Mapping multiple QTL in experimental crosses

Karl W Broman
Department of Biostatistics and Medical Informatics
University of Wisconsin – Madison

www.biostat.wisc.edu/~kbroman

[→ Teaching → Miscellaneous lectures]

Example


250 male mice from the backcross (A × B) × B
Blood pressure after two weeks drinking water with 1% NaCl
Goals

- Identify quantitative trait loci (QTL) (and interactions among QTL)
- Interval estimates of QTL location
- Estimated QTL effects

LOD curves
Estimated effects

Chr 1 @ 48 cM

Chr 4 @ 30 cM

Chr 6 @ 24 cM

Chr 15 @ 20 cM

Genotype

blood pressure

Modeling multiple QTL

• Reduce residual variation $\rightarrow$ increased power

• Separate linked QTL

• Identify interactions among QTL (epistasis)
For all pairs of positions, fit the following models:

\[ H_f : y = \mu + \beta_1 q_1 + \beta_2 q_2 + \gamma q_1 q_2 + \epsilon \]
\[ H_a : y = \mu + \beta_1 q_1 + \beta_2 q_2 + \epsilon \]
\[ H_1 : y = \mu + \beta_1 q_1 + \epsilon \]
\[ H_0 : y = \mu + \epsilon \]

\( \log_{10} \) likelihoods:

\[ l_f(s, t) \quad l_a(s, t) \quad l_1(s) \quad l_0 \]

LOD scores:

\[ \text{LOD}_f(s, t) = l_f(s, t) - l_0 \]
\[ \text{LOD}_a(s, t) = l_a(s, t) - l_0 \]
\[ \text{LOD}_i(s, t) = l_f(s, t) - l_a(s, t) \]
\[ \text{LOD}_1(s) = l_1(s) - l_0 \]
Results: $\text{LOD}_i$ and $\text{LOD}_f$

![Heatmap showing LOD values for different chromosomes](image-url)
Consider each pair of chromosomes, \((j, k)\), and let \(c(s)\) denote the chromosome for position \(s\).

\[
M_f(j, k) = \max_{c(s) = j, c(t) = k} \text{LOD}_f(s, t)
\]

\[
M_a(j, k) = \max_{c(s) = j, c(t) = k} \text{LOD}_a(s, t)
\]

\[
M_1(j, k) = \max_{c(s) = j \text{ or } k} \text{LOD}_1(s)
\]

\[
M_i(j, k) = M_f(j, k) - M_a(j, k)
\]

\[
M_{fv1}(j, k) = M_f(j, k) - M_1(j, k)
\]

\[
M_{av1}(j, k) = M_a(j, k) - M_1(j, k)
\]

Results: LOD\(_i\) and LOD\(_{fv1}\)
Thresholds

A pair of chromosomes \((j, k)\) is considered interesting if:

\[
M_f(j, k) > T_f \quad \text{and} \quad \{ M_{fv1}(j, k) > T_{fv1} \text{ or } M_i(j, k) > T_i \}
\]

or

\[
M_a(j, k) > T_a \quad \text{and} \quad M_{av1}(j, k) > T_{av1}
\]

where the thresholds \((T_f, T_{fv1}, T_i, T_a, T_{av1})\) are determined by a permutation test with a 2d scan.

Estimated effects

1 x 4

6 x 15
Hypothesis testing?

- In the past, QTL mapping has been regarded as a task of hypothesis testing.

  Is this a QTL?

  Much of the focus has been on adjusting for test multiplicity.

- It is better to view the problem as one of model selection.

  What set of QTL are well supported?
  Is there evidence for QTL-QTL interactions?

Model = a defined set of QTL and QTL-QTL interactions (and possibly covariates and QTL-covariate interactions).
Model selection

- Class of models
  - Additive models
  - + pairwise interactions
  - + higher-order interactions
  - Regression trees

- Model fit
  - Maximum likelihood
  - Haley-Knott regression
  - extended Haley-Knott
  - Multiple imputation
  - MCMC

- Model comparison
  - Estimated prediction error
  - AIC, BIC, penalized likelihood
  - Bayes

- Model search
  - Forward selection
  - Backward elimination
  - Stepwise selection
  - Randomized algorithms

Target

- Selection of a model includes two types of errors:
  - Miss important terms (QTLs or interactions)
  - Include extraneous terms

- Unlike in hypothesis testing, we can make both errors at the same time.

- Identify as many correct terms as possible, while controlling the rate of inclusion of extraneous terms.
What is special here?

• Goal: identify the major players

• A continuum of ordinal-valued covariates (the genetic loci)

• Association among the covariates
  – Loci on different chromosomes are independent
  – Along chromosome, a very simple (and known) correlation structure

Exploratory methods

• Condition on a large-effect QTL
  – Reduce residual variation
  – Conditional LOD score:

\[
\text{LOD}(q_2 \mid q_1) = \log_{10} \left\{ \frac{\Pr(\text{data} \mid q_1, q_2)}{\Pr(\text{data} \mid q_1)} \right\}
\]

• Piece together the putative QTL from the 1d and 2d scans
  – Omit loci that no longer look interesting (drop-one-at-a-time analysis)
  – Study potential interactions among the identified loci
  – Scan for additional loci (perhaps allowing interactions), conditional on these
Controlling for chr 4

Automation

- Assistance to the masses
- Understanding performance
- Many phenotypes
Additive QTL

Simple situation:

- Dense markers
- Complete genotype data
- No epistasis

\[ y = \mu + \sum \beta_j q_j + \epsilon \quad \text{which } \beta_j \neq 0? \]

\[ pLOD(\gamma) = LOD(\gamma) - T |\gamma| \]

0 vs 1 QTL: \( pLOD(\emptyset) = 0 \)

\[ pLOD(\{\lambda\}) = LOD(\lambda) - T \]

Experience

- Controls rate of inclusion of extraneous terms
- Forward selection over-selects
- Forward selection followed by backward elimination works as well as MCMC
- Need to define performance criteria
- Need large-scale simulations

Broman & Speed, JRSS B 64:641-656, 2002
Epistasis

\[ y = \mu + \sum \beta_j q_j + \sum \gamma_{jk} q_j q_k + \epsilon \]

\[ pLOD(\gamma) = LOD(\gamma) - T_m |\gamma|_m - T_i |\gamma|_i \]

\[ T_m = \text{as chosen previously} \]

\[ T_i = ? \]

Idea 1

Imagine there are two additive QTL and consider a 2d, 2-QTL scan.

\[ T_i = 95\text{th percentile of the distribution of} \]
\[ \max \text{LOD}_f(s, t) - \max \text{LOD}_a(s, t) \]

For the mouse genome:

\[ T_m = 2.69 \text{ (BC) or 3.52 (F}_2\text{)} \]

\[ T_i^H = 2.62 \text{ (BC) or 4.28 (F}_2\text{)} \]
Imagine there is one QTL and consider a 2d, 2-QTL scan.

\[ T_m + T_i = 95\text{th percentile of the distribution of } \max \text{LOD}_f(s, t) - \max \text{LOD}_1(s) \]

For the mouse genome:

\[ T_m = 2.69 \text{ (BC) or 3.52 (F}_2) \]
\[ T_i^H = 2.62 \text{ (BC) or 4.28 (F}_2) \]
\[ T_i^L = 1.19 \text{ (BC) or 2.69 (F}_2) \]
Results

\[ T_m = 2.69 \quad T_i^H = 2.62 \quad T_i^L = 1.19 \quad T_m + T_i^H = 5.31 \quad T_m + T_i^L = 3.88 \quad 2T_m = 5.38 \]

Profile LOD curves
Add an interaction?

Add another QTL?

$T_m = 2.69 \quad T_i^H = 2.62 \quad T_i^L = 1.19 \quad T_m + T_i^H = 5.31 \quad T_m + T_i^L = 3.88 \quad 2T_m = 5.38$
Add a pair of QTL?

\[ T_m = 2.69 \quad T^H_i = 2.62 \quad T^L_i = 1.19 \quad T_m + T^H_i = 5.31 \quad T_m + T^L_i = 3.88 \quad 2T_m = 5.38 \]

Summary

- QTL mapping is a model selection problem
- The criterion for comparing models is most important
- We’re focusing on a penalized likelihood method, with penalties derived from permutation tests with 1d and 2d scans
- Manichaikul et al., Genetics 181:1077–1086, 2009