Identifying QTLs in experimental crosses

Karl W. Broman

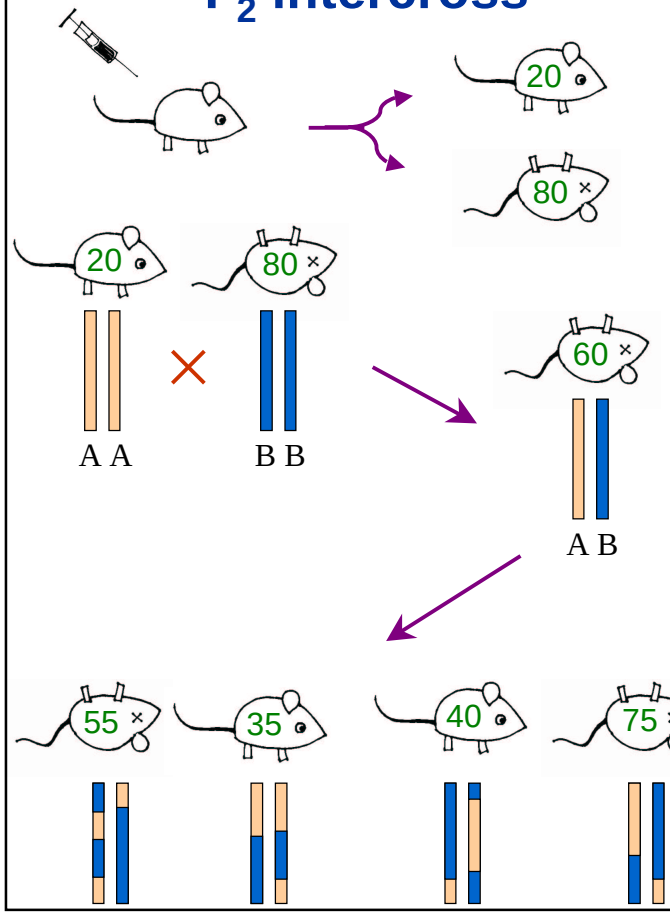
Department of Biostatistics
Johns Hopkins School of Public Health

kbroman@jhsph.edu

http://kbroman.homepage.com

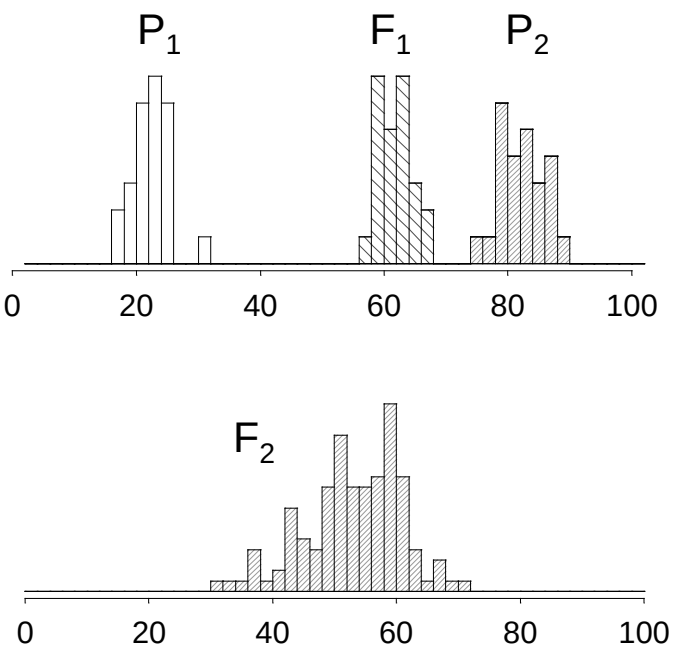
1

F_2 intercross



2

Distribution of the QT



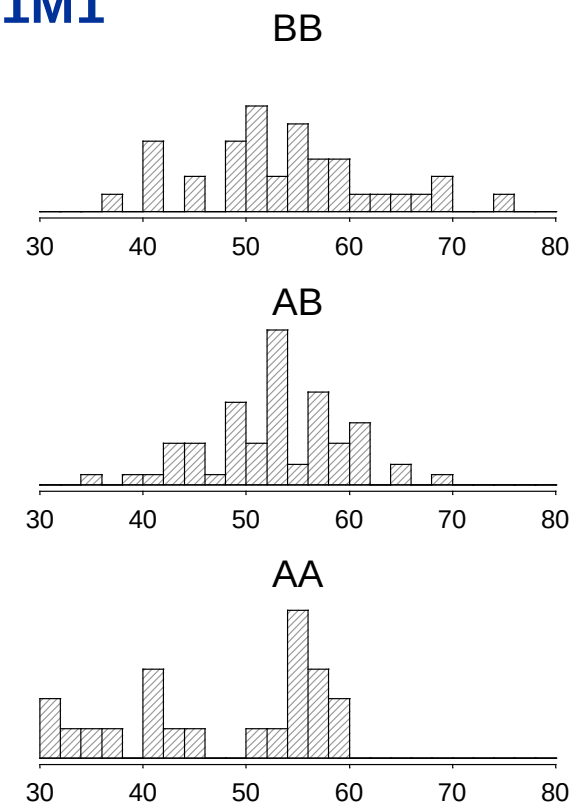
3

Data

- n (100–1000) F_2 progeny
- y_i = phenotype for individual i
- g_{ij} = genotype for indiv i at marker j
  (AA, AB or BB)
- Genetic map of the markers
- Phenotypes of parentals and F_1

4

D1M1



5

Single QTL analysis

Marker D1M1

genotype	Ave (SE)	SD	N
BB	54.0 (1.4)	8.7	22
AB	52.9 (0.8)	6.5	63
AA	47.5 (2.1)	9.8	37
Overall	52.3 (0.7)	8.1	122

Additive effect $a = [\mu(\text{BB}) - \mu(\text{AA})]/2 \approx 3.3\ (1.3)$

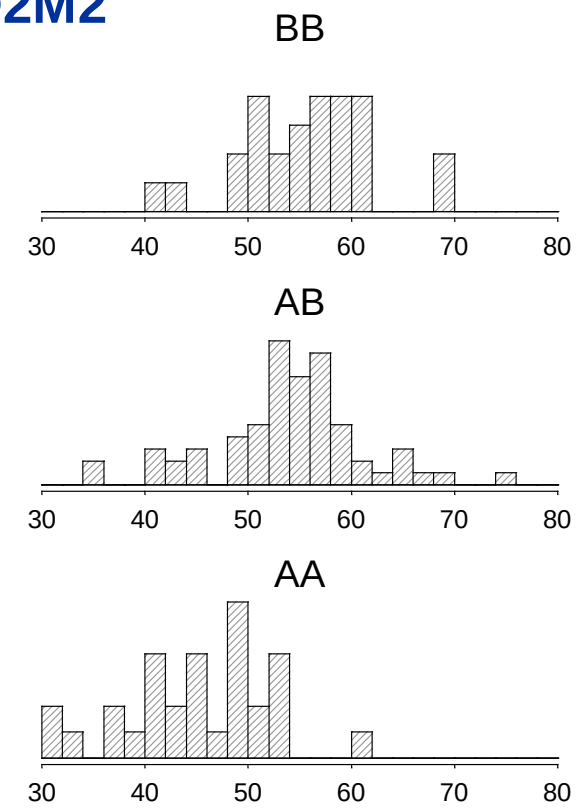
Dominance deviation $d = \mu(\text{AB}) - \frac{\mu(\text{BB})+\mu(\text{AA})}{2} \approx 2.2\ (1.5)$

Prop'n var explained $h^2 = \frac{\text{var in } F_2 - \text{residual var}}{\text{var in } F_2}$

$\approx \frac{\frac{1}{2}a^2 + \frac{1}{4}d^2}{\text{var in } F_2} \approx 8\%$

6

D2M2



7

Single QTL analysis

Marker D2M2

genotype	Ave (SE)	SD	N
BB	55.8 (1.2)	6.4	27
AB	54.0 (0.9)	7.6	65
AA	45.3 (1.3)	7.2	30
Overall	52.3 (0.7)	8.1	122

$a \approx 5.2\ (0.9)$

$d \approx 3.5\ (1.3)$

$h^2 \approx 25\%$

8

Is it real?

$$LOD = \log_{10}\left\{\frac{\Pr(\text{data} \mid 1\ \text{QTL})}{\Pr(\text{data} \mid \text{no QTL})}\right\}$$

$$LOD(D1M1) \approx 2.22$$

$$LOD(D2M2) \approx 7.54$$

Hypothesis testing

Null hypothesis, H_0: no QTL

P-value = Pr(LOD > observed | no QTL)

Small P (large LOD) $\Rightarrow$ Reject H_0 (Good)
Large P (small LOD) $\Rightarrow$ Fail to reject H_0 (Bad)

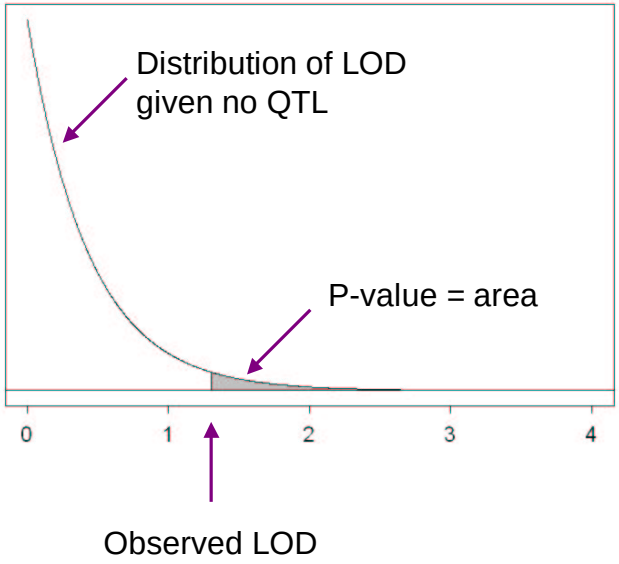
Generally want P < 0.05 or < 0.01

$$P \approx 0.049 \equiv P \approx 0.051$$

$$LOD \approx 3.01 \equiv LOD \approx 2.99$$

9

A picture



10

Multiple testing

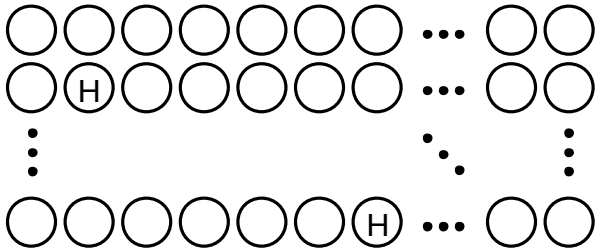
→ We're doing ~ 200 tests (one at each marker; correlated due to linkage)

→ Imagine the tests were uncorrelated, and that H_0 is true (there is no QTL)

Toss 200 **biased** coins

Heads ≡ Reject H_0 (falsely conclude that there is a QTL)

Pr(Heads) = 5%

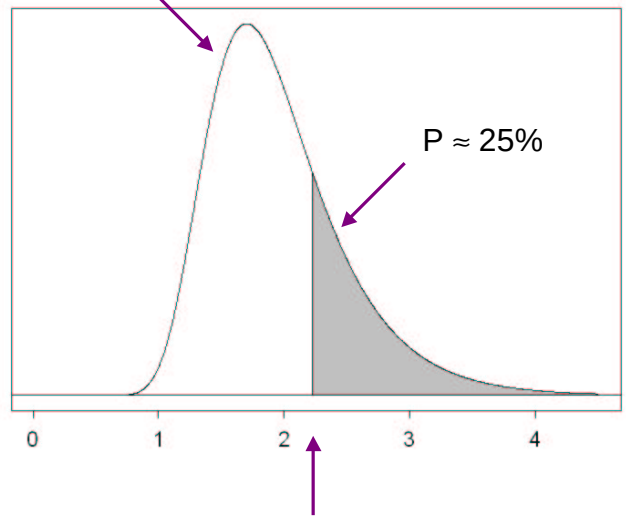


Ave no. heads in 200 tosses = 10

Pr(at least one head in 200 tosses) ≈ 100%

A new picture

Distribution of max LOD given no QTL



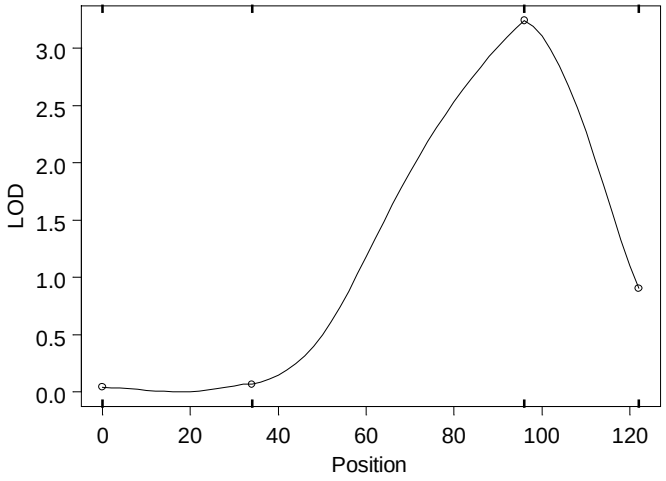
Interval mapping

(Lander and Botstein 1989)

Interpolation between markers

At each point, imagine a putative QTL and maximize Pr(data | QTL, parameters)

Great for dealing with missing genotype data; important for widely spaced markers



13

Power

Power = Pr(Identify a QTL | there is a QTL)

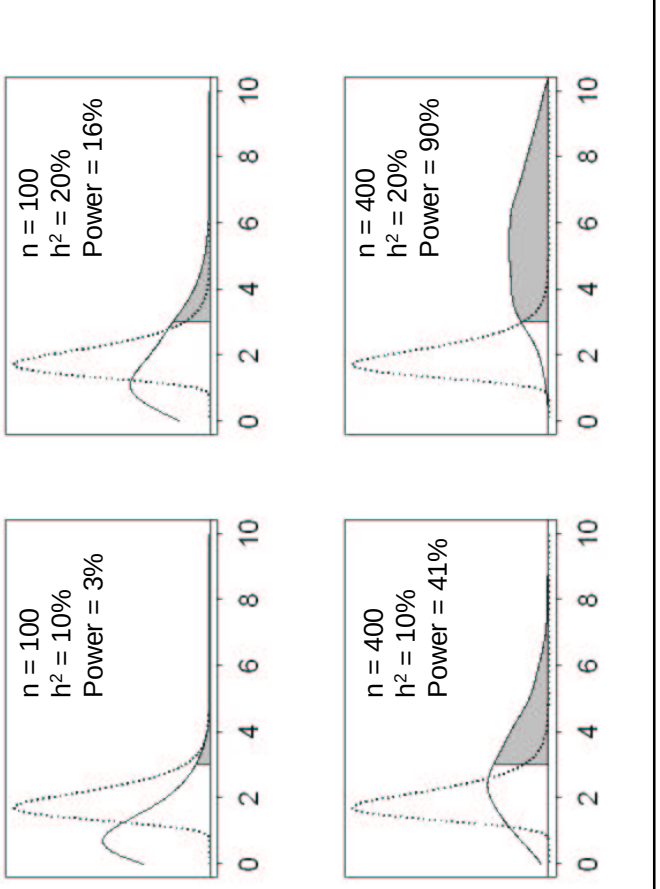
Power depends on

- Sample size
- Size of QTL effect (relative to resid. var.)
- Marker density
- Level of statistical significance

Consider

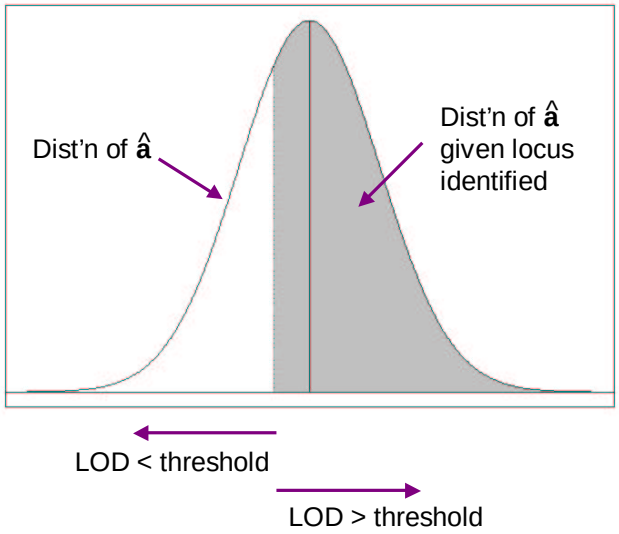
- Pr(detect a particular locus)
- Pr(detect at least one locus)

14



15

Selection bias



If the power to detect a particular locus is not super high, its estimated effect (when it is identified) will be biased

Power ≈ 90% ⇒ Bias ≈ 2%
Power ≈ 45% ⇒ Bias ≈ 20%
Power ≈ 5% ⇒ Bias ≈ 100%

16

Multiple QTLs

It is often important to consider multiple QTLs simultaneously

- Increase power by reducing residual variation
- Separate linked loci
- Estimate epistatic effects

Analysis of single QTL:

analysis of variance (ANOVA) or simple linear regression

Analysis of multiple QTL:

multiple linear regression, possibly with interaction terms; possibly using tree-based models

A key issue: Things are more complicated than "Is there a QTL here or not?"

17

An example

Full model

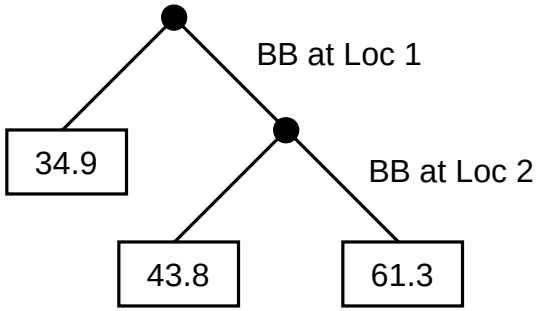
Locus 1 \ Locus 2	AA	AB	BB	
AA	34.7	36.4	33.7	35.2
AB	34.6	35.0	34.4	34.8
BB	43.3	43.9	61.3	50.4
	36.3	37.3	43.7	38.8

Additive model

Locus 1 \ Locus 2	AA	AB	BB
AA	33.7	34.2	38.6
AB	33.4	33.9	38.4
BB	48.3	48.8	53.3

18

A tree-based model



Locus 2

Locus 1	AA	AB	BB
AA	34.9		
AB			
BB	43.8		61.3

Summary

- LOD scores
- Hypothesis testing
  - Null hypothesis
  - P-values
  - Significance levels
- Adjustment for multiple tests
- Power
  - To identify a particular locus
  - To identify at least one locus
- Selection bias
- Multiple QTLs
  - Increase power
  - Separate linked loci
  - Estimate epistasis
  - Things get complicated