EXPERIMENTAL DESIGNS FOR MAPPING QUANTITATIVE TRAIT LOCI IN SEGREGATING POPULATIONS

J. I. Weller
Institute of Animal Science, A. R. O., The Volcani Center
P. O. Box 6, Bet Dagan, 50250, Israel

SUMMARY

Two designs, denoted the "daughter" and "granddaughter" designs, are presented to detect QTL of moderate magnitude in outcrossing populations. The granddaughter design was twice as powerful as the "daughter" design, but still requires genotype determination on several hundred to several thousand individuals to achieve power of .5 to detect a quantitative trait loci responsible for 1% of the phenotypic variance. Previous studies have shown that in crosses between inbred lines, power can be increased by selective genotyping, marker brackets, and analysis by maximum likelihood techniques. The likelihood function is presented for the daughter design, and methods are discussed for ML hypothesis testing and parameter solution. The effects of these techniques on analysis of segregating populations has not yet been studied.

INTRODUCTION

Genetic markers can be used to map loci affecting quantitative traits (QTL) in agricultural species. Most studies have been based on the analysis of F-2 or backcross populations produced by crossing two inbred lines, each homozygous for different marker alleles (Edwards et al., 1987; Paterson et al., 1988; Tanksley et al., 1982; Thoday, 1961; Weller, Brody, and Soller; 1988). If the parental lines are also homozygous for alternative alleles of a genetic marker and a genetically linked QTL, the effect of the QTL can be detected in the F-2 or backcross populations by comparing mean trait values of the different marker genotypes. These analyses are not practical for livestock, because suitable inbred lines are not available, and the time and expense involved in producing the required crosses are prohibitive. Various studies have therefore attempted to detect linkage between genetic markers and QTL by analysis of livestock field data (Arave et al., 1971; Brum et al., 1968). However, even if a segregating QTL is linked to a genetic marker segregating in the population, a significant difference between marker genotype means for the quantitative trait will not be detected, if the population is at linkage equilibrium between the two loci.

In dairy cattle, the number of progeny per male is often quite large. Thus it should be possible to detect linkage within the progeny of a sire heterozygous for a genetic marker, if the particular sire is also heterozygous for a linked QTL (Neimann-Sorressen and Roberson, 1961). The goals of this study are to describe the specific experimental designs and statistical methods for detecting segregating QTL in outbreeding populations, and sampling and statistical techniques that may increase the power of QTL detection.

EXPERIMENTAL DESIGNS

Two basic experimental designs have been suggested. In the "daughter" design, marker genotype and trait value are both scored on the daughters of sires heterozygous for the genetic markers. If more than one sire is included in the analysis, linkage between QTL and the genetic markers can be determined by a linear model analysis of the quantitative trait. The analysis model will include the effects of sire and genotype nested within sires. A significant genotype within sire effect is indicative of linkage. In the "granddaughter"

design, sons of a sire heterozygous for the genetic marker are scored for marker genotype, while the quantitative traits are measured on the granddaughters. If progeny of several heterozygous grandsires are analyzed, the analysis model will include the effects of grandsire, marker genotype nested within sire, and sire nested within marker genotype and grandsire. Although the QTL effect detected in the grandsire design is only half the magnitude of the daughter design effect, the number of individuals with trait records may be much greater.

Estimating power for anova tests is generally quite laborious (Soller and Genizi, 1978). Neimann-Soressen and Roberson (1961), and Gelderman (1975) noted for the daughter design that the sum of squared differences between the means of the within-sire marker genotypes divided by the standard deviations of the difference will have a Chi-squared distribution, with degrees of freedom equal to the number of sires. Weller, Kashi, and Soller (1988, 1990) estimated power of both the daughter and granddaughter designs by Chi-squared. For the granddaughter design, the differences between the within-grandsire genotype means, with genotypes determined on the son, but traits measured on the daughters, divided by their standard errors will also have a Chi-squared distribution, with degrees of freedom equal to the number of grandsires. For a trait with heritability of .2, a QTL effect of .2 phenotypic standard deviations, and Type I error of .01; power was .70 if 400 daughters of 10 sires were scored for both the genetic markers and the quantitative traits; and .95 if 100 sons of each of 20 grandsires were scored for the genetic markers, with quantitative traits recorded on 50 granddaughters/son. Thus even though only half as many individuals were assayed for the genetic markers in the granddaughter design, power was greater than in the daughter design. Since RFLP analysis is still relatively expensive, and large data banks on quantitative traits are available for many populations, the granddaughter design is preferable, provided that the pedigree structure is appropriate, i. e., a number of sires with many sons per sire, each with a relative large number of milk-recorded daughters. The U.S. dairy cow population is suitable for analysis by the granddaughter design (Weller, Kashi, and Soller; 1990).

METHODS TO INCREASE POWER OF DETECTION

Three techniques have been suggested to increase the power of QTL detection in crosses between inbred lines: selective genotyping, marker brackets, and analysis by maximum likelihood. Lebowitz, Soller and Beckmann (1987); and Lander and Botstein (1989) found that power is increased if only individuals with extreme values for the quantitative trait were selected for genotyping for the genetic marker. Selective genotyping is probably more viable for the daughter design, where an elite sire may have thousands of daughters, and the highest and lowest few hundred can be selected for genotyping. For the granddaughter design this would entail selecting sons for genotyping based on their genetic evaluations. Since the number of sons per sire is unlikely to be more than a few hundred, there is not much possibility of selection. The additional power obtainable by selective genotyping for either the daughter or granddaughter design has not been estimated. Selective genotyping can only be applied for one or two traits. Selecting individuals with extreme values for one trait will yield a random sample with respect to other uncorrelated traits.

If a single marker is used to detect linked QTL, then the magnitude of the effect measured will be equal to 1 - 2r of the actual QTL effect for both the daughter and granddaughter designs, where r is the proportion of recombination. Thus for .2 recombination between the QTL and the genetic marker, the effect measured will be only 0.6 of the actual QTL effect. However, if the putative QTL lies between a pair of segregating markers, and zero interference is assumed, then the estimated QTL effect will be equal to the actual effect times $1 - 2r_1r_1$

for both designs, where r_1 and r_2 are the recombination frequencies between the QTL and each of the genetic markers. For marker brackets of less than .4 recombination between the genetic markers, the effect measured will always be greater than 0.9 of the actual effect. Saturated restriction fragment length polymorphism (RFLP) maps have already been produced for several plant species (Paterson et al., 1988).

MAXIMUM LIKELIHOOD METHODOLOGY

Weller (1986, 1987) used maximum likelihood (ML) to estimate linkage distances between QTL and genetic markers in an F-2 cross between inbred lines. Simpson (1989) showed that power to detect a QTL was greater by a likelihood ratio test, in which the maximum likelihood was compared to the maximum likelihood obtainable with .5 recombination between the genetic marker and the putative QTL, as compared to anova. Lander and Botstein (1989) applied ML to the case of a QTL located between marker brackets, although it would seem that in this case the possible gain by ML would be marginal. For crosses between inbred lines, ML estimates were obtained for the marker genotype means and variances, and for recombination frequency between the QTL and genetic marker. Thus even assuming unequal variances for the different genotypes, solutions must be obtained for at most seven parameters.

In the daughter design additional parameters would be the probabilities of the different linkage relationships between the QTL and the genetic marker in the sample of sires. This problem is simplified somewhat if four assumptions are employed: 1) Hardy-Weinberg equilibrium at the QTL, 2) linkage equilibrium between the QTL and the genetic marker, 3) only two alleles present at the QTL, and 4) equal variances within QTL genotype. Then it is necessary to estimate only one additional parameter, the frequency of one of the QTL alleles in the sample of sires or grandsires. For the daughter design, the log of the likelihood, log L, can be constructed as follows:

$$\log L = \sum_{N}\sum_{n}\{\log[f(a,b|x)]\} = \sum_{N}\{\log[\Sigma P(a)\prod_{4}\prod_{n}[f(b|a,x)]]\}$$

Where $f(a,b|x)$ is the density function of the distribution the daughters' records for the quantitative trait, conditional on the trait value x; N is the number of sires; n is the number of daughters/sire; $\prod$ represents the product of a series; $P(a)$ is the probability that the sire will have a given genotype for the QTL; and $f(b|a,x)$ is the density of the daughter distribution conditional on the sire's QTL genotype and the records for the quantitative traits. Under the assumptions of only two alleles for the QTL, and both Hardy-Weinberg and linkage equilibrium, the probabilities for the sires' QTL genotype will be direct functions of the frequency of either QTL allele.

The alleles for the genetic markers will be denoted as M and m, the alleles for the QTL as A and a, the means of these alleles in the daughter population as μ_A and μ_a, and the probability of A as p. Then p^2, $2p(1-p)$, and $(1-p)^2$ are the frequency of the sire's genotype being AA, Aa and aa, respectively. For either the first or the last case, $f(b|a,x)$ will be the normal density function, the only difference being the value of the mean. The conditional probability of daughters of QTL-heterozygous sires will be a mixture of two distributions. If the two loci are in coupling mode, (M and A on the same chromosome) then, except for recombinants, daughters that received M will also have received A. If the two loci were in repulsion, then the opposite is true. The trait values will be denoted as x_M and x_m for daughters that received the alleles M and m, respectively, the number of daughters that received each allele as n_M and n_m,

respectively, and the normal density function for a variable with a value of x and a mean of μ as $f(x,\mu)$. Then log L can be then be written as follows:

$$L = \sum^{N} \log\{[p^2 \prod^{n} [f(x,\mu_1)] + p(1-p)\{\prod^{n_{12}} [(1-r)f(x_{12},\mu_1)+(r)f(x_{12},\mu_2)] \prod^{n_{12}} [(r)f(x_{12},\mu_1)+(1-r)f(x_{12},\mu_2)] + \prod^{n_{12}} [(1-r)f(x_{12},\mu_2)+(r)f(x_{12},\mu_1)] \prod^{n_{12}} [(r)f(x_{12},\mu_2)+(1-r)f(x_{12},\mu_1)]\} + [(1-p)^2] \prod^{n} [f(x,\mu_2)]\}$$

Thus it is necessary to maximize the likelihood for five parameters: μ_1, μ_2, r, p and the within genotype variance. This likelihood can then be compared to ML under the null hypothesis of a single normal distribution, and tested as described by Simpson (1989).

Even for crosses between inbred lines, direct solutions could not be derived for the ML parameter estimates. Weller (1986) used a combination of moments methods and scanning of the restricted parameter space, Lander and Botstein (1989) used a combination of the EM algorithm (Dempster et al., 1977) and scanning for recombination frequency, while Simpson (1989) used a general purpose algorithm for function maximization. Darvasi, Soller and Weller (1990, in preparation) used Newton-Ralphson iteration to solve for the parameters of a backcross with marker brackets. Newton-Ralphson can be applied to cases where EM cannot be utilized, it converges more rapidly than EM, and also yields the standard errors of the estimates. The disadvantages of this algorithm are that all the first and second derivatives with respect to all parameters must be computed, which is not trivial for the situations presented; and Newton-Ralphson may not converge to the ML solution.

REFERENCES

ARAVE, C. W., LAMB, R. C. and HINES, H. C. 1971. J. Dairy Sci. 54:106-112.
BRUM, E. W., RAUSCH, W. H., HINES, H. C. and LUDWICK, T. M. 1968. J. Dairy Sci. 51:1031-1038.
DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. 1977. J. Roy. Stat. Soc. (series B) 39:1-22.
EDWARDS, M. D., STUBER, C. W. and WENDEL, J. F. 1987. Genetics 116:113-125.
GELDERMAN, H. 1975. Theor. Appl. Genet. 46:300-319.
LANDER, E. S. and BOTSTEIN, D. 1989. Genetics 121:185-199.
LEBOWITZ, R. J., SOLLER, M. and BECKMANN, J. S.. 1987. Theor. Appl. Genet. 73:556-562.
NEIMANN-SORESSEN, A. and ROBERSON, A.. 1961. Acta. Agr. Scand. 11:163-196.
PATERSON, A. H., LANDER, E. S., HEWITT, J. D., PETERSON, S., LINCOLN, S. E. and TANKSLEY, S. D. 1988. Nature 335:721-726.
SIMPSON, S. P. 1989. Theor. Appl. Genet. 77:815-819.
SOLLER, M. and GENIZI, A. 1978. Biometrics. 34:47-55.
TANKSLEY, S. D., MEDINA-FILHO, R. and RICK, C. M. 1982. Heredity 49:11-25.
THODAY, J. M. 1961. Nature 191:368-370.
WELLER, J. I. 1986. Biometrics 42:627-640.
WELLER, J. I. 1987. Heredity 59:413-421.
WELLER, J. I., KASHI, Y. and SOLLER, M. 1988. Dairy Sci. 71(Suppl. 1) 142.
WELLER, J. I., KASHI, Y. and SOLLER, M. 1990. Dairy Sci. 73 (In press).
WELLER, J. I., SOLLER, M. and BRODY, T. 1988. Genetics 118:329-339.