# Quantitative Trait Loci in *Brassica rapa*

Brian S. Yandell*
Departments of Horticulture and Statistics, University of Wisconsin–Madison

## Abstract

This paper briefly examines current methodology for developing genetic linkage maps and using them to find loci for quantitative traits (QTL). Maximum likelihood interval mapping is viewed as an extension of classical least squares methods when the trait of interest is normally distributed and located near a genetic marker. Some problems in finding multiple loci for a quantitative trait are examined for days to budding as measured on F2 plants from a *Brassica rapa* cross. Relevant design aspects of molecular biology experiments are briefly noted.

## Introduction

Last year colleagues in the plant sciences approached me with questions about recently developed programs for locating quantitative traits on genetic linkage maps (Lander and Botstein, 1989). They wanted to know how this related to "classical" approaches, and whether this new maximum likelihood approach was more appropriate.

The present study concerns the cross of two varieties of *Brassica rapa*, a Michihili Chinese cabbage (M) female with a Spring broccoli (S) male plant, producing a single F1 offspring which was self-pollinated. The resultant F2 seeds were germinated, yielding 95 plants which were measured for various phenotypic traits (observable characteristics such as days to first flower, days to budding, etc.). The F2s were assayed by 297 restriction fragment length polymorphisms, or RFLPs, which were drawn from previous studies of *Brassica*, DNA from both grandparents or the F1 parent, or selfed progeny of same.

A genetic linkage map was constructed (Song et al., 1990) which locates markers relative to one another based on the frequency of genetic recombination (crossover of chromosome pairs during meiosis) on the ten chromosome pairs in this genome. They measured the association between the RFLP patterns for each marker and the phenotypic traits and used the genetic linkage map to find probable sites, or loci, of genes controlling those traits (Song, Slocum and Osborn, 1991).

The purpose of this paper is to examine the statistical properties of such quantitative trait loci (QTL). In particular we show the connection between the classical regression model at markers and the maximum likelihood interval mapping method presented in Lander and Botstein (1989) and discuss some inferential questions concerning confidence intervals and finding multiple loci (major and minor genes) which control days to budding.

## RFLPs and Linkage Maps

Chromosomes come in pairs, and offspring inherit one of a pair from each parent. Any locus on a chromosome pair has two "alleles," or forms of DNA, one from each parent. The F1 was "heterozygous" (had two different alleles) at each marker locus and presumably at all QTL. F2 plants inherit (via F1) both alleles at a locus from one grandparent (MM or SS) or one from each (MS). Genetic recombination can lead to different allele types along the same chromosome, which is exploited to generate RFLP linkage maps (Lander and Botstein, 1989).

RFLP involves digesting DNA with an enzyme and using discrepant fragment lengths as markers for genetic differences among individuals. The enzyme cuts DNA adjacent to a specific base pair pattern, say `ACGTAT`. A change (mutation or recombination) in this restriction site for one variety (say `ACTTAT`) would be missed by the enzyme, resulting in one long fragment rather than two shorter ones—a polymorphism. Other forms of DNA rearrangement between restriction sites (e.g. insertion/deletion/transposition) can also create polymorphisms. DNA fragments are separated by size on a Southern blot and "probed" by $^{32}P$-labelled DNA pieces which bond to homologous

*Research supported by USDA-CSRS grant 511-100. Conference accomodations supported in part by Interface Foundation of North America. Special thanks to Keming Song and Mary Slocum for providing data.

DNA fragments. Ideal genetic markers are probes which highlight exactly one RFLP, scoring F2s at this marker as MM, SS (parent types) or MS (hybrid, having both length fragments). However, RFLP patterns may be difficult to align (Branscomb, 1991) or distinguish (Figure 1).

Nearly adjacent genetic markers should largely agree in allele type across the F2s, with differences probably due to recombination between the markers. Distance is roughly proportional to the frequency of recombination: $1\% \approx 1$ centi-Morgan (cM) $\approx 10^5 - 10^6$ DNA base pairs (may vary along a chromosome and between species). Genetic linkage maps are currently constructed by examining pairs of markers, then triplets, then piecing together whole chromosomes (Lander and Green, 1987; Song et al. 1991).

Lander's MAPMAKER program provides a user-friendly environment for this empirical maximization of the joint likelihood of all marker loci across the genome. Interesting questions remain about further optimizing the search algorithm and ascertaining that it converges to a unique global maximum.

## QTL, LOD and MLE

Genetic linkage maps are used to find quantitative traits loci, or QTL. The mean for a single locus quantitative trait $y$ depends on the allele type, i.e.,

$$E(y) = \mu + ax + d(1 - |x|),\ V(y) = \sigma^2,$$

with $x = 1, -1, 0$ if the locus is of grandparent type MM, SS, or hybrid (MS), respectively. Here $\mu$ is the reference mean, $a$ the additive allelic effect and $d$ the dominance effect of allele type M. Under the null hypothesis of no QTL ($a = d = 0$), the $F$-statistic

$$F = [\sum(\bar{y} - \hat{y})^2/\nu_1]/\hat{\sigma}^2 \text{ is distributed as } F_{\nu_1,\nu_2},$$

with $\bar{y}$ the sample mean, $\hat{y}$ the least squares estimate, $\hat{\sigma}^2$ the variance estimate and degrees of freedom $\nu_1 = 2$ and $\nu_2 = n - 3$ for $n$ F2s scored at this locus. For normal $y$, this is equivalent to the likelihood ratio statistic, typically presented in human genetics as

$$\begin{aligned} LOD &= \log_{10}(\textit{likelihood ratio}) \\ &= [0.5\sum(\bar{y} - \hat{y})^2/\hat{\sigma}^2]/\log(10) \\ &= \nu_1 F/2\log(10)\ . \end{aligned}$$

For normally distributed traits, these two approaches are equivalent and exact. Transformations toward normality are used in practice. Qualitative traits (counts and $+/-$) should use the "deviance" (McCullagh and Nelder, 1983) instead of the sum of squares. In some cases, this reduces to a $\chi^2$ test on two-way frequency tables at each marker locus.

The "classical approach" computes the $F$-statistic at all markers, concluding that a QTL is near the marker locus with the most significant value. Lander and Botstein (1989) expanded the normal model to examine intervals between marker loci. Consider markers $m$ and $m'$ with recombinant frequency $r$ and indicators $x$ and $x'$. A QTL with $\rho r$ recombination with $m$ has conditional expectation

$$E(y|r, m, m') = \mu + a[(1 - \rho)x + \rho x'] + df(x, x'; \rho, r),$$

where $f$ is complicated but tractable (Knapp, Bridges and Birkes, 1990). However, conditional on position ($\rho$ and $r$) the model is linear in parameters $a$ and $d$.

Lander's MAPMAKER/QTL program profiles the likelihood (*cf.* Kalbfleisch and Sprott, 1970; Bates and Watts, 1988) across intervals for adjacent markers on the linkage map, with the maximum likelihood estimator (MLE) corresponding to the highest peak. At the MLE, if $a = d = 0$ then

$$\max(LOD) \approx [\nu_1/2\log(10)]F_{\nu_1,\nu_2} \approx \chi^2_{\nu_1}/2\log(10),$$

with the latter approximation used in practice, ignoring the extra variation of the estimate $\hat{\sigma}^2$.

## Confidence Regions for QTL

Confidence regions arise by inverting the probability statement $Pr\{\max(LOD) \le c_\alpha\} = 1 - \alpha$. The 99% theoretical confidence region for the major QTL of the phenotypic trait "days to budding" lies on chromosome 3 (Figure 2a), primarily around 100 cM but with small intervals around 60 and 80 cM. These intervals have $LOD$ scores at least $\max(LOD) - 2$, with $c_{.01} \approx \chi^2_{2;.01}/2log(10) \approx 2$. Some QTL had confidence regions spanning intervals on several chromosomes. Beware that such regions may be too narrow, having much smaller coverage probabilty than expected (Terry Speed, pers. comm.).

The $LOD$ score should be roughly quadratic near the true locus. In practice, the profile is quite irregular (Figure 2a) and the profile traces (Bates and Watts, 1988; Ritter, Bisgaard and Bates, 1991) for $a$ and $d$ exhibit strong nonlinearity and some numerical problems (Figure 3). This suggests caution in interpreting the parameter estimates from current methods, and a need for some refinement.

## Major and Minor QTL

Finding multiple loci which control a quantitative trait is a stepwise process in which one identifies the major QTL, removes its effect, then proceeds to the most important minor QTL, and so on. For the two-loci additive model (ignoring interval mapping),

$$E(y) = \mu + a_1 x_1 + d_1(1 - |x_1|) + a_2 x_2 + d_2(2 - |x_2|),$$

where the major (1) and minor (2) loci may be on different chromosomes. The *LOD* can be decomposed, $LOD(1,2) = LOD(1) + LOD(2|1)$, suggesting that one fit the major locus model (as $\hat{y}_1$) and then conditionally fit the minor locus,

$$E(y - \hat{y}_1|x_1) = a_2 x_2 + d_2(2 - |x_2|) \ .$$

If the two loci were on separate chromosomes, one would expect estimates of $a_2$ and $d_2$ to be independent of $x_1$ and $LOD(2|1) = LOD(2)$. However, the profile likelihoods for possible minor loci for days to budding on chromosomes 6 and 7 changed substantially after removing a major QTL on chromosome 3 (Figures 2 and 4). Further, the MLE for the first minor QTL is at one end of chromosome 7, not in the middle of chromosome 6 as Figure 2 implies. These discrepancies may be due to epistasis (interaction), cosegregation of chromosomes during meiosis, or to a problem with modest sample size imbalance.

## Discussion

Maximum likelihood interval mapping of QTL builds naturally on classical approaches. Important computational and theoretical issues remain in linkage map construction and finding QTL.

Several sources of variation arise in building linkage maps. "Riflotyping" of polymorphisms involves a visual assay of thousands of columns on blots, although these may soon be scanned by computer. Riflotype errors of RFLP patterns along linkage maps may affect estimates of map distance and marker loci order (Steve Knapp, Tom Osborn, pers. comm.).

The interval mapping approach to QTL assumes independence between marker intervals and that epistasis (interaction) between loci is negligible (Steve Knapp, Terry Speed, pers. comm.). Variation in estimated marker location on the linkage map may affect QTL peaks and parameter estimates (Figure 3). Further, the empirical distribution of *LOD* scores needs investigation under varied conditions.

As markers become more closely spaced, one wonders how information from neighboring regions could be effectively included in the estimation of QTL, particularly when there may be multiple loci. Present technology allows closely spaced markers (1–2 cM), increasing the problems of riflotyping. This raises both estimation and design questions: should one gather more F2s or more markers? How can one account for riflotype and other errors in the estimation procedure? Finally, how can one efficiently use information in the local neighborhood of QTL to smooth the likelihood surface by appropriate penalization?

## References


Bates, D. M., and Watts, D. G. (1988) Graphical summaries of nonlinear inference regions. *Nonlinear Regression Analysis & Its Applications*, ch. 6. Wiley: New York.

Branscomb, E. (1991) Building physical genome maps by random clone overlap; a progress assessment of work on human chromosome 19. Proc. Workshop on Computational Molecular Biology (this conference).

Kalbfleisch, J. D., and Sprott, D. A. (1970) Application of likelihood methods to models involving large numbers of parameters. *J. Roy. Statist. Soc. B32*: 175-194.

Knapp, S. J., Bridges, Jr., W. C., and Birkes, D. (1990) Mapping quantitative trait loci using molecular marker linkage maps. *Theor. Appl. Genet. 79*: 583-592.

Lander, E. S., and Botstein, D. (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics 121*: 185-199.

Lander, E. S., and Green, P. (1987) Construction of multilocus genetic linkage maps in humans. *Proc. Natl. Acad. Sci. USA 84*: 2363-2367.

McCullagh, P., and Nelder, J. A. (1987) *Generalized Linear Models.* Chapman and Hall: New York.

Ritter, C., Bisgaard, S., and Bates, D. M. (1991) A comparison of approaches to inference for nonlinear models. (this conference).

Song, K., Slocum, M. K., and Osborn, T. C. (1991) Use of RFLP markers for locating genes controlling morphological traits in *Brassica rapa. Ms. in prep.*

Song, K., Suzuki, J. Y., Slocum, M. K., Williams, P. H., and Osborn, T. C. (1991) A linkage map of *Brassica rapa* (syn. *campestris*) based on restriction fragment length polymorphism loci. *Theor. Appl. Genet.* (in press).