Mapping multiple QTL in experimental crosses

Karl W Broman
Biostatistics & Medical Informatics
University of Wisconsin – Madison

www.biostat.wisc.edu/~kbroman
Backcross

P₁ × P₂ → F₁

F₁ → BC

250 male mice from the backcross \((A \times B) \times B\)

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
Two problems

The missing data problem:  
Markers $\leftrightarrow$ QTL

The model selection problem:  
QTL, covariates $\rightarrow$ phenotype
ANOVA at marker loci

- Split mice into groups according to genotype at a marker.
- Do a t-test / ANOVA.
- Repeat for each marker.
Interval mapping

Lander & Botstein (1989)

• Assume a single QTL model.

• Consider each position in the genome, one at a time, as the location of the putative QTL.

• Let $q = 0/1$ if the (unobserved) QTL genotype is BB/AB. (Or 0/1/2 if the QTL genotype is AA/AB/BB in an intercross.)

Assume $y \mid q \sim N(\mu_q, \sigma)$

• Calculate $p_q = \Pr(q \mid \text{marker data})$.

$$y \mid \text{marker data} \sim \sum_q p_q \phi(y \mid \mu_q, \sigma)$$

• $\text{LOD}(\lambda) = \log_{10} \left\{ \frac{\Pr(y|\text{QTL at } \lambda, \hat{\mu}_q, \hat{\sigma}_\lambda)}{\Pr(y|\text{no QTL}, \hat{\mu}, \hat{\sigma})} \right\}$
LOD curves
Permutation results

Genome-wide maximum LOD score
LOD curves
Modeling multiple QTL

- Reduce residual variation $\rightarrow$ increased power
- Separate linked QTL
- Identify interactions among QTL (epistasis)
Estimated effects

1 x 4

Chr 1 genotype

6 x 15

Chr 6 genotype

Chr 4 genotype

Chr 15 Genotype

Blood pressure

BB

BA

BB

BA

Estimated effects
Model selection

• Class of models
  – Additive models
  – + pairwise interactions
  – + higher-order interactions
  – Impose hierarchy on interactions?
  – Don’t allow QTL to be too close

• 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
Selection of a model includes two types of errors:

- Miss important terms (QTL or interactions)
- Include extraneous terms

Identify as many correct terms as possible, while controlling the rate of inclusion of extraneous terms.

Want the major players; correct identification of interactions is of secondary importance.
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
Additive QTL

Simple situation:

- Dense markers
- Complete genotype data
- No epistasis

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

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

Simple situation:

- Dense markers
- Complete genotype data
- No epistasis

\[ y = \mu + \sum \beta_j q_j + \epsilon \] which \( \beta_j \neq 0 \)?

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

0 vs 1 QTL: \( \text{pLOD}(\emptyset) = 0 \)

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

Simple situation:
- Dense markers
- Complete genotype data
- No epistasis

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

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

For the mouse genome:

\[ T = 2.69 \ (\text{BC}) \text{ or } 3.52 \ (F_2) \]
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

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

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

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

\[ T_i = ? \]
Imagine there are two additive QTL and consider a 2d, 2-QTL scan.

\[ T_i = 95\text{th percentile of the distribution of} \]
\[ \max \text{ LOD}_f(\lambda_1, \lambda_2) - \max \text{ LOD}_a(\lambda_1, \lambda_2) \]
Imagine there are two additive QTL and consider a 2d, 2-QTL scan.

\[ T_i = 95\text{th percentile of the distribution of} \]
\[ \max \text{LOD}_f(\lambda_1, \lambda_2) - \max \text{LOD}_a(\lambda_1, \lambda_2) \]

For the mouse genome:

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

\[ T_i^H = 2.62 \text{ (BC) or 4.28 (F}_2) \]
Models as graphs

A

B

C

D
Imagine there is one QTL and consider a 2d, 2-QTL scan.

\[ T_m + T_i = 95 \text{th percentile of the distribution of} \]
\[ \max \text{LOD}_f(\lambda_1, \lambda_2) - \max \text{LOD}_1(\lambda) \]
Imagine there is one QTL and consider a 2d, 2-QTL scan.

\[ T_m + T_i = 95\text{th percentile of the distribution of} \]
\[ \max \text{LOD}_f(\lambda_1, \lambda_2) - \max \text{LOD}_1(\lambda) \]

For the mouse genome:

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

A

B

C

D
Results

LOD = 23.1
Drop one term?

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

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

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

\[ T_m = 2.69 \quad T_H^i = 2.62 \quad T_L^i = 1.19 \quad T_m + T_H^i = 5.31 \quad T_m + T_L^i = 3.88 \quad 2T_m = 5.38 \]
Drop one term?

$T_m = 2.69 \quad T_i^H = 2.62 \quad T_i^L = 1.19 \quad T_m + T_i^H = 5.31 \quad T_m + T_i^L = 3.88 \quad 2T_m = 5.38$
Drop one at time

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

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

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

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

$T_m = 2.69 \quad T_i^H = 2.62 \quad T_i^L = 1.19 \quad T_m + T_i^H = 5.31 \quad T_m + T_i^L = 3.88 \quad 2T_m = 5.38$
Add an interaction?

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

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

- Study performance
  (especially relative to other approaches)
- Improve search procedures
- Measuring model uncertainty
- Measuring uncertainty in QTL location
- Covariates and QTL × covariate interactions
- That evil X chromosome
- Treat linked QTL differently?
• QTL mapping is a model selection problem

• The criterion for comparing models is most important

• We’re focusing on a penalized likelihood method and believe we have a practiceable solution
<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ani Manichaikul</td>
<td>Johns Hopkins University</td>
</tr>
<tr>
<td></td>
<td>(now at University of Virginia)</td>
</tr>
<tr>
<td>Gary Churchill</td>
<td>Jackson Laboratory</td>
</tr>
<tr>
<td>Saunak Sen</td>
<td>University of California, San Francisco</td>
</tr>
<tr>
<td>Terry Speed</td>
<td>University of California, Berkeley</td>
</tr>
<tr>
<td>Brian Yandell</td>
<td>University of Wisconsin – Madison</td>
</tr>
<tr>
<td>Fumihiro Sugiyama</td>
<td>now at University of Tsukuba, Japan</td>
</tr>
<tr>
<td>Bev Paigen</td>
<td>Jackson Laboratory</td>
</tr>
</tbody>
</table>