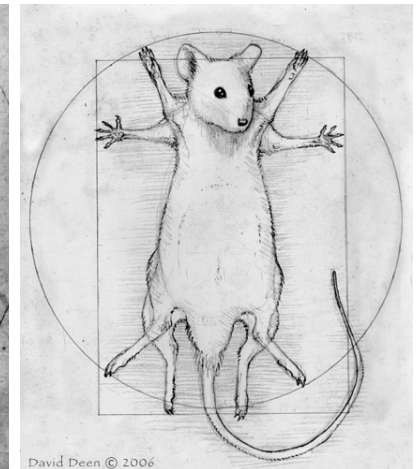
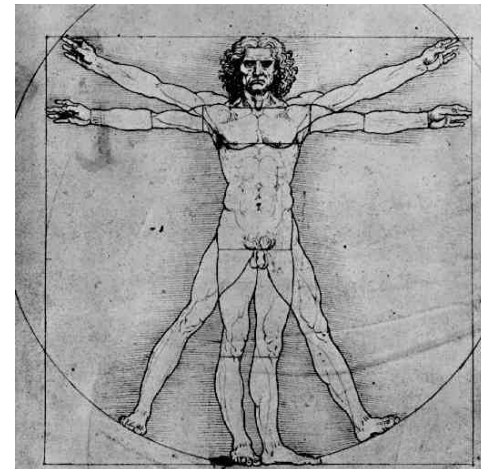


# Mapping multiple QTL in experimental crosses

Karl W Broman

Department of Biostatistics & Medical Informatics  
University of Wisconsin – Madison

[www.biostat.wisc.edu/~kbroman](http://www.biostat.wisc.edu/~kbroman)

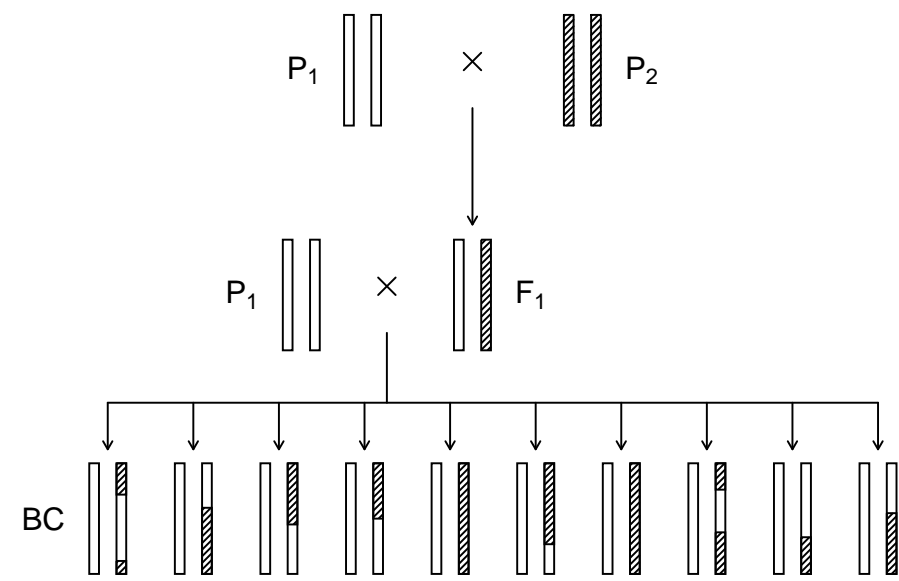


[www.daviddeen.com](http://www.daviddeen.com)

3

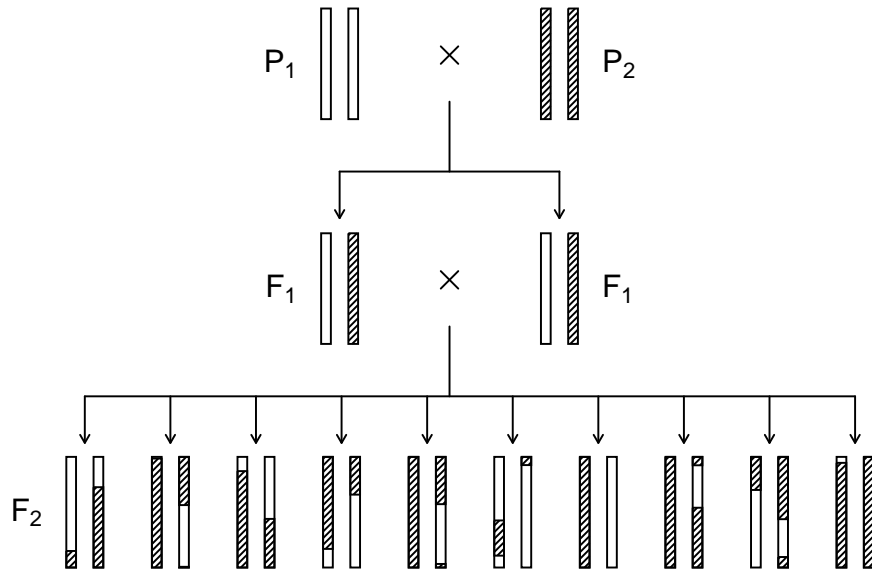


## Backcross



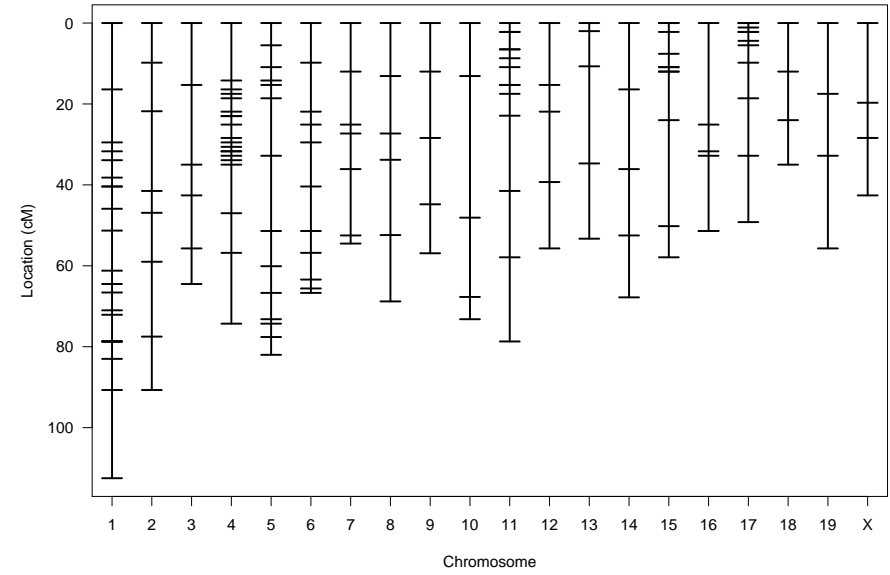
4

## Intercross



5

## Genetic map



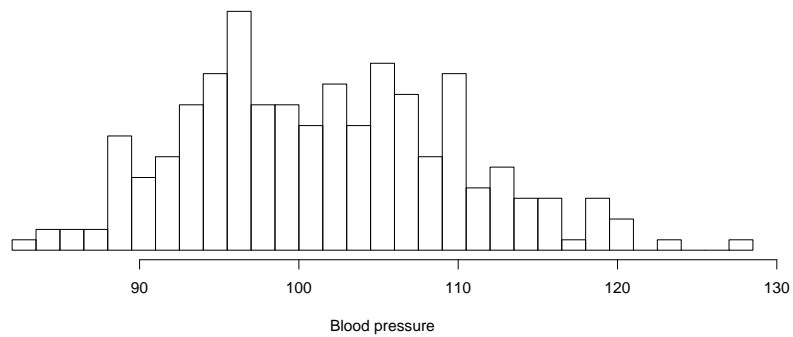
7

## Phenotype data

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

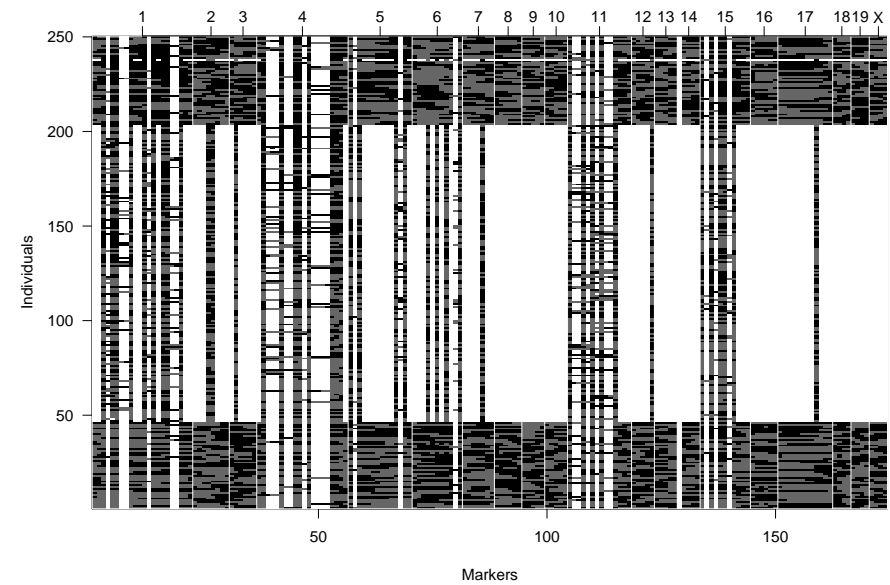
250 male mice from the backcross (A × B) × B

Blood pressure after two weeks drinking water with 1% NaCl



6

## Genotype data



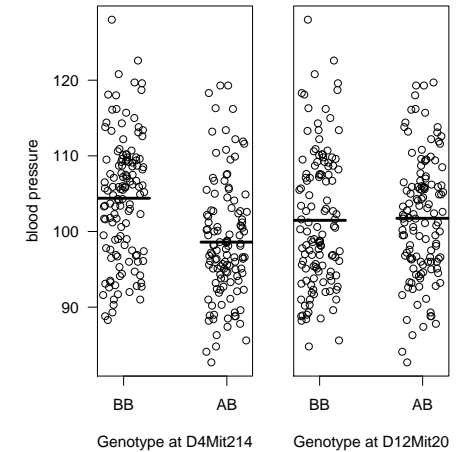
8

# Goals

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

# ANOVA at marker loci

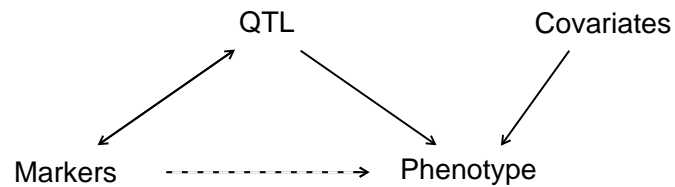
- Split mice into groups according to genotype at a marker.
- Do a t-test / ANOVA.
- Repeat for each marker.



9

11

# Statistical structure



# 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  $y | q \sim N(\mu_q, \sigma)$

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

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

The missing data problem:

Markers  $\longleftrightarrow$  QTL

The model selection problem:

QTL, covariates  $\longrightarrow$  phenotype

10

12

## LOD scores

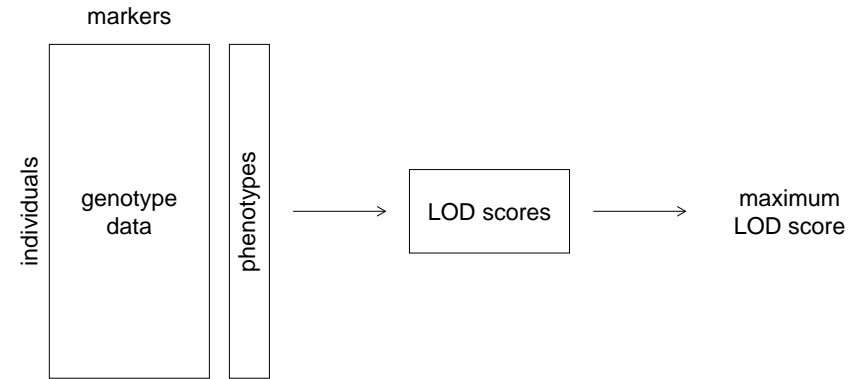
$\text{LOD}(\lambda) = \log_{10}$  likelihood ratio comparing the hypothesis of a QTL at position  $\lambda$  versus that of no QTL

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

$\hat{\mu}_{q\lambda}, \hat{\sigma}_\lambda$  are the MLEs, assuming a single QTL at position  $\lambda$ .

No QTL model: The phenotypes are iid  $N(\mu, \sigma^2)$ .

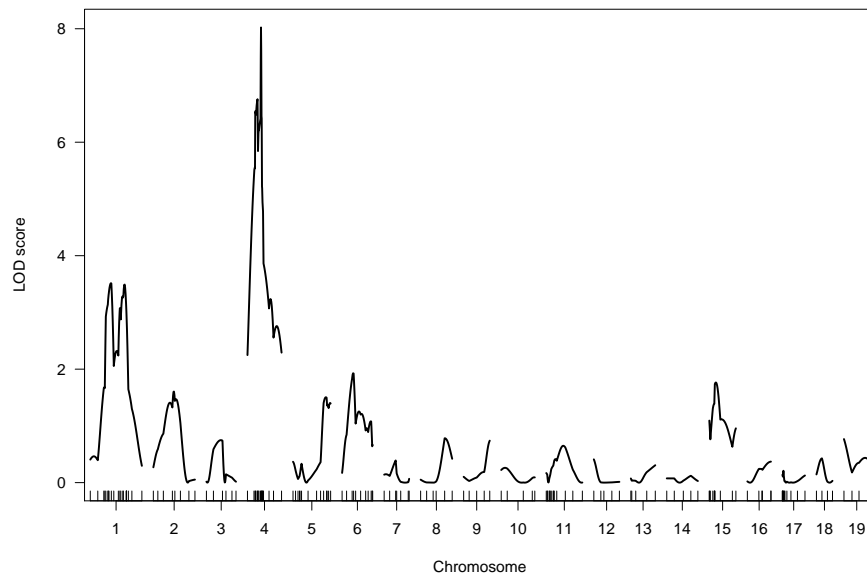
## Permutation test



13

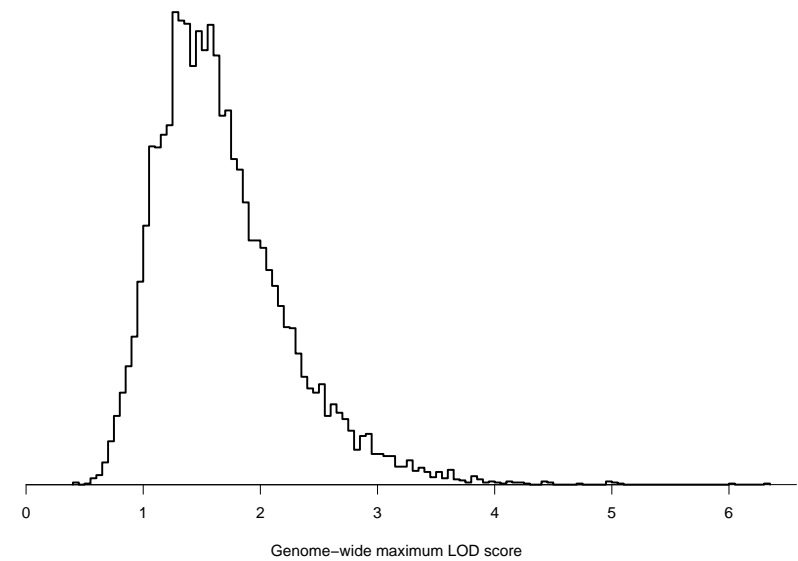
15

## LOD curves



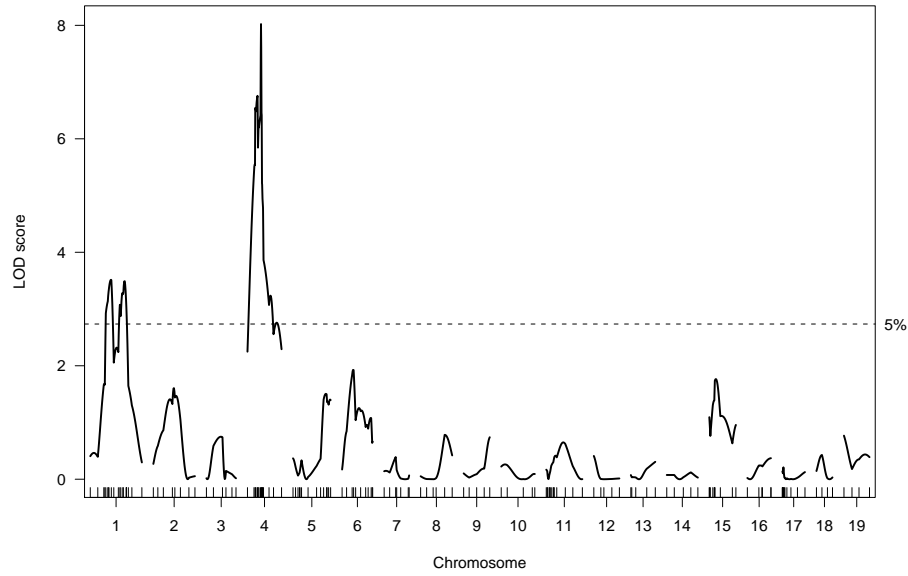
14

## Permutation results



16

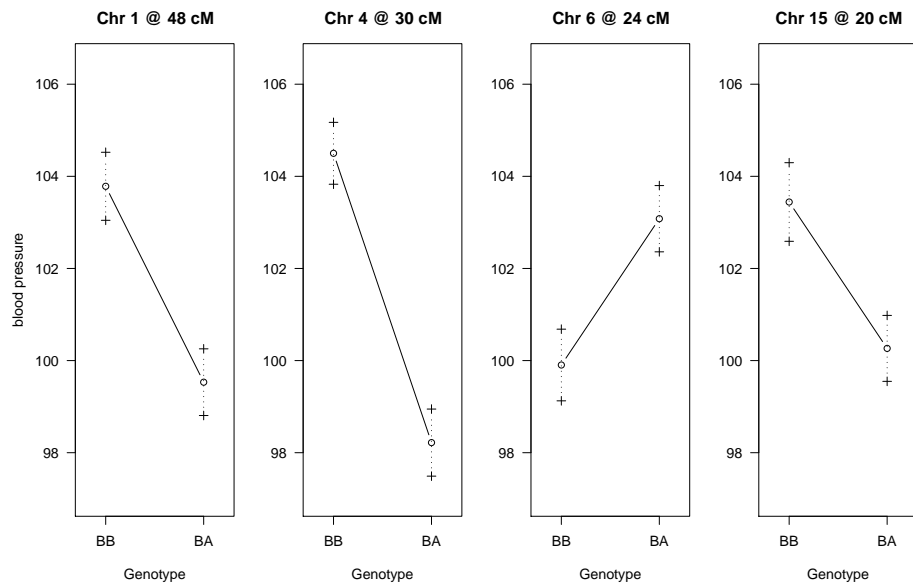
# LOD curves



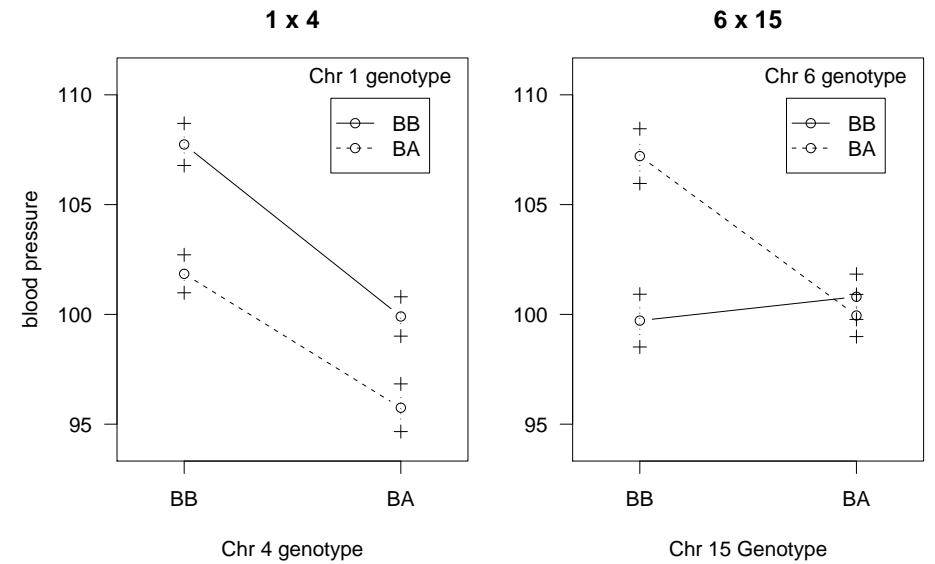
# Modeling multiple QTL

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

# Estimated effects



# Estimated effects



# 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).

22

# 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.

24

# 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

23

# 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

25

# Exploratory methods

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

$$\text{LOD}(q_2 | q_1) = \log_{10} \left\{ \frac{\text{Pr}(\text{data} | q_1, q_2)}{\text{Pr}(\text{data} | q_1)} \right\}$$

- 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

26

# Automation

- Assistance to the masses
- Understanding performance
- Many phenotypes

27

# Additive QTL

Simple situation:

- Dense markers
- Complete genotype data
- No epistasis

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

$$\text{LOD}_\delta(\gamma) = \text{LOD}(\gamma) - T |\gamma|$$

$$0 \text{ vs } 1 \text{ QTL: } \text{LOD}_\delta(\emptyset) = 0$$

$$\text{LOD}_\delta(\{\lambda\}) = \text{LOD}(\{\lambda\}) - T$$

32

# 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

34

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

$$\text{LOD}_{\delta\epsilon}(\gamma) = \text{LOD}(\gamma) - T_m |\gamma|_m + T_i |\gamma|_i$$

$T_m$  = as chosen previously

$T_i$  = ?

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:

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

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

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

35

39

## Idea 1

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

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

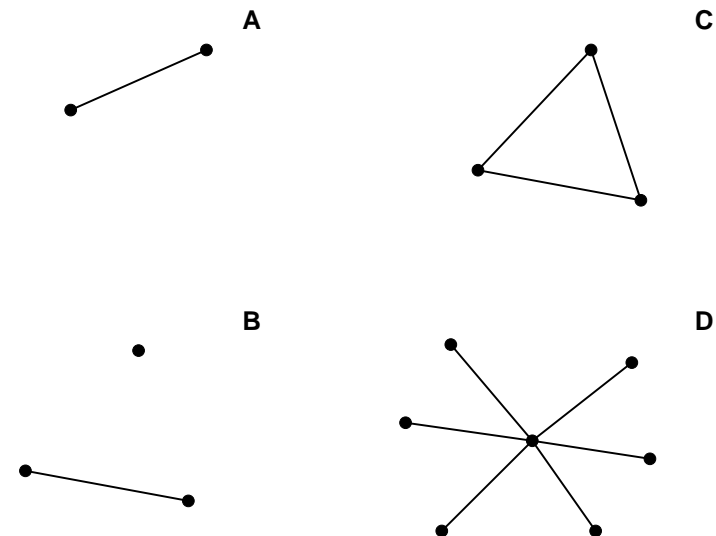
For the mouse genome:

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

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

37

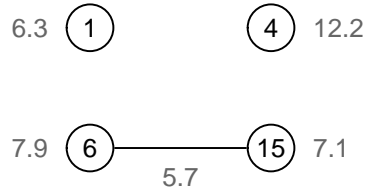
## Models as graphs



40



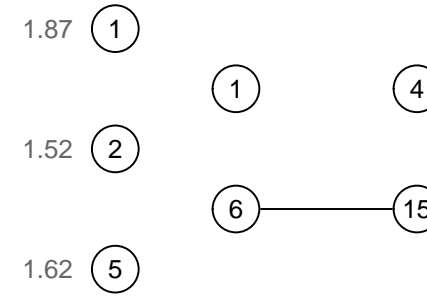
# Results



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

43

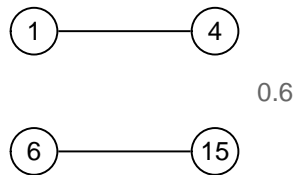
# Add another QTL?



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

49

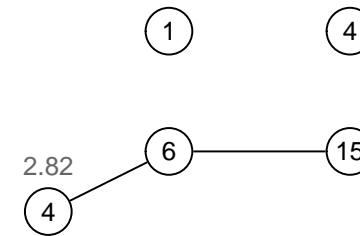
# Add an interaction?



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

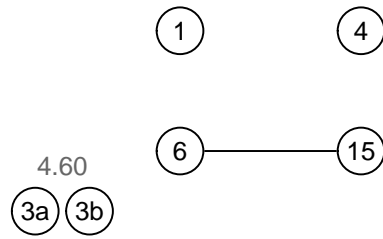
44

# Add another QTL?



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

51



- QTL mapping is a model selection problem
- The criterion for comparing models is most important
- We're focusing on a penalized likelihood method and are close to a practiceable solution

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

52

54

## To do

- Improve search procedures
- Study performance (especially relative to other approaches)
- Measuring model uncertainty
- Measuring uncertainty in QTL location

## Acknowledgments

Ani Manichaikul	Johns Hopkins University
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