QTL Model Search

Brian S. Yandell, UW-Madison

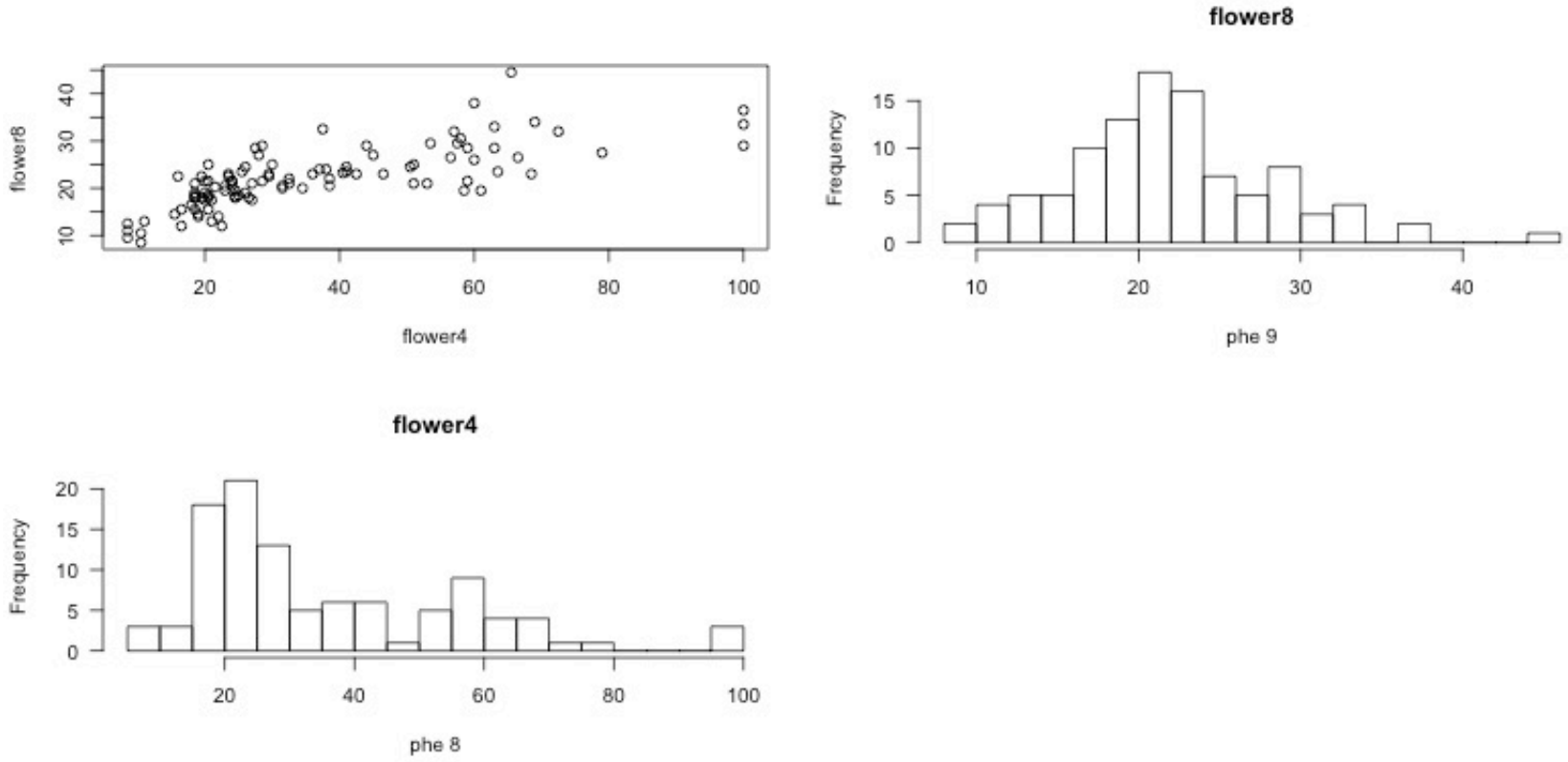
January 2017

evolution of QTL models

original ideas focused on rare & costly markers
models & methods refined as technology advanced

- single marker regression
- QTL (quantitative trait loci)
  - single locus models: interval mapping for QTL
  - **QTL model search: QTLs & epistasis**
- GWA (genome-wide association mapping)
  - adjust for population structure
  - capture "missing heritability"
  - genome-wide selection

phenotype data: flowering time



Satagopan JM, Yandell BS, Newton MA, Osborn TC (1996) Genetics

covariates

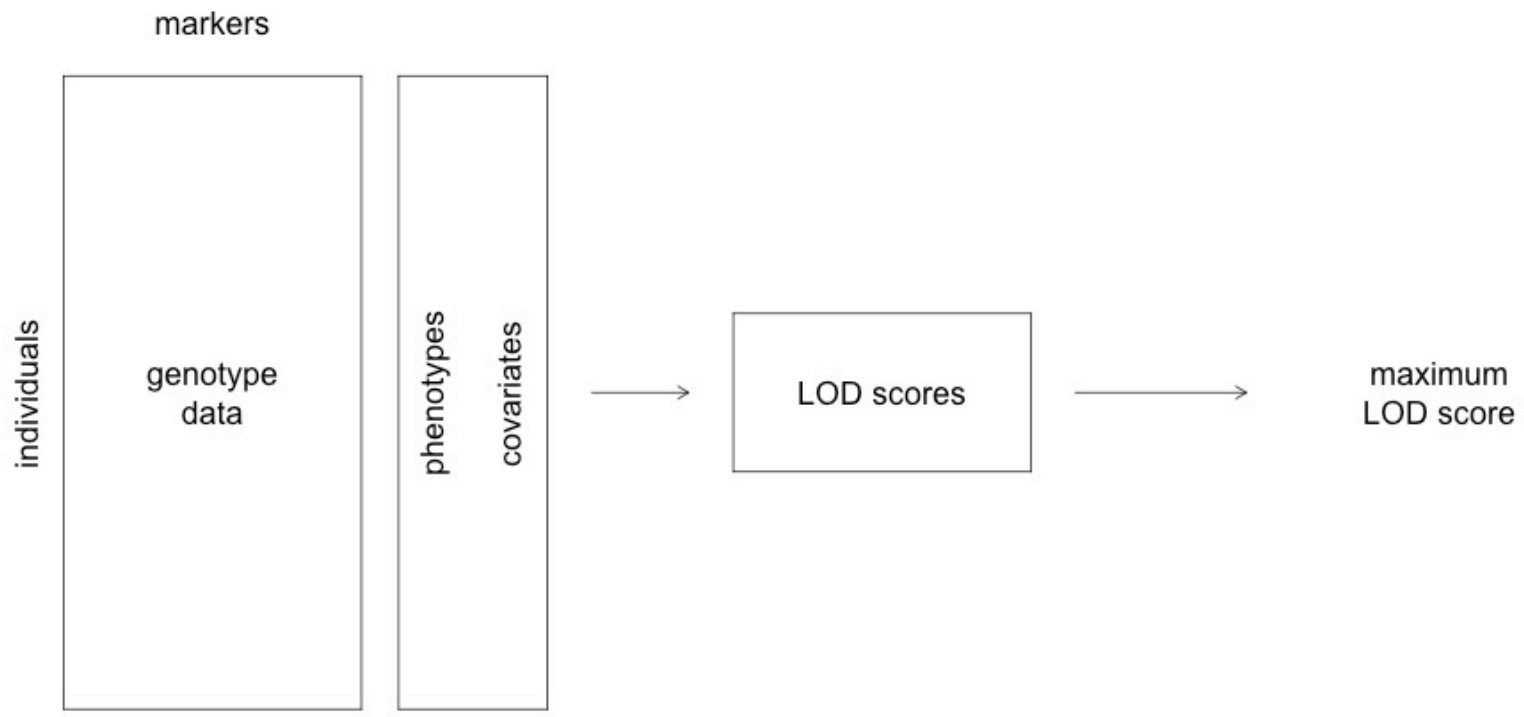
- examples: treatment, location, age, weight, height
- account for important design structure (location)
- adjust for important predictors
- reduce residual variation to increase power
    - covariate with strong effect

$$\mathrm{H}_0 : y = \mu + \beta_x x + e$$
$$\mathrm{H}_a : y = \mu_q + \beta_x x + e$$

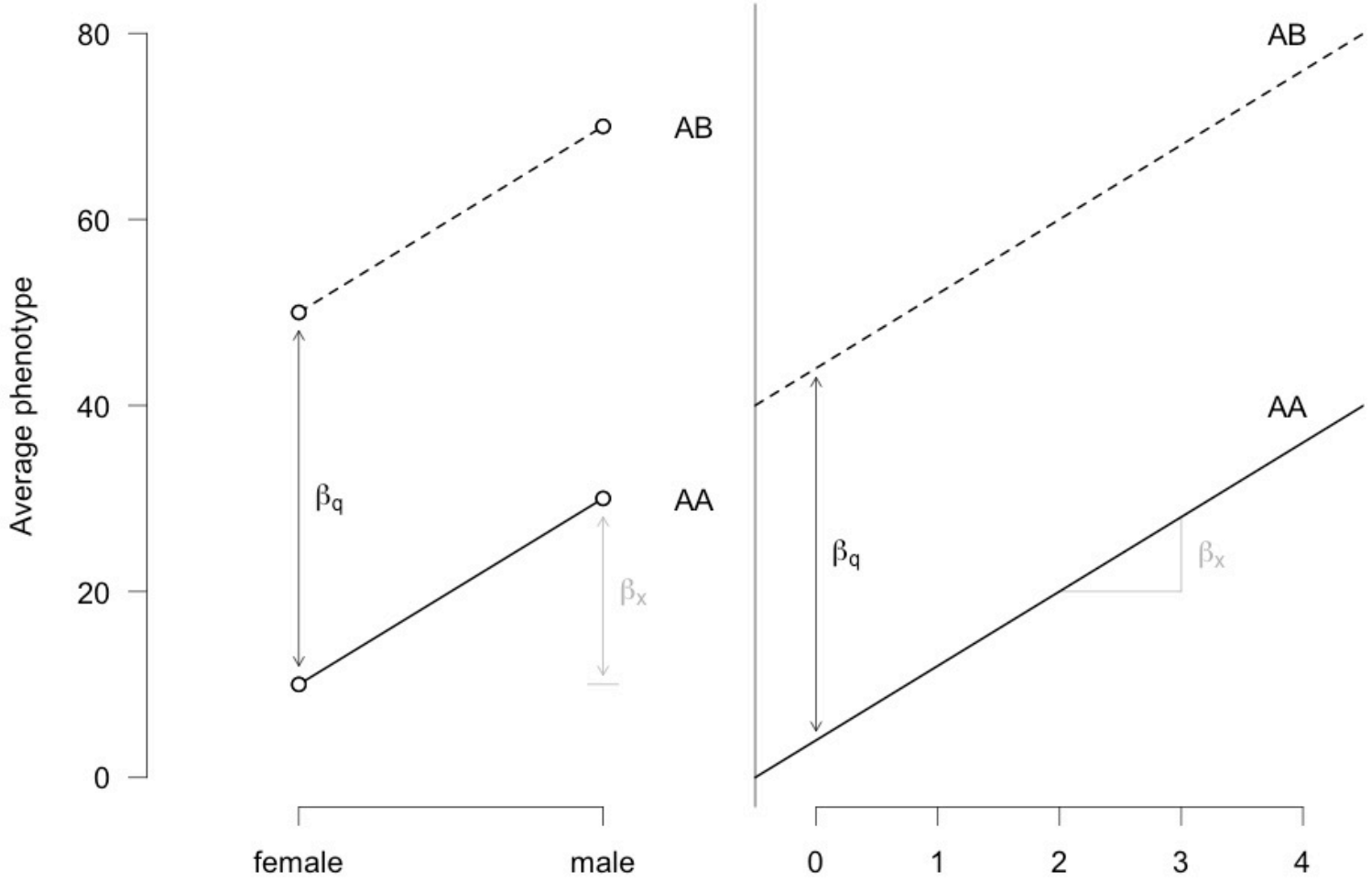
- more complicated (ignored here)
    - QTL * covariate interactions
    - covariate as ratio y/x

covariate cautions

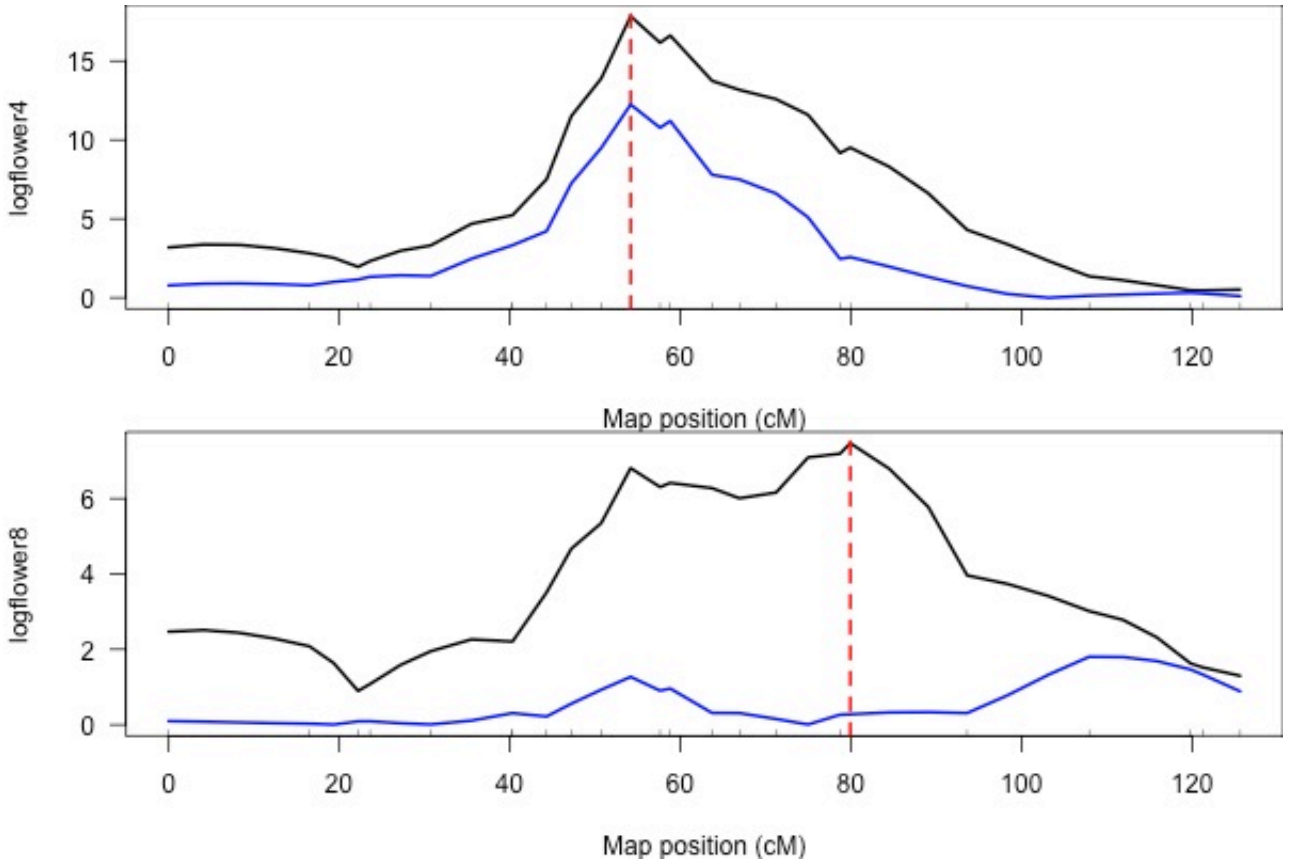
- use care when the covariate is another phenotype
- permutations: keep phenotype & covariates together



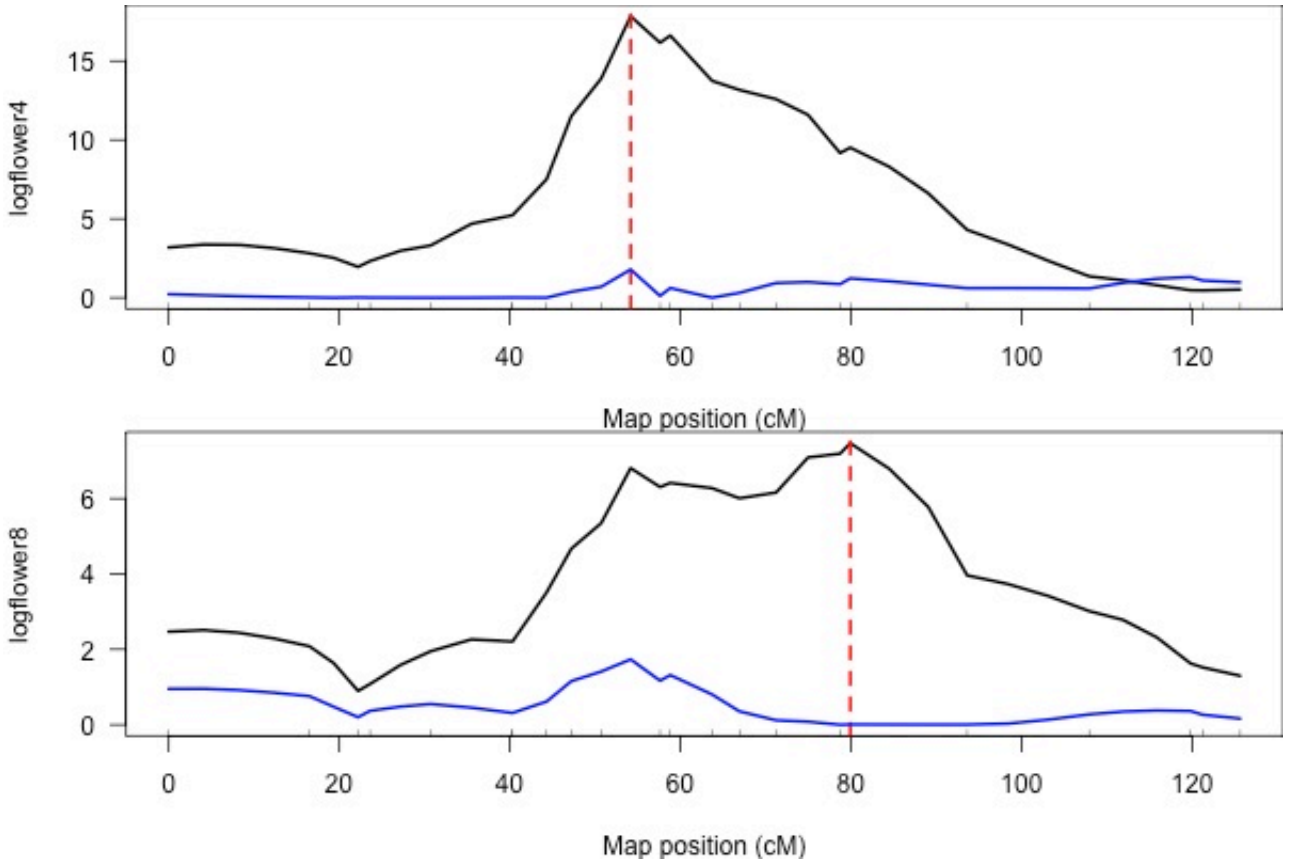
additive covariate



other phenotype as covariate



flowering time: QTL as covariate



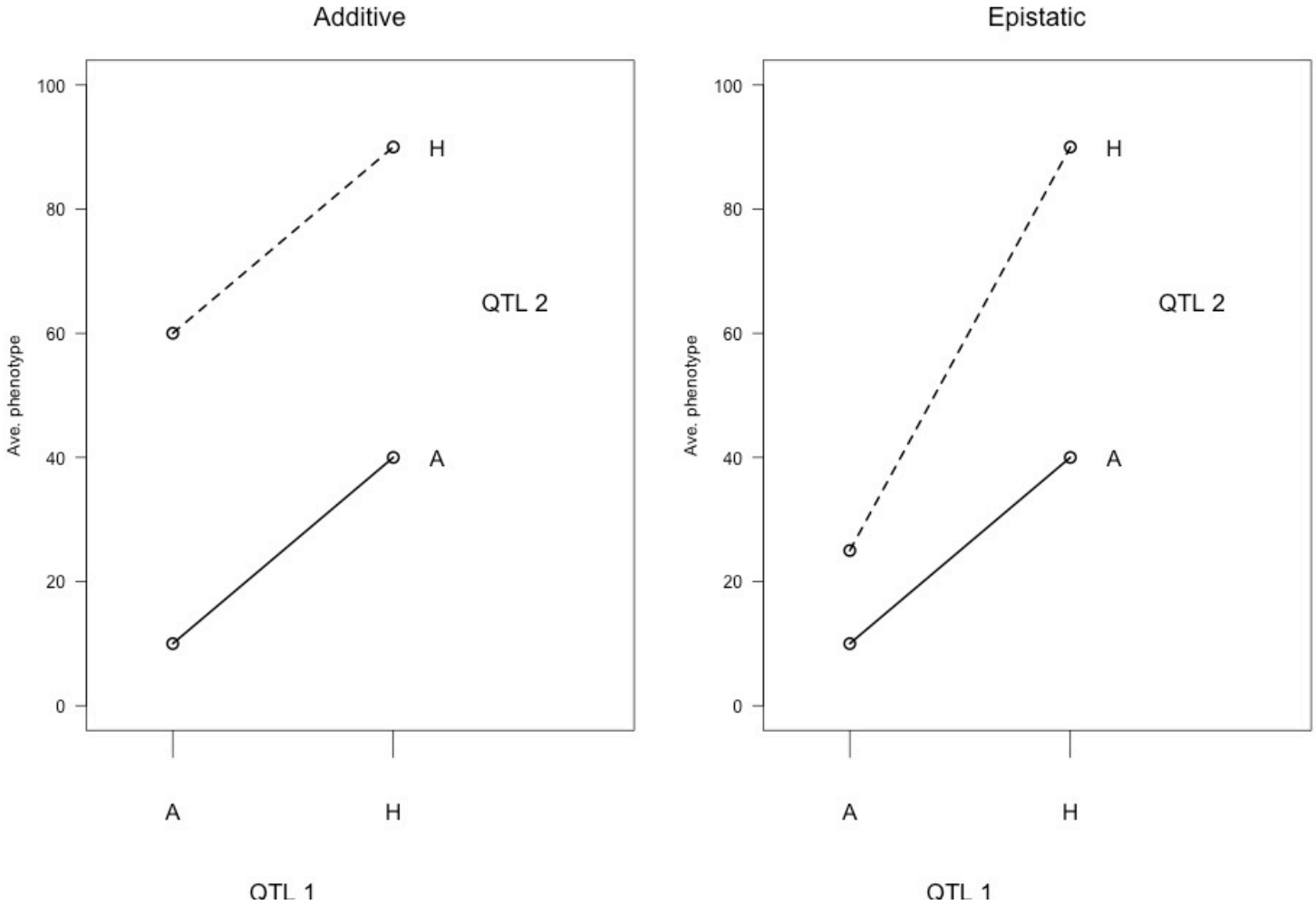
QTL model search

- Goals
  - identify QTL (and possible interactions among QTL)
  - estimate interval for QTL location
  - estimate QTL effects
- Challenges
  - how many QTL? which ones?
  - more complicated to fit each multiple QTL model
  - need rules to search across many QTL models

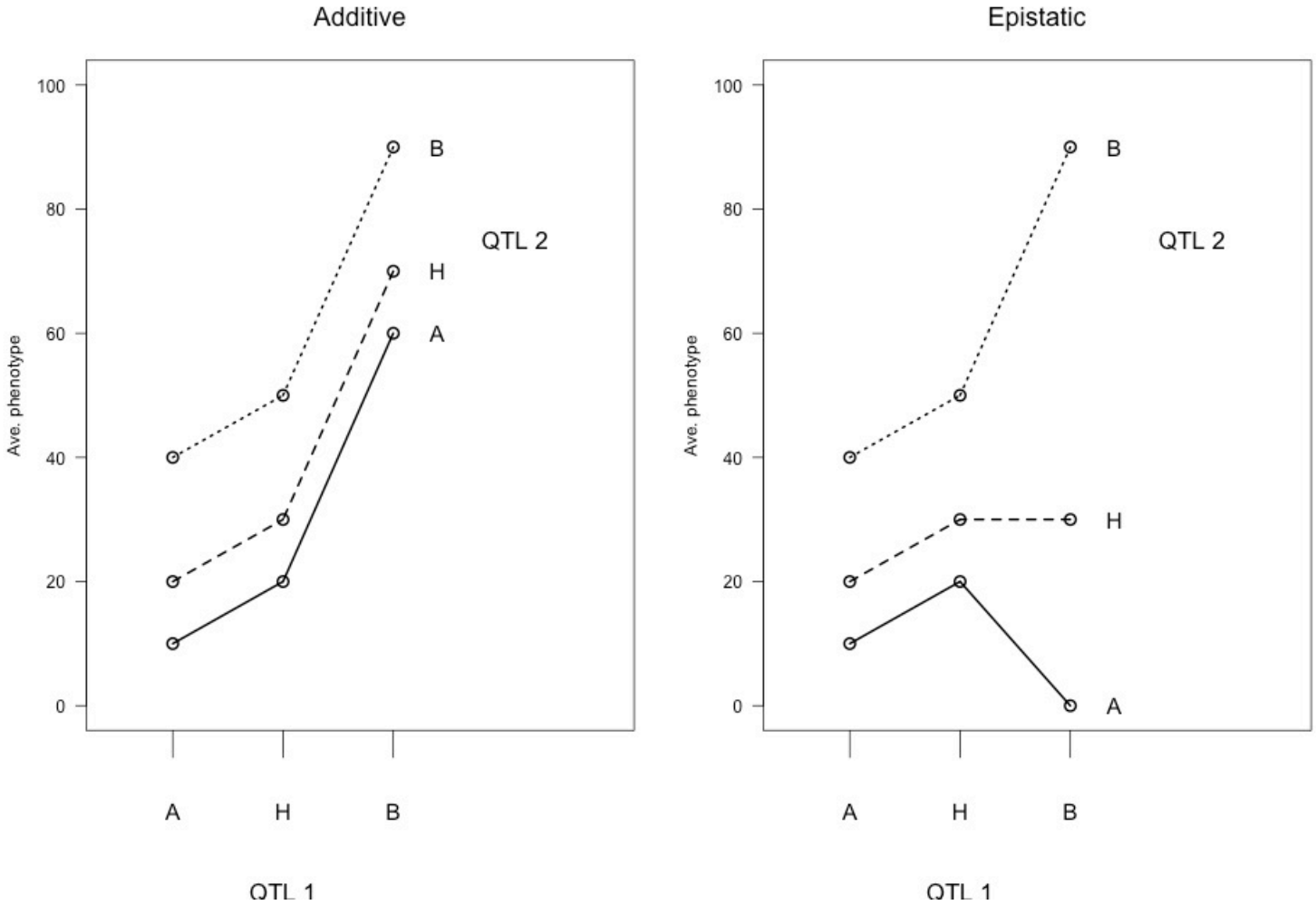
pros & cons of multiple QTL models

- benefits
  - reduce residual variation
  - increased power
  - separate linked QTL
  - identify interactions among QTL (epistasis)
- shortcomings
  - only includes significant loci
  - gets complicated very quickly
  - selection bias: overestimate effects of included loci
  - many loci of small effect ignored …

epistasis in BC or DH



epistasis in F2



multiple loci models

basic model looks the same

$$y = \mu_q + e$$

but now QTL has parts: $q = (q_1, q_2, \cdots)$

$$\mu_q = \mu(q_1, q_2, \cdots) = \mu + q_1\beta_1 + q_2\beta_2 + \cdots$$

- allows for multiple loci
- can add epistasis (here for BC)

$$\mu_q = \mu + \beta_1 q_1 + \beta_2 q_2 + \gamma q_1 q_2$$

- more terms for F2 …

LOD-based tests for 2 QTL

For all pairs of positions, fit the following models:

$$\mathrm{H}_f : y = \mu + \beta_1 q_1 + \beta_2 q_2 + \gamma q_1 q_2 + e$$
$$\mathrm{H}_a : y = \mu + \beta_1 q_1 + \beta_2 q_2 + e$$
$$\mathrm{H}_1 : y = \mu + \beta_1 q_1 + e$$
$$\mathrm{H}_0 : y = \mu + e$$

$\log_{10}$ likelihoods for QTL positions λ_1 (for q_1) and λ_2 (for q_2)

$l_f(\lambda_1, \lambda_2)$
$l_a(\lambda_1, \lambda_2)$
$l_1(\lambda_1)$
l_0

LOD scores for 2-QTL scan

full (interactive) vs no QTL ($f - 0$):

$$\texttt{lod}_f(\lambda_1, \lambda_2) = l_f(\lambda_1, \lambda_2) - l_0$$

additive vs no QTL ($a - 0$):

$$\texttt{lod}_a(\lambda_1, \lambda_2) = l_a(\lambda_1, \lambda_2) - l_0$$

interaction, or full vs additive ($f - a$):

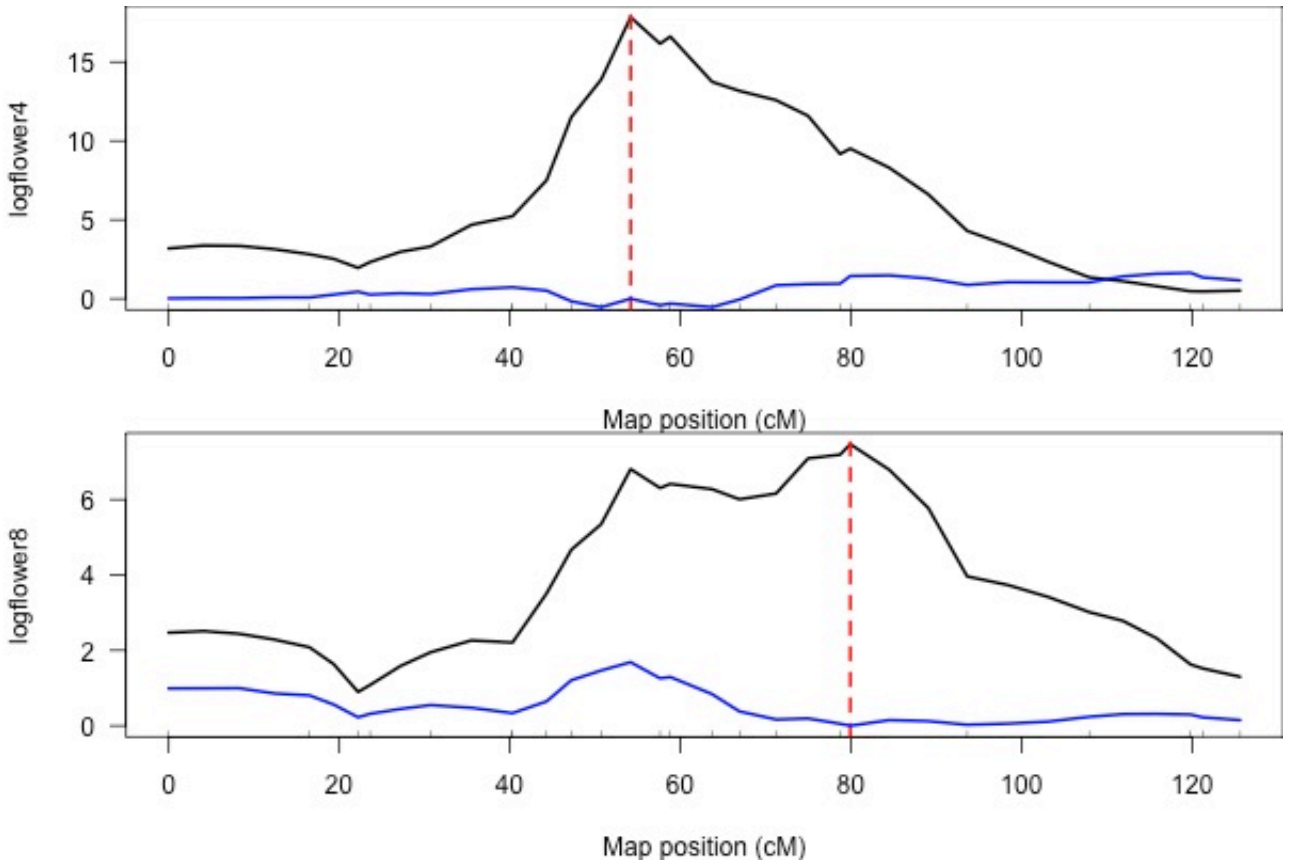
$$\texttt{lod}_i(\lambda_1, \lambda_2) = l_f(\lambda_1, \lambda_2) - l_a(\lambda_1, \lambda_2)$$

usual 1 QTL vs no QTL ($1 - 0$):

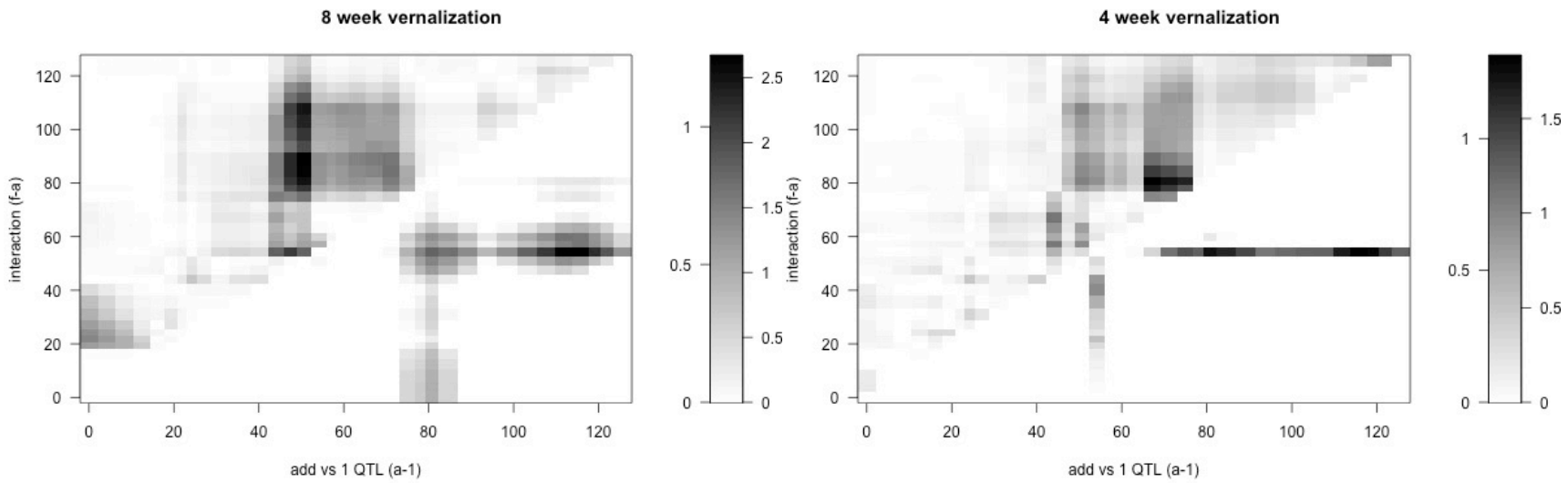
$$\texttt{lod}_1(\lambda_1) = l_1(\lambda_1) - l_0$$

flowering time: add QTL?

`R/qtl` tools: `sim.geno`, `makeqtl`, `addqtl`
fancier form of using first QTL as covariate

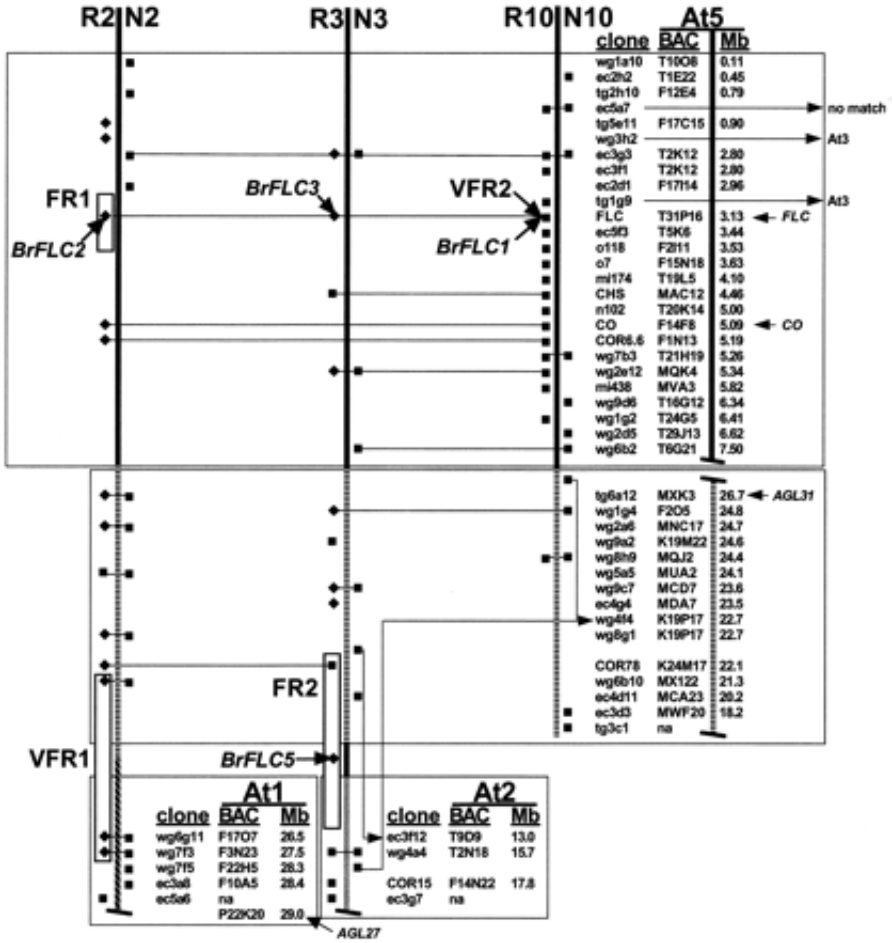


flowering time 2-D search

- 

  no evidence for 2 QTL at 4 week
- modest evidencde for 2 QTL at 8 week
- no evidence for interaction

Brassica FLC homologs (6 years later)



Schranz et al. Osborn (2002)

model search and selection

- QTL mapping began as hypothesis testing: is this a QTL?
  - much focus on adjusting for multiple testing
- better to view problem as model selection
  - what set of QTL are well supported?
  - is there evidence for QTL-QTL interactions?

Model = an identified set of QTL and QTL-QTL interactions
(and possibly covariates and QTL-covariate interactions)

QTL model search and selection

- Class of models: begin with additive models
  - add pairwise or higher interactions?
  - other approaches?
- Model fit (MLE, Haley-Knott, …)
- Model comparison
  - estimated prediction error
  - model effects criterion: AIC, BIC, penalized likelihood
  - Bayes method (prior across model space)
- Model search
  - forward, backward, stepwise selection
  - randomized algorithm

QTL model search goal

Goal: identify major players

- selected model has two types of errors:
  - miss important terms (QTLs or interactions)
  - include extraneous terms
- both errors likely at the same time
  - identify as many correct terms as possible
  - while controlling rate of inclusion of extra terms
- hypothesis testing only has one error at a time
  - pick no QTL model, but there is really a QTL at λ_1
  - pick 1 QTL model, but there is really no QTL

special nature of QTL models

What is special here?

- continuum of ordinal-valued predictors (the genetic loci)
- association among these QTL predictors
- loci on different chromosomes are independent
- along chromosome:
    - simple (and known) correlation structure

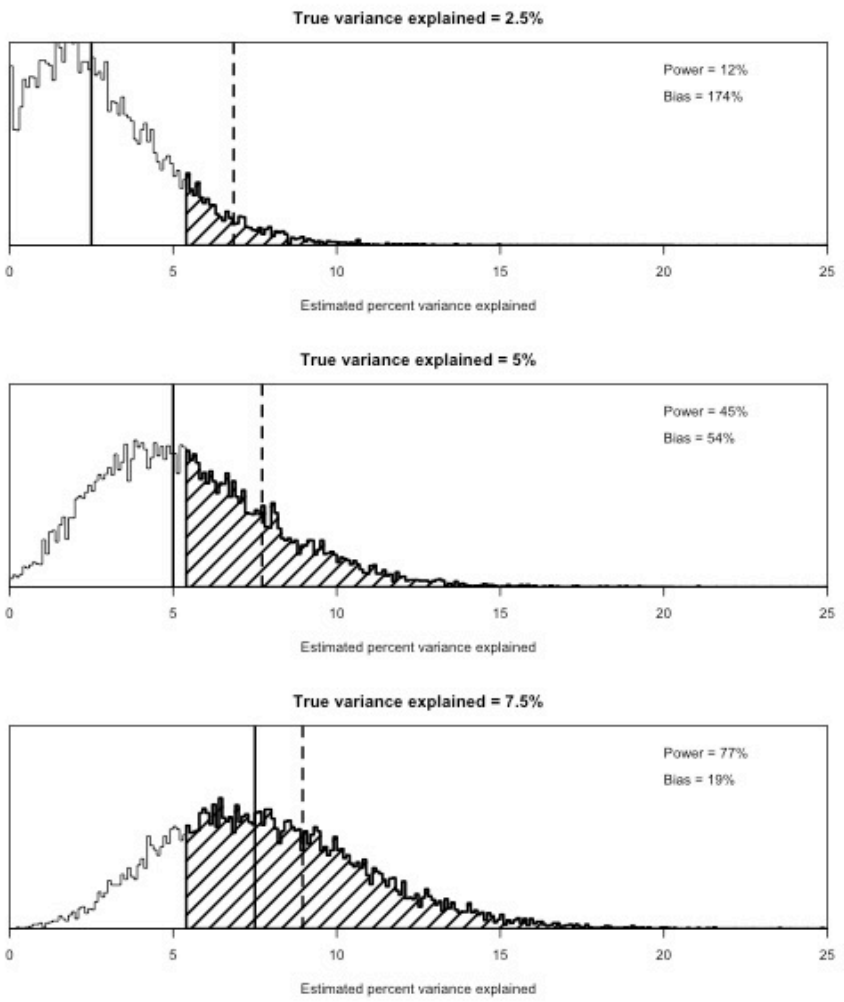
See Broman MultiQTL talk for more details

pros & cons of multiple QTL revisted

- Benefits
  - reduce residual variation
  - increased power
  - separate linked QTL
  - identify interactions among QTL (epistasis)
- Shortcomings
  - only includes significant loci
  - gets complicated very quickly
  - selection bias: overestimate effects of included loci
  - many loci of small effect ignored …

selection bias

- estimated QTL effect QTL varies from true effect
- detect QTL when estimated effect is large
- experiments with detected QTL often have larger estimated than true effect
- selection bias largest in QTLs with small or moderate effects
- true QTL effects smaller than those observed



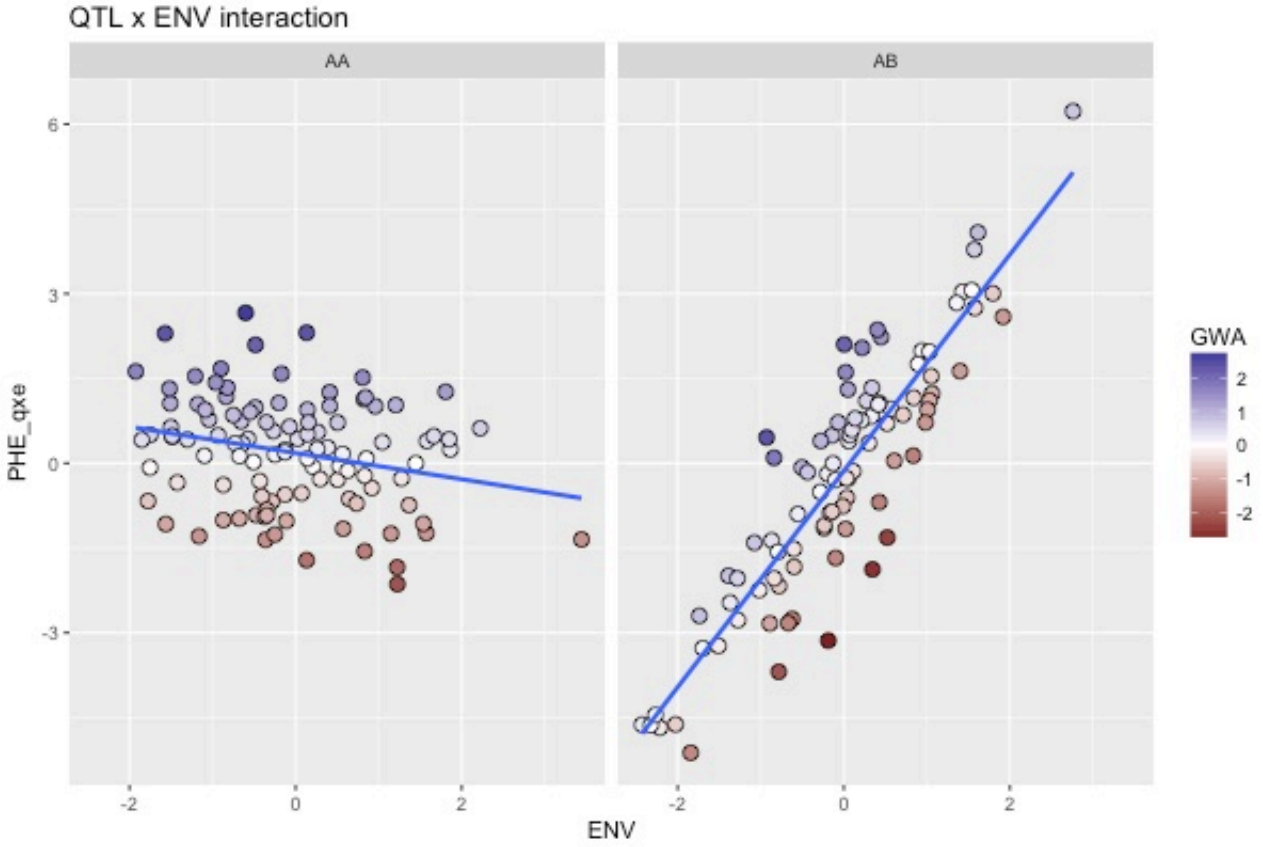
implications of selection bias

- estimated % variance explained by identified QTLs: too high
- repeating an experiment: different QTL (Beavis effect)
- congenics (or near isogenic lines): off base
- marker-assisted selection: missed effect

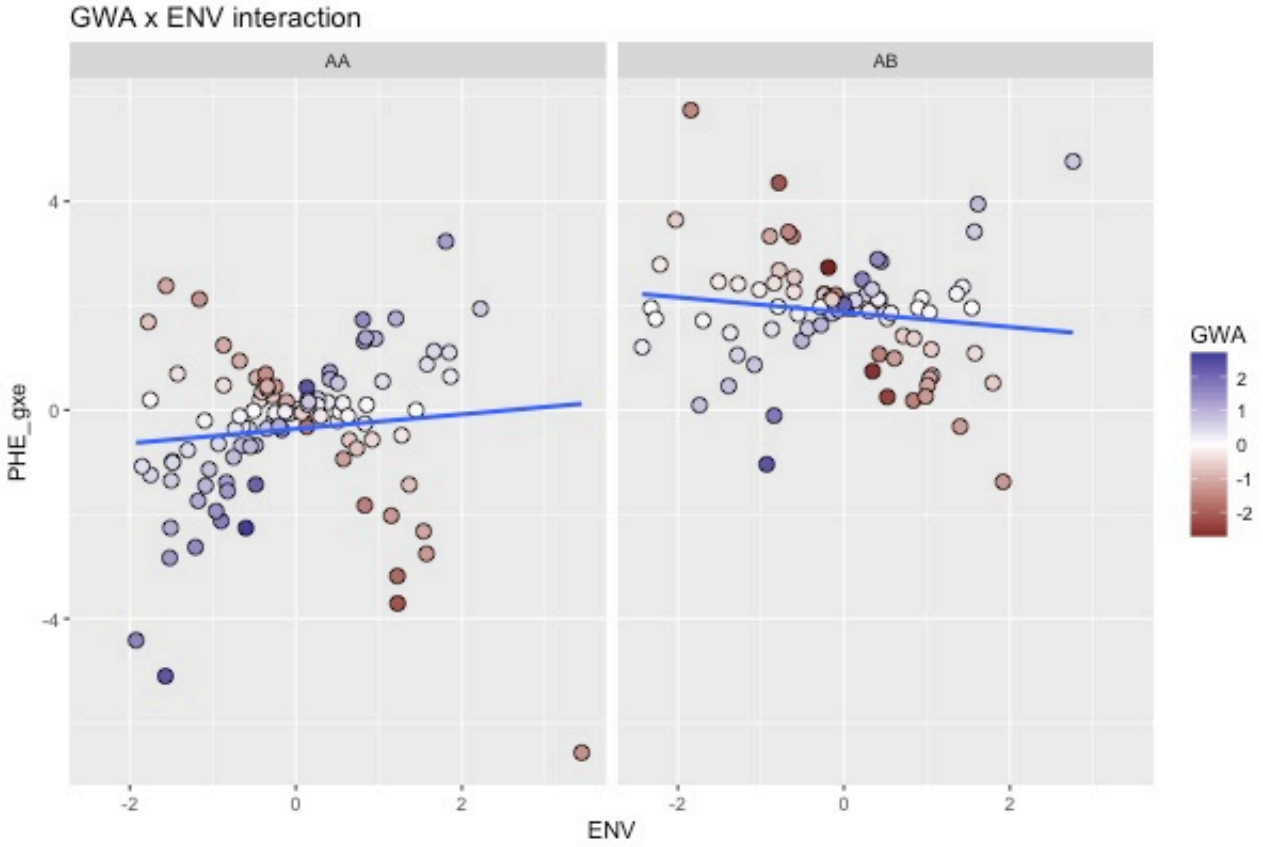
See Broman (2003) and Haley, Knott (1992).

Beavis WD (1994). The power and deceit of QTL experiments: Lessons from comparative QTL studies. In DB Wilkinson, (ed) 49th Ann Corn Sorghum Res Conf, pp 252–268. Amer Seed Trade Asso, Washington, DC.

PHE = GWA + QTL * ENV example



PHE = QTL + GWA * ENV example



Pareto chart: from QTL to GWA