Discussion of "A model selection approach for the identification of quantitative trait loci in experimental crosses" by Broman and Speed

Dr Brian S Yandell, Ms Chunfang Jin,
University of Wisconsin–Madison,
and Dr Jaya M Satagopan,
Memorial Sloan Kettering Cancer Center

May 30, 2002

The balance of model fit and complexity central to assessment is captured in BIC_δ, but complementary instruments have great value. Empirical studies of complex traits may detect 'major QTL', overlooking modifier genes that cannnot be localized. Model assessment guides the discovery process, selecting 'better' models without fixating on one 'best' model.

We find closely related Bayes factors effective despite recent criticism. Judicious choice of robust priors and empirical Bayes reduces influence of priors on Bayes factors (Gaffney 2001). Averaging over nuisance parameters can stabilize Bayes factors (Satagopan *et al.* 2000). Semi-log plots of posterior/prior against model identifier (number and chromosome pattern of QTL) provide useful graphical guides.

We differ on prediction. Model-averaged posteriors over the 'better' models of QTL loci and effects along the genome reveal genetic architecture (*cf.* Ball 2001). Further, agricultural breeding studies use predicted breeding values to 'select' individuals for future crosses. Marker assisted selection alone ignores important modifiers that become fixed in a few generations (Edwards and Page 1994).

Model search on a simulated framework map with 10cM spacing is revealing, but markers are clearly in a model or not. Practical QTL search spans a genome continuum, with two closely linked loci easily confused as one, depending on sample size and marker spacing. Model search with reversible jump MCMC allows joint sampling in the neighborhood to distinguish them. Whole genome RJ-MCMC differs from the authors' marker regression-based MCMC (*cf.* Satagopan and Yandell 1996; Silanpää and Arjas 1998; Stephens and Fisch 1998; Gaffney 2001).

Forward selection is biased, and backward elimination is impossible on the whole genome. Gaffney (2001) used a 'pre-burnin' phase beginning with no QTL and a high prior mean to aggressively build large initial models for subsequent

sampling with RJ-MCMC. This combined with block updates of effects and long-range position updates improves efficacy of multiple QTL searches (Gaffney 2001).

Finally, thresholds for model assessment criteria should be used with extreme caution. Thresholds were developed to test a single QTL against no QTL. Simulations (Goffinet and Mangin 1998) show empirical dependence on the size of other linked QTL.

References

[1] Ball RD (2001) Bayesian methods for quantitative trait loci mapping based on model selection: approximate analysis using the Bayesian information criterion. *Genetics 159*: 1351–1364.

[2] Edwards MD, Page NJ (1994) Evaluation of marker-assisted selection through computer-simulation. *Theor. Appl. Genet. 88*: 376–382.

[3] Gaffney PJ (2001) An efficient reversible jump Markov chain Monte Carlo approach to detect multiple loci and their effects in inbred crosses. PhD dissertation, Department of Statistics, University of Wisconsin–Madison.

[4] Goffinet B, Mangin B (1998) Comparing methods to detect more than one QTL on a chromosome. *Theor. Appl. Genet. 96*: 628-633.

[5] Satagopan JM, Newton MA, Raftery AE (2000) Easy estimation of normalizing constants and Bayes factors from posterior simulation: Stabilizing the harmonic mean estimator. Technical Report 1028, Department of Statistics, University of Wisconsin.

[6] Satagopan JM, Yandell BS (1996) Estimating the number of quantitative trait loci via Bayesian model determination. Special Contributed Paper Session on Genetic Analysis of Quantitative Traits and Complex Diseases, Biometrics Section, Joint Statistical Meetings, Chicago, IL.

[7] Stephens DA, Fisch, RD (1998) Bayesian analysis of quantitative trait locus data using reversible jump Markov chain Monte Carlo, *Biometrics 54*: 1334-1347.

[8] Sillanpää MJ, Arjas E (1998) Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data, *Genetics 148*: 1373-1388.