

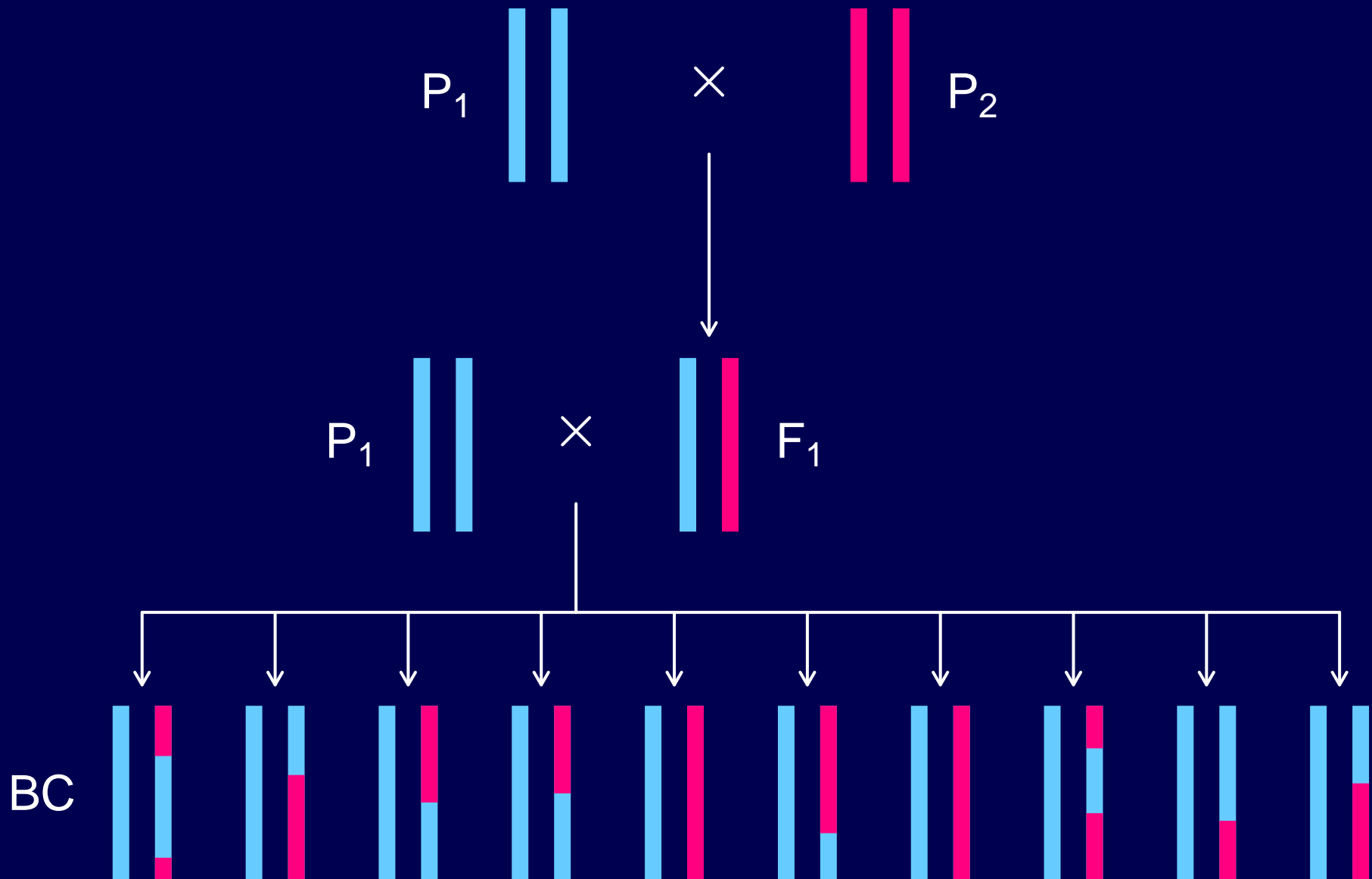
Mapping multiple QTL in experimental crosses

Karl W Broman

Biostatistics & Medical Informatics
University of Wisconsin – Madison

www.biostat.wisc.edu/~kbroman

Backcross

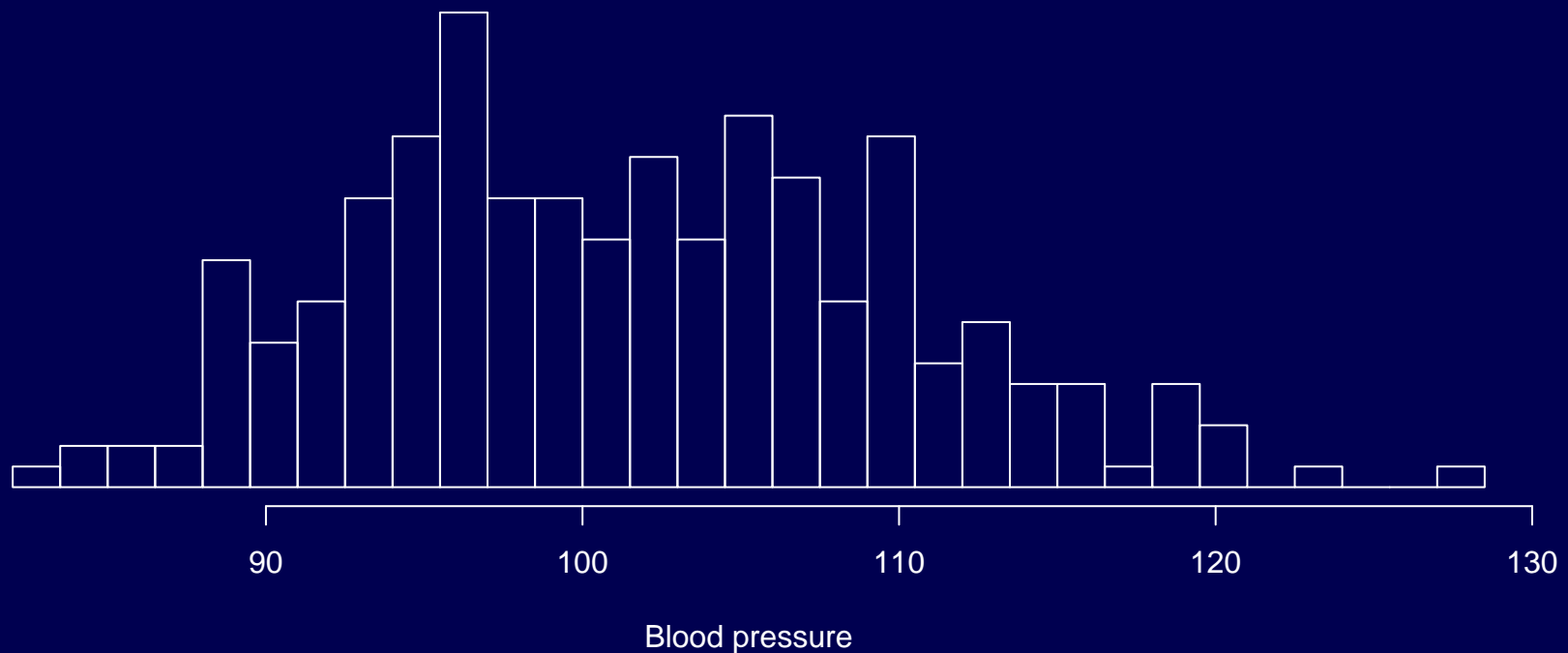


Phenotype data

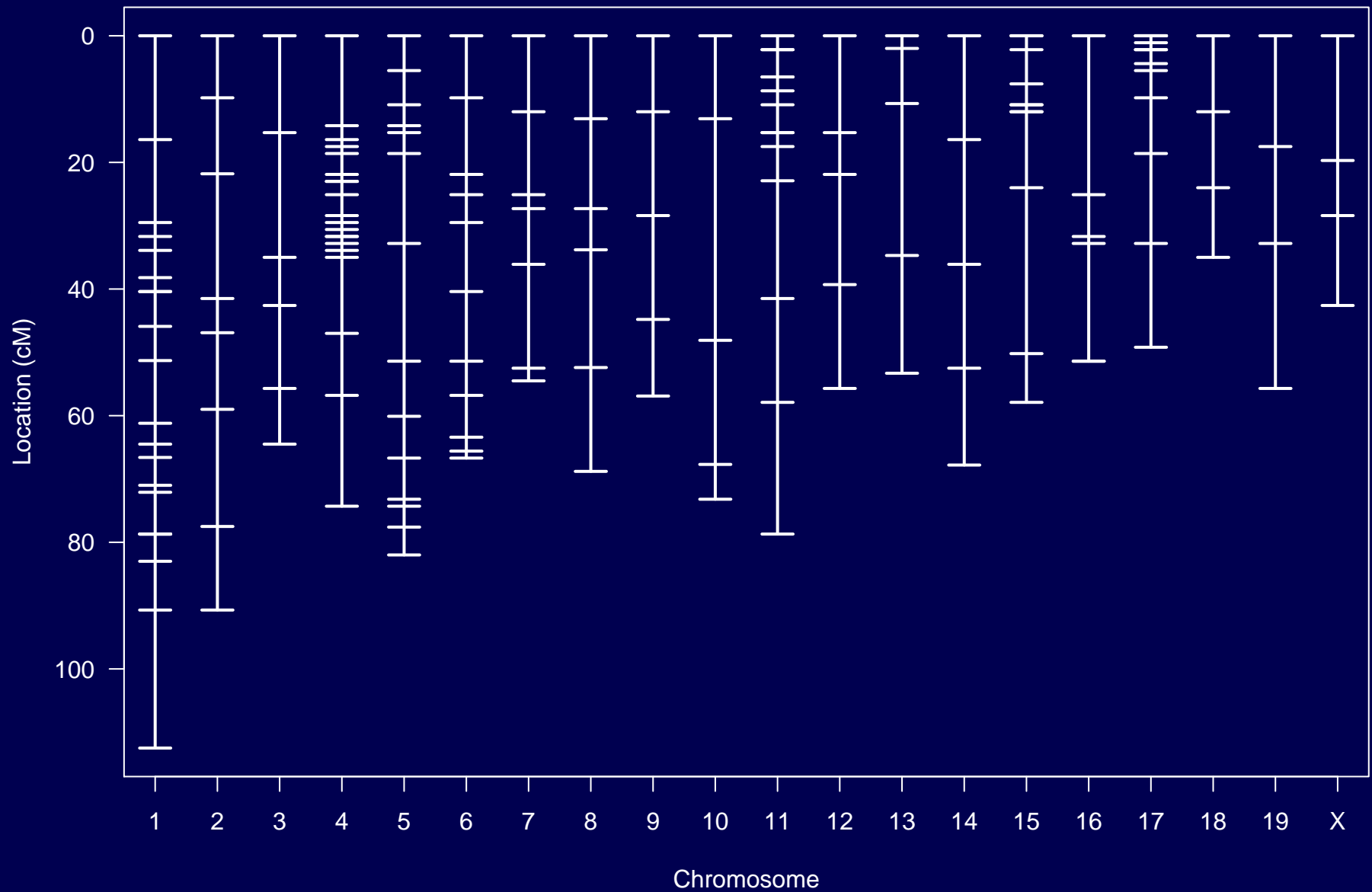
Sugiyama et al. Genomics 71:70-77, 2001

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

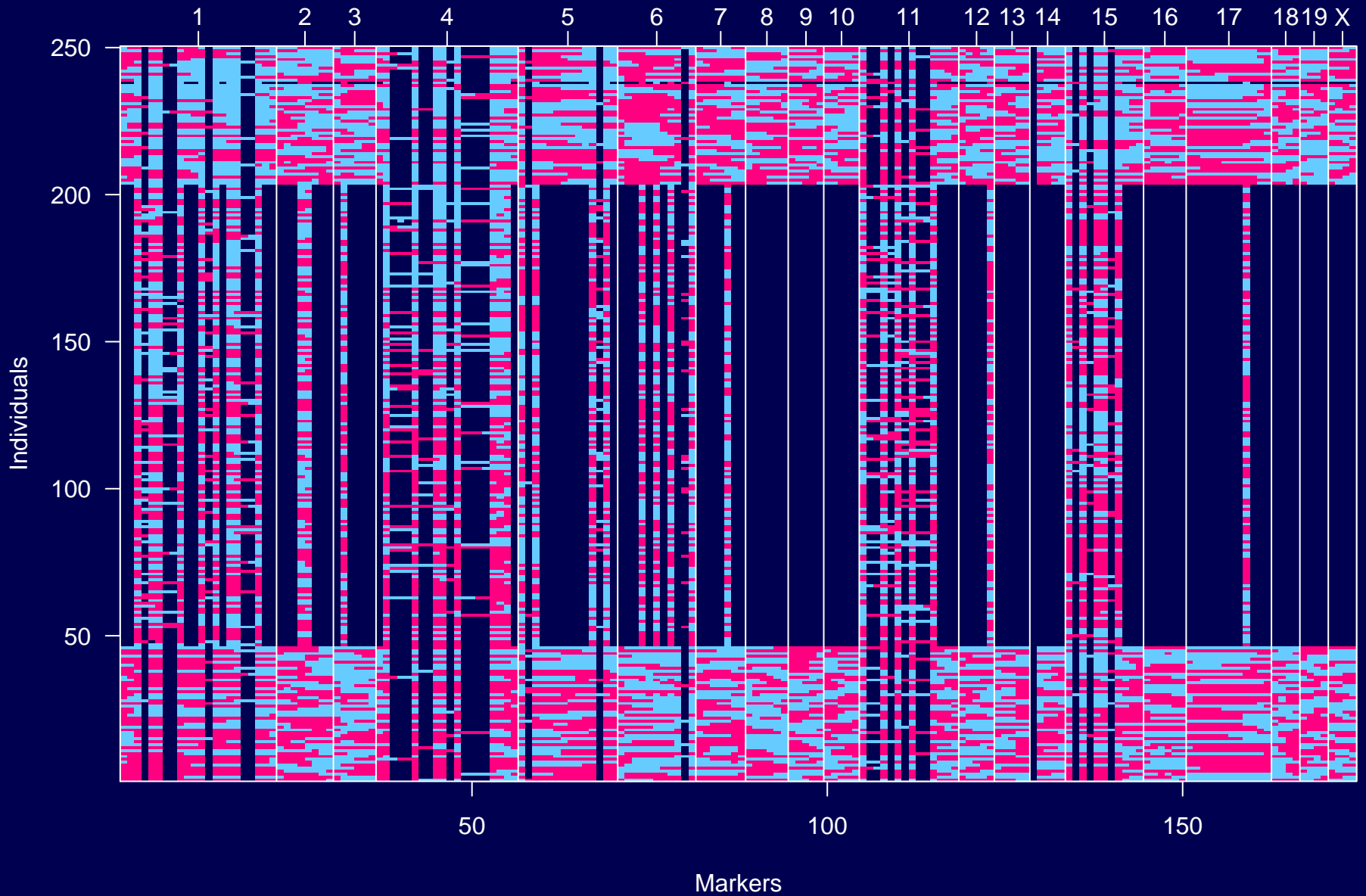
Blood pressure after two weeks drinking water with 1% NaCl



Genetic map



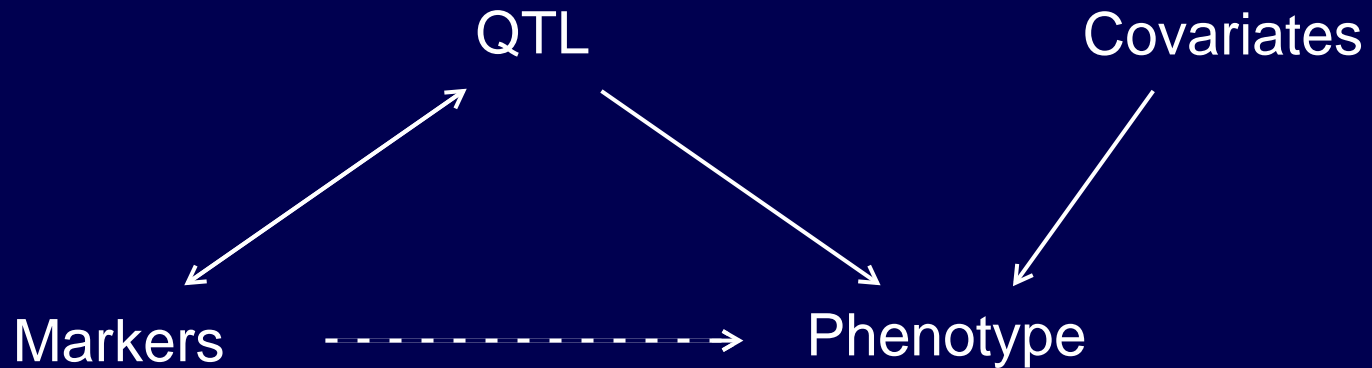
Genotype data



Goals

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

Statistical structure



The missing data problem:

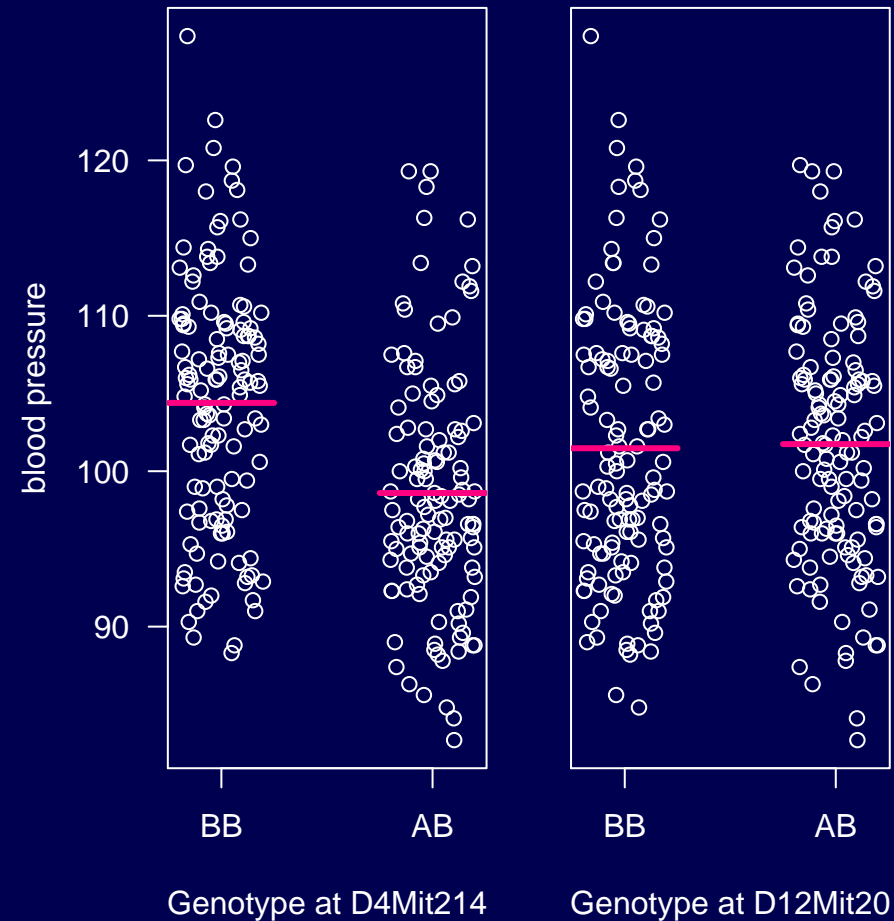
Markers \longleftrightarrow QTL

The model selection problem:

QTL, covariates \longrightarrow 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.)

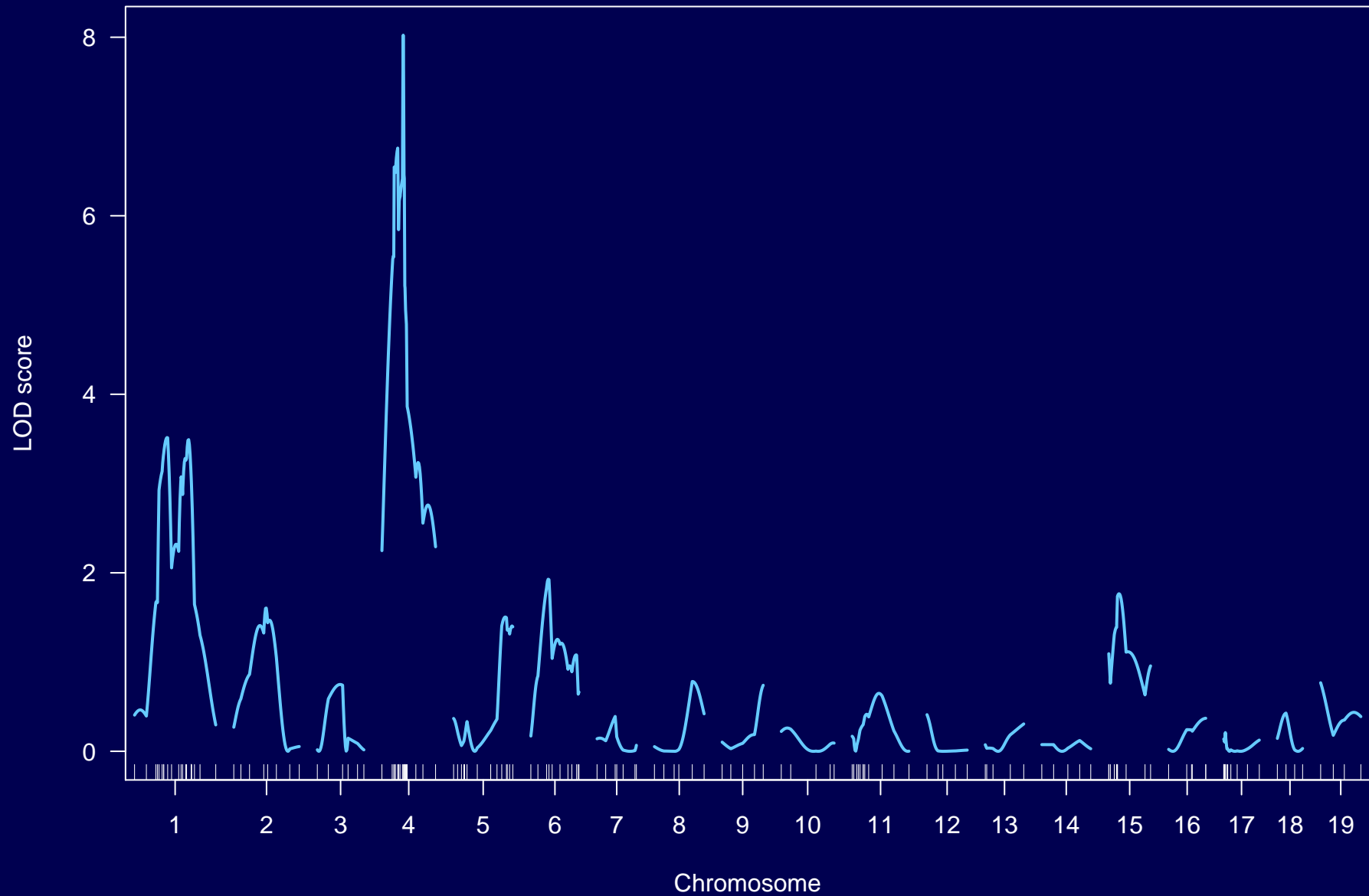
Assume $y \mid q \sim N(\mu_q, \sigma)$

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

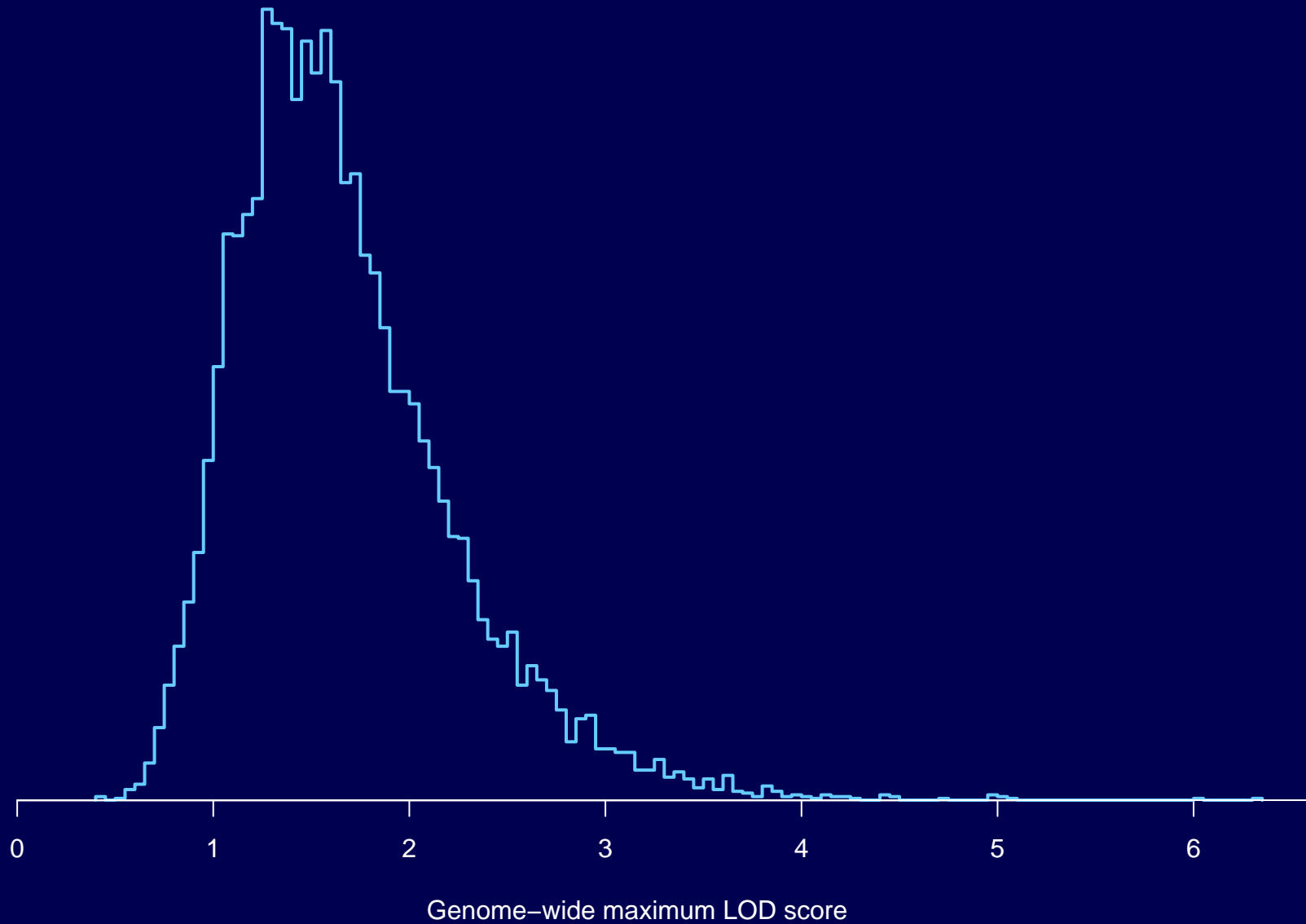
$$y \mid \text{marker data} \sim \sum_q p_q \phi(y \mid \mu_q, \sigma)$$

- $\text{LOD}(\lambda) = \log_{10} \left\{ \frac{\Pr(y \mid \text{QTL at } \lambda, \hat{\mu}_{q\lambda}, \hat{\sigma}_\lambda)}{\Pr(y \mid \text{no QTL}, \hat{\mu}, \hat{\sigma})} \right\}$

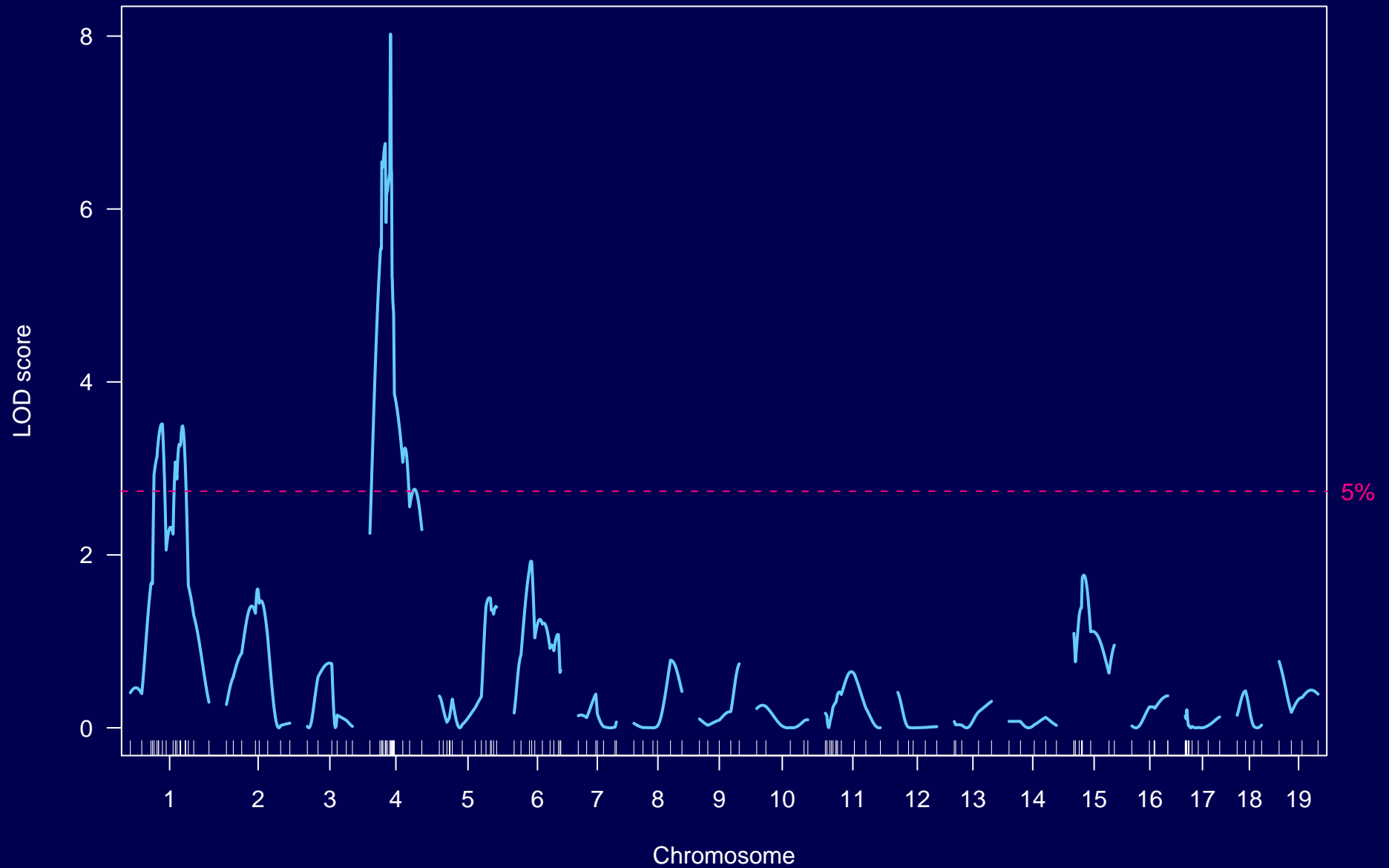
LOD curves



Permutation results



LOD curves

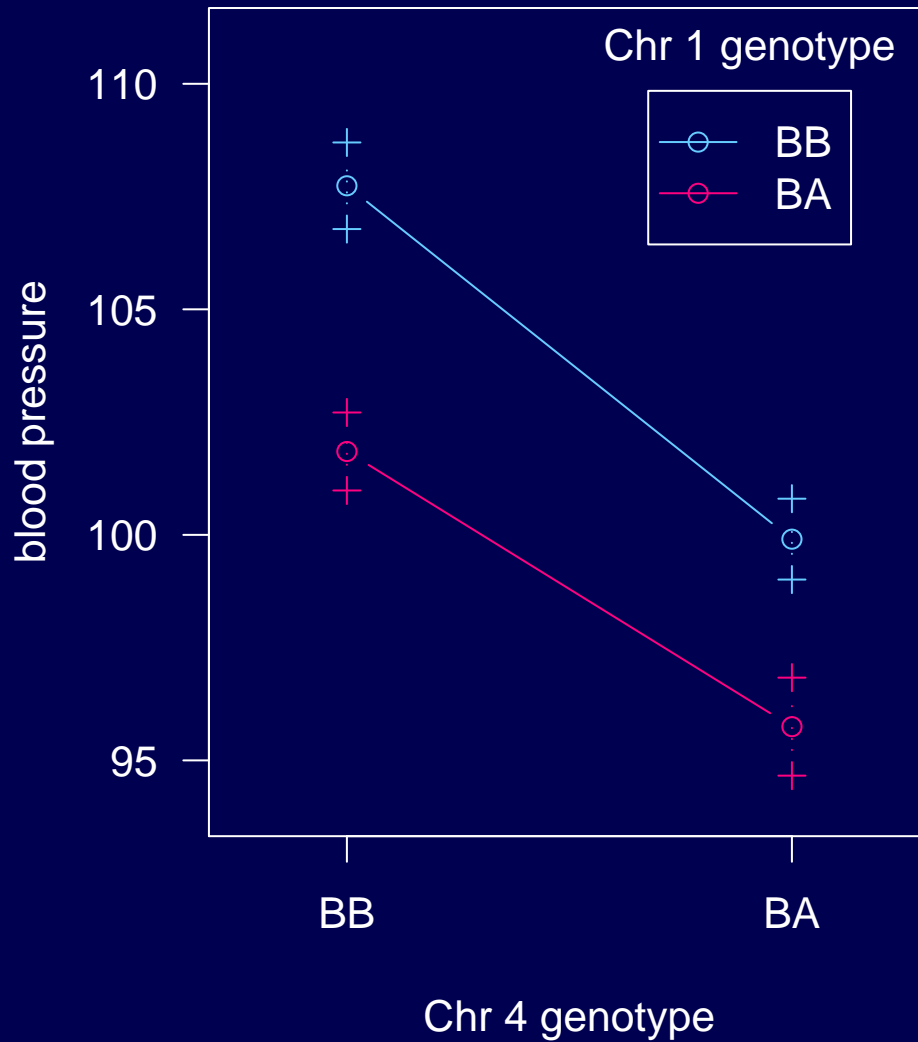


Modeling multiple QTL

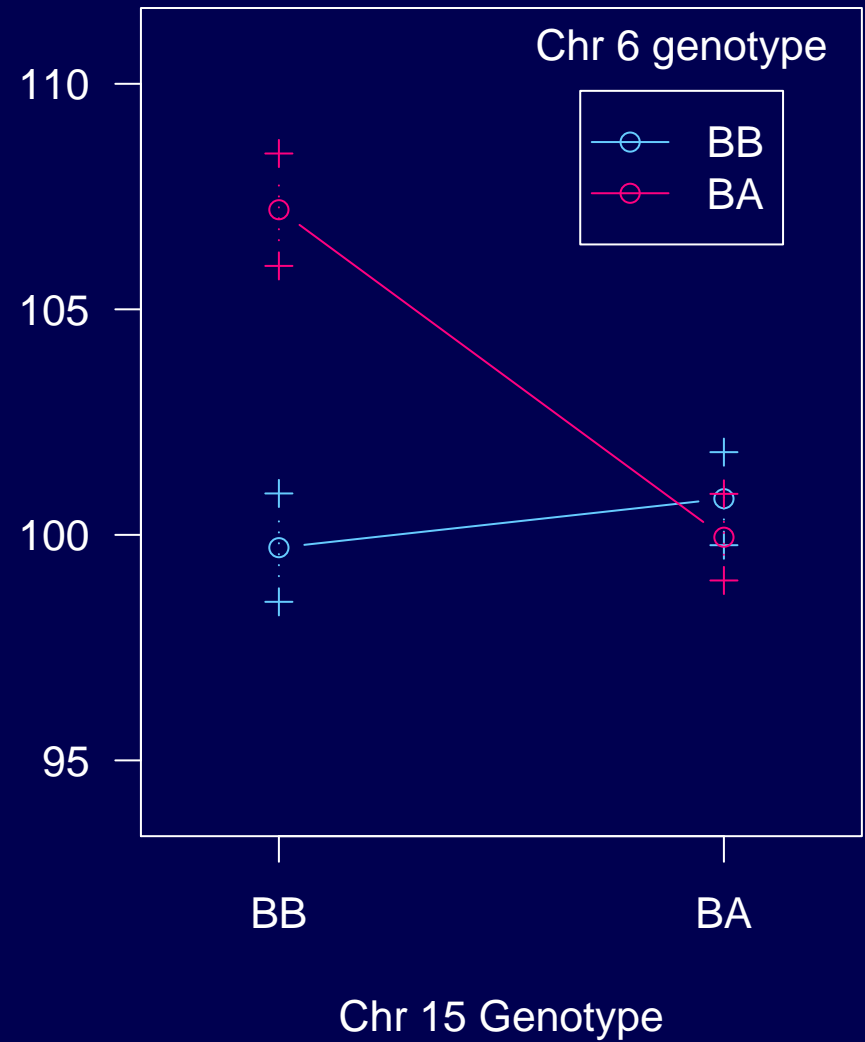
- Reduce residual variation \longrightarrow increased power
- Separate linked QTL
- Identify interactions among QTL (epistasis)

Estimated effects

1 x 4



6 x 15



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

Target

- 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

$$y = \mu + \sum \beta_j \mathbf{q}_j + \epsilon \quad \text{which } \beta_j \neq 0?$$

$$\text{pLOD}(\gamma) = \text{LOD}(\gamma) - \mathbf{T} |\gamma|$$

Additive QTL

Simple situation:

- Dense markers
- Complete genotype data
- No epistasis

$$y = \mu + \sum \beta_j \mathbf{q}_j + \epsilon \quad \text{which } \beta_j \neq 0?$$

$$\text{pLOD}(\gamma) = \text{LOD}(\gamma) - \mathbf{T} |\gamma|$$

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

$$\text{pLOD}(\{\lambda\}) = \text{LOD}(\{\lambda\}) - \mathbf{T}$$

Additive QTL

Simple situation:

- Dense markers
- Complete genotype data
- No epistasis

$$y = \mu + \sum \beta_j \mathbf{q}_j + \epsilon \quad \text{which } \beta_j \neq 0?$$

$$\text{pLOD}(\gamma) = \text{LOD}(\gamma) - \mathbf{T} |\gamma|$$

For the mouse genome:

$$\mathbf{T} = 2.69 \text{ (BC) or } 3.52 \text{ (F}_2\text{)}$$

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

Epistasis

$$y = \mu + \sum \beta_j \mathbf{q}_j + \sum \gamma_{jk} \mathbf{q}_j \mathbf{q}_k + \epsilon$$

$$\text{pLOD}(\gamma) = \text{LOD}(\gamma) - T_m |\gamma|_m - T_i |\gamma|_i$$

T_m = as chosen previously

T_i = ?

Idea 1

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

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

Idea 1

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

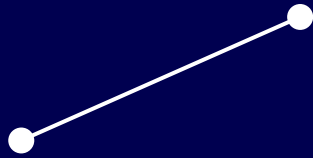
$$T_i = 95\text{th percentile of the distribution of} \\ \max \text{LOD}_f(\lambda_1, \lambda_2) - \max \text{LOD}_a(\lambda_1, \lambda_2)$$

For the mouse genome:

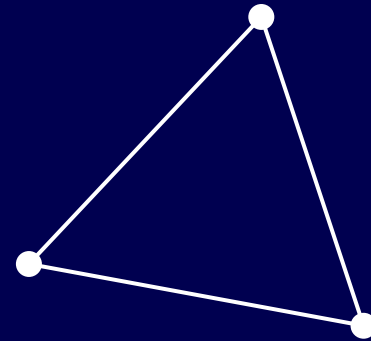
$$T_m = 2.69 \text{ (BC) or } 3.52 \text{ (F}_2\text{)}$$

$$T_i^H = 2.62 \text{ (BC) or } 4.28 \text{ (F}_2\text{)}$$

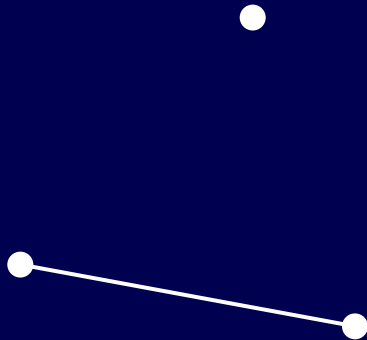
Models as graphs



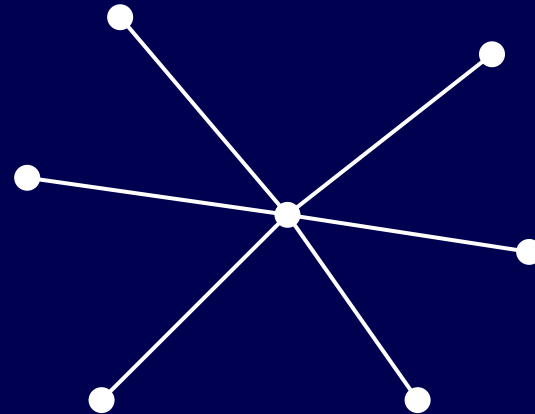
A



C



B



D

Idea 2

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

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

Idea 2

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

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

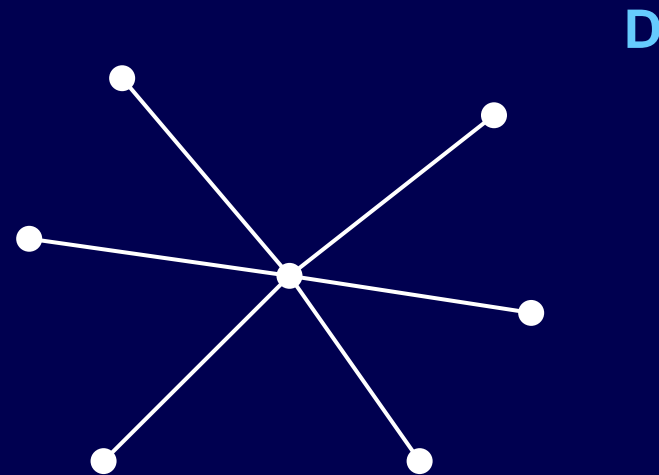
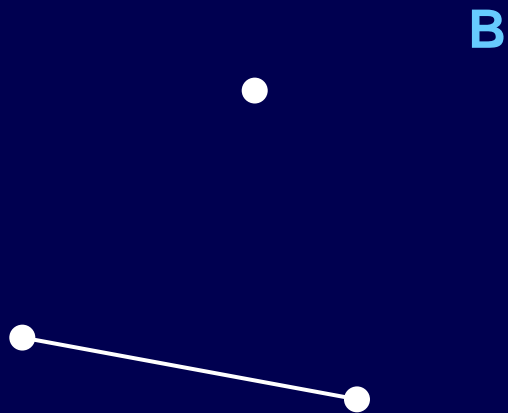
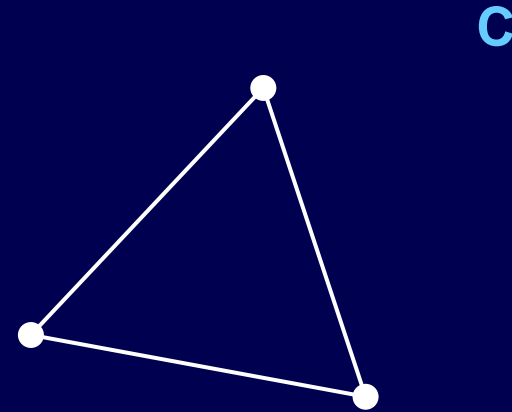
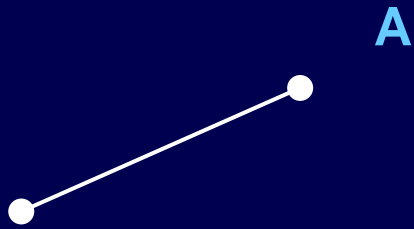
For the mouse genome:

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

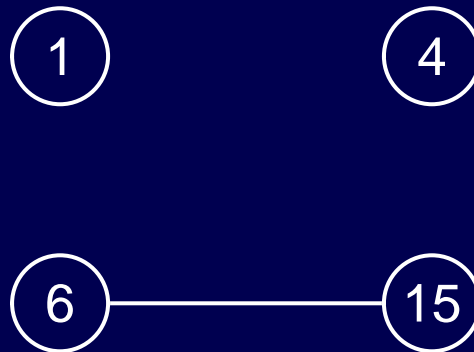
$$T_i^H = 2.62 \text{ (BC) or } 4.28 \text{ (F}_2\text{)}$$

$$T_i^L = 1.19 \text{ (BC) or } 2.69 \text{ (F}_2\text{)}$$

Models as graphs

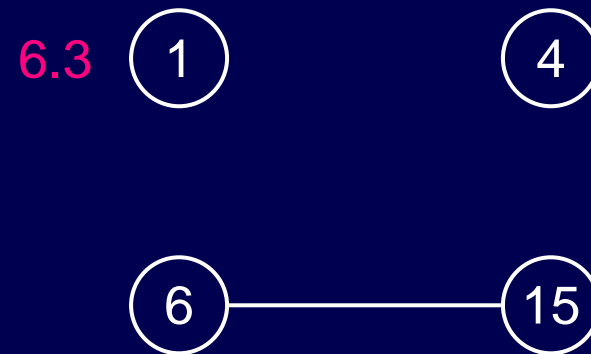


Results



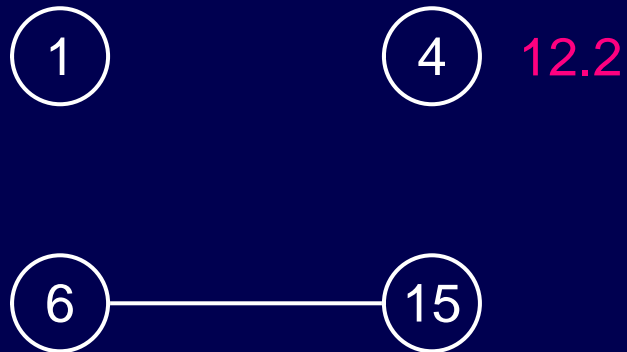
LOD = 23.1

Drop one term?



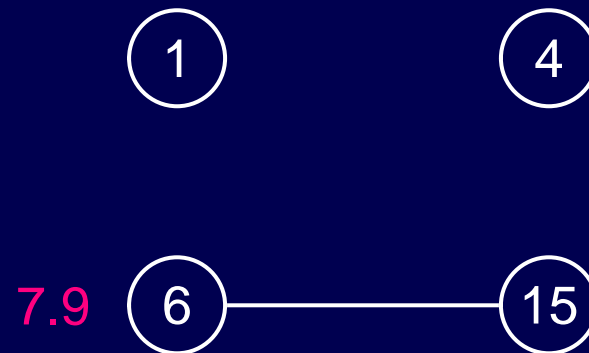
$$T_m = 2.69 \quad T_i^H = 2.62 \quad T_i^L = 1.19 \quad T_m + T_i^H = 5.31 \quad T_m + T_i^L = 3.88 \quad 2T_m = 5.38$$

Drop one term?



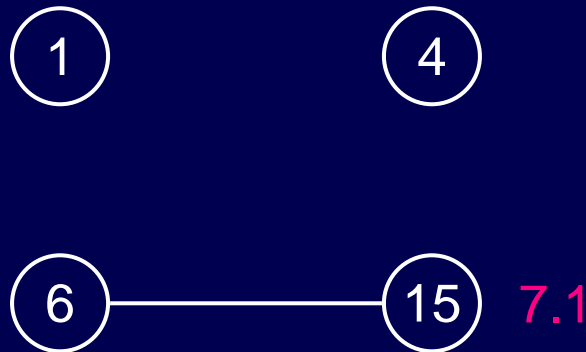
$$T_m = 2.69 \quad T_i^H = 2.62 \quad T_i^L = 1.19 \quad T_m + T_i^H = 5.31 \quad T_m + T_i^L = 3.88 \quad 2T_m = 5.38$$

Drop one term?



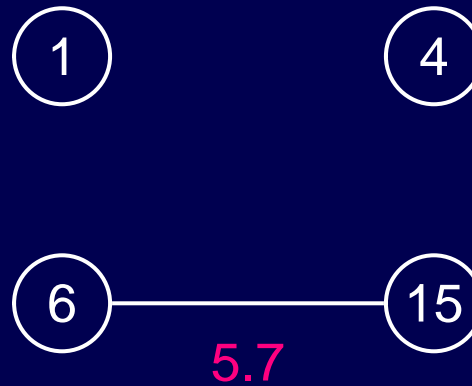
$$T_m = 2.69 \quad T_i^H = 2.62 \quad T_i^L = 1.19 \quad T_m + T_i^H = 5.31 \quad T_m + T_i^L = 3.88 \quad 2T_m = 5.38$$

Drop one term?



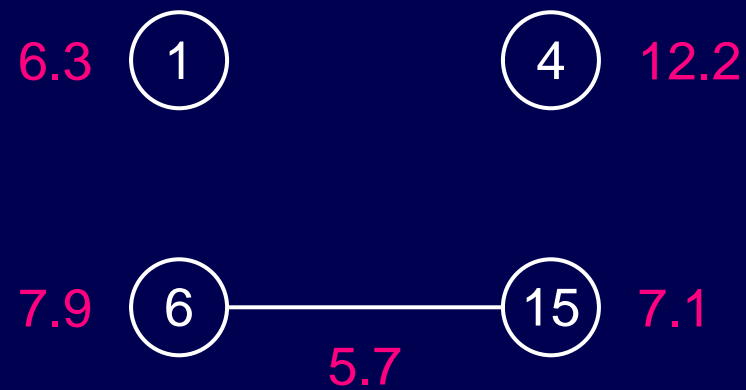
$$T_m = 2.69 \quad T_i^H = 2.62 \quad T_i^L = 1.19 \quad T_m + T_i^H = 5.31 \quad T_m + T_i^L = 3.88 \quad 2T_m = 5.38$$

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 at time



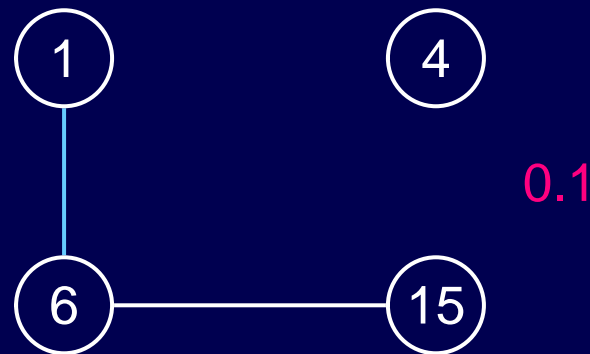
$$T_m = 2.69 \quad T_i^H = 2.62 \quad T_i^L = 1.19 \quad T_m + T_i^H = 5.31 \quad T_m + T_i^L = 3.88 \quad 2T_m = 5.38$$

Add an interaction?



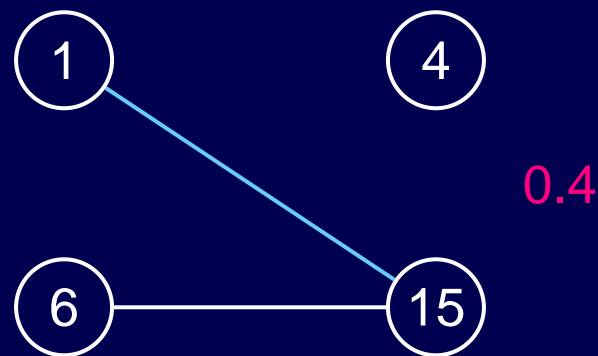
$$T_m = 2.69 \quad T_i^H = 2.62 \quad T_i^L = 1.19 \quad T_m + T_i^H = 5.31 \quad T_m + T_i^L = 3.88 \quad 2T_m = 5.38$$

Add an interaction?



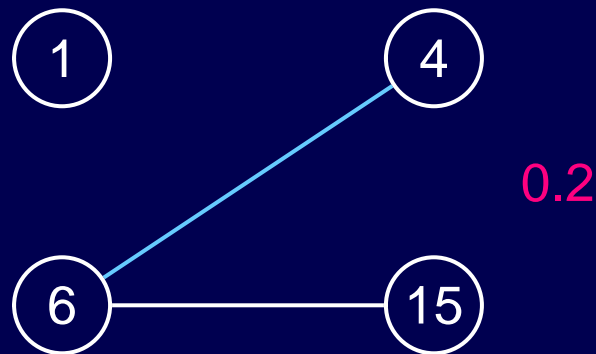
$$T_m = 2.69 \quad T_i^H = 2.62 \quad T_i^L = 1.19 \quad T_m + T_i^H = 5.31 \quad T_m + T_i^L = 3.88 \quad 2T_m = 5.38$$

Add an interaction?



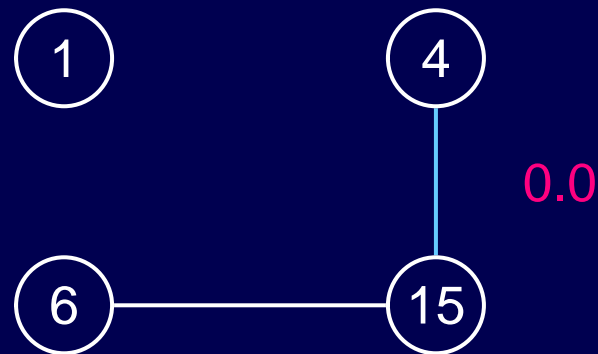
$$T_m = 2.69 \quad T_i^H = 2.62 \quad T_i^L = 1.19 \quad T_m + T_i^H = 5.31 \quad T_m + T_i^L = 3.88 \quad 2T_m = 5.38$$

Add an interaction?



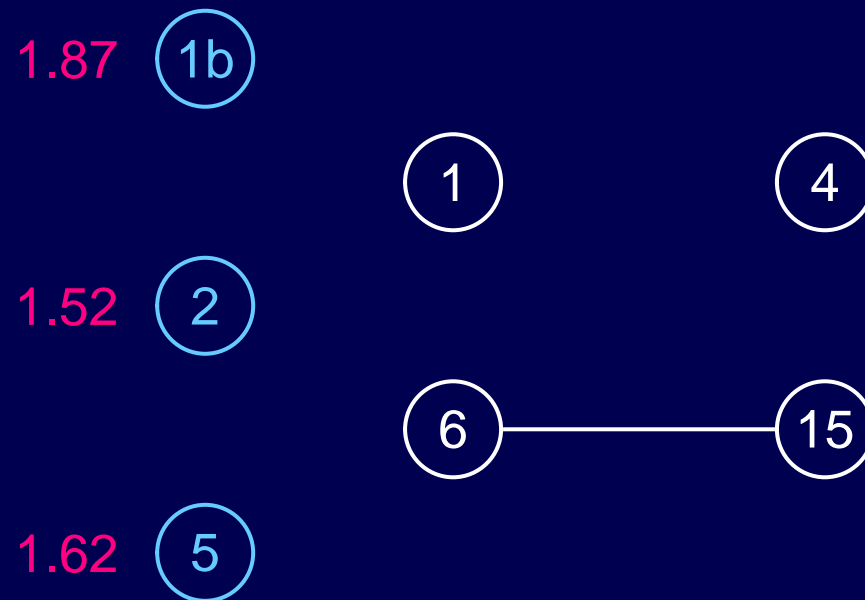
$$T_m = 2.69 \quad T_i^H = 2.62 \quad T_i^L = 1.19 \quad T_m + T_i^H = 5.31 \quad T_m + T_i^L = 3.88 \quad 2T_m = 5.38$$

Add an interaction?



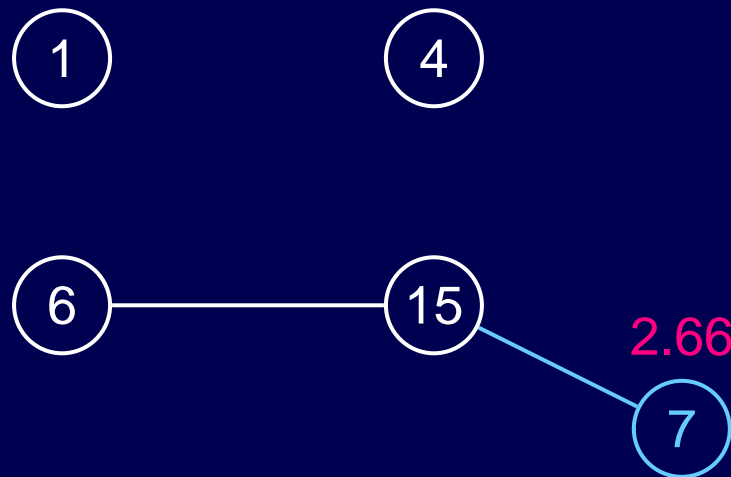
$$T_m = 2.69 \quad T_i^H = 2.62 \quad T_i^L = 1.19 \quad T_m + T_i^H = 5.31 \quad T_m + T_i^L = 3.88 \quad 2T_m = 5.38$$

Add another QTL?



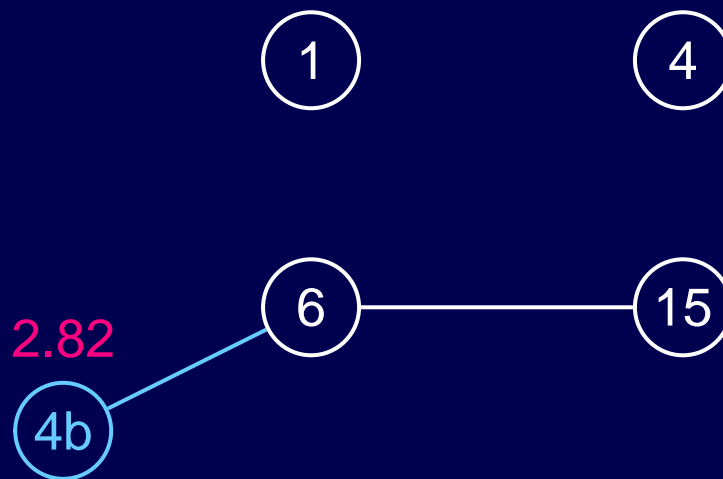
$$T_m = 2.69 \quad T_i^H = 2.62 \quad T_i^L = 1.19 \quad T_m + T_i^H = 5.31 \quad T_m + T_i^L = 3.88 \quad 2T_m = 5.38$$

Add another QTL?



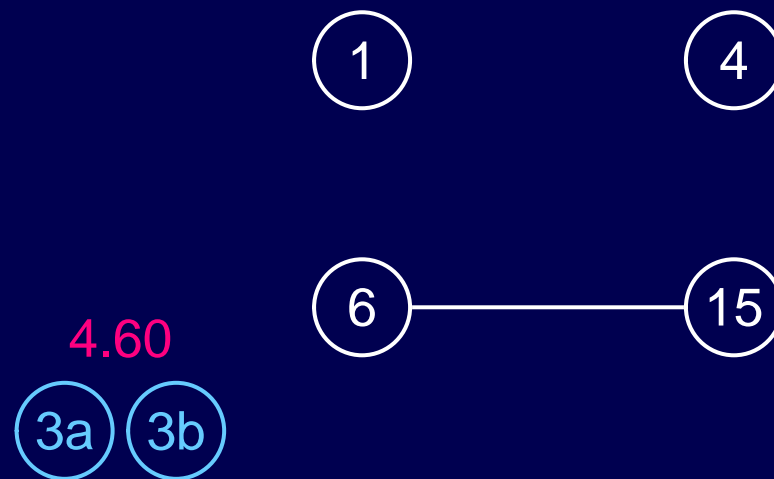
$$T_m = 2.69 \quad T_i^H = 2.62 \quad T_i^L = 1.19 \quad T_m + T_i^H = 5.31 \quad T_m + T_i^L = 3.88 \quad 2T_m = 5.38$$

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 \times covariate interactions
- That evil X chromosome
- Treat linked QTL differently?

Summary

- 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

Acknowledgments

Ani Manichaikul	Johns Hopkins University (now at University of Virginia)
Gary Churchill	Jackson Laboratory
Śaunak Sen	University of California, San Francisco
Terry Speed	University of California, Berkeley
Brian Yandell	University of Wisconsin – Madison
Fumihiko Sugiyama	now at University of Tsukuba, Japan
Bev Paigen	Jackson Laboratory