

EMPIRICAL BEST LINEAR UNBIASED PREDICTION TO MAP QTL

F. E. Grignola and I. Hoeschele
Department of Dairy Science, Virginia Polytechnic Institute and
State University, Blacksburg, VA 24061-0315, U.S.A.

K. Meyer
Animal Genetics and Breeding Unit, University of New England,
Armidale, NSW 2351, Australia

SUMMARY

A derivative-free Residual Maximum Likelihood analysis was developed to map a QTL within a marker brackett and to simultaneously predict QTL effects and estimate QTL, additive genetic, and residual variances. The analysis was applied to a simulated granddaughter design typical for dairy cattle.

INTRODUCTION

Traditional methods for the mapping of Quantitative Trait Loci (QTL) include linear regression or ANOVA (e.g., Weller et al., 1990; Dentine and Cowan, 1990) and Maximum Likelihood (ML) (e.g., Lander and Botstein, 1989; Knott and Haley, 1992). These methods treat QTL substitution effects (α) as fixed. Then, the α of identified QTL are overestimated as predicted by Smith and Simpson (1986) and by Hoeschele and VanRaden (1993a,b) and as verified via simulation by Georges et al. (1994). In the simulation, overestimation increased with decreasing family size and decreasing true α . If QTL gene effects were treated as random rather than fixed, estimates would be shrunken toward a prior mean more strongly for those QTL mapped in smaller families and with smaller true α .

Two methods that treat QTL gene effects as random are Best Linear Unbiased Prediction (BLUP) of QTL effects (Fernando and Grossman, 1989; Goddard, 1992; Hoeschele, 1993) and Bayesian linkage analysis (Hoeschele and VanRaden, 1993a,b; Hoeschele, 1994). The former method can be viewed as a modification of linear regression assuming gene effects to be normally rather than uniformly distributed a priori. The latter method can be viewed as a modification of ML assuming an exponential rather than uniform prior distribution of α at biallelic QTL and incorporating a prior probability of linkage depending on heritability.

Van Arendonk et al. (1993) found that the contribution of a QTL to additive genetic variance and its recombination rate (r) with a single marker were not separately estimable from a halfsib design in a BLUP/Residual ML (REML) analysis. Goddard (1992) presented BLUP for multiple linked markers, but assumed location and variances of QTL to be known. Here we show that QTL can be mapped and their variances estimated simultaneously from halfsib designs, when marker intervals rather than singletons are analyzed.

MATERIALS AND METHODS

The halfsib design considered here was a granddaughter design

(GDD) consisting of several sires with many progeny-tested sons. In a GDD, marker genotypes are available on sires and sons, and phenotypes of sons are preadjusted daughter averages, termed daughter yield deviations (dyd) (VanRaden and Wiggans, 1991). The linear mixed model

$$dyd_i = x_i' \beta + s_i + e_i \quad [1]$$

was an animal model (AM) except that s was half the son's breeding value, β was vector of fixed effects, x_i' row i of design matrix X , and e a residual. Model [1] represents the AM incorporating marker data of Hoeschele (1993) (model H), where data are linked to breeding values instead of to polygenic and QTL effects in the equivalent model of Fernando and Grossman (1989) and of Goddard (1992) (model FG-G). Hence, vector s was augmented by those s effects of ancestors and by those QTL gene effects of sons and ancestors needed to correctly account for additive genetic relationships among sons.

Variance-covariance matrices were $G = \text{Var}(s)$ and $\text{Var}(e) = D\sigma_E^2$. Matrix D was diagonal with element $(1 - \text{rel}_{PT0})/\text{rel}_{PT0}$, where rel_{PT} was reliability of a son's genetic evaluations contributed by its progeny test. rel_{PT} was computed by converting total rel to total number of daughter equivalents (de), subtracting de due to parent average to obtain de due to progeny test and converting these back to rel (VanRaden and Wiggans, 1991). $\text{Var}(G)$ was not the usual additive genetic relationship matrix times σ_s^2 , but rather a variance-covariance matrix of transmitting abilities and QTL gene effects in s , which depended on the marker haplotypes of sons inherited from their sires. $\text{Var}(G)$ was a function of three parameters, σ_s^2 , r_1 or the recombination rate between the left marker (assuming only one marker interval is considered) and the QTL, and σ_v^2 or the variance due to the QTL.

These three parameters and σ_E^2 were estimated via REML by modifying the DFREML package of Meyer (1989). Model [1] or model H was chosen over FG-G, because it allowed using the single record - AM option of DFREML and because it minimizes the number of QTL equations needed, which is important when many animals without marker genotypes must be included in the analysis. Inclusion of such animals is necessary to provide relationship ties (ancestors) and to avoid selection biases, because in a real GDD (Georges et al., 1994), many sons culled after progeny were not genotyped due to unavailability of semen samples. The log likelihood was (in notation of Meyer, 1989, p. 320 & p. 323)

$$\log L = -.5 \log |G| - .5(N - NF^* - NR) \log(\sigma_E^2) - .5 y' P y / \sigma_E^2 \quad [2]$$

with $\log |C^*|$ and $y' P y$ evaluated using Gaussian elimination applied to augmented mixed model equations (AMME) as shown in Meyer (1989).

The main modifications required in DFREML were the inclusion of D when forming the AMME, the evaluation of G^{-1} and of $\log |G|$ in

each round of iteration using the algorithm of Hoeschele (1993), and the augmentation and reparameterization of the parameter list to include total narrow sense heritability h^2 , $\{v_i, i=1,q\}$, $\{r_i, i=1,q\}$, and σ_e^2 , where v_i is fraction of total additive genetic variance σ_A^2 explained by QTL i with position r_i within marker bracket.

The DFREML analysis was used to demonstrate that QTL location and variance are separately estimable from halfsib designs with a marker interval, although they are confounded with a single marker (van Arendonk et al., 1993). This result was expected because for a halfsib family and a marker interval, a reduced AM (RAM) requires fitting two random marker effects representing $(1 - r_1/r_m)\alpha$ and $(r_1/r_m)\alpha$ (Goddard, 1992), where r_m is marker recombination rate. However, for a single marker a RAM requires fitting only one marker effect, $m = (1 - 2r)\alpha$, indicating that r and α are not separately estimable.

RESULTS AND DISCUSSION

A GDD was simulated which contained 20 sires with 50 sons each. One biallelic QTL was simulated inside a marker interval, with gene frequency of .5 and α equal to one additive genetic standard deviation. Trait heritability was $h^2=.3$, and ratio of QTL to total additive genetic variance was $v^2=.1$. Value of r_m was .165 and r_1 was .052. The dyd were generated assuming constant rel_{PT} of .7. Marker haplotype inherited from the sire was known for each son. For dyd, true parameter values were $h^2=.7$, $v^2=.1$, $\sigma_e^2=355.$, and $r_1=.05$; estimates were .721, .097, 421., and .04, and these were obtained with very different sets of starting values, using the Simplex algorithm of DFREML. The point was to verify estimability of all parameters with a halfsib design and marker bracket, not to evaluate accuracy achievable with existing designs at this stage.

Future work will expand this analysis to the simultaneous mapping of several QTL in different intervals and on flanks (chromosomal ends not bracketed by markers), as well as to the estimation of QTL variances by a Bayesian approach with an informative prior distribution of QTL variances rather than by REML with a uniform prior. A comparison of this empirical BLUP analysis with the Bayesian analysis of Hoeschele and VanRaden (1993a,b) and of Hoeschele (1994) and ML interval mapping will be conducted with simulated GDDs.

REFERENCES

- FERNANDO, R.L. and GROSSMAN, M. (1989) Genet. Sel. Evol.21:467-477.
GEORGES, M., NIELSEN, D, MACKINNON, M., MISHRA, A., OKIMOTO, R., PASQUINO, A.T., SARGEANT, L.S., SORENSEN, A., STEELE, M.R., ZHAO, X., WOMACK, J.E., and HOESCHELE, I. (1994) Science (submitted).
GODDARD, M. (1992) Theor. Appl. Genet. 83:878-886.
HOESCHELE, I. (1994) Bayesian QTL mapping via the Gibbs sampler. Proc. 5th World Congr. Genet. Appl. Livestock Prod. (submitted).
HOESCHELE, I. and VANRADEN, P.M. (1993a) Theor. Appl. Genet.

85:953-960.

HOESCHELE, I. and VANRADEN, P.M. (1993b) Theor. Appl. Genet. 85:946-952.

KNOTT, S. and HAILEY, C.S. (1992) Genetics 132:1211-1222.

LANDER, E.S. and BOTSTEIN, D.S. (1989) Genetics 121:185-199.

SMITH, C. AND S.P. SIMPSON (1986) J. Anim. Breed. Genet. 103:205-217.

VAN ARENDONK, J.A.M., TIER, B. and KINGHORN, B. (1993) Proc. 17th Int. Congr. Genet., Birmingham, p.192.

WELLER, J.I., KASHI, Y. and SOLLER, M. (1990) J. Dairy Sci. 73:2525-2537.