An Integrated approach to detecting QTL in practical QTL mapping applications

Ruming Li, Ph.D.
Louisiana State University

Abstract:

Since Lander and Botstein (1989) proposed quantitative trait locus (QTL) mapping using a molecular marker, it has received much attention of researchers and spawned a variety of statistical approaches to detecting a putative QTL that is otherwise hard to localize. Among these approaches is the typical simple interval mapping, multiple interval mapping, composite interval mapping, etc. and more advanced ones would be expected. Essentially, however, these approaches just provide statistical or logic association of markers with the QTL of interest, rather than a genuine or physical one, and the claimed establishment of a marker-QTL linkage is based totally on statistic inference. One issue is that statistic inference could run a false positive/negative risk, depending upon significance thresholds. The other issue is the fact that maximum likelihood (ML) ratio from EM methodology is deemed to follow an "approximate" chi-squared distribution and it is not the exact one to follow, which would lead to more uncertainty, or even worse, arbitrary. Therefore, to minimize the uncertainty brought by a statistical inference and increase robustness of a QTL mapping, using just one of these approaches should be judicious and adopting joint approaches is recommended. One more issue is that some of approaches apply only to specific biological settings (e.g., in plant and animal breeding), which may not be universally applicable (e.g., in human genetics and medicine). To extend to a generality one must work out separate solutions in various contexts. This study proposes an integrated way to perform ML and regression approaches simultaneously, ended with a hypothesis test to detect if they are significantly different, and we will choose whichever is of nonsignificant difference as the acceptable outcome. Otherwise, it is treated as a failure and has to follow up with a next mapping scanning along a genome using more polymorphism markers until it arrives at an agreement between two approaches. The reason is that some of their results are claimed to have similarity but may not be true across varying species and research projects, and that this joint analysis is applied to the plants, animals, or humans if its pedigree data are available and meaningful. For human genetics and medical study, a model emphasizing on using regression of marker(s) on a phenotype is introduced here to be tailored to their contexts. In this model, each marker variable contributes to the entire dependence of a phenotype on at least one marker if more polymorphism associated with a target QTL may not be available or acquiring such multiple markers is technically unfeasible or operationally expensive for a large population size. The model is linearly built, as advanced interactive terms are not significant to detecting a QTL. The phenotypic data will be classified as within-family analysis for pedigree data and as among-family analysis for non-pedigree data but their postulated marker genotypes are same. The classified results are still subject to statistical testing as our ultimate resort that makes them agreeable and hence enhance the power of statistic inferences. A Windows-based software will be developed for implementing and demonstrating this integrated approach.