

Identifying QTLs in experimental crosses

Karl W Broman



Quantitative traits

e.g. yield of a tomato crop
number of bristles on a fly
blood pressure

- Many genes (QTLs)
- Environmental variation

Why?

- Biochemical basis of trait
- Selection experiments
- Evolution

Goal: find some of the genes

Humans vs. model organisms

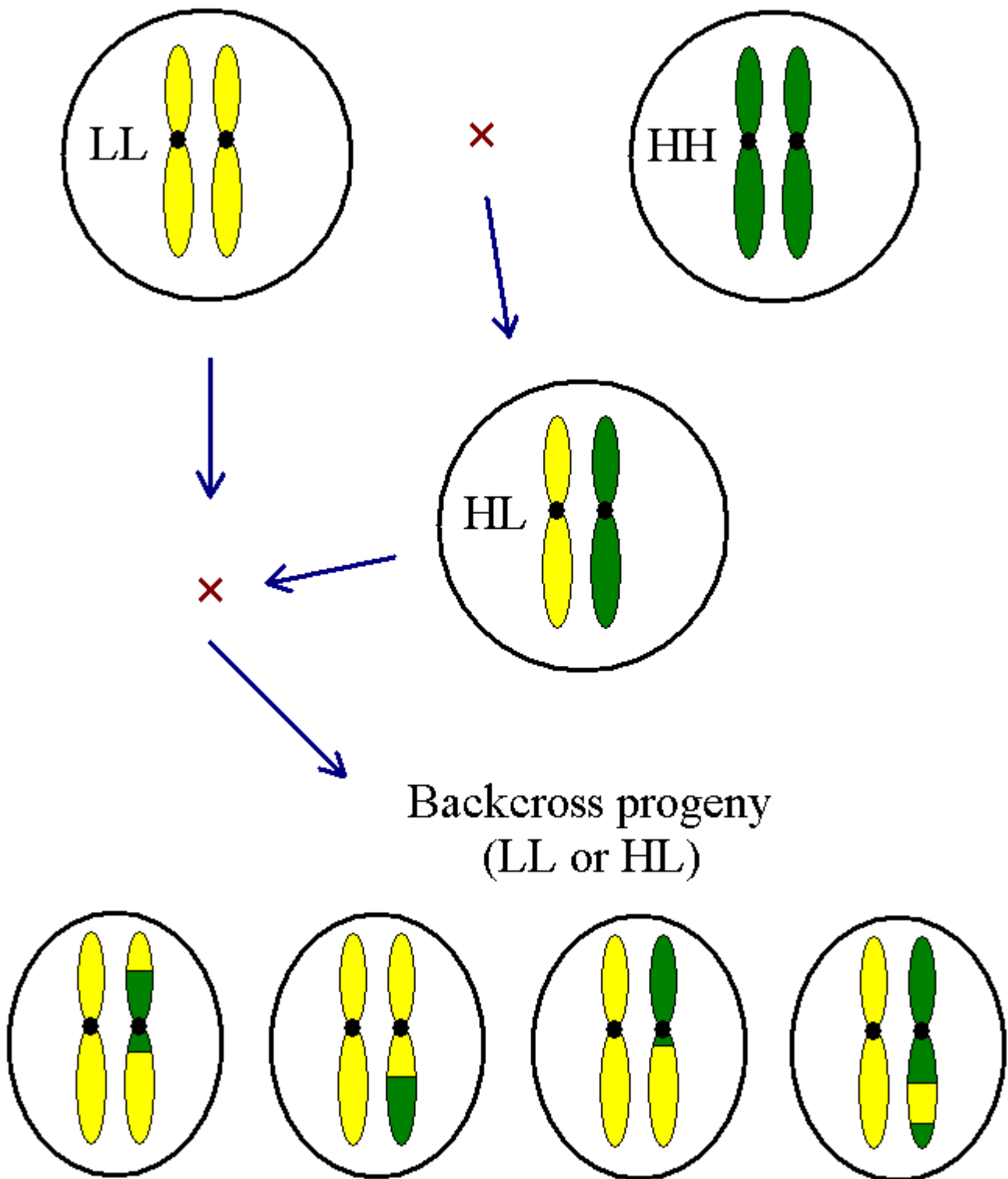
“Model”:

- Genes
- Genetic architecture
- Analysis methods

Differences:

- Complexity of pedigrees
- Complexity of genet. architec.
- Environmental variation

Backcross experiment



Data

Phenotypes (trait values)

y_i = phenotype for individual i

Marker genotypes

$x_{ij} = 1/0$ if i is HL/LL at marker j

Genetic map

Locations of markers

Models

Recombination: No interference

Phenotype/genotype connection

$$y = \mu + \sum \beta_j z_j + \varepsilon$$

$$\varepsilon \sim \text{Normal}(0, \sigma^2)$$

Problem

100 to 1000 backcross progeny

100 to 400 markers

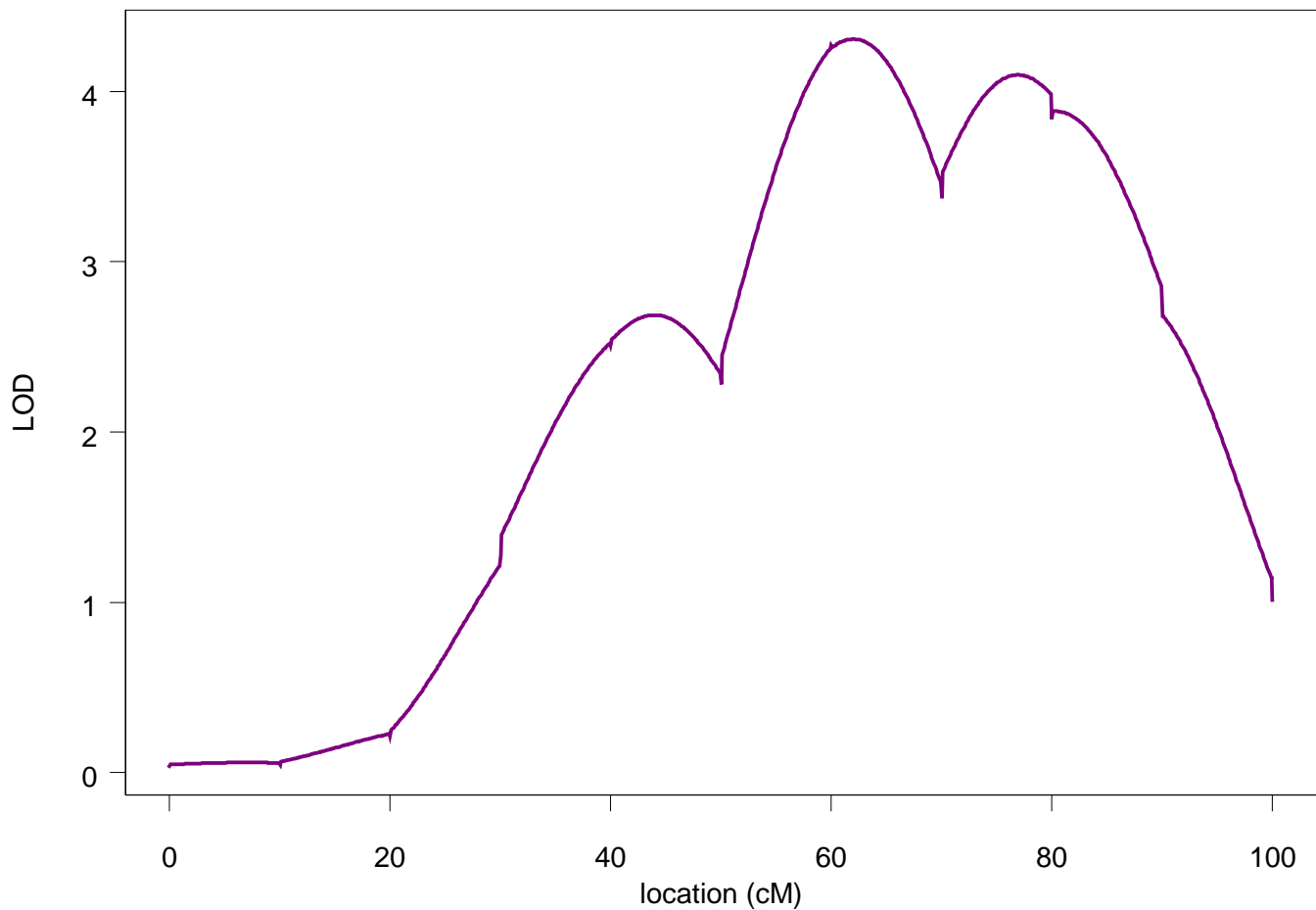
$$y = \mu + \sum \beta_j x_j + \varepsilon$$

Find the x 's with $\beta_j \neq 0$

Errors:

- Miss important loci
- Include extraneous loci

The usual method



At each location:

- Imagine a single QTL
- Infer genotypes
- Regression of phenotype on genotype

Model selection

- **Minimize prediction error**

- **Two problems**

- **Comparing models**

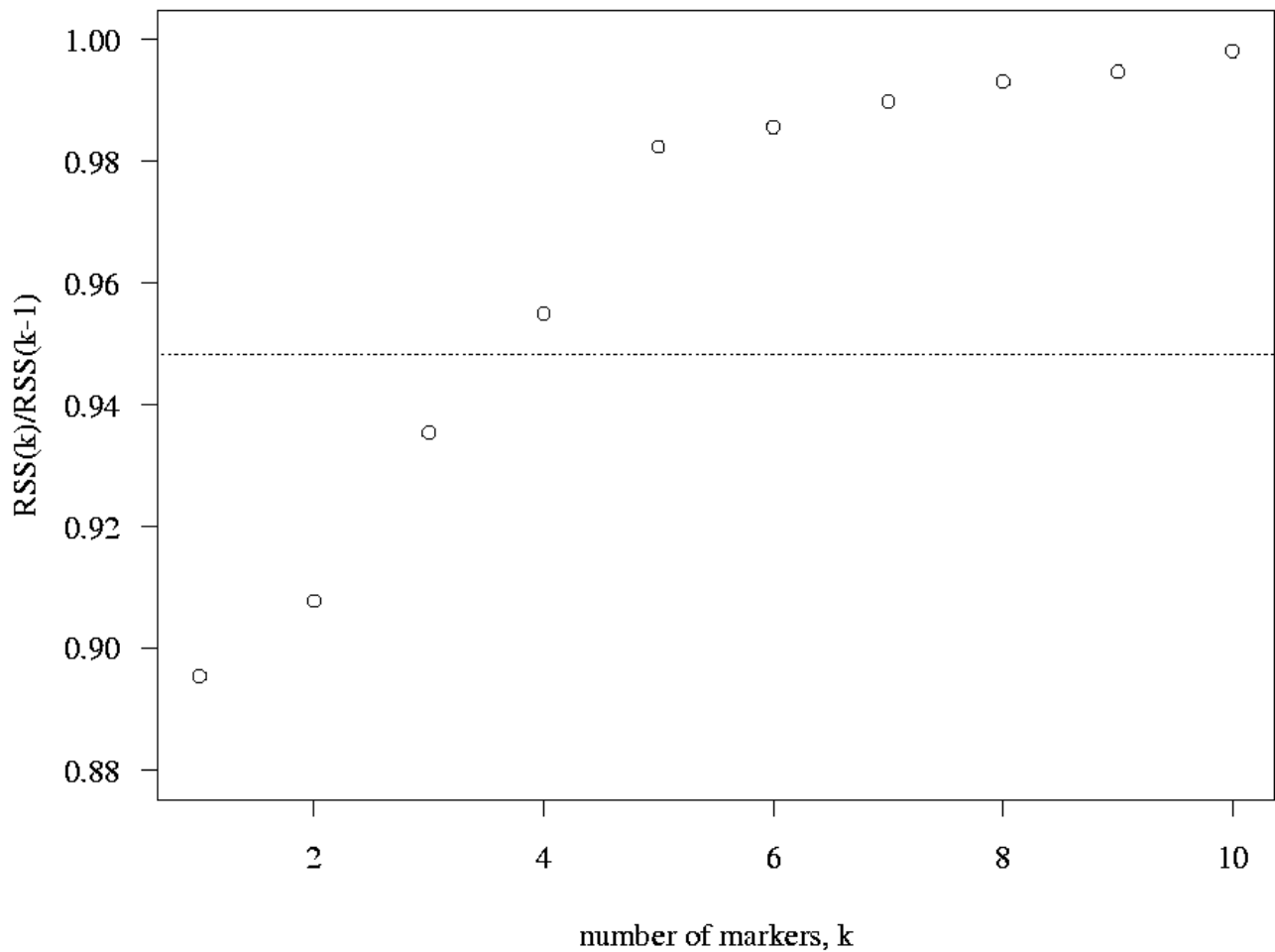
C_p	adj- R^2
cross-val	bootstrap
AIC	BIC
prior + Bayes	

- **Searching through models**

branch-and-bound
forward selection
backward elimination
stepwise selection
MCMC, etc.

BIC

$$\text{BIC} - \delta = \log \text{RSS}_k + \delta k \log(n) / n$$



$$\Delta \text{LOD} > \delta \log(n)$$

Simulation

- 250 progeny
- 9 chromosomes, each 100 cM
- 11 markers per chr. (10 cM sp.)
- 5 QTLs
- 2 linked pairs (sep. by 40 cM)
- Coupling or repulsion
- Effects = 0.5σ

Model searching

linked QTLs in coupling

cor.	forw	back	branch-and-bound		MCMC
			forw	back	
5	13	9	13	13	12
4	120	129	105	126	125
3	502	478	488	769	499
2	356	369	386	371	355
1	9	14	8	20	9
0	0	1	0	1	0
# incor.	59	56	52	66	59

Comparison of methods

linked QTLs in coupling

cor.	IM	CIM		BIC- δ	
		3	5	2	2.5
5	4	0	0	36	10
4	84	70	42	258	110
3	504	428	236	502	470
2	396	436	412	204	396
1	12	66	252	0	12
0	0	0	58	0	2
# incor.	56	30	22	114	40

Comparison of methods

linked QTLs in repulsion

cor.	IM	CIM		BIC- δ	
		3	5	2	2.5
5	0	1	7	68	15
4	0	3	39	59	17
3	22	136	175	216	107
2	215	305	267	165	112
1	505	388	363	332	418
0	258	167	149	160	331
# incor.	75	47	38	135	53

Discussion

- **Identifying QTLs is model selection**
- **Simulation studies are necessary**
 - compare procedures
 - understand a procedure's performance
- **Different situations will require different procedures**
- **Model organisms:**
 - underlying genes
 - genetic architecture
 - analytical methods
- **Gene to pathway; multiple loci to gene**
- **Other issues: epistasis, multiple traits**