Heredity **68** (1992) 43–46
©The Genetical Society of Great Britain

Received 8 February 1991

On the use of the moments method of estimation to obtain approximate maximum likelihood estimates of linkage between a genetic marker and a quantitative locus

A. DARVASI & J. I. WELLER*

*Department of Genetics, The Hebrew University, Jerusalem, 91904, and *Institute of Animal Sciences, ARO, The Volcani Center, P.O. Box 6, Bet Dagan, 50250, Israel*

The approximate maximum likelihood method of Luo & Kearsey (1989) to determine the parameters of a segregating quantitative trait locus (QTL) in an F_2 population by linkage to a genetic marker is compared to the 'true' maximum likelihood estimates derived by scanning the seven parameter likelihood space. The two methods did not yield identical results, and diverged markedly for a dominant QTL loosely linked to the genetic marker. Further study is suggested to evaluate other methods that do not simultaneously maximize the likelihood for all parameters.

Keywords: genetic markers, moments method, maximum likelihood, quantitative trait lóci.

Introduction

Numerous studies have shown that the individual loci that affect quantitative traits (henceforth QTL) can be detected via linkage to genetic markers (Sax, 1923; Neimann-Sorensen & Robertson, 1961; Thoday, 1961; Brum *et al.*, 1968; Hines *et al.*, 1969; Arave *et al.*, 1971; Zhuchenko *et al.*, 1979; Tanksley *et al.*, 1982; Gelderman *et al.*, 1985; Kahler & Wherhahn, 1986; Edwards *et al.*, 1987; Goyon *et al.*, 1987; Haenlein *et al.*, 1987; Stuber *et al.*, 1987; Weller, 1987; Paterson *et al.*, 1988; Weller *et al.*, 1988; Jensen, 1989; Beever *et al.*, 1990; Hines, 1990). Studies of this nature have been facilitated in the last decade by the development of methods to detect polymorphisms at the DNA level (Soller & Beckmann, 1982, 1983, 1985; Beckmann & Soller, 1983, 1987).

Although standard linear model methodology can be used to detect a marker-linked QTL, it cannot be used to derive an estimate of the recombination frequency (r) between the QTL and the genetic marker, unless the QTL is bracketed by two markers. Furthermore, the linear model estimate of the effect will be biased by recombination between the two loci (Knapp *et al.*, 1990). Several studies have shown that by maximum likelihood methodology, the means and variances of the QTL genotypes, and the recombination frequency between the QTL and the marker can be estimated simultaneously (Weller, 1986, 1987; Jensen, 1989; Simpson, 1989; Darvasi, 1990). Although it is generally possile to write the likelihood function, L, and compute the partial differentials of log L with respect to the different parameters, it is often not possible to derive an analytical solution to the resultant system of equations.

A variety of iterative methods have been developed to derive ML multiparameter estimates, including Fisher's method of scoring (Bailey, 1961), expectation–maximization (EM), (Dempster *et al.*,, 1977) and Newton–Raphson (Dahlquist & Bjorck, 1974). The parameter estimates of the ith iterative, for all of these methods, are computed by solving a system of equations equal in number to the number of parameters being estimated. These reduced equations are themselves functions of the parameter estimates from the previous iteration. Thus, it is necessary to continue iteration until changes between rounds fall below a sufficiently small value. Furthermore, all of these methods converge to a local maximum. Thus if the likelihood function has more than one local maximum, it is possible that these algorithms will converge to different local maxima, depending on the initial parameter estimates.

As an alternative to these methods, Weller (1986) presented a combination of moments method and ML for analysis of an F_2 population. Assuming that the three F_2 QTL genotypes have different means and variances for the quantitative trait, it is necessary to estimate seven parameters for this population; the three means, the three variances, and *r*. The likelihood function was scanned by this method for three dimensions; the mean and variance of the QTL heterozygote, and *r*. For each combination of these three parameters, the remaining four parameters were estimated as functions of these parameters, and the first and second moments of the marker-genotype distributions. Scanning of the three-dimensional space is well within current computing capability.

Weller (1986) assumed that the three-dimensional *ML* values were not identical to the seven-dimensional *ML* solution. Based on the invariant property of *ML* estimators, Luo & Kearsey (1989) proposed that: (a) the three-dimensional *ML* estimate is in fact equivalent to the complete *ML* solution, and (b) rather than search a three-dimensional space, equivalent *ML* solutions could be derived by searching a one-dimensional space for *r*, with the other six parameters estimated from *r* and the first and second moments of the marker-genotype distributions. Estimates obtained in this manner will be termed 'pseudo' *ML* (*PML*) estimates.

In this study we consider the theoretical basis of *PML* estimates, and demonstrate by computer simulation that *PML* estimates do not necessarily approximate the 'true' *ML* solutions.

Theoretical approach

The derivation of *PML* estimates is based on the invariant property of maximum likelihood estimators. Application of this property is based on the assumption that equations (12–14) in Luo & Kearsey (1989) are transformations of the parameter space. However, the functions on the right-hand side of these equation are *expectations* of the parameters that appear on the left-hand sides. Thus, strictly speaking, these equations are in error. Although replacing the 'true' parameter values with the expectations would correct the equations, it is then obvious that the invariant property cannot be applied because the 'true' parameter values

Table 1 Simulated quantitative locus parameters, marker locus statistics, maximum likelihood estimates by two methods, and log likelihood at *ML* for two simulated populations

Simulation*	Parameter†	Simulated value	Marker statistic	*ML* estimate‡ *TML*	*PML*
1 (500)	μ_{11}	25.0	25.69	24.47	24.69
	μ_{12}	30.0	30.42	30.63	30.51
	μ_{22}	35.0	34.36	35.22	35.19
	σ_{11}	5.0	5.24	4.34	4.54
	σ_{12}	5.5	5.85	5.50	5.46
	σ_{22}	6.0	6.45	6.09	6.29
	r	0.1	—	0.106	0.087
	Log likelihood§			−1593.05	−1593.24
2 (1000)	μ_{11}	25.0	31.90	24.81	31.90
	μ_{12}	35.0	32.42	35.00	32.42
	μ_{22}	35.0	33.72	35.69	33.72
	σ_{11}	5.0	7.84	4.81	7.84
	σ_{12}	7.0	7.69	6.06	7.69
	σ_{22}	7.0	7.73	7.94	7.73
	r	0.4	—	0.419	0.000
	Log likelihood			−3459.94	−3464.90

*Number of individuals in each population is given in parenthesis.

†μ_{11}, μ_{12}, μ_{22} = Means for the three QTL genotypes.
σ_{11}, σ_{12}, σ_{22} = standard deviations for the three QTL genotypes.
r = Recombination frequency between the QTL and the genetic marker.

‡*TML* = *ML* estimate derived by scanning the seven-dimension likelihood space.
PML = *ML* estimate derived by the method of Luo & Kearsey (1989).

§Natural base.

cannot be expressed as functions of the statistics that appear on the right-hand sides of these equations.

Simulations

The F_2 populations were simulated, as described by Weller (1986), consisting of 500 and 1000 individuals, respectively. The simulation values are given in Table 1. In the first population, a co-dominant QTL was simulated, with $r=0.1$. In the second population, the homozygote with the higher trait value was dominant, with $r=0.4$. *PML* estimates were computed with *r* scanned over the range of 0–0.5. 'True' *ML* estimates (*TML*) were computed by scanning the seven-dimensional space for all parameters. The scanning interval for the original search was ±5 units of the marker-genotype mean for the corresponding QTL genotype, and ±2 units of the marker-genotype standard deviation of the corresponding QTL genotype. In each round of scanning, five points were tested, thus the number of combinations tested were $5^7=78{,}125$. In subsequent scanning rounds for both methods, the interval was decreased by half, centred on the values with the highest likelihood. Iteration was continued until the difference in log *L* (natural base) between the highest combination of parameter values and the next highest fell below 10^{-3}.

Results and discussion

The marker-genotype means and standard deviations, and the *PML* and *TML* parameter estimates are presented in Table 1. For Simulation 1, a co-dominant QTL, the *PML* and *TML* estimates were quite similar for all seven parameters. The difference of 0.2 between the log likelihoods means that the likelihood of the *TML* solution is 1.23 times the likelihood of the *PML* solution. Both methods give results that are quite close to the true values, and within the approximate standard errors given by Weller (1986) for a population of this size. For Simulation 2, a QTL with complete dominance, the TML estimates were again close to the true values, and within the approximate standard errors; while for PML, maximum likelihood was obtained with complete linkage between the genetic marker and the QTL. Consequently, the *PML* estimates for the other parameters were equal to the market locus values. In addition, the difference between log *L* at the solution values was 5.0 for the second simulation, that is the likelihood of the *TML* solution is 148 times the likelihood of the *PML* solution.

One of the authors (J. I. Weller, unpublished data) attempted to derive *PML* solutions on the field data presented in Weller (1986). Results similar to those presented in this study for the dominant locus were obtained. This method was therefore deemed unsuitable as a short-cut *ML* algorithm. Some of the biases evident in the *PML* estimates for the dominant locus, i.e. underestimation of *r* with complete dominance and high recombination values, and underestimation of the heterozygote mean, were also found by Luo & Kearsey (1989). They also found that the QTL variances were consistently underestimated by their method, while in the present study, the *PML* variance estimates were greater than the simulated values for the dominant locus.

Luo & Kearsey (1989) did not compare the *PML* estimates to *TML* estimates. A single simulation, as performed in the present study, is sufficient to demonstrate the the *PML* and *TML* estimates are not identical. Even if their proof of equivalence is not correct, it is still possible that, under conditions of interest, the two methods may be approximately equal. As shown by the results in Table 1, this is not necessarily the case. Clearly a far more extensive analysis would be necessary to determine the criteria as to which populations would be amenable to *PML* estimation, if this is at all possible. Even for the two examples given, no major differences are apparent from inspection of the marker genotype statistics. Both loci appear to be co-dominant.

Lander & Botstein (1989) used the *EM* algorithm on a backcross population to derive '*ML*' estimates for the means and variances of a QTL with recombination frequency assumed to be constant, and then scanned the one-dimensional space for *r*. Thus, a '*ML*' value was computed by *EM* for each *r*-value. The combination of parameter estimates that gave the highest likelihood was then taken as the final *ML* estimate. Although the *EM* algorithm is guaranteed to converge to a maximum, provided one exists within the parameter space (Dempster *et al.*, 1977), this method is not 'true' *EM* because solutions were not estimated simultaneously for all five parameters. Further study is suggested to determine whether the method of Lander & Botstein (1989) does in fact yield 'true' *ML* solutions.

Acknowledgements

This research was supported in part by a grant from the Israeli Milk Marketing Board.

References

ARAVE, C. W., LAMB, R. C. AND HINES, H. C. 1971. Blood and milk protein polymorphisms in relation to feed efficiency and production traits of dairy cattle. *J. Dairy Sci.*, **54**, 106–112.

BAILEY, N. T. 1961. *Introduction to the Mathematical Theory of Genetical Linkage*. Clarendon Press, Oxford.

BECKMANN, J. S. AND SOLLER, M. 1983. Restriction fragment length polymorphisms in genetic improvement, methodologies, mapping and costs. *Theor. Appl. Genet.*, **67**, 35–43.

BECKMANN, J. S. AND SOLLER, M. 1987. Molecular markers in animal genetic improvement. *Bio/Technol.*, **5**, 573–576.

BEEVER, J. S., GEORGE, P. D., FERNANDO, R. L., STORMONT, C. J. AND LEWIN, H. A. 1990. Associations between genetic markers and growth and carcass traits in a paternal half-sib family of Angus cattle. *J. Anim. Sci.*, **68**, 337–344.

BRUM, E. W., RAUSCH, W. H., HINES, H. C. AND LUDWICK, T. M. 1968. Association between milk and blood polymorphism types and lactation traits of Holstein cattle. *J. Dairy Sci.*, **51**, 1031–1038.

DAHLQUIST, G. AND BJORCK, A. 1974. *Numerical Methods*. Prentice Hall, Englewood Cliffs.

DARVASI, A. 1990. *Analysis of genes affecting quantitative traits with the aid of genetic marker brackets and maximum likelihood methodology.* MSc thesis, The Hebrew University, Jerusalem.

DEMPSTER, A. P., LAIRD, N. M. AND RUBIN, D. B. 1977. Maximum likelihood from incomplete data via the *EM* algorithm. *J. Roy. Stat. Sco. (Series B)*, **39**, 1–38.

EDWARDS, M. D., STUBER, C. W. AND WENDEL, J. F. 1987. Molecular-marker-facilitated investigations of quantitative trait loci in maize. I. Numbers, genomic distribution and types of gene action. *Genetics*, **116**, 113–125.

GELDERMAN, H., PEIPER, U. AND ROTH, B. 1985. Effects of marker chromosome sections on milk performance in cattle. *Theor. Appl. Genet.*, **70**, 138–146.

GOYON, D. S., MATHER, R. E., HINES, H. C., HAENLEIN, G. F. W., ARAVE, C. W. AND GAUNT, S. N. 1987. Associations of bovine blood and milk polymorphisms with lactation traits, Holsteins. *J. Dairy Sci.*, **70**, 2585–2598.

HAENLEIN, G. F. W., GOYON, D. S., MATHER, R. E. AND HINES, H. C. 1987. Associations of bovine blood and milk polymorphisms with lactation traits, Guernseys. *J. Dairy Sci.*, **70**, 2599–2609.

HINES, H. C. 1990. Genetic markers for quantitative trait loci in dairy cattle. In: *Proceedings of the 4th World Congress on Genetics Applied to Livestock Production*, Edinburgh, **13**, 121–124.

HINES, H. C., KIDDY, C. A., BRUM, E. W. AND ARAVE, C. W. 1969. Linkage among cattle blood and milk polymorphisms. *Genetics*, **62**, 401–412.

JENSEN, J. 1989. Estimation of recombination parameters between a quantitative trait locus (QTL) and two marker gene loci. *Theor. Appl. Genet.*, **78**, 613–618.

KAHLER, A. L. AND WHERHAHN, C. F. 1986. Associations between quantitative traits and enzyme loci in the F2 population of a maize hybrid. *Theor. Appl. Genet.*, **72**, 15–26.

KNAPP, S. J., BRIDGES, W. C. AND BIRKES, D. 1990. Quasi-Mendelian analyses of quantitative trait loci using molecular marker linkage maps. *Theor. Appl. Genet.*, **79**, 582–592.

LANDER, E. S. AND BOTSTEIN, D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, **121**, 185–199.

LUO, Z. W. AND KEARSEY, M. J. 1989. Maximum likelihood estimation of linkage between a marker gene and a quantitative locus. *Heredity*, **63**, 401–408.

NEIMANN-SORENSEN, A. AND ROBERTSON, A. 1961. The association between blood groups and several production characters in three Danish cattle breeds. *Acta Agric. Scand.*, **11**, 163–196.

PATERSON, A. H., LANDER, E. S., HEWITT, J. D., PETERSON, S., LINCOLN, S. E. AND TANKSLEY, S. D. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature*, **335**, 721–726.

SAX, K. 1923. The association of size differences with seedcoat pattern and pigmentation in *Phaseeolus vulgaris*. *Genetics*, **8**, 552–560.

SIMPSON, S. P. 1989. Detection of linkage between quantitative trait loci and restriction fragment length polymorphisms using inbred lines. *Theor. Appl. Genet.*, **77**, 815–819.

SOLLER, M. AND BECKMANN, J. S. 1982. Restriction fragment length polymorphisms and genetic improvement. In: *Proceedings of the 2nd World Congress on Genetics Applied to Livestock Production*, Madrid, **6**, 396–404.

SOLLER, M. AND BECKMANN, J. S. 1983. Genetic polymorphisms in varietal identification and genetic improvement. *Theor. Appl. Genet.*, **67**, 25–33.

SOLLER, M. AND BECKMANN, J. S. 1985. Restriction fragment length polymorphisms and animal genetic improvement. *Rural Rev.*, **6**, 10–18.

STUBER, C. W., EDWARDS, M. D. AND WENDEL, J. F. 1987. Molecular marker-facilitated investigations of quantitative trait loci. II. Factors influencing yield and its component traits. *Crop Sci.*, **27**, 639–648.

TANKSLEY, S. D., MEDINA-FILHO, H. AND RICK, C. M. 1982. Use of naturally-occurring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. *Heredity*, **49**, 11–25.

THODAY, J. M. 1961. Location of polygenes. *Nature*, **191**, 368–370.

WELLER, J.I. 1986. Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. *Biometrics*, **42**, 627–640.

WELLER, J. I. 1987. Mapping and analysis of quantitative trait loci in Lycopersicon. *Heredity*, **59**, 413–421.

WELLER, J. I., SOLLER, M. AND BRODY, T. 1988. Linkage analysis of quantitative traits in an Interspecific cross of tomato (*Lycopersicon esculentum* × *Lycopersicon pimpinellifolium*) by means of genetic markers. *Genetics*, **118**, 329–339.

ZHUCHENKO, A. A., SAMOVOL, A. P., KOROL, A. B. AND ANDRYUSHCHENKO, V. K. 1979. Linkage between loci of quantitative characters and marker loci. II. Influence of three tomato chromosomes on variability of five quantitative characters in backcross progenies. *Genetika*, **15**, 672–683.