MARKER ASSISTED COMPLEX SEGREGATION ANALYSIS OF MILK PRODUCTION TRAITS IN DAIRY CATTLE

Z. Liu *, H. Simianer † and H. Haussmann †

* Department of Animal and Poultry Science, University of Guelph, Guelph, Ontario, N1G 2W1, Canada

† Department of Animal Husbandry and Animal Breeding, University of Hohenheim (470/HG), 70593 Stuttgart, Germany

SUMMARY

A marker assisted complex segregation analysis (MACSA) method was developed to estimate the effects and position of a putative quantitative trait locus (QTL) in addition to the polygenic parameters using marker information. Genetic parameters were estimated with the maximum likelihood method. Hypotheses were tested with the likelihood ratio test. Using the MACSA method, five milk production traits (milk, fat and protein yields, fat and protein contents) from 1163 first lactations were analysed with the marker information of the four milk proteins α_{S_1}-casein, β-casein, κ-casein and β-lactoglobulin. The MACSA was carried out both for single markers (one-marker-analyses) and two markers simultaneously (two-marker-analyses). For all the five traits, the polygenic variation was found to be significant. Significant QTL effects could only be detected for milk yield and protein yield. No strong evidence was found for a linkage between the QTL of milk production traits and the milk protein markers.

INTRODUCTION

The classical complex segregation analysis (Elston and Stewart, 1971; Morton and MacLean, 1974) has been proved to be useful for the detection of a segregating major gene. The utilization of marker information to detect a QTL provides greater power than segregation analysis without marker information (Knott and Haley, 1992; Simianer, 1993).

MATERIAL

1163 Holstein-Friesian dairy heifers of 268 sires have calved between 1979 and 1988 in 33 herds in Schleswig-Holstein, Germany[1]. Individual information on five milk production traits (milk, fat and protein yields, fat and protein contents) and four milk protein genotypes (α_{S_1}-casein, β-casein, κ-casein and β-lactoglobulin) was available[1]. Since the MACSA method can hardly account for fixed environmental effects (Haussmann and Liu, 1992), the original performance data were corrected with BLU-estimates (Henderson, 1973) of herd-year-season effects estimated from the whole population. Only two alleles for each marker (B and C for α_{S_1}-casein, A_1 and A_2 for β-casein, A and B for κ-casein and β-lactoglobulin) were chosen for this study, which were all shown to be in Hardy-Weinberg equilibrium.

METHODS

The population is assumed to be composed of a number of unrelated half-sib families whose parents mated at random. Marker and phenotypic information is only available for offspring but not for the parents. It is assumed that QTL be in linkage equilibrium with all markers in the parental population. The genetic model for the corrected observations is $y_{ij} = \mu + u_i + g_{Q_{ij}} + e_{ij}$, where y_{ij} is the corrected observation of the j-th daughter of

1. Data were provided by Dr. G. Erhardt, Giessen, which is gratefully acknowledged.

the i-th sire, μ is the general mean, u_i is the breeding value of the i-th sire, $g_{Q_{ij}}$ is the QTL effect of the ij-th daughter ($g_{Q_{ij}} = a$, d or $-a$) and e_{ij} is the residual effect. The sire and residual effects are assumed to be normally and independently distributed with the variances $\frac{1}{4}h^2\sigma_p^2$ and $(1-\frac{1}{4}h^2)\sigma_p^2$, where σ_p^2 is the sum of polygenic and residual variances and h^2 is the classical polygenic heritability. Let q be the frequency of the QTL allele with the larger effect, m the frequency of one of the marker alleles and r the recombination rate between QTL and marker locus. The parameter vector under the model above is $\theta' = [\mu, q, a, d, m, r, h^2, \sigma_p^2]$. Given the phenotypes and information on one marker locus, the likelihood of n half-sib families with d_i daughters ($i = 1, ..., n$) can be written as:

$$L(\theta|\, y, M) = \prod_{i=1}^{n} \sum_{i_m=1}^{3} P(M_{ii_m}^S |\, M_{i1}^O, ..., M_{id_i}^O) \sum_{i_q=1}^{n_q} P(Q_{ii_q}^S |\, M_{ii_m}^S) \int_{-\infty}^{+\infty} \phi(u_i, \tfrac{1}{4}h^2\sigma_p^2)$$

$$\prod_{j=1}^{d_i} P(M_{ij}^O |\, M_{ii_m}^S) \sum_{h=1}^{3} P(Q_h^O |\, M_{ij}^O, M_{ii_m}^S, Q_{ii_q}^S)\ \phi(y_{ij} - u_i - g_{Q_h},\ (1-\tfrac{1}{4}h^2)\sigma_p^2)\, du_i$$

where $P(M_{ii_m}^S |\, M_{i1}^O, ..., M_{id_i}^O)$ is the probability of sire's marker genotype given its daughters' marker genotypes, n_q is the number of sire's marker-QTL genotype combinations given its marker genotype, $P(Q_{ii_q}^S |\, M_{ii_m}^S)$ is the probability of sire's QTL genotype given its marker genotype, $P(M_{ij}^O |\, M_{ii_m}^S)$ is the probability of a daughter's marker genotype given its sire's marker genotype and $P(Q_h^O |\, M_{ij}^O, M_{ii_m}^S, Q_{ii_q}^S)$ is the probability of daughter's QTL genotype given its marker genotype and sire's QTL and marker genotypes. $\phi(x, v)$ is the density in x of a normal distribution with mean 0 and variance v.

The likelihood function for two-marker-analyses is almost the same as above, except that the number of sire's marker genotypes is 10 instead of 3. In addition to that, three possible QTL positions were taken into account for two linked marker loci, not only the case of QTL bracketed by two marker loci. The integration over sire's breeding value was approximated with the Gauss-Hermite quadrature with 10 points. The Powell method and Brent method were applied to maximize the likelihood. Because of possible multimodality of the likelihood function, 24 sets of starting points were chosen to search for the global maximum of the likelihood. The allele frequencies of marker loci, m for one-marker-analysis or m_1 and m_2 for two-marker-analysis, were estimated from the sample and kept constant in the likelihood. The recombination rates between casein markers for two-marker-analyses were fixed to $r_{\alpha_{s_1}-\beta} = 0.05$, $r_{\beta-\kappa} = 0.1$ and $r_{\alpha_{s_1}-\kappa} = 0.14$. From the seven estimated parameters $(\mu, q, a, d, r, h^2, \sigma_p^2)$, the additive genetic variance at QTL (σ_q^2), dominance variance (σ_d^2) and the total variance ($\sigma_T^2 = \sigma_q^2+\sigma_d^2+\sigma_p^2$) were calculated. The large sample variances and covariances were approximated for the maximum likelihood estimates of the parameters.

In order to investigate the inheritance mode, five genetic models were established: mixed model 1 (the full model), mixed model 2 with free recombination ($r = 0.5$), major gene model 3 ($h^2 = 0$), major gene model 4 with free recombination ($h^2 = 0$ and $r = 0.5$) and polygenic model 5 ($q = a = d = 0$). Three hypotheses were tested by the comparisons of nested models: the major gene model 3 versus the mixed model 1 for the hypothesis $h^2 = 0$, the polygenic model 5 versus the mixed model 1 for the hypothesis $q = a = d = 0$, the mixed model 2 versus the mixed model 1 for the hypothesis $r = 0.5$.

RESULTS

Only a part of the results is presented here (for a complete report, see Liu, 1994). In table 1, the maximum likelihood estimates and their large sample standard errors under the model 1 were averaged from the results of the seven MACSA (four one-marker-analyses with α_{S_1}-casein, β-casein, κ-casein and β-lactoglobulin and three two-marker-analyses with α_{S_1} – β-casein, α_{S_1} – κ-casein, β – κ-casein). Due to lack of heterozygosities at the marker loci, $\hat{r}$ tended to 0 or 0.5 in most cases, therefore no standard error could be approximated. From the MACSA of the three marker systems β-casein, α_{S_1} – β-caseins and β – κ-caseins, the estimates of the recombination rate $\hat{r}_{\beta-QTL}$ under model 1 for milk yield did agree very well: $\hat{r}_{\beta-QTL}$ = 0.26, 0.26 and 0.24. The standard errors were generally large. The standard errors of the polygenic parameters ($\hat{h}^2$, $\hat{\sigma}_p^2$) were smaller than those of the major gene parameters ($\hat{a}$, $\hat{d}$, $\hat{\mu}$). The estimates $\hat{q}$ and $\hat{a}$ were highly correlated. The dominance variance at QTL (σ_d^2) was relatively small (σ_d^2/σ_T^2 = 1% - 4%). σ_q^2/σ_T^2 was about 20%, if the QTL was significant; otherwise, σ_q^2/σ_T^2 accounted for 7% to 20% of the total variance.

Table 1: Means of the ML estimates and the averaged large sample standard errors (italic, in parenthesis) of all parameters but the recombination rate (r) from seven marker systems under the full model 1 for the five corrected milk production traits

	q	a	d	h^2	σ_p^2	μ	σ_q^2/σ_T^2
Milk yield	**.19** (*.10*)	**1053** (*289*)	**-378** (*218*)	**.32** (*.13*)	**494602** (*69402*)	**1342** (*296*)	**28.7%**
Fat yield	**.02** (*.01*)	**53** (*77*)	**28** (*77*)	**.27** (*.09*)	**1134** (*66*)	**79** (*77*)	**10.7%**
Protein yield	**.04** (*.02*)	**13** (*37*)	**34** (*38*)	**.39** (*.13*)	**533** (*39*)	**24** (*38*)	**20.6%**
Fat content	**.09** (*.04*)	**.10** (*.22*)	**.50** (*.23*)	**.28** (*.10*)	**.1367** (*.0122*)	**4.350** (*.225*)	**22.4%**
Protein content	**.03** (*.01*)	**.15** (*.32*)	**.19** (*.33*)	**.54** (*.17*)	**.0299** (*.0020*)	**3.472** (*.318*)	**16.0%**

According to the Akaike Information Content, mixed model 2 with an unlinked QTL was the best among all models for all seven marker systems, i.e. the mixed model with an unlinked QTL (model 2) explained the data better than the other models.

The hypothesis $h^2 = 0$ was rejected in all 35 MACSA analyses (five traits with seven marker systems each) at the significance level $\alpha = 0.1$ (Simianer, 1993) according to the likelihood ratio tests of the model 3 against the model 1 (d.f. = 1). A significant QTL could only be detected for milk yield and protein yield at the significance level $\alpha = 0.001$ with the likelihood ratio test statistics of the model 5 against the model 1 (d.f. = 4, see table 2). The linkage between the QTL of the five milk production traits and the four milk protein markers could not be confirmed at $\alpha = 0.001$ as the result of the comparisons of the model 2 against the model 1 (d.f. =1, see table 3).

DISCUSSION

A significant QTL was detected for milk and protein yield. Linkage between the QTL of the milk production traits and the milk protein markers was not confirmed. Consistent results were found using all the milk protein

markers. On the contrary, "Almost every marker locus investigated has been found in one or more studies to be associated with lactation trait differences (Hines, 1990)." The conflicting results in the literature were presumably due to the use of inappropriate statistical methods or the choice of too high type I error probabilities. MACSA analyses of larger data sets will allow a better understanding of the true state of nature.

Table 2: Error probabilities obtained when testing the hypothesis $q = a = d = 0$ (model 5 compared with model 1, significant results printed in boldface, $\alpha = 0.001$)

	α_{S_1}-Cn	β-Cn	κ-Cn	β-Lg	$\alpha_{S_1} - \beta$	$\alpha_{S_1} - \kappa$	$\beta - \kappa$
Milk yield	**.0003**	**.0001**	**.0002**	**<.0001**	**.0001**	**.0003**	**.0001**
Fat yield	.0031	.0020	.0029	.0029	.0022	.0030	.0021
Protein yield	.0012	.0016	**.0006**	**.0008**	.0016	**.0006**	**.0007**
Fat content	.0185	.0203	.0087	.0180	.0197	.0084	.0179
Protein content	.0108	.0290	.0124	.0123	.0256	.0111	.0339

Table 3: Error probabilities obtained when testing the hypothesis $r = 0.5$ (model 2 compared with model 1, significant results printed in boldface, $\alpha = 0.001$)

	α_{S_1}-Cn	β-Cn	κ-Cn	β-Lg	$\alpha_{S_1} - \beta$	$\alpha_{S_1} - \kappa$	$\beta - \kappa$
Milk yield	1	.6437	.9165	.0587	.6547	1	.4795
Fat yield	1	1	1	.9563	1	.8478	1
Protein yield	1	.9496	.2351	.4187	.8080	.2341	.1592
Fat content	1	1	.1758	.7123	1	.1770	.5839
Protein content	.8097	.5441	1	1	.5164	1	1

REFERENCES

ELSTON, R.C. and STEWART, J. (1971): Human Heredity **21**: 523 - 542.

HAUSSMANN, H. and LIU, Z. (1992): Attempts to incorporate the numerator relationship matrix and fixed environmental effects into segregation analysis for continuous traits. Workshop on advanced biometrical methods in animal breeding, Flawil (Switzerland), October 9-15, 1992, Polycopy.

HENDERSON, C. R. (1973): Sire evaluation and genetic trends. Champaign, Il.

HINES, C. (1990): The 4th WCGALP. Edinburgh.

KNOTT, S. A. and HALEY, C. S. (1992): Genetics **132**:1211-1222.

LIU, Z. (1994): Markergestützte komplexe Segregationsanalyse der Milchleistungsmerkmale beim Rind. University of Hohenheim, Doctoral thesis.

MORTON, N. E. and MACLEAN, C. J. (1974): Am. J. Hum. Genet. **26**, 489-503.

SIMIANER, H. (1993): Maximum likelihood estimation of the additive genetic variance at single loci using marker information - With an application to milk protein polymorphisms and milk production traits in dairy cattle. University of Hohenheim, Habilitation thesis.