Quantitative Trait Loci (QTL) Mapping in Experimental Crosses

Karl Broman
Lab Animal 30(7):44-52,2001

Presented by:

Yan Wang

Outline

- **Introduction**
  - Terminology, data, model and assumptions
- **Single QTL analysis**
  - Estimation of QTL effect
  - Inference of QTL mapping – hypothesis testing
    - ANOVA
    - Interval mapping
- **Multiple QTL mapping – model selection**

Phenotypic outcomes

- Dichotomous trait
  - presence / absence of a disease
- Quantitative trait
  - blood pressure
  - survival time
  - Tumor mass
- No absolute distinction

Quantitative trait loci (QTLs)

- QTLs determine the genetic component of variation in quantitative traits.

- Quantitative traits are usually encoded by many genes (polygenes).
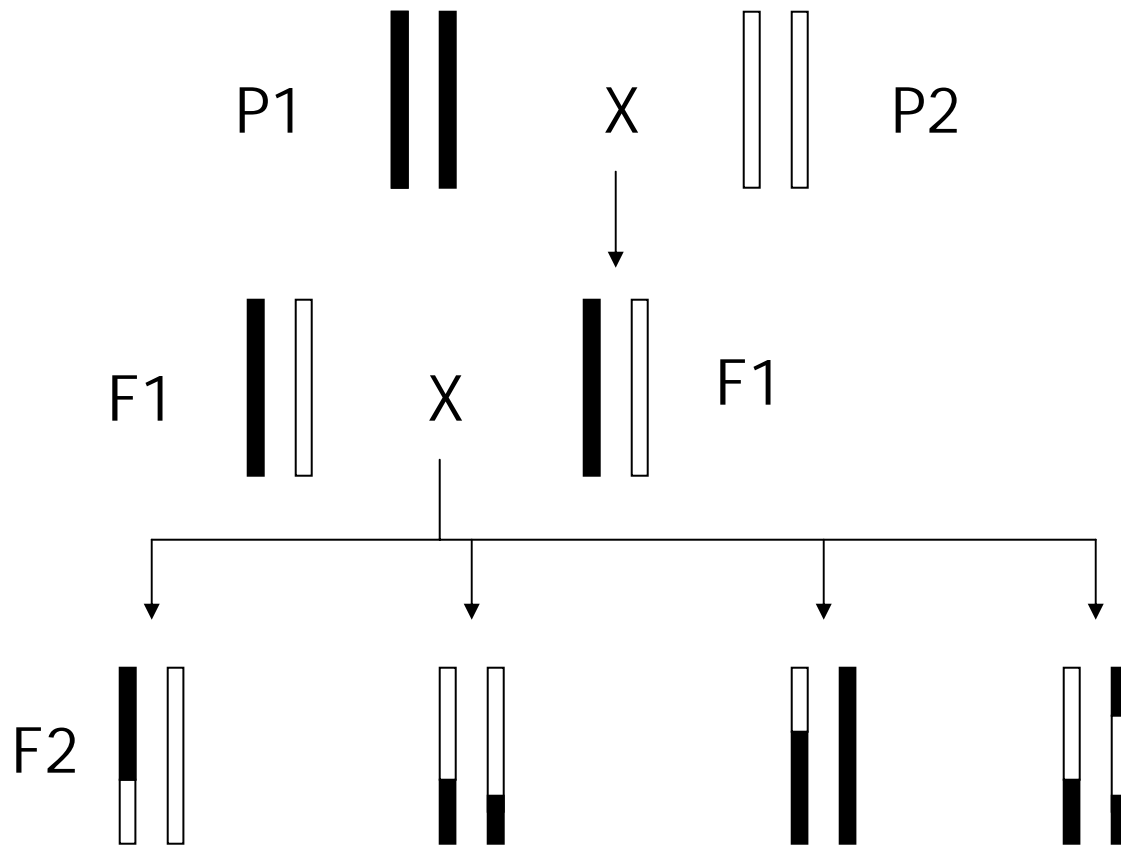
Experimental crosses

- Model organisms
  - quickly breed
  - extensively studied
  - E. coli, Drosophila, mouse, etc.
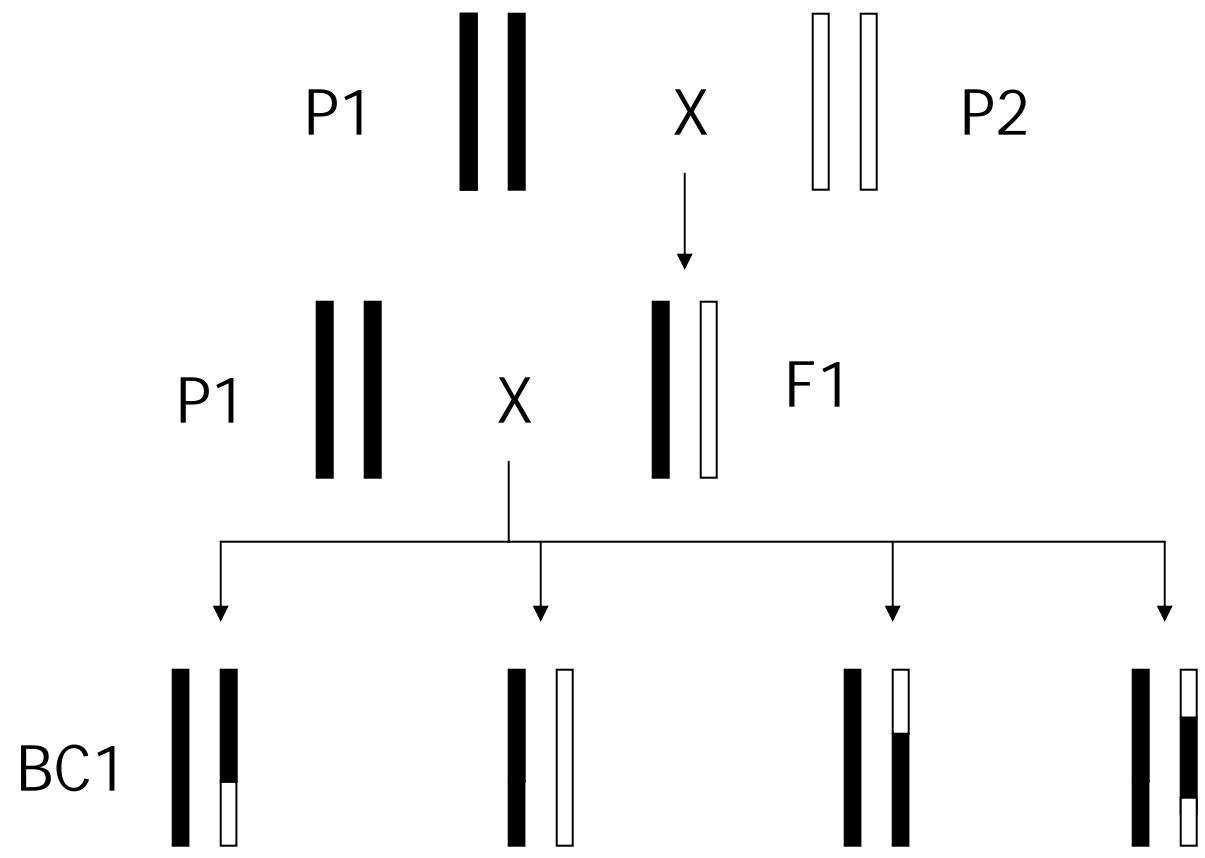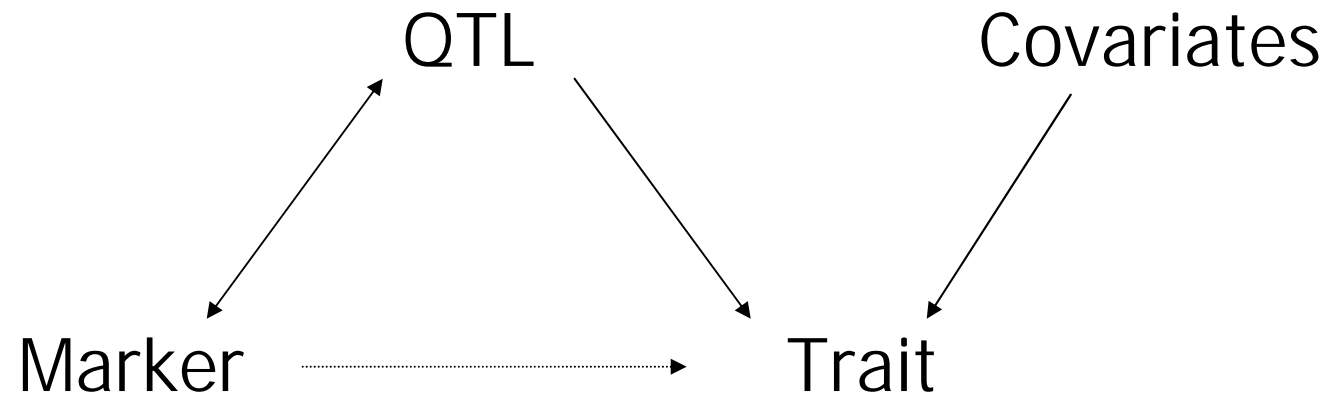- Intercross
- Backcross

Intercross

P1   **||**   X || P2

F1 | | X | | F1

F2 || || || ||

Backcross

QTL mapping
in experimental crosses

Experimental crossing creates associations between genetic marker loci and traits to allow localization of QTL.

QTL   Covariates

Marker ----→ Trait

Data structure
for a backcross experiment

- Phenotypes:

  y_i = quantitative measurement of trait

- Genotypes:

  x_{ij} = 0/1 coded for AA/AB at marker j

- Covariates:

  $\mathbf{Z}_i$ = environmental factors, demographics, etc.

where $i = 1, ..., n$; $j = 1, ..., M$.

Goals of QTL analysis

- Detect genetic effects
- QTL mapping: inference of the QTL location on chromosome
- Estimate the effects of allelic substitution

Model and Assumptions

- No interference in recombination process
- Independence
- Normality

$$y_i|X \sim N(\mu_X, \sigma_X^2)$$

- Homoscedasticity

$$\sigma_X^2 = \sigma^2$$

QTL effect from backcross

A   Q   a   q

- QTL effect $\Delta = \mu_{QQ} - \mu_{Qq}$
- $E[y_{AA}] = \mu_{QQ} \times Pr\{QQ|AA\} + \mu_{Qq} \times Pr\{Qq|AA\}$
  $$= \mu_{QQ} \times (1-r) + \mu_{Qq} \times r$$
- $E[y_{Aa}] = \mu_{QQ} \times Pr\{QQ|Aa\} + \mu_{Qq} \times Pr\{Qq|Aa\}$
  $$= \mu_{QQ} \times r + \mu_{Qq} \times (1-r)$$
- $E[y_{AA}] - E[y_{Aa}] = \Delta \times (1-2r)$

ANOVA (Marker regression)

- Split mice into groups according to genotypes at a marker
  - backcross: AA, Aa (two groups)
  - intercross: AA, Aa, aa (three groups)
- ANOVA/t-test
- Repeat for each marker j = 1, ..., M
- Adjust for multiple testing

ANOVA Table

Source of Variation	SS	DF	MS	F
Between groups	SSA	k-1	SSA/(k-1)	MST/MSE
Within groups	SSE	N-k	SSE/(N-k)	

Where k is the number of groups, N is the total sample size.

ANOVA (cont'd)

Advantages

- Simple
- Easily incorporates covariates Z
- Doesn't require a genetic map of markers
- Easily extended to multiple regression to account for multiple loci

Disadvantages

- Imperfect information about QTL location
- Individuals with missing genotype are excluded
- Power is small when linkage between marker and QTL is weak (sparse marker data)

Interval Mapping

L   Q   R   l   q   r

$Pr\{QQ|LL,RR\}=(1-r_L)(1-r_R)/(1-r)$

$Pr\{QQ|LL,Rr\}=(1-r_L)r_R/r$

$Pr\{QQ|Ll,RR\}=r_L(1-r_R)/r$

$Pr\{QQ|Ll,Rr\}=r_L r_R/(1-r)$

$Pr\{Qq|LL,RR\}=r_L r_R/(1-r)$

$Pr\{Qq|LL,Rr\}=r_L(1-r_R)/r$

$Pr\{Qq|Ll,RR\}=(1-r_L)r_R/r$

$Pr\{Qq|Ll,Rr\}=(1-r_L)(1-r_R)/(1-r)$

LOD Score – a likelihood ratio statistic

$$LOD(z) = \log_{10} \frac{L_1(\hat{\theta}_1)}{L_0(\hat{\theta}_0)}$$

where z is a location on the chromosome, and

$$
\begin{aligned}
L_1(\hat{\theta}_1) &= Pr\{y_i | \text{QTL at} z, \hat{\mu}_g, \hat{\sigma}_g^2\} \\
&= \prod_{i=1}^{n} \sum_g Pr\{y_i | G = g_i\} Pr\{G = g_i | x_{ij}, x_{i,j+1}\} \\
&= \prod_{i=1}^{n} \sum_g f_{\hat{\mu}_g, \hat{\sigma}_g^2}(y_i) Pr\{G = g_i | x_{ij}, x_{i,j+1}\} \\
L_0(\hat{\theta}_0) &= Pr\{y_i | \text{no QTL at} z, \hat{\mu}, \hat{\sigma}^2\} \\
&= \prod_{i=1}^{n} f_{\hat{\mu}, \hat{\sigma}^2}(y_i)
\end{aligned}
$$

LOD curve

- Likelihood profile
- A clear peak is taken as the QTL
- 1.5-LOD support interval

Null distribution of LOD score

- Computer simulations
    - Type of cross
    - Size of the genome
    - Number and spacing of genetic markers
    - Amount and pattern of missing genotypes
    - True phenotype distribution
- Permutation or bootstrap

Interval mapping (cont'd)

Advantages

- Takes proper account of missing data
- Interpolate positions between markers
- Provide a support interval
- Provide more accurate estimate of QTL effect

Disadvantages

- Intense computation
- Rely on a genetic map with good quality
- Difficult to incorporate covariate

Multiple QTLs

- Extension from ANOVA – multiple regression

- Extension from interval mapping
  - Composite Interval Mapping (CIM)
  - Multiple Interval Mapping (MIM)

Model selection

- Forward selection
- Backward deletion
- Stepwise selection
- Randomized search

Forward Selection

Backward Deletion

Stepwise Selection

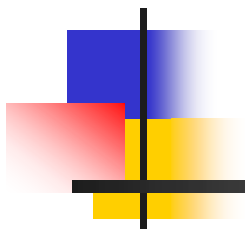# Model Selection in Interval Mapping -- Multiple Interval Mapping (MIM)

Forward selection:

- Assumption: QTLs are acting additively.

$$y = \mu + \Sigma \, \Delta_i x_i + \varepsilon$$

- $LOD(x|M) = \log_{10}\{Pr(data|M+x)/ Pr(data|M)\}$

Thank you!