

Estimating allelic number and identity-in-state of QTLs: Bayesian modeling, sampling strategy, and computational implementation*

Xiao-Lin Wu, Ph.D.
Washington State University

Abstract:

Several circumstances lead to the need for QTL analysis using crosses derived from multiple inbred lines. Analyses exist that assume all founders carry unique alleles. In practice, however, many configurations are possible and some founders may carry QTL alleles that are identical in state (IIS). Though the identity of alleles is not observable, the phenotypic effect of alleles provides information that may guide the specification of statistical models estimating the correct number of alleles and inferring identity-in-state at a QTL among parents.

Variable-allele model Bayesian analysis is proposed that selects among the possible identity-in-state configurations and jointly estimates the allelic effects of identical-in-state parents. Changing the number of alleles is carried out via a reversible jump algorithm, which leads to necessary changes in the dimensions of the allelic configuration matrix in the model. The results from analyzing simulation data indicated that the variable model Bayesian analysis consistently estimated allelic effects with lower mean-squared errors than the analysis assuming a fixed number of alleles. Estimated number of alleles actually reflected groups of parents carrying alleles with similar effects. When a QTL consisted of three or four alleles, their effects diverged sufficiently to be distinguished by the analysis. But with ≥ 20 alleles segregating, the analysis would lump alleles with similar effects into a single group. The results also showed a large influence of priors on posterior allelic number in small-sized families. With two or three progeny per family, estimated number of alleles was dominated by the prior information. The likelihood of configurations that falsely grouped parents carrying different alleles decreased as family size increased. Unless there were many (i.e. >100) progeny in each mapping family, the information from DNA marker and phenotypic data was insufficient to determine with high probability the allelic number. Further, sampling strategy played an important role in determining the power of the analysis and its accuracy in estimating the allelic number. Uncertainty in sampling QTL alleles took place at two levels: when sampling founders to produce the F1 parents, and during Mendelian segregation of the F1 parents to produce mapping families. Sampling an incomplete pool of alleles was significantly associated with biased estimation of QTL alleles and variance. This was a typical situation that challenged QTL mapping using single families. The present results, however, did not point to an unequivocal optimal sampling but instead suggested that the strategy be tailored to the objectives at hand. A software package has been developed using C++ programming to implement the present analysis, based on one-dimensional search of the allelic number given a pre-determined number of QTLs.

* The research was funded by USDA-NRI CSREES Project Award No. 2001-35301-10848, which was awarded to Dr. Jean-Luc Jannink (Iowa State University).