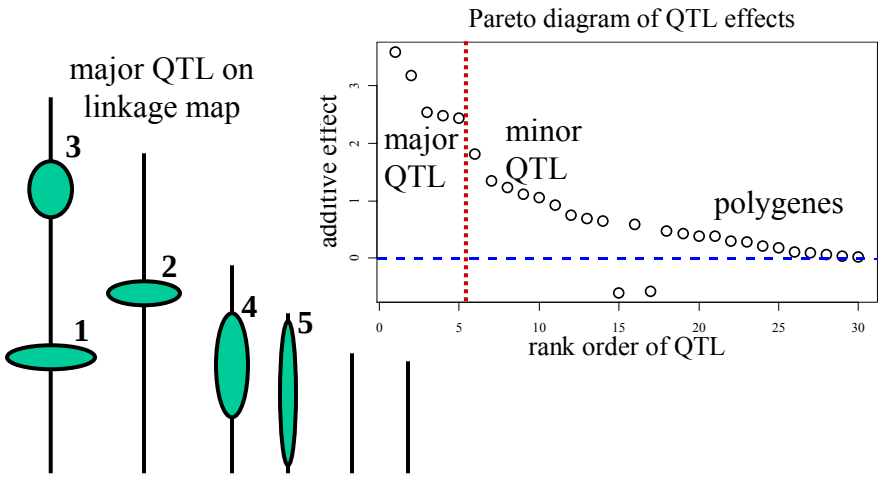
Bayesian Model Selection for Multiple QTL

Brian S. Yandell
University of Wisconsin-Madison
www.stat.wisc.edu/~yandell/statgen
Jackson Laboratory, October 2005

outline

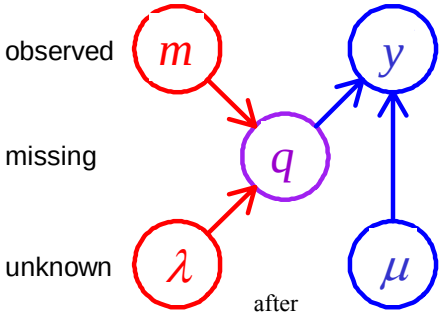
1. what is the goal of QTL study?
2. Bayesian priors & posteriors
3. model search using MCMC
   - Gibbs sampler and Metropolis-Hastings
4. model assessment
   - Bayes factors & model averaging
5. data examples in detail
   - plants & animals

1. what is the goal of QTL study?



interval mapping basics

- observed measurements
  - y = phenotypic trait
  - m = markers & linkage map
  - i = individual index (1,…,n)
- missing data
  - missing marker data
  - q = QT genotypes
    - alleles QQ, Qq, or qq at locus
- unknown quantities
  - λ = QT locus (or loci)
  - μ = phenotype model parameters
  - H = QTL model/genetic architecture
- pr($q|m,\lambda,H$) genotype model
  - grounded by linkage map, experimental cross
  - recombination yields multinomial for q given m
- pr($y|q,\mu,H$) phenotype model
  - distribution shape (assumed normal here)
  - unknown parameters μ (could be non-parametric)



after Sen Churchill (2001)

2. Bayesian priors & posteriors

- augment data (y,m) with missing genotypes q
- study model parameters (μ,λ) given data (y,m,q)
  - properties of posterior pr(μ,λ,q | y,m)
  - average posterior over missing genotypes q
    - pr(μ,λ | y,m) = sum$_q$ posterior
- sample from posterior in some clever way
  - multiple imputation (Sen Churchill 2002)
  - Markov chain Monte Carlo (MCMC)

posterior = likelihood * prior / constant

$$\text{pr}(\mu,\lambda,q \mid y,m) = \frac{\text{pr}(y \mid q,\mu)\text{pr}(q \mid m,\lambda)\text{pr}(\mu)\text{pr}(\lambda \mid m)}{\text{pr}(y \mid m)}$$

Bayes posterior vs. maximum likelihood

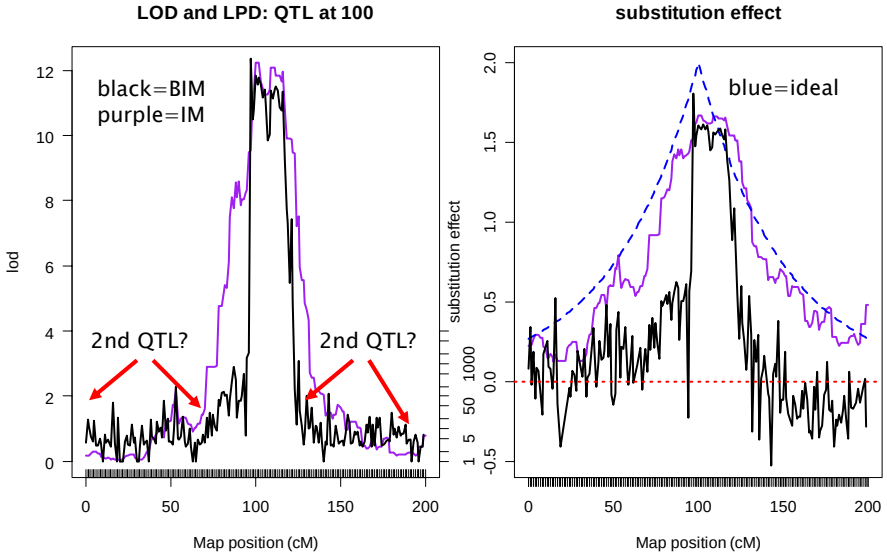
- classical approach maximizes likelihood
- Bayesian posterior averages over other parameters

$$\text{LOD}(\lambda) = \log_{10}\{\max_{\mu} \text{pr}(y \mid m,\mu,\lambda)\} + c$$

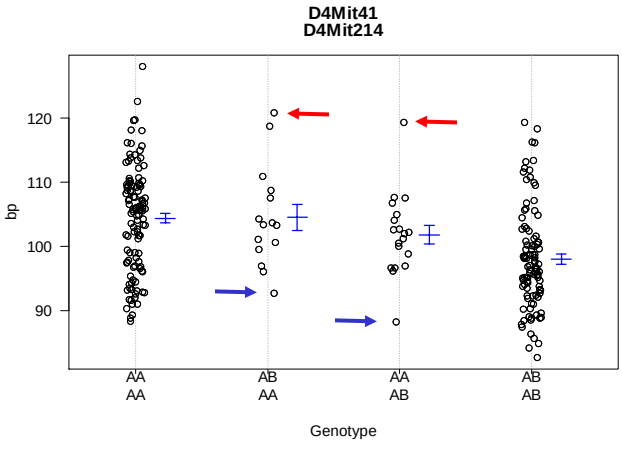
$$\text{LPD}(\lambda) = \log_{10}\{\text{pr}(\lambda \mid m)\int \text{pr}(y \mid m,\mu,\lambda)\text{pr}(\mu)d\mu\} + C$$

$$\text{pr}(y \mid m,\mu,\lambda) = \sum_q \text{pr}(y \mid q,\mu)\text{pr}(q \mid m,\lambda)$$

BC with 1 QTL: IM vs. BIM



how does phenotype *y* improve posterior for genotype *q*?



what are probabilities for genotype *q* between markers?

recombinants AA:AB

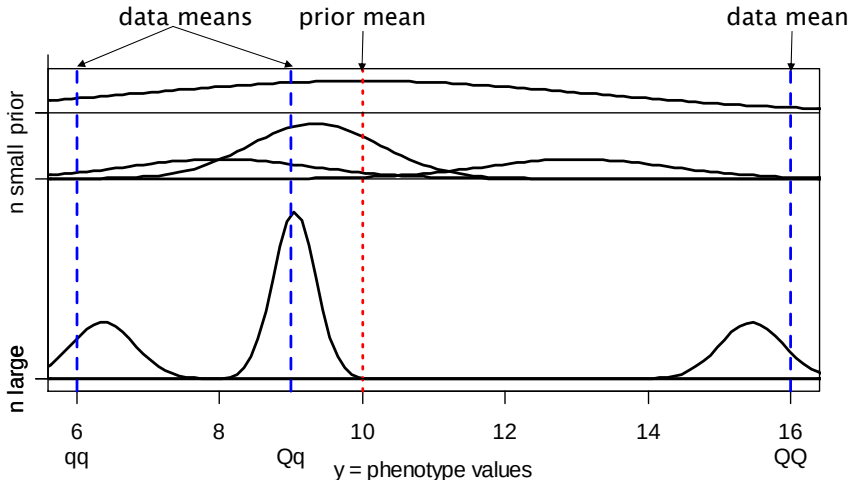
all 1:1 if ignore *y* and if we use *y*?

posterior on QTL genotypes

- full conditional of q given data, parameters
  - proportional to prior pr($q \mid m, \lambda$)
    - weight toward q that agrees with flanking markers
  - proportional to likelihood pr($y|q,\mu$)
    - weight toward q with similar phenotype values
- phenotype and flanking markers may conflict
  - posterior recombination balances these two weights
- this *is* the E-step of EM computations

$$\text{pr}(q \mid y, m, \mu, \lambda) = \frac{\text{pr}(y \mid q, \mu)\text{pr}(q \mid m, \lambda)}{\text{pr}(y \mid m, \mu, \lambda)}$$

prior & posteriors: genotypic means μ_q



prior & posteriors: genotypic means μ_q

shrink posterior from sample genotypic mean toward overall mean
partition individuals into sets S_q by genotype q
hyperparameter κ is related to heritability

prior: $$\mu_q \sim N\left(\bar{y}_\bullet, \kappa\sigma^2\right)$$

posterior given data: $$\mu_q \sim N\left(b_q\bar{y}_q + (1-b_q)\bar{y}_\bullet, b_q\frac{\sigma^2}{n_q}\right)$$

$$\bar{y}_q = \underset{S_q}{\text{sum}}\frac{y}{n_q}, n_q = \text{count}\{S_q\}$$

shrinkage factor: $$b_q = \frac{\kappa n_q}{\kappa n_q + 1} \to 1$$

what if variance σ^2 is unknown?

- sample variance is proportional to chi-square
  - $ns^2 / \sigma^2 \sim \chi^2(n)$
  - likelihood of sample variance s^2 given n, σ^2
- conjugate prior is inverse chi-square
  - $\nu\tau^2 / \sigma^2 \sim \chi^2(\nu)$
  - prior of population variance σ^2 given ν, τ^2
- posterior is weighted average of likelihood and prior
  - $(\nu\tau^2 + ns^2) / \sigma^2 \sim \chi^2(\nu + n)$
  - posterior of population variance σ^2 given n, s^2, ν, τ^2
- empirical choice of hyper-parameters
  - $\tau^2 = s^2/3$, $\nu = 6$
  - $E(\sigma^2 \mid \nu,\tau^2) = s^2/2$, $Var(\sigma^2 \mid \nu,\tau^2) = s^4/4$

multiple QTL phenotype model

- phenotype affected by genotype & environment

  $\text{pr}(y|q,\mu) \sim N(\mu_q, \sigma^2)$

  $y = \mu_q$ + environment

  $\mu_q = \beta_0 + \text{sum}_{\{j \text{ in } H\}} \beta_j(q)$

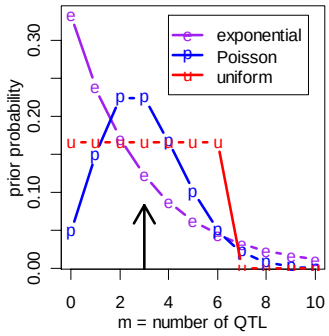
  number of terms in QTL model $H \le 2^{nqtl}$ (3^{nqtl} for F_2)
- partition genotypic mean into QTL effects

  $\mu_q = \beta_0 + \beta_1(q_1) + \beta_2(q_2) + \beta_{12}(q_1,q_2)$

  μ_q = mean + main effects + epistatic interactions
- partition prior and posterior (details omitted)

prior & posterior on QTL model *H*

- index model *H* by number of QTL
- what prior on number of QTL?
  - uniform over some range
  - Poisson with prior mean
  - geometric with prior mean
- prior influences posterior
  - good: reflects prior belief
    - push data in discovery process
  - bad: skeptic revolts!
    - "answer" depends on "guess"



epistatic interactions

- model space issues
  - 2-QTL interactions only?
  - Fisher-Cockerham partition vs. tree-structured?
  - general interactions among multiple QTL
- model search issues
  - epistasis between significant QTL
    - check all possible pairs when QTL included?
    - allow higher order epistasis?
  - epistasis with non-significant QTL
    - whole genome paired with each significant QTL?
    - pairs of non-significant QTL?
- Yi Xu (2000) *Genetics;* Yi, Xu, Allison (2003) *Genetics;* Yi (2004)

3. QTL Model Search using MCMC

- construct Markov chain around posterior
  - want posterior as stable distribution of Markov chain
  - in practice, the chain tends toward stable distribution
    - initial values may have low posterior probability
    - burn-in period to get chain mixing well
- update QTL model components from full conditionals
  - update locus λ given q,H (using Metropolis-Hastings step)
  - update genotypes q given λ,μ,y,H (using Gibbs sampler)
  - update effects μ given q,y,H (using Gibbs sampler)
  - update QTL model H given λ,μ,y,q (using Gibbs or M-H)

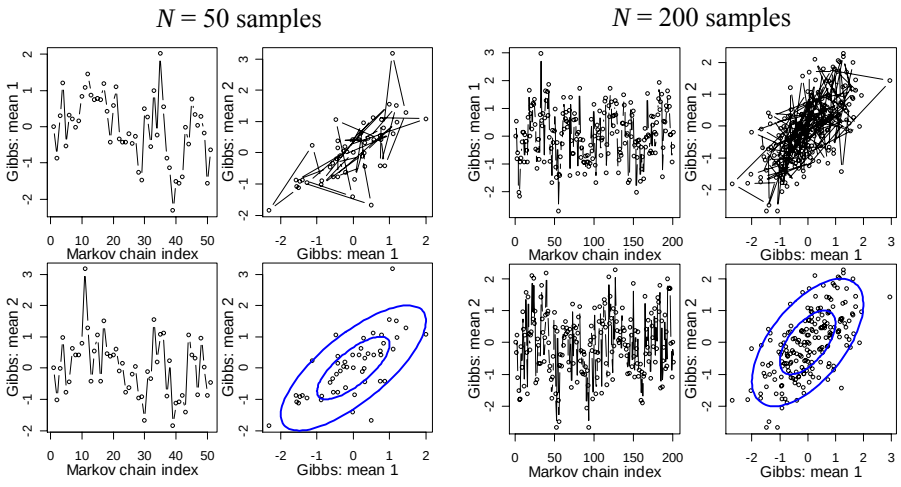
$$(\lambda, q, \mu, H) \sim \text{pr}(\lambda, q, \mu, H \mid Y, X)$$
$$(\lambda, q, \mu, H)_1 \rightarrow (\lambda, q, \mu, H)_2 \rightarrow \cdots \rightarrow (\lambda, q, \mu, H)_N$$

Gibbs sampler idea

- two correlated normals (genotypic means in BC)
  - could draw samples from both together
  - but easier to sample one at a time
- Gibbs sampler:
  - sample each from its full conditional
  - pick order of sampling at random
  - repeat *N* times

$$\mu_{QQ}, \mu_{Qq} \sim N(0,1) \text{ but } cor(\mu_{QQ}, \mu_{Qq}) = \rho$$
$$\mu_{QQ} \text{ given } \mu_{Qq} \sim N\left(\rho\mu_{Qq}, 1-\rho^2\right)$$
$$\mu_{Qq} \text{ given } \mu_{QQ} \sim N\left(\rho\mu_{QQ}, 1-\rho^2\right)$$

Gibbs sampler samples: $\rho = 0.6$



How to sample a locus λ?

- cannot easily sample from locus full conditional

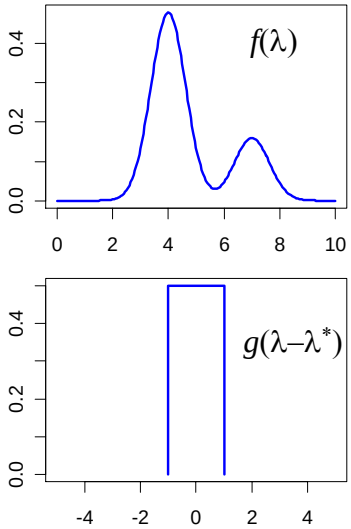
  $\text{pr}(\lambda \mid m,q) = \text{pr}(\lambda)\ \text{pr}(q \mid m,\lambda)$ / constant
- constant determined by averaging
  - over all possible genotypes
  - over entire map
- Gibbs sampler will not work in general
  - but can use method based on ratios of probabilities
  - Metropolis-Hastings is extension of Gibbs sampler

Metropolis-Hastings idea

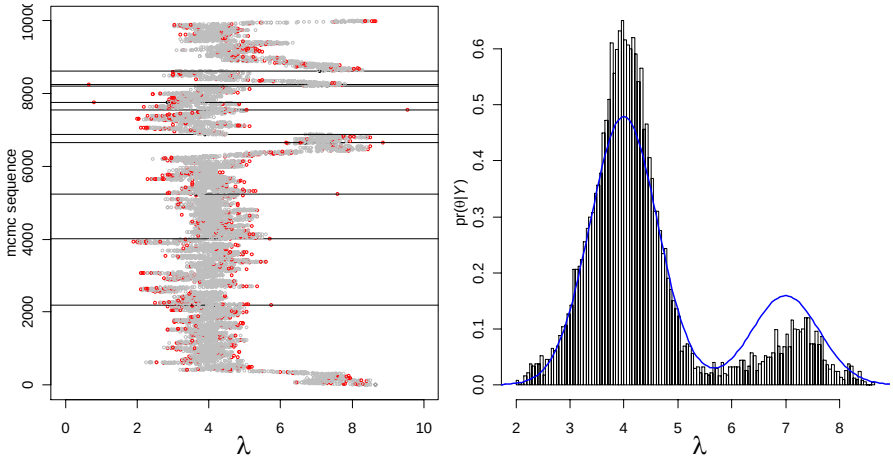
- want to study distribution $f(\lambda)$
- take Monte Carlo samples
  - unless too complicated
- pick an *arbitrary* distribution $g(\lambda, \lambda^*)$
- draw Metropolis-Hastings samples:
  - current sample value λ
  - propose new value λ^* from $g(\lambda, \lambda^*)$
  - accept new value with prob A

$$A = \min\left(1, \frac{f(\lambda^*)g(\lambda^*,\lambda)}{f(\lambda)g(\lambda,\lambda^*)}\right)$$

- Gibbs sampler is special case
  - $g(\lambda, \lambda^*) = f(\lambda^*)$
  - always accept new proposal: $A = 1$

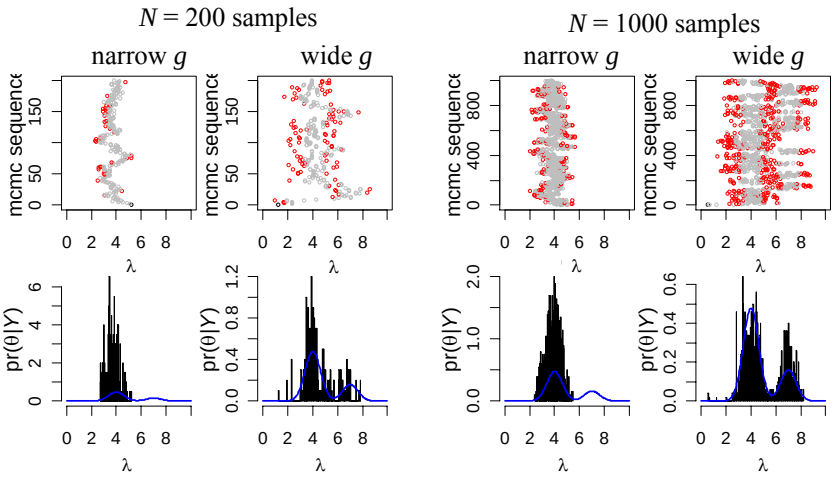


MCMC realization

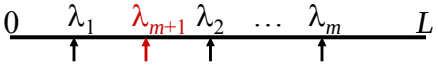


added twist: occasionally propose from whole domain

Metropolis-Hastings samples



sampling across QTL models *H*



action steps: draw one of three choices

- update QTL model *H* with probability 1-*b*(*H*)-*d*(*H*)
  - update current model using full conditionals
  - sample QTL loci, effects, and genotypes
- add a locus with probability *b*(*H*)
  - propose a new locus along genome
  - innovate new genotypes at locus and phenotype effect
  - decide whether to accept the "birth" of new locus
- drop a locus with probability *d*(*H*)
  - propose dropping one of existing loci
  - decide whether to accept the "death" of locus

reversible jump MCMC

- consider known genotypes *q* at 2 known loci λ
  - models with 1 or 2 QTL
- M-H step between 1-QTL and 2-QTL models
  - model changes dimension (via careful bookkeeping)
  - consider mixture over QTL models *H*

$$nqtl = 1 : Y = \beta_0 + \beta_1(q_1) + e$$

$$nqtl = 2 : Y = \beta_0 + \beta_1(q_1) + \beta_2(q_2) + e$$

Gibbs sampler with loci indicators

- partition genome into intervals
  - at most one QTL per interval
  - interval = 1 cM in length
  - assume QTL in middle of interval
- use loci to indicate presence/absence of QTL in each interval
  - $\gamma = 1$ if QTL in interval
  - $\gamma = 0$ if no QTL
- Gibbs sampler on loci indicators
  - see work of Nengjun Yi (and earlier work of Ina Hoeschele)

$$Y = \beta_0 + \gamma_1 \beta_1(q_1) + \gamma_2 \beta_2(q_1) + e$$

Bayesian shrinkage estimation

- soft loci indicators
  - strength of evidence for λ_j depends on variance of β_j
  - similar to $\gamma > 0$ on grey scale
- include all possible loci in model
  - pseudo-markers at 1cM intervals
- Wang et al. (2005 *Genetics*)
  - Shizhong Xu group at U CA Riverside

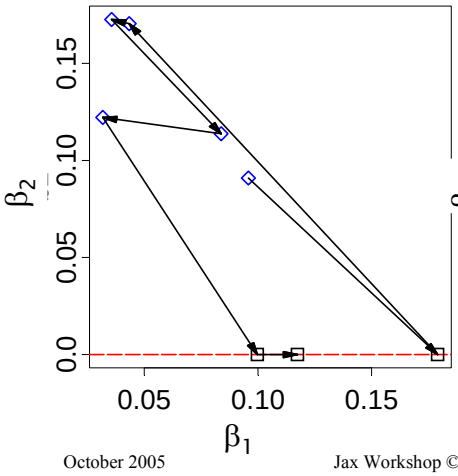
$$Y = \beta_0 + \beta_1(q_1) + \beta_2(q_1) + \ldots + e$$

$$\beta_j(q_j) \sim N(0, \sigma_j^2), \sigma_j^2 \sim \text{inverse - chisquare}$$

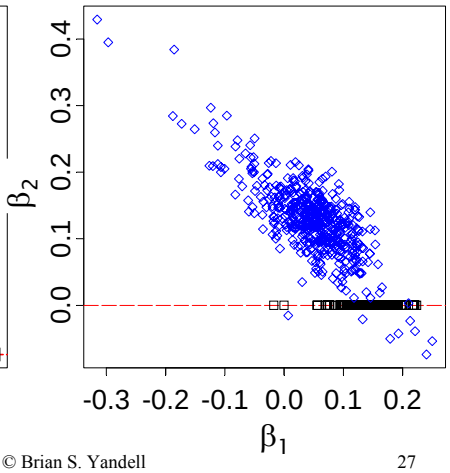
geometry of effect samples

(collinearity of QTL yields correlated estimates)

a short sequence



first 1000 with m<3



4. Model Assessment

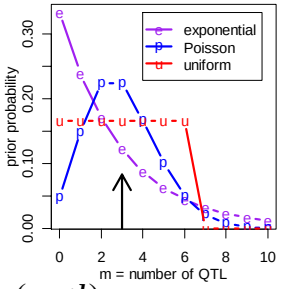
- balance model fit against model complexity

	smaller model	bigger model
model fit	miss key features	fits better
prediction	may be biased	no bias
interpretation	easier	more complicated
parameters	low variance	high variance

- information criteria: penalize L by model size $|H|$
  - compare IC = $-2 \log L(H \mid y) + \text{penalty}(H)$
- Bayes factors: balance posterior by prior choice
  - compare pr(data y | model H)

QTL Bayes factors

- BF = posterior odds / prior odds
- BF equivalent to BIC
  - simple comparison: 1 vs 2 QTL
    - same as LOD test
  - general comparison of models
  - want Bayes factor >> 1
- *nqtl* = number of QTL
  - indexes model complexity
  - genetic architecture also important



$$BF_{nqtl,nqtl+1} = \frac{\text{pr}(nqtl|\text{data})/\text{pr}(nqtl)}{\text{pr}(nqtl+1|\text{data})/\text{pr}(nqtl+1)}$$

Bayes factors to assess models

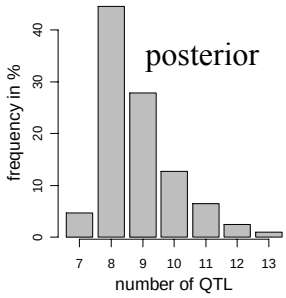
- Bayes factor: which model best supports the data?
  - ratio of posterior odds to prior odds
  - ratio of model likelihoods
- equivalent to *LR* statistic when
  - comparing two nested models
  - simple hypotheses (e.g. 1 vs 2 QTL)
- Bayes Information Criteria (BIC)
  - Schwartz introduced for model selection in general settings
  - penalty to balance model size (p = number of parameters)

$$B_{12} = \frac{\text{pr}(\text{model}_1|Y)/\text{pr}(\text{model}_2|Y)}{\text{pr}(\text{model}_1)/\text{pr}(\text{model}_2)} = \frac{\text{pr}(Y|\text{model}_1)}{\text{pr}(Y|\text{model}_2)}$$

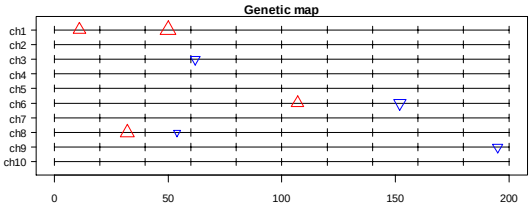
$$-2\log(B_{12}) = -2\log(LR) - (p_2 - p_1)\log(n)$$

simulations and data studies

- simulated F2 intercross, 8 QTL
  - (Stephens, Fisch 1998)
  - *n*=200, heritability = 50%
  - detected 3 QTL
- increase to detect all 8
  - *n*=500, heritability to 97%



QTL	chr	loci	effect
1	1	11	–3
2	1	50	–5
3	3	62	+2
4	6	107	–3
5	6	152	+3
6	8	32	–4
7	8	54	+1
8	9	195	+2



loci pattern across genome

- notice which chromosomes have persistent loci
- best pattern found 42% of the time

Chromosome

m	**1**	2	**3**	4	5	**6**	7	**8**	**9**	10	**Count of 8000**
8	**2**	**0**	**1**	**0**	**0**	**2**	**0**	**2**	**1**	**0**	3371
9	*3*	0	1	0	0	2	0	2	1	0	751
7	2	0	1	0	0	2	0	*1*	1	0	377
9	2	0	1	0	0	2	0	2	1	0	218
9	2	0	1	0	0	*3*	0	2	1	0	218
9	2	0	1	0	0	2	0	2	*2*	0	198

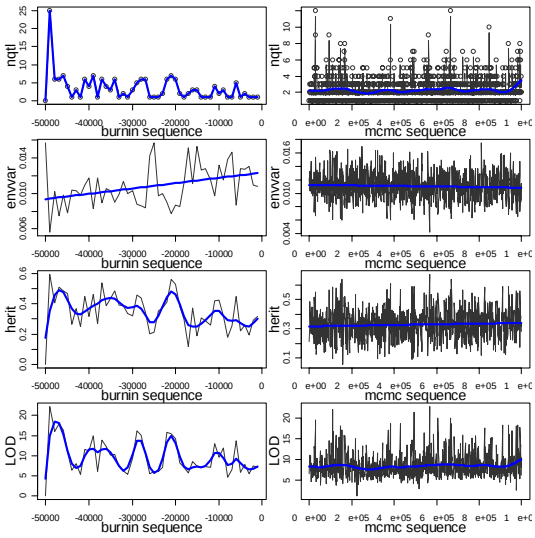
B. napus 8-week vernalization whole genome study

- 108 plants from double haploid
  - similar genetics to backcross: follow 1 gamete
  - parents are Major (biennial) and Stellar (annual)
- 300 markers across genome
  - 19 chromosomes
  - average 6cM between markers
    - median 3.8cM, max 34cM
  - 83% markers genotyped
- phenotype is days to flowering
  - after 8 weeks of vernalization (cooling)
  - Stellar parent requires vernalization to flower
- available in R/bim package
- Ferreira et al. (1994); Kole et al. (2001); Schranz et al. (2002)

Markov chain Monte Carlo sequence

burnin (sets up chain)
mcmc sequence

number of QTL
environmental variance
h^2 = heritability
(genetic/total variance)
LOD = likelihood



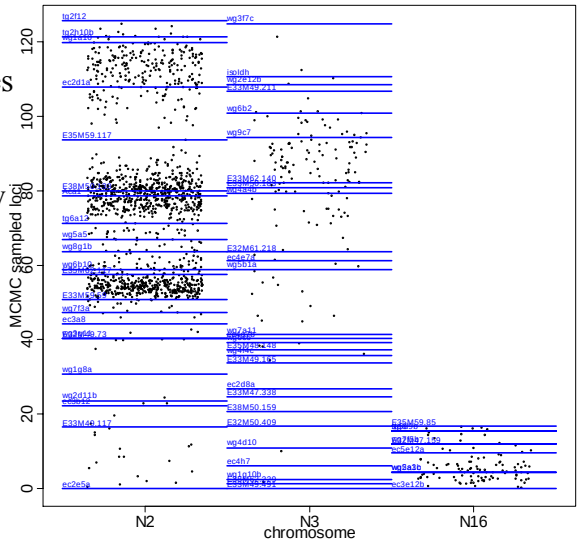
MCMC sampled loci

subset of chromosomes
N2, N3, N16

points jittered for view
blue lines at markers

note concentration
on chromosome N2

includes all models

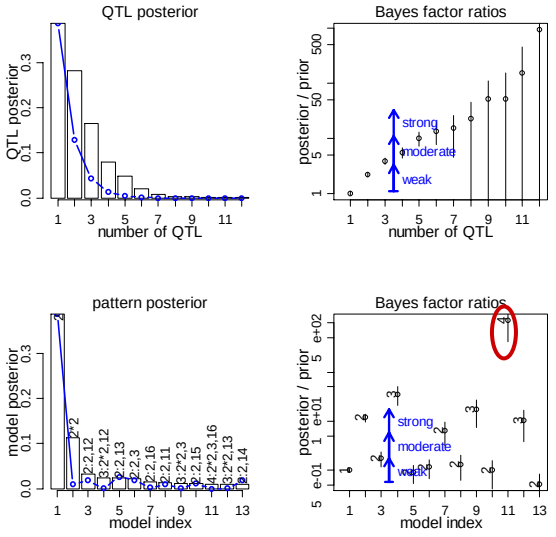


Bayesian model assessment

row 1: # QTL
row 2: pattern

col 1: posterior
col 2: Bayes factor
note error bars on bf

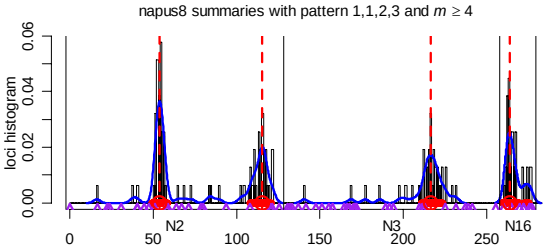
evidence suggests
4-5 QTL
N2(2-3),N3,N16



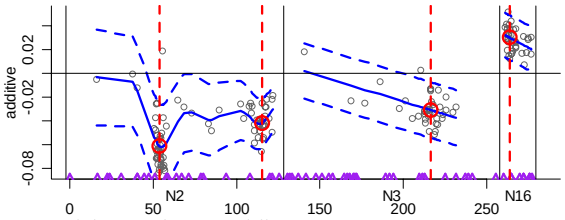
Bayesian estimates of loci & effects

model averaging: at least 4 QTL

histogram of loci
blue line is density
red lines at estimates



estimate additive effects
(red circles)
grey points sampled
from posterior
blue line is cubic spline
dashed line for 2 SD

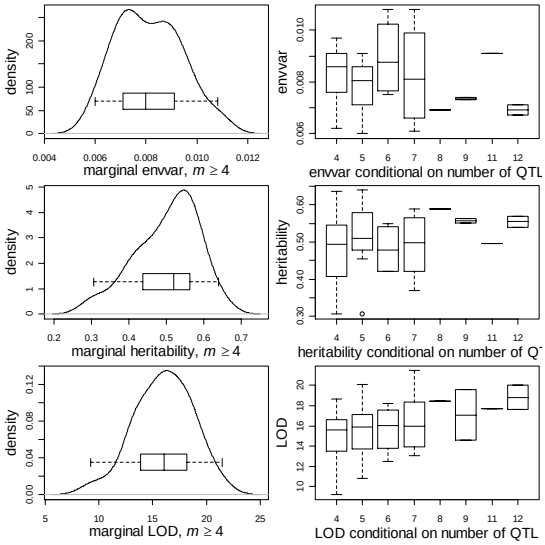


Bayesian model diagnostics

pattern: N2(2),N3,N16
col 1: density
col 2: boxplots by *m*

environmental variance
$\sigma^2 = .008$, $\sigma = .09$
heritability
$h^2 = 52\%$
LOD = 16
(highly significant)

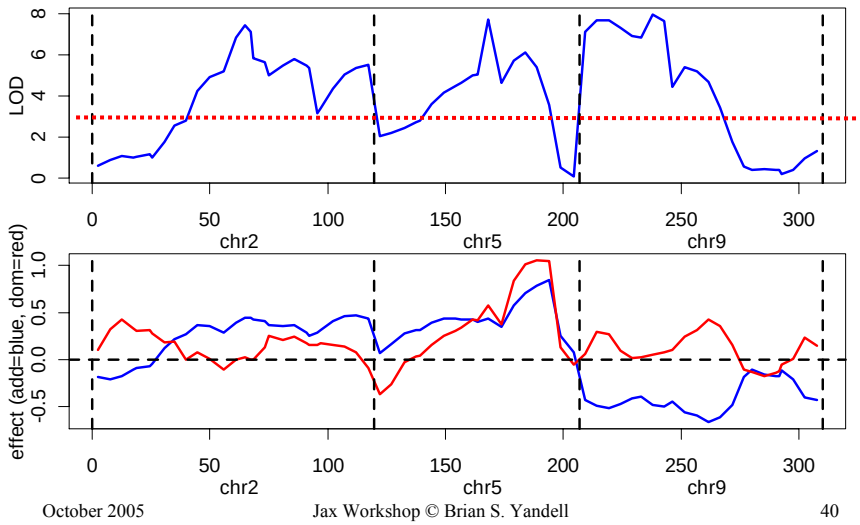
but note change with *m*



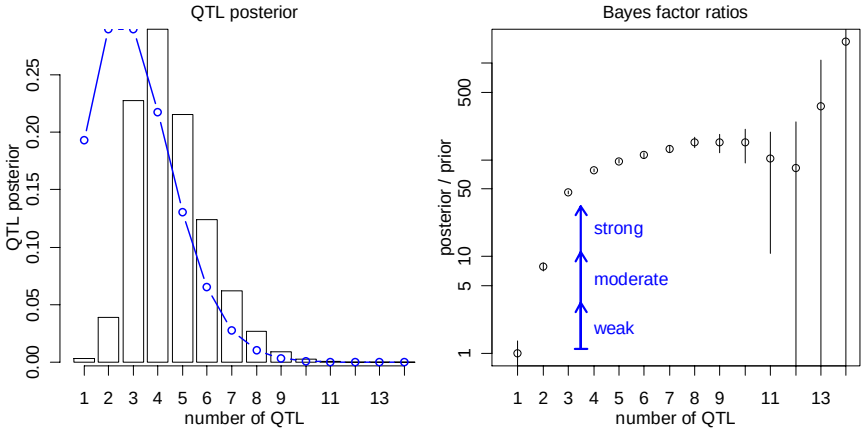
studying diabetes in an F2

- segregating cross of inbred lines
  - B6.ob x BTBR.ob → F1 → F2
  - selected mice with ob/ob alleles at leptin gene (chr 6)
  - measured and mapped body weight, insulin, glucose at various ages (Stoehr et al. 2000 Diabetes)
  - sacrificed at 14 weeks, tissues preserved
- gene expression data
  - Affymetrix microarrays on parental strains, F1
    - key tissues: adipose, liver, muscle, β-cells
    - novel discoveries of differential expression (Nadler et al. 2000 PNAS; Lan et al. 2002 in review; Ntambi et al. 2002 PNAS)
  - RT-PCR on 108 F2 mice liver tissues
    - 15 genes, selected as important in diabetes pathways
    - SCD1, PEPCK, ACO, FAS, GPAT, PPARgamma, PPARalpha, G6Pase, PDI,…

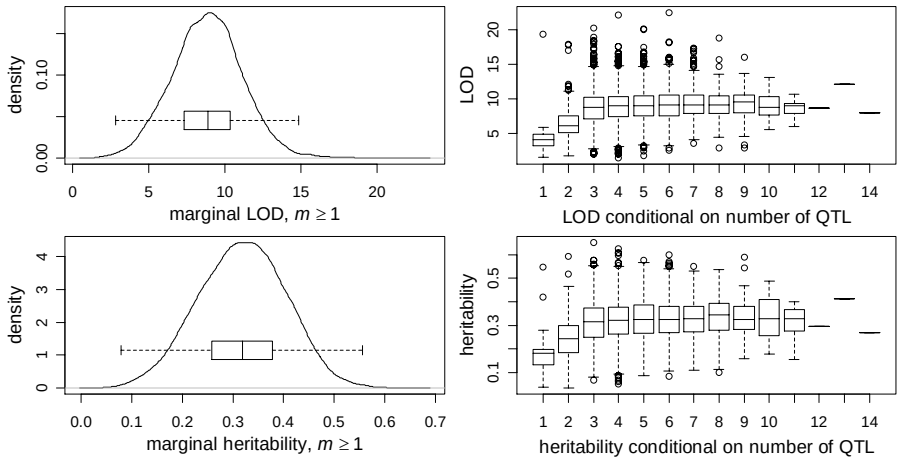
Multiple Interval Mapping (QTLCart)
SCD1: multiple QTL plus epistasis!



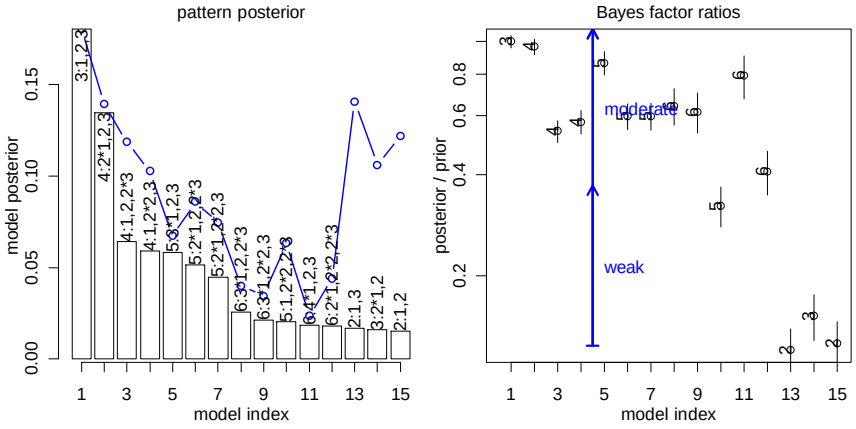
Bayesian model assessment: number of QTL for SCD1



Bayesian LOD and h^2 for SCD1

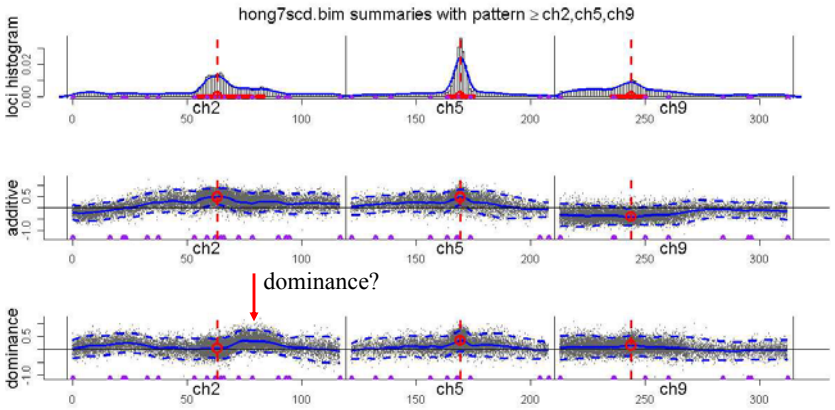


Bayesian model assessment: chromosome QTL pattern for SCD1

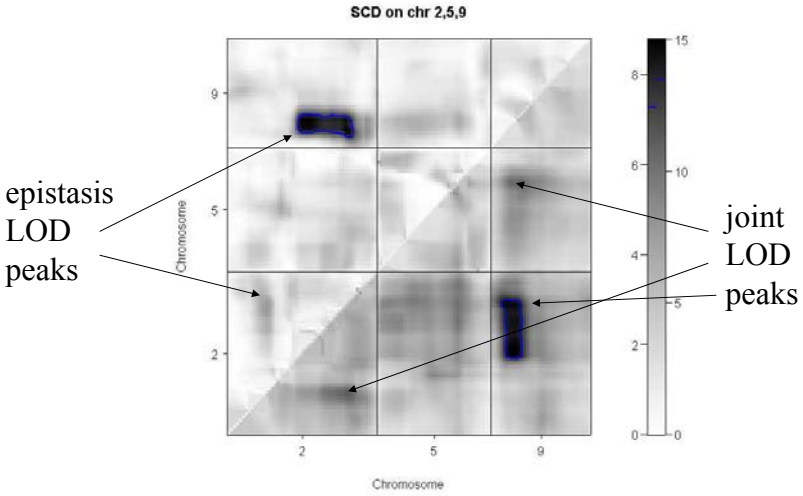


trans-acting QTL for SCD1

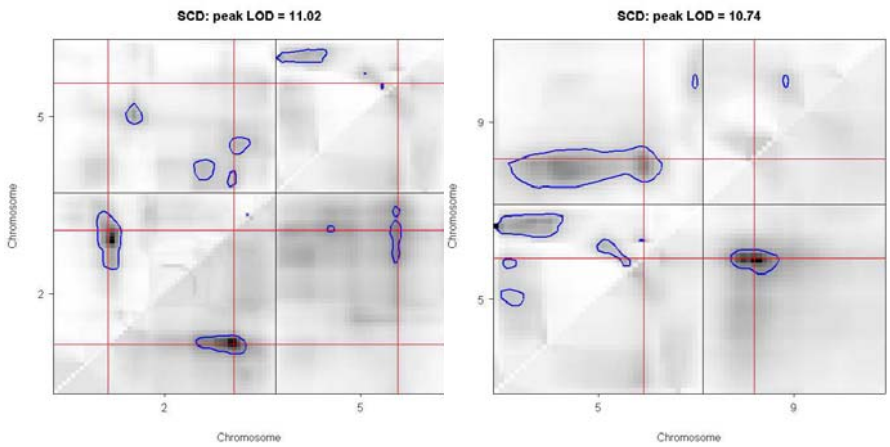
(no epistasis yet: see Yi, Xu, Allison 2003)



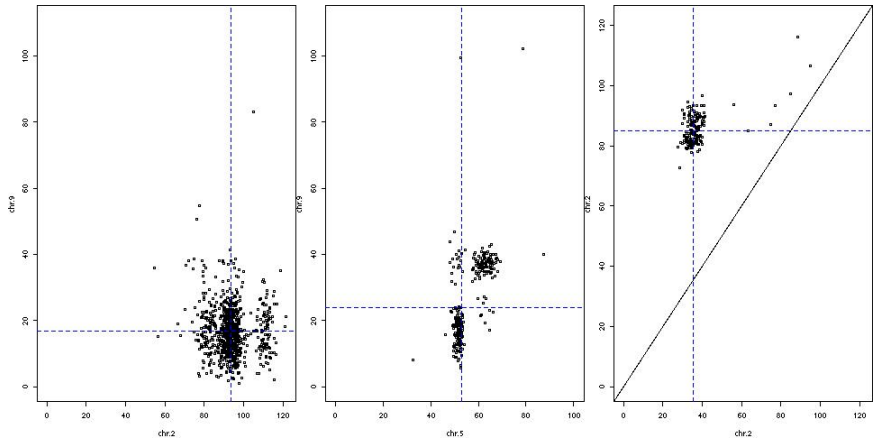
2-D scan: assumes only 2 QTL!



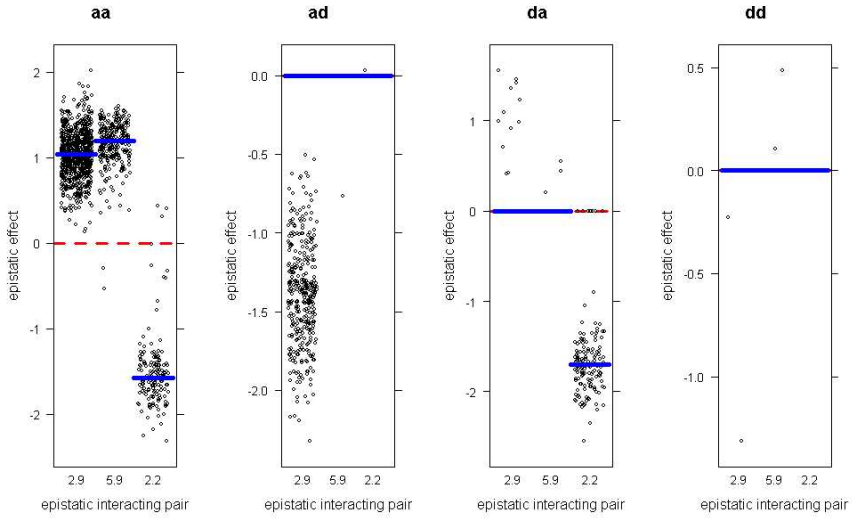
sub-peaks can be easily overlooked!



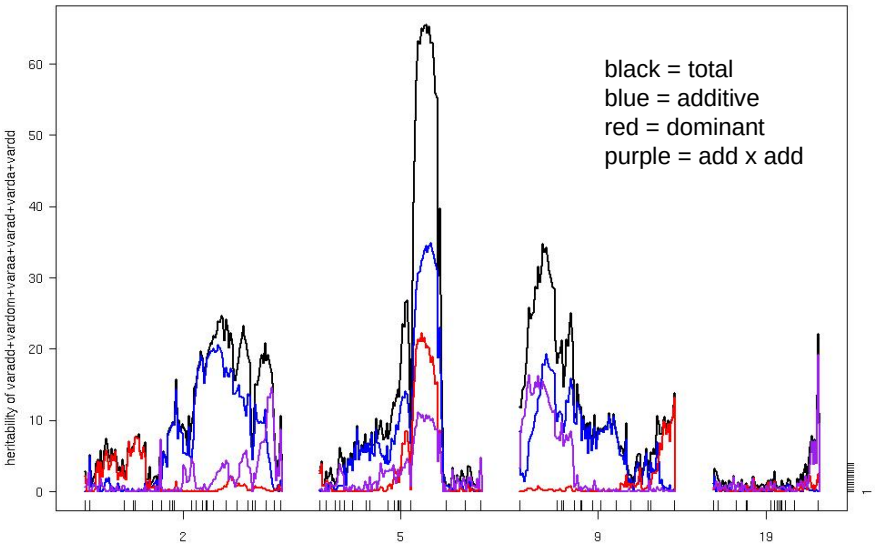
epistatic model fit



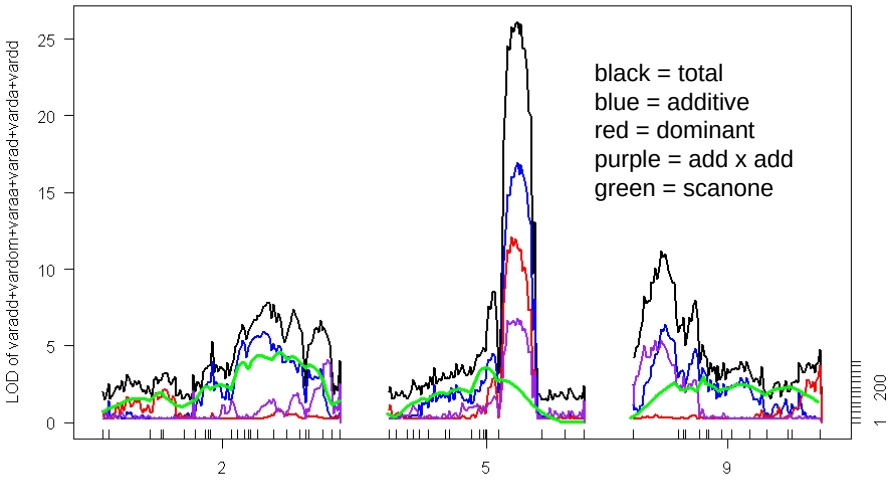
Cockerham epistatic effects



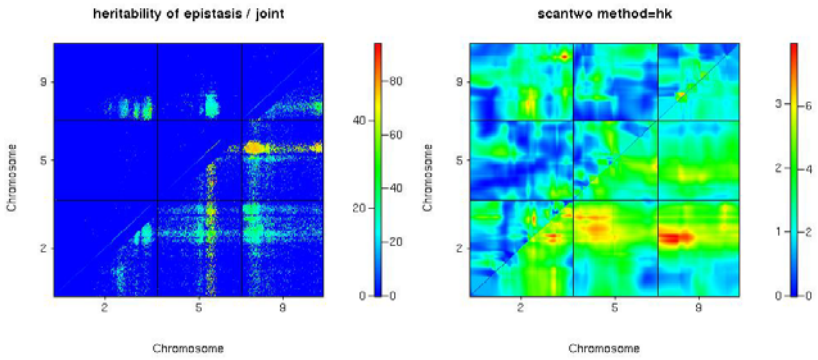
marginal heritability by locus



marginal LOD by locus



Bayesian & classical (HK)



sex-adjusted anova for SCD1

(only 3-way sex interactions significant)

```
Summary for fit QTL  Method is:  imp
Number of observations:  107

Full model result
----------------------------------
Model formula is:  y ~ sex * (Q1 + Q2 + Q3 + Q1:Q3 + Q2:Q3)

      df        SS       MS      LOD     %var Pvalue(Chi2)    Pvalue(F)
Model  29 109.02204 3.7593806 25.79574 67.05143 7.869261e-13 1.592058e-09
Error  77  53.57261 0.6957482
Total 106 162.59465


Drop one QTL at a time ANOVA table:
----------------------------------
                    df Type III SS    LOD   %var F value Pvalue
sex                 15      36.369 12.039 22.368   3.485 0.000154
Chr2@105            12      41.248 13.266 25.369   4.940 5.38e-06
Chr5@67             12      39.771 12.901 24.460   4.764 8.93e-06
Chr9@15             20      59.542 17.365 36.620   4.279 1.87e-06
Chr2@105:Chr9@15     8      33.637 11.322 20.687   6.043 5.11e-06
Chr5@67:Chr9@15      8      30.042 10.344 18.476   5.397 2.12e-05
sex:Chr2@105         6      18.499  6.892 11.377   4.431 0.000669
sex:Chr5@67          6      18.130  6.773 11.151   4.343 0.000793
sex:Chr9@15         10      25.945  9.176 15.957   3.729 0.000413
sex:Chr2@105:Chr9@15  4      17.444  6.549 10.729   6.268 0.000202
sex:Chr5@67:Chr9@15   4      15.728  5.981  9.673   5.652 0.000483
```

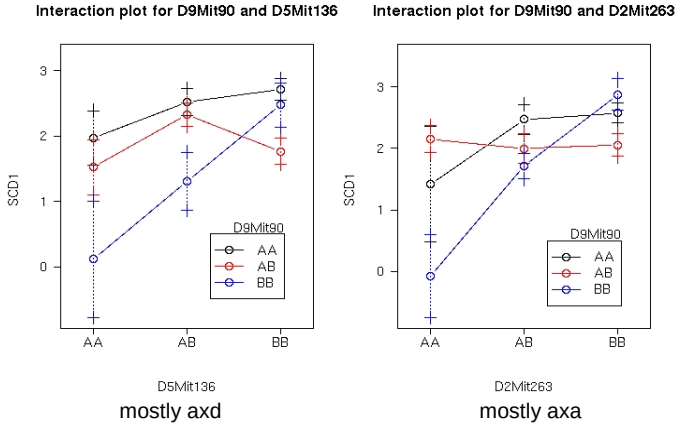
anova for SCD1

(sex terms n.s.)

```
Model:
y ~ Chr2@105 + Chr5@67 + Chr9@67 + Chr2@80 + Chr9@67^2
    + Chr2@105:Chr9@67 + Chr5@67:Chr9@67^2

       Df  Sum Sq Mean Sq F value    Pr(>F)
Model   7  69.060  69.060  73.095 < 2.2e-16 ***
Error  99  93.535   0.945
Total 106 162.595  70.005

Single term deletions
                Df Sum of Sq     RSS     AIC F value    Pr(F)
<none>                93.535  22.992
Chr2@80          1     9.073 102.608  28.225  9.6036 0.002528 **

Chr2@105         1     0.073  93.607  18.402  0.0769 0.782140
Chr5@67          1     0.218  93.753  18.568  0.2307 0.632084
Chr9@67          1     7.156 100.691  26.207  7.5746 0.007042 **
Chr9@67^2        1     0.106  93.641  18.440  0.1123 0.738200

Chr2@105:Chr9@67   1    15.612 109.147  34.836 16.5246 9.64e-05 ***
Chr5@67:Chr9@67^2  1     7.211 100.745  26.265  7.6318 0.006838 **
```

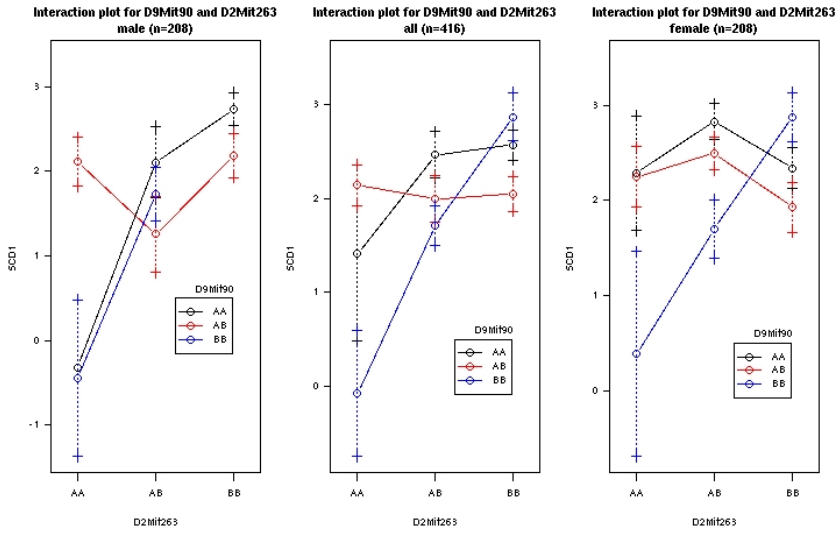

epistatic interaction plots

(chr 5x9 p=0.007; chr 2x9 p<0.0001)



mostly axd mostly axa

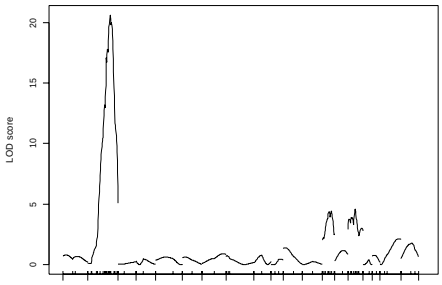
SCD1 on chr2 @ 0 by sex



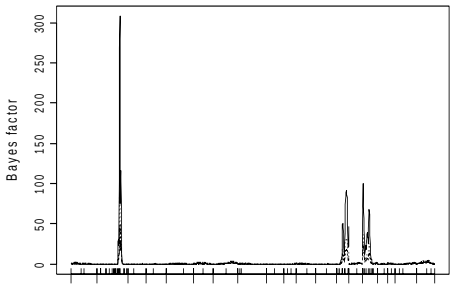
obesity in CAST/Ei BC onto M16i

- 421 mice (Daniel Pomp)
  - (213 male, 208 female)
- 92 microsatellites on 19 chromosomes
  - 1214 cM map
- subcutaneous fat pads
  - pre-adjusted for sex and dam effects
- Yi, Yandell, Churchill, Allison, Eisen, Pomp (2005) *Genetics* (in press)

non-epistatic analysis

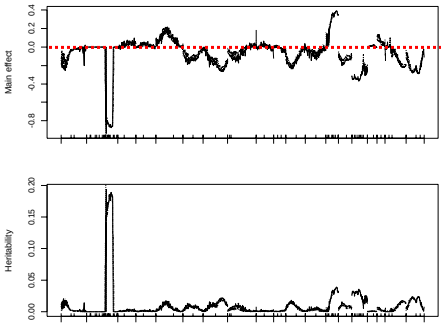


single QTL LOD profile

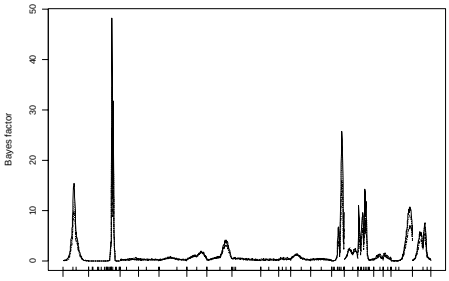


multiple QTL
Bayes factor profile

posterior profile of main effects in epistatic analysis

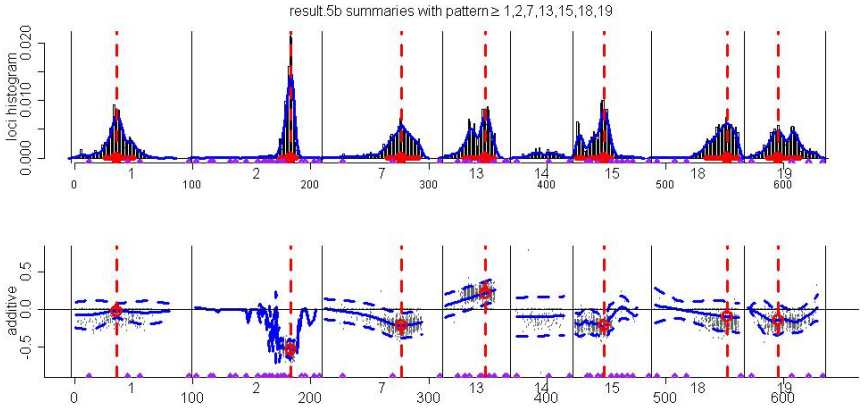


main effects & heritability profile

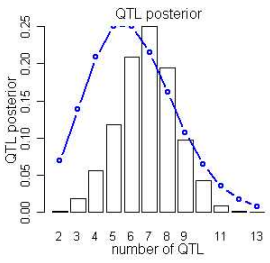


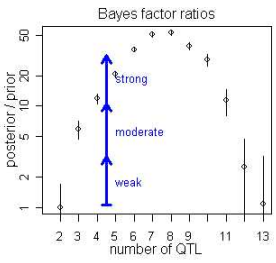
Bayes factor profile

posterior profile of main effects in epistatic analysis



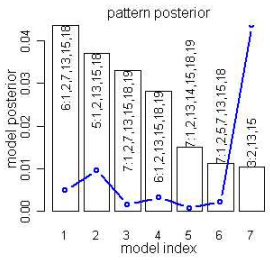
model selection via Bayes factors for epistatic model

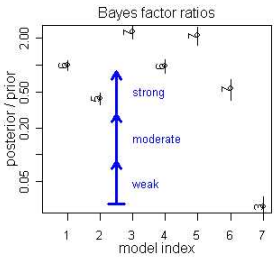




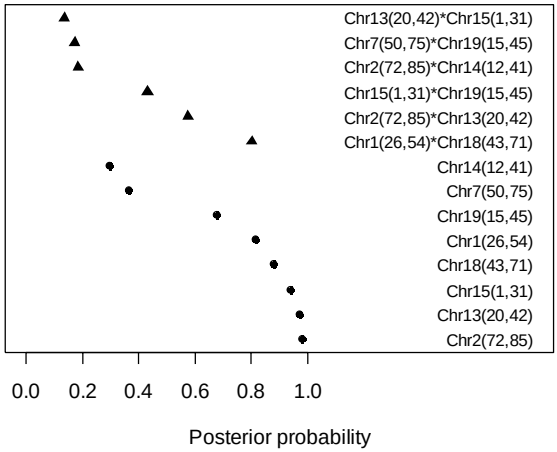
number of QTL

QTL pattern

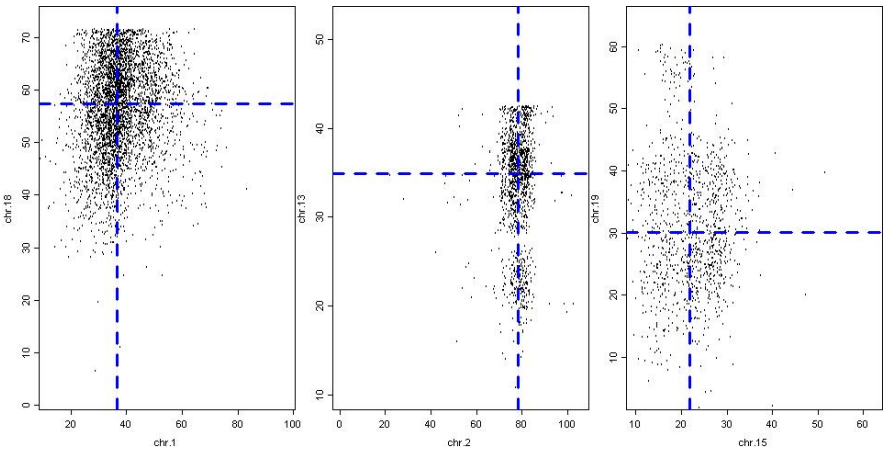




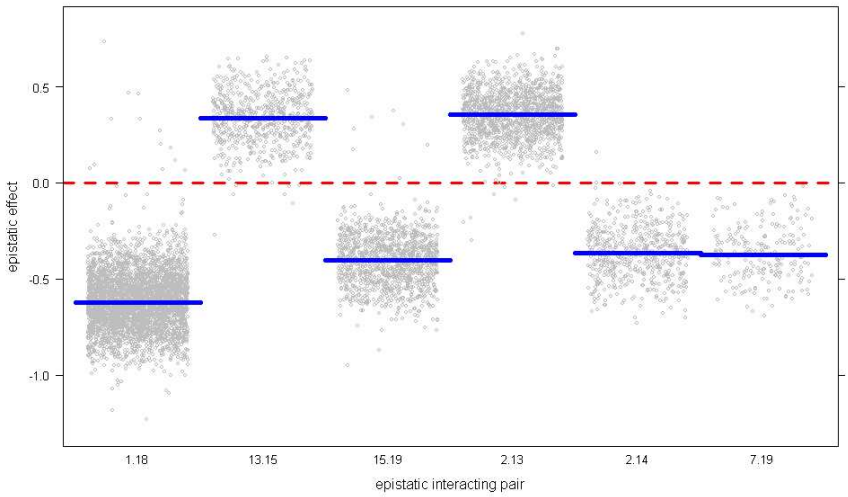
posterior probability of effects



scatterplot estimates of epistatic loci



stronger epistatic effects



Bayesian software for QTLs

- R/bim: Bayesian IM (Satagopan Yandell 1996; Gaffney 2001)*
  - www.stat.wisc.edu/~yandell/qtl/software/Bmapqtl
  - www.r-project.org contributed package to CRAN
  - version available within WinQTLCart (statgen.ncsu.edu/qtlcart)
- R/bmqtl: Bayesian Multiple QTL (N Yi, T Mehta & BS Yandell)*
  - epistasis, new MCMC algorithms, extensive graphics
  - extension of R/bim; due out Fall 2005
- R/qtl (Broman et al. 2003 Bioinformatics)*
  - biosun01.biostat.jhsph.edu/~kbroman/software
  - www.r-project.org contributed package
- Pseudomarker (Sen, Churchill 2002 Genetics)
  - www.jax.org/staff/churchill/labsite/software
- Bayesian QTL / Multimapper
  - Sillanpää Arjas (1998)
  - www.rni.helsinki.fi/~mjs
- Stephens & Fisch (1998 Biometrics)
- R/bqtl (C Berry, hacuna.ucsd.edu/bqtl)

* Hao Wu, Jackson Labs, provide crucial computing support

R/bim: our software

- www.stat.wisc.edu/~yandell/qtl/software/Bmapqtl
  - R contributed library (www.r-project.org)
    - library(bim) is cross-compatible with library(qtl)
  - Bayesian module within WinQTLCart
    - WinQTLCart output can be processed using R library
- Software history
  - initially designed by JM Satagopan (1996)
  - major revision and extension by PJ Gaffney (2001)
    - whole genome
    - multivariate update of effects; long range position updates
    - substantial improvements in speed, efficiency
    - pre-burnin: initial prior number of QTL very large
  - upgrade (H Wu, PJ Gaffney, CF Jin, BS Yandell 2003)
  - epistasis in progress (H Wu, BS Yandell, N Yi 2004)

many thanks

Jackson Labs
Gary Churchill
Hao Wu

U AL Birmingham
David Allison
Nengjun Yi
Tapan Mehta

Tom Osborn
David Butruille
Marcio Ferrera
Josh Udahl
Pablo Quijada

Alan Attie
Jonathan Stoehr
Hong Lan
Susie Clee
Jessica Byers

Michael Newton
Daniel Sorensen
Daniel Gianola
Liang Li

my students
Jaya Satagopan
Fei Zou
Patrick Gaffney
Chunfang Jin
Elias Chaibub Neto

USDA Hatch, NIH/NIDDK (Attie), NIH/R01 (Yi)