. Nevertheless, once linkage has been detected, random samples can be used for the estimation of the QTL variance and position in order to achieve better results and avoid biases due to selection (Tab. II).

Selective genotyping can be applied in populations with hierarchical structures for QTL detection with almost no negative effect on the power of detection. This makes sib pair methods attractive to use for QTL detection in a variety of populations where large full and half sib families exist (*e.g.*, plant, fish, chicken, pig etc.) [10]. This includes dairy cattle populations where large half-sib families exist, since the use of half-sib pair analysis is more powerful than the full-sib pair analysis when the same number of progeny is available [9]. Moreover, sib pair methods are very flexible since full sib and half-sib information can be used independently or combined in one single analysis in selected or unselected samples with very good results in power and parameter estimates.

Selective genotyping for a single trait is considered here. If more than one correlated traits are considered some decrease in the selection intensity of the samples may secure sufficient power of detection for all traits. Unfortunately, when the traits are uncorrelated then the applicability of any selective genotyping scheme might be reduced. However, this could be an interesting area for deterministic simulation studies concerning the applicability as well as the cost effectiveness of selective genotyping using sib pair analysis in different populations.

ACKNOWLEDGEMENTS

D.G. Chatziplis and C.S. Haley are grateful for support from the State Scholarship Foundation of Greece and the Biotechnology and Biological Sciences Research Council of the United Kingdom, respectively.

REFERENCES

[1] Amos C.I., Elston R.C., Robust methods for the detection of genetic linkage for quantitative data from pedigrees, Genet. Epidem. 6 (1989) 349–360.
[2] Blackwelder W.C., Elston R.C., Power and robustness of sib pair linkage tests and extension to larger sibships, Commun. Statist. Theor. Meth. 11 (1982) 449–484.
[3] Cardon L.R., Fulker D.W., The power of interval mapping of quantitative trait loci, using selected sib pairs, Am. J. Hum. Genet. 55 (1994) 825–833.
[4] Chatziplis D.G., The use of selective genotyping in the detection of quantitative trait loci (QTL) by sib pair analysis, Ph.D. thesis, University of Edinburgh, 1998.
[5] Chatziplis D.G., Hamann H., Haley C.S., Selection and subsequent analysis of sib-pair data for QTL detection (2000) Genet. Res. (in press).
[6] Chatziplis D.G., Hill W.G., Haley C.S., Selective genotyping for QTL detection by sib pair analysis, 6th World Cong. Genet. Appl. Livest. Prod., 11–16 January 1998, University of New England, Armidale, Vol. 26, pp. 249–252.
[7] Eaves L., Meyer J., Locating human quantitative trait loci: Guidelines for the selection of sibling pairs for genotyping, Behavior. Genetics 24 (1994) 443–455.
[8] Fulker D.W., Cardon L.R., A sib-pair approach to interval mapping of quantitative trait loci, Am. J. Hum. Genet. 54 (1994) 1092–1103.
[9] Götz K.U., Hamann H., Detection of QTL's in populations with hierarchical structures, in: Proceedings of the 2nd European Workshop on Advanced Biometrical Methods in Animal Breeding, 12–20 June 1995, Salzburg, pp. 1–11.
[10] Götz K.U., Ollivier L., Theoretical aspects of applying sib-pair linkage tests to livestock species, Genet. Sel. Evol. 24 (1992) 29–42.
[11] Hamann H., Götz K.U., Use of sib-pair linkage methods for the estimation of the genetic variance at a quantitative trait locus, Genet. Sel. Evol.27 (1995) 97–110.
[12] Haseman J.K., Elston R.C., The investigation of linkage between a quantitative trait and a marker locus, Behav. Genet. 2 (1972) 3–19.
[13] Knott S.A., Elsen J.M., Haley C.S., Methods for multiple marker mapping of quantitative trait loci in half-sib populations, Theor. Appl. Genet. 93 (1996) 71–80.
[14] Risch N., Zhang H., Mapping quantitative trait loci with extreme discordant sib pairs: Sampling considerations, Am. J. Hum. Genet. 58 (1996) 836–843.
[15] Soller M., Genizi A., The efficiency of experimental designs for the detection of linkage between marker locus and a locus affecting a quantitative trait in segregating populations, Biometrics 34 (1978) 47–58.

To access this journal on line:
www.edpsciences.org