Bayesian Model Selection for Multiple QTL

Jackson Laboratory, October 2009
Brian S. Yandell, UW-Madison
www.stat.wisc.edu/~yandell/statgen

Real knowledge is to know the extent of one's ignorance.
Confucius (on a bench in Seattle)

outline

1. What is the goal of QTL study?
2. Bayesian vs. classical QTL study
3. Bayesian strategy for QTLs
4. model search using MCMC
5. model assessment
6. analysis of hyper data
7. software for Bayesian QTLs

1. what is the goal of QTL study?

- uncover underlying biochemistry
  - identify how networks function, break down
  - find useful candidates for (medical) intervention
  - epistasis may play key role
  - statistical goal: maximize number of correctly identified QTL
- basic science/evolution
  - how is the genome organized?
  - identify units of natural selection
  - additive effects may be most important (Wright/Fisher debate)
  - statistical goal: maximize number of correctly identified QTL
- select “elite” individuals
  - predict phenotype (breeding value) using suite of characteristics (phenotypes) translated into a few QTL
  - statistical goal: mimimize prediction error

problems of single QTL approach

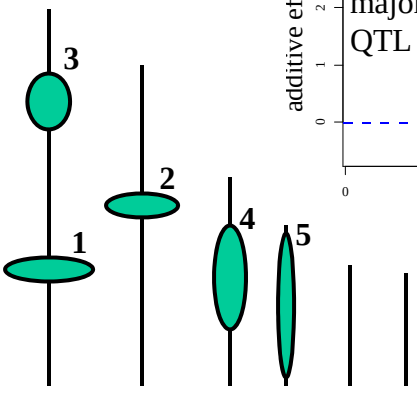
- wrong model: biased view
  - fool yourself: bad guess at locations, effects
  - detect ghost QTL between linked loci
  - miss epistasis completely
- low power
- bad science
  - use best tools for the job
  - maximize scarce research resources
  - leverage already big investment in experiment

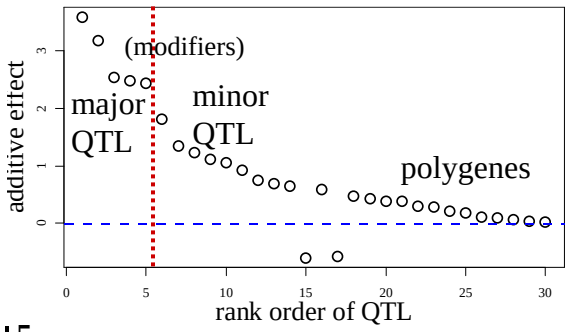
advantages of multiple QTL approach

- improve statistical power, precision
  - increase number of QTL detected
  - better estimates of loci: less bias, smaller intervals
- improve inference of complex genetic architecture
  - patterns and individual elements of epistasis
  - appropriate estimates of means, variances, covariances
    - asymptotically unbiased, efficient
  - assess relative contributions of different QTL
- improve estimates of genotypic values
  - less bias (more accurate) and smaller variance (more precise)
  - mean squared error = MSE = $(\text{bias})^2$ + variance

Pareto diagram of QTL effects

major QTL on linkage map





limits of multiple QTL?

- limits of statistical inference
  - power depends on sample size, heritability, environmental variation
  - "best" model balances fit to data and complexity (model size)
  - genetic linkage = correlated estimates of gene effects
- limits of biological utility
  - sampling: only see some patterns with many QTL
  - marker assisted selection (Bernardo 2001 *Crop Sci*)
    - 10 QTL ok, 50 QTL are too many
    - phenotype better predictor than genotype when too many QTL
    - increasing sample size may not give multiple QTL any advantage
  - hard to select many QTL simultaneously

QTL below detection level?

- problem of selection bias
  - QTL of modest effect only detected sometimes
  - effects overestimated when detected
  - repeat studies may fail to detect these QTL
- think of probability of detecting QTL
  - avoids sharp in/out dichotomy
  - avoid pitfalls of one "best" model
  - examine "better" models with more probable QTL
- rethink formal approach for QTL
  - directly allow uncertainty in genetic architecture
  - QTL model selection over genetic architecture

check QTL in context of genetic architecture

- scan for each QTL adjusting for all others
  - adjust for linked and unlinked QTL
    - adjust for linked QTL: reduce bias
    - adjust for unlinked QTL: reduce variance
  - adjust for environment/covariates
- examine entire genetic architecture
  - number and location of QTL, epistasis, GxE
  - model selection for best genetic architecture

2. Bayesian vs. classical QTL study

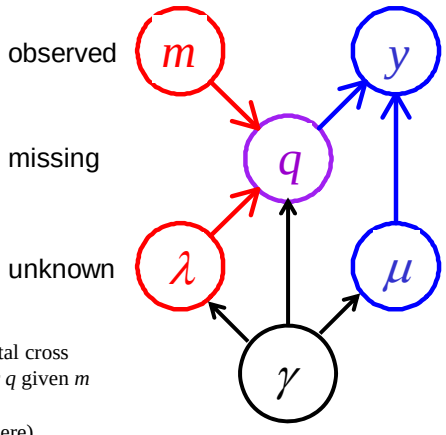
- classical study
  - ***maximize*** over unknown effects
  - ***test*** for detection of QTL at loci
  - model selection in stepwise fashion
- Bayesian study
  - ***average*** over unknown effects
  - ***estimate*** chance of detecting QTL
  - sample all possible models
- both approaches
  - average over missing QTL genotypes
  - scan over possible loci

Bayesian idea

- Reverend Thomas Bayes (1702-1761)
  - part-time mathematician
  - buried in Bunhill Cemetary, Moongate, London
  - famous paper in 1763 *Phil Trans Roy Soc London*
  - was Bayes the first with this idea? (Laplace?)
- basic idea (from Bayes' original example)
  - two billiard balls tossed at random (uniform) on table
  - where is first ball if the second is to its left?
    - prior: anywhere on the table
    - posterior: more likely toward right end of table

QTL model selection: key players

- 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
  - γ = QTL model/genetic architecture
- pr($q|m,\lambda,\gamma$) genotype model
  - grounded by linkage map, experimental cross
  - recombination yields multinomial for q given m
- pr($y|q,\mu,\gamma$) phenotype model
  - distribution shape (assumed normal here)
  - unknown parameters μ (could be non-parametric)



after
Sen Churchill (2001)

likelihood and posterior

- likelihood relates "known" data (y,m,q) to unknown values of interest (μ,λ,γ)
  - pr(y,q|m,μ,λ,γ) = pr(y|q,μ,γ) pr(q|m,λ,γ)
  - mix over unknown genotypes (q)
- posterior turns likelihood into a distribution
  - weight likelihood by priors
  - rescale to sum to 1.0
  - posterior = likelihood * prior / constant

Bayes posterior vs. maximum likelihood

- LOD: classical Log ODds
  - maximize likelihood over effects μ
  - `R/qtl scanone/scantwo: method = "em"`
- *LPD*: Bayesian *L*og *P*osterior *D*ensity
  - average posterior over effects μ
  - `R/qtl scanone/scantwo: method = "imp"`

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

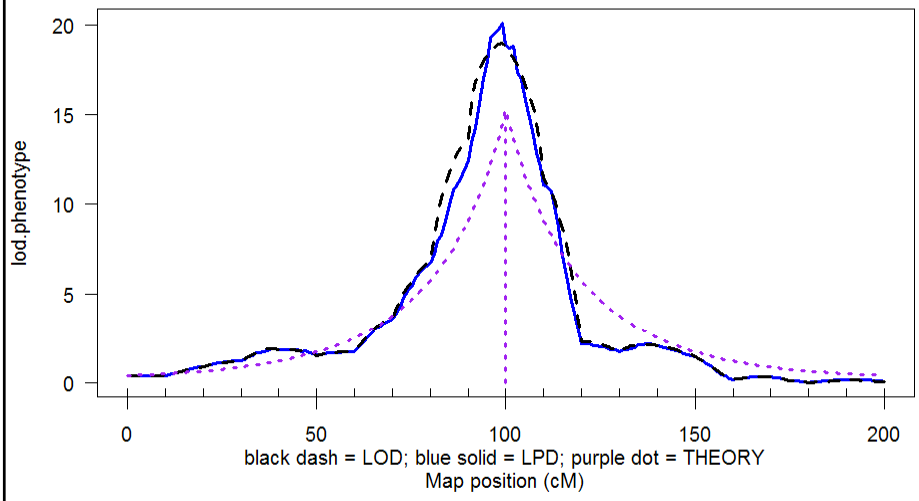
$$\mathrm{LPD}(\lambda) = \log_{10}\{\mathrm{pr}(\lambda \mid m)\int \mathrm{pr}(y \mid m, \mu, \lambda)\mathrm{pr}(\mu)d\mu\} + C$$

likelihood mixes over missing QTL genotypes:

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

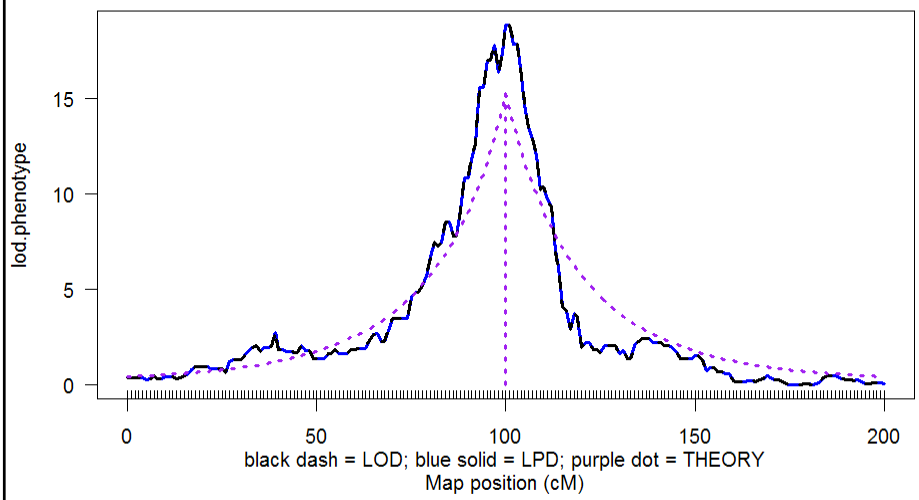
LOD & LPD: 1 QTL

n.ind = 100, 10 cM marker spacing



LOD & LPD: 1 QTL

n.ind = 100, 1 cM marker spacing



marginal LOD or LPD

- What is contribution of a QTL adjusting for all others?
  - improvement in LPD due to QTL at locus λ
  - contribution due to main effects, epistasis, GxE?
- How does adjusted LPD *differ* from unadjusted LPD?
  - raised by removing variance due to unlinked QTL
  - raised or lowered due to bias of linked QTL
  - analogous to Type III adjusted ANOVA tests
- can ask these same questions using classical LOD
  - see Broman's newer tools for multiple QTL inference

marginal LOD or LPD

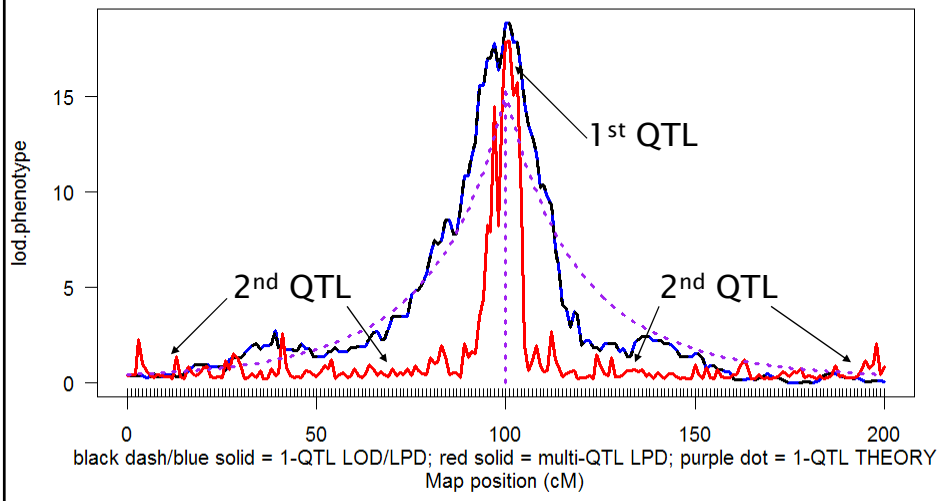
- compare two genetic architectures (γ_2,γ_1) at each locus
  - with (γ_2) or without (γ_1) another QTL at locus λ
    - preserve model hierarchy (e.g. drop any epistasis with QTL at λ)
  - with (γ_2) or without (γ_1) epistasis with QTL at locus λ
  - γ_2 contains γ_1 as a sub-architecture
- allow for multiple QTL besides locus being scanned
  - architectures γ_1 and γ_2 may have QTL at several other loci
  - use marginal LOD, LPD or other diagnostic
  - posterior, Bayes factor, heritability

$$\mathrm{LOD}(\lambda \mid \gamma_2) - \mathrm{LOD}(\lambda \mid \gamma_1)$$

$$\mathrm{LPD}(\lambda \mid \gamma_2) - \mathrm{LPD}(\lambda \mid \gamma_1)$$

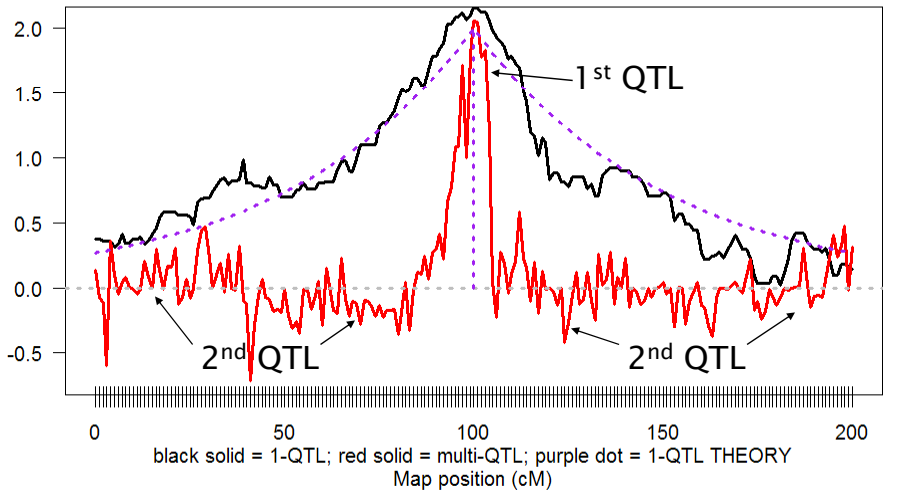
LPD: 1 QTL vs. multi-QTL

marginal contribution to LPD from QTL at λ



substitution effect: 1 QTL vs. multi-QTL

single QTL effect vs. marginal effect from QTL at λ



why use a Bayesian approach?

- first, do *both* classical and Bayesian
  - always nice to have a separate validation
  - each approach has its strengths and weaknesses
- classical approach works quite well
  - selects large effect QTL easily
  - directly builds on regression ideas for model selection
- Bayesian approach is comprehensive
  - samples most probable genetic architectures
  - formalizes model selection within one framework
  - readily (!) extends to more complicated problems

3. Bayesian strategy for QTL study

- augment data (y,m) with missing genotypes q
- study unknowns (μ,λ,γ) given augmented data (y,m,q)
  - find better genetic architectures γ
  - find most likely genomic regions = QTL = λ
  - estimate phenotype parameters = genotype means = μ
- sample from posterior in some clever way
  - multiple imputation (Sen Churchill 2002)
  - Markov chain Monte Carlo (MCMC)
    - (Satagopan et al. 1996; Yi et al. 2005, 2007)

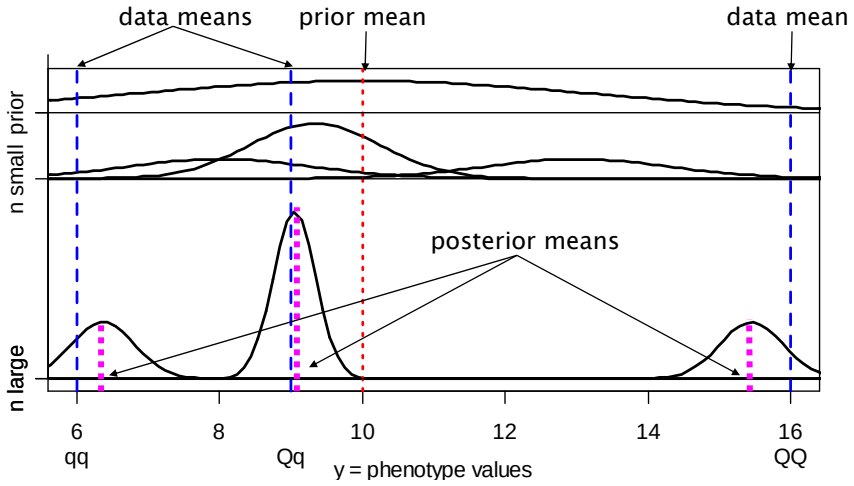
$$\text{posterior} = \frac{\text{likelihood} * \text{prior}}{\text{constant}}$$

$$\text{posterior for } q,\mu,\lambda,\gamma = \frac{\text{phenotype likelihood} * [\text{prior for } q,\mu,\lambda,\gamma]}{\text{constant}}$$

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

what values are the genotypic means?

phenotype model pr($y|q,\mu$)



Bayes posterior QTL means

posterior centered on sample genotypic mean
but shrunken slightly toward overall mean

phenotype mean: $E(y \mid q) = \mu_q \qquad V(y \mid q) = \sigma^2$

genotypic prior: $E(\mu_q) = \bar{y}_{\bullet} \qquad V(\mu_q) = \kappa\sigma^2$

posterior: $E(\mu_q \mid y) = b_q \bar{y}_q + (1-b_q)\bar{y}_{\bullet} \qquad V(\mu_q \mid y) = b_q \sigma^2 / n_q$

$n_q = \text{count}\{q_i = q\} \qquad \bar{y}_q = \underset{\{q_i=q\}}{\text{sum}}\, y_i / n_q$

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

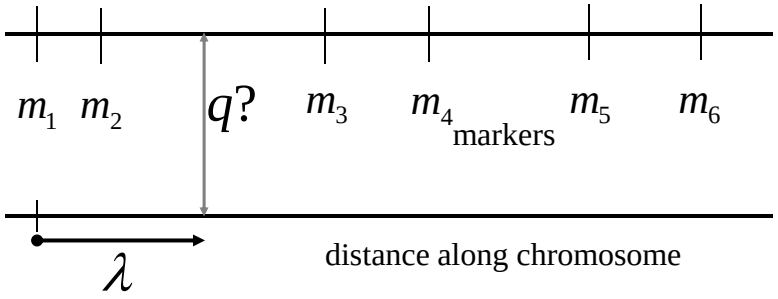
partition genotypic effects on phenotype

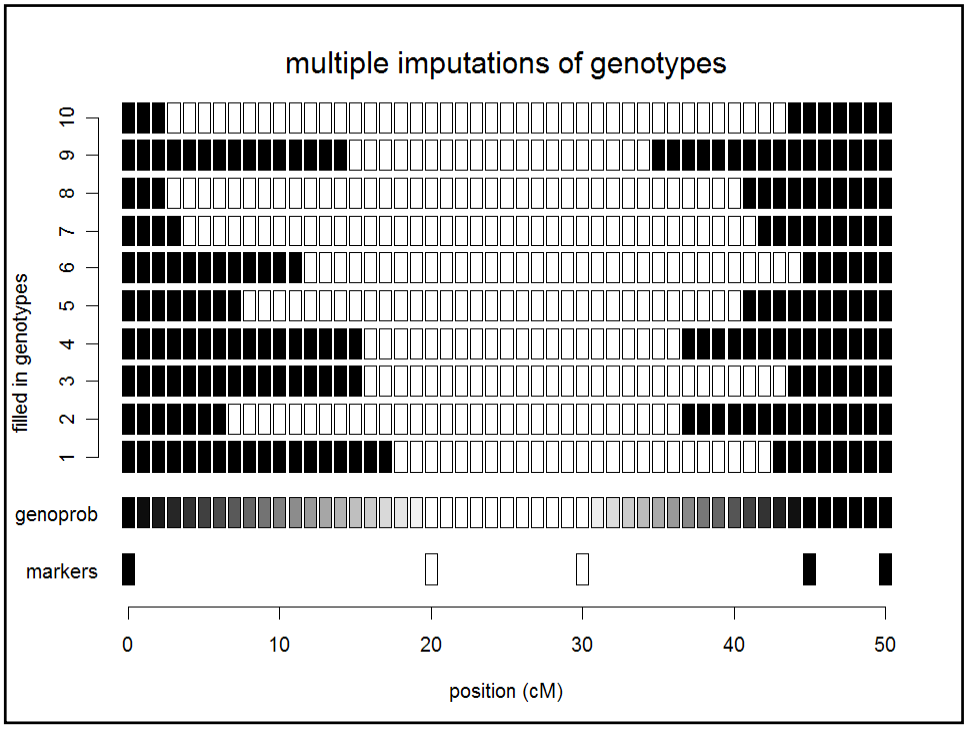
- phenotype depends on genotype
- genotypic value partitioned into
  - main effects of single QTL
  - epistasis (interaction) between pairs of QTL

$$\mu_q = \beta_0 + \beta_q = E(Y;q)$$
$$\beta_q = \beta(q_2) + \beta(q_2) + \beta(q_1, q_2)$$

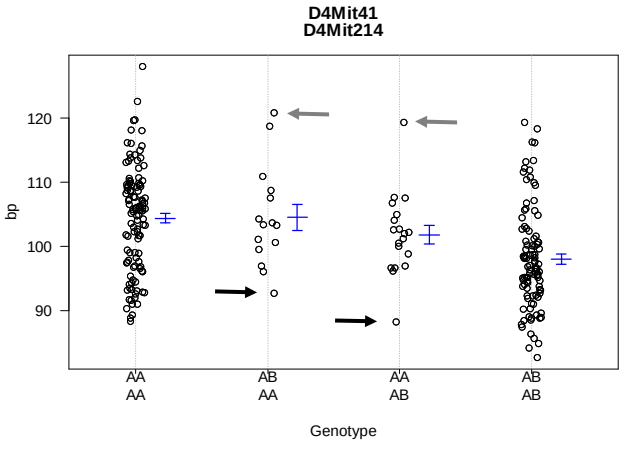
pr($q|m,\lambda$) recombination model

pr($q|m,\lambda$) = pr(geno | map, locus) ≈
pr(geno | flanking markers, locus)





how does phenotype *y* improve guess of QTL genotypes *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 q

- full conditional of q given data, parameters
  - proportional to prior pr(q | m, λ)
    - weight toward q that agrees with flanking markers
  - proportional to likelihood pr(y | q, μ)
    - weight toward q with similar phenotype values
  - posterior recombination model balances these two
- this *is* the E-step of EM computations

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

what is the genetic architecture γ?

- which positions correspond to QTLs?
  - priors on loci (previous slide)
- which QTL have main effects?
  - priors for presence/absence of main effects
    - same prior for all QTL
    - can put prior on each d.f. (1 for BC, 2 for F2)
- which pairs of QTL have epistatic interactions?
  - prior for presence/absence of epistatic pairs
    - depends on whether 0,1,2 QTL have main effects
    - epistatic effects less probable than main effects

γ = genetic architecture:

loci:

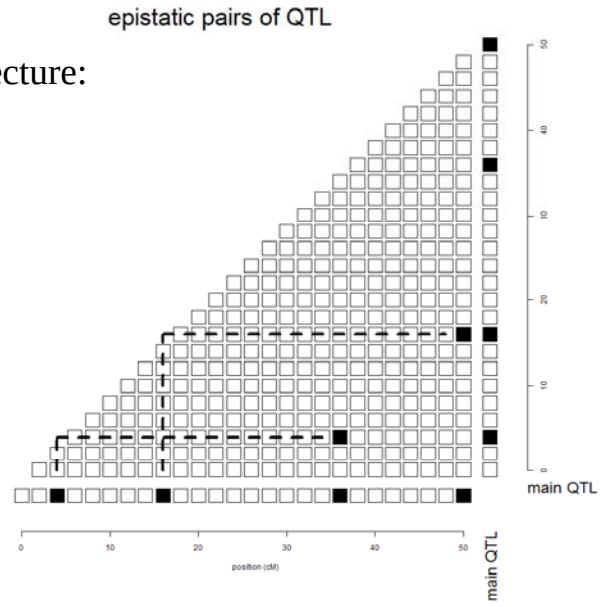
main QTL

epistatic pairs

effects:

add, dom

aa, ad, dd



4. Markov chain sampling

- 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
- sample QTL model components from full conditionals
  - sample locus λ given q,γ (using Metropolis-Hastings step)
  - sample genotypes q given λ,μ,y,γ (using Gibbs sampler)
  - sample effects μ given q,y,γ (using Gibbs sampler)
  - sample QTL model γ given λ,μ,y,q (using Gibbs or M-H)

$$(\lambda,q,\mu,\gamma) \sim \text{pr}(\lambda,q,\mu,\gamma \mid y,m)$$

$$(\lambda,q,\mu,\gamma)_1 \to (\lambda,q,\mu,\gamma)_2 \to \cdots \to (\lambda,q,\mu,\gamma)_N$$

MCMC sampling of unknowns (μ,q,λ)

for given genetic architecture γ

$$\mu \sim \frac{\text{pr}(y \mid q, \mu)\text{pr}(\mu)}{\text{pr}(y \mid q)}$$

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

$$\lambda \sim \frac{\text{pr}(q \mid m, \lambda)\text{pr}(\lambda \mid m)}{\text{pr}(q \mid m)}$$

Gibbs sampler

for two genotypic means

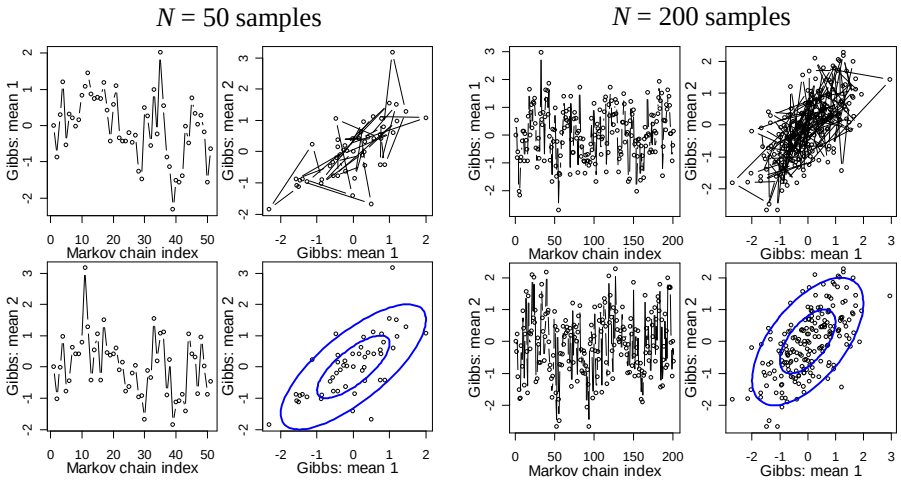
- want to study two correlated effects β_1, β_2
  - assume correlation ρ is known
- sample from full distribution?
- or use Gibbs sampler:
  - sample each effect from its full conditional given the other
  - pick order of sampling at random
  - repeat many times

$$\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$$

$$\beta_1 \sim N\left(\rho\beta_2, 1-\rho^2\right)$$

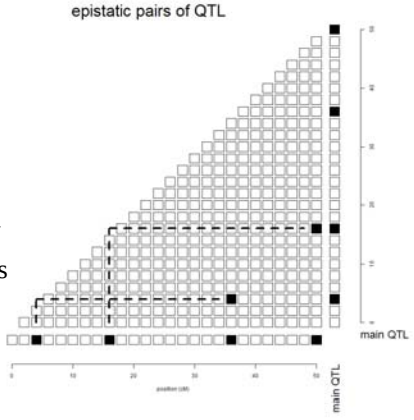
$$\beta_2 \sim N\left(\rho\beta_1, 1-\rho^2\right)$$

Gibbs sampler samples: $\rho = 0.6$



Gibbs sampler for loci indicators

- QTL at pseudomarkers
- loci indicators γ
  - $\gamma = 1$ if QTL present
  - $\gamma = 0$ if no QTL present
- Gibbs sampler on loci indicators γ
  - relatively easy to incorporate epistasis
  - Yi *et al.* (2005 *Genetics*)
    - (earlier work of Yi, Ina Hoeschele)



$$\mu_q = \mu + \gamma_1 \beta(q_1) + \gamma_2 \beta(q_2), \quad \gamma_k = 0,1$$

epistatic interactions

- model space issues
  - partition QTL effects? (additive, dominance, etc.)
  - 2-QTL interactions only?
  - 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
    - pairs with one significant QTL?
    - pairs of non-significant QTL?
- Yi et al. (2005, 2007)

5. 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 likelihood by model size
  - compare IC = $-2 \log L(\text{ model} \mid \text{data}) + \text{penalty(model size)}$
- Bayes factors: balance posterior by prior choice
  - compare pr(data | model)

Bayes factors

- ratio of model likelihoods
  - ratio of posterior to prior odds for architectures
  - average over unknown effects (μ) and loci (λ)

$$BF = \frac{\text{pr}(\text{data} \mid \text{model } \gamma_1)}{\text{pr}(\text{data} \mid \text{model } \gamma_2)}$$

- roughly equivalent to BIC
  - BIC maximizes over unknowns
  - BF averages over unknowns

$$2\log_{10}(BF) = 2LOD + (\text{change in model size})\log_{10}(n)$$

marginal BF scan by QTL

- compare models with and without QTL at λ
  - find frequency of MCMC samples with (without) λ
    - averages over all models with (without) QTL at λ
  - BF = ratio of frequencies with and without QTL at λ
- scan over genome for peaks
  - 2log(BF) has similar behavior to LPD

$$BF_\lambda = \frac{\text{pr}(y \mid m, \text{model with } \lambda)}{\text{pr}(y \mid m, \text{model without } \lambda)}$$

6. analysis of hyper data

- marginal scans of genome
  - detect significant loci
  - infer main and epistatic QTL, GxE
- infer most probable genetic architecture
  - number of QTL
  - chromosome pattern of QTL with epistasis
- diagnostic summaries
  - heritability, unexplained variation

R/qtlbim: tutorial

(www.stat.wisc.edu/~yandell/qtlbim)

```
> data(hyper)
## Drop X chromosome (for now).
> hyper <- subset(hyper, chr=1:19)
> hyper <- qb.genoprob(hyper, step=2)

## This is the time-consuming step:
> qbHyper <- qb.mcmc(hyper, pheno.col = 1)

## Here we get pre-stored samples.
> data(qbHyper)

## Summary printing and plots
> summary(qbHyper)
> plot(qbHyper)
```

R/qtlbim: initial summaries

```
> summary(qbHyper)

Bayesian model selection QTL mapping object qbHyper on cross object hyper
had 3000 iterations recorded at each 40 steps with 1200 burn-in steps.

Diagnostic summaries:
         nqtl   mean envvar varadd  varaa    var
Min.     2.000  97.42  28.07  5.112  0.000  5.112
1st Qu.  5.000 101.00  44.33 17.010  1.639 20.180
Median   7.000 101.30  48.57 20.060  4.580 25.160
Mean     6.543 101.30  48.80 20.310  5.321 25.630
3rd Qu.  8.000 101.70  53.11 23.480  7.862 30.370
Max.    13.000 103.90  74.03 51.730 34.940 65.220

Percentages for number of QTL detected:
 2  3  4  5  6  7  8  9 10 11 12 13
 2  3  9 14 21 19 17 10  4  1  0  0

Percentages for number of epistatic pairs detected:
pairs
 1  2  3  4  5  6
29 31 23 11  5  1

Percentages for common epistatic pairs:
 6.15  4.15   4.6   1.7 15.15   1.4   1.6   4.9  1.15  1.17   1.5  5.11   1.2  7.15   1.1
   63    18    10     6     6     5     4     4     3     3     3     2     2     2     2

> plot(qb.diag(qbHyper, items = c("herit", "envvar")))
```

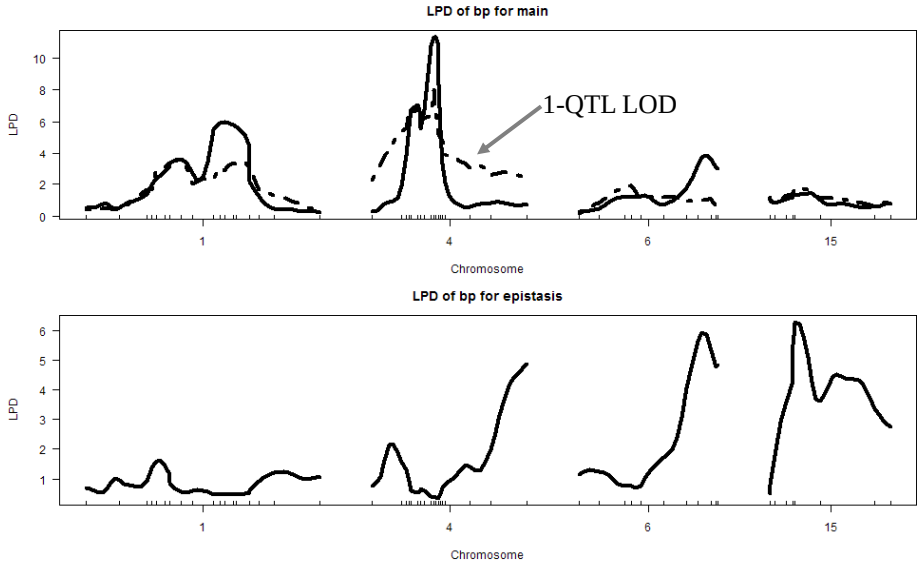

marginal scans of genome

- LPD and 2log(BF) “tests” for each locus
- estimates of QTL effects at each locus
- separately infer main effects and epistasis
  - main effect for each locus (blue)
  - epistasis for loci paired with another (purple)
    - identify epistatic QTL in 1-D scan
    - infer pairing in 2-D scan

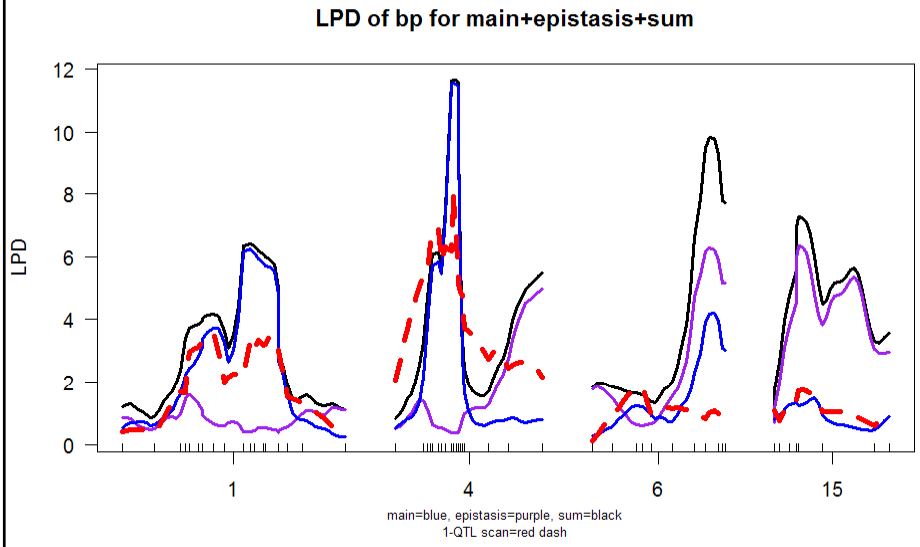
R/qtlbim: 1-D (*not* 1-QTL!) scan

```
> one <- qb.scanone(qbHyper, chr = c(1,4,6,15), type = "LPD")
> summary(one)

LPD of bp for main,epistasis,sum

    n.qtl  pos m.pos e.pos  main epistasis   sum
c1  1.331 64.5  64.5  67.8  6.10     0.442  6.27
c4  1.377 29.5  29.5  29.5 11.49     0.375 11.61
c6  0.838 59.0  59.0  59.0  3.99     6.265  9.60
c15 0.961 17.5  17.5  17.5  1.30     6.325  7.28

> plot(one, scan = "main")
> plot(out.em, chr=c(1,4,6,15), add = TRUE, lty = 2)
> plot(one, scan = "epistasis")
```


1-QTL LOD vs. marginal LPD

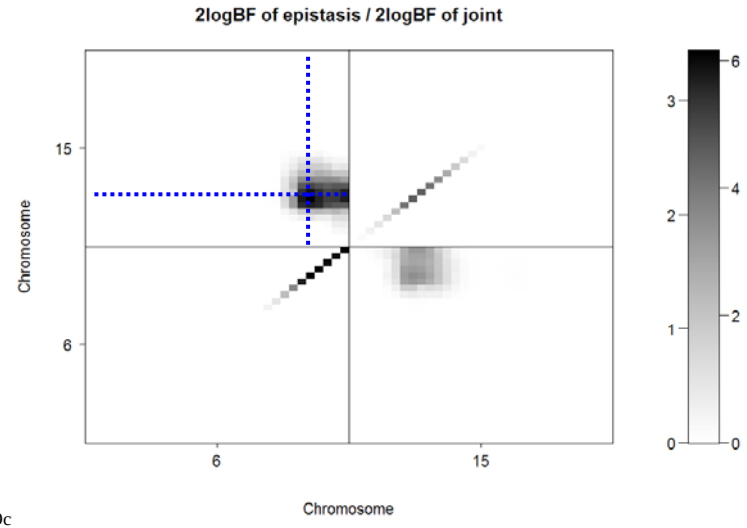


hyper data: scanone



2-D plot of 2logBF: chr 6 & 15

> plot(qb.scantwo(qbHyper, chr = c(6,16), type = "2logBF")



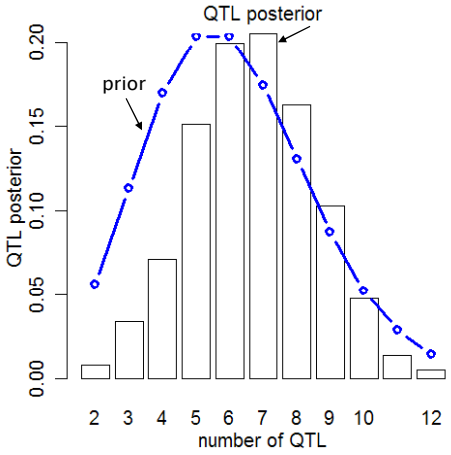
Bayes Factor ratios

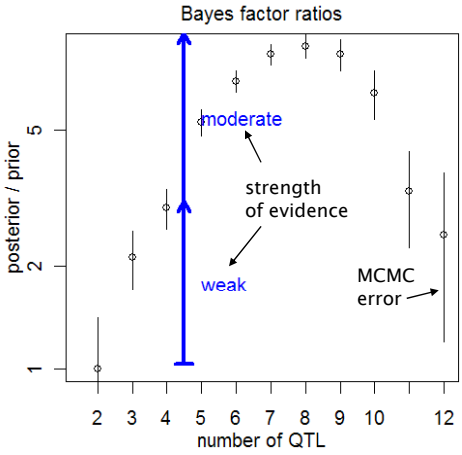
- BF = ratios of pr(data|model)
  - pr(data|model) = pr(model|data) / pr(model)
  - use ruler on log scale to compare models
- BF for quantities of interest
  - how many QTL?
  - what is pattern across chromosomes?

most probable patterns

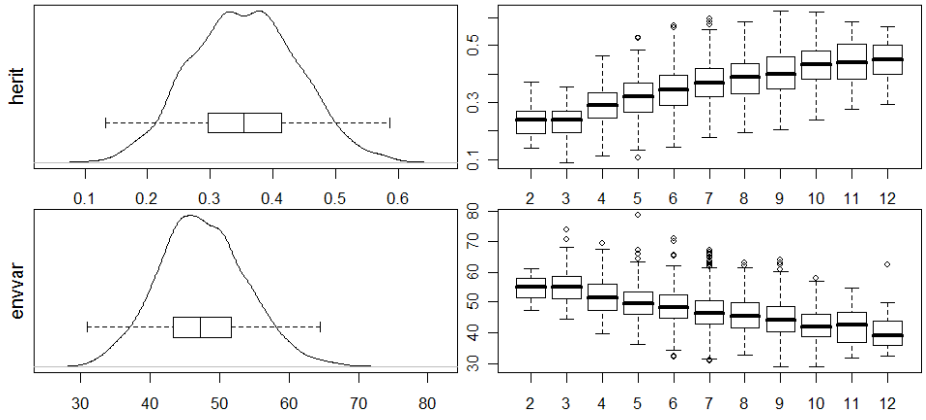
```
> summary(qb.BayesFactor(qbHyper, item = "pattern"))

                  nqtl posterior    prior    bf  bfse
1,4,6,15,6:15        5   0.03400 2.71e-05 24.30 2.360
1,4,6,6,15,6:15      6   0.00467 5.22e-06 17.40 4.630
1,1,4,6,15,6:15      6   0.00600 9.05e-06 12.80 3.020
1,1,4,5,6,15,6:15    7   0.00267 4.11e-06 12.60 4.450
1,4,6,15,15,6:15     6   0.00300 4.96e-06 11.70 3.910
1,4,4,6,15,6:15      6   0.00300 5.81e-06 10.00 3.330
1,2,4,6,15,6:15      6   0.00767 1.54e-05  9.66 2.010
1,4,5,6,15,6:15      6   0.00500 1.28e-05  7.56 1.950
1,2,4,5,6,15,6:15    7   0.00267 6.98e-06  7.41 2.620
1,4                  2   0.01430 1.51e-04  1.84 0.279
1,1,2,4              4   0.00300 3.66e-05  1.59 0.529
1,2,4                3   0.00733 1.03e-04  1.38 0.294
1,1,4                3   0.00400 6.05e-05  1.28 0.370
1,4,19               3   0.00300 5.82e-05  1.00 0.333

> plot(qb.BayesFactor(qbHyper, item = "nqtl"))
```

How many QTL?
posterior, prior, Bayes factor ratios





diagnostic summaries
> plot(qb.diag(qbHyper))

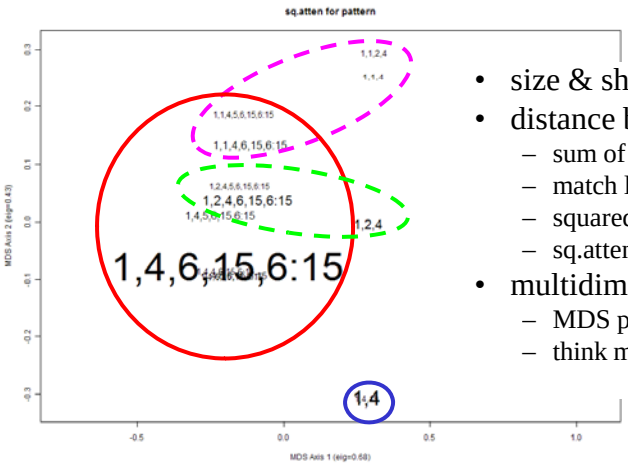


what is best estimate of QTL?

- find most probable pattern
  - 1,4,6,15,6:15 has posterior of 3.4%
- estimate locus across all nested patterns
  - Exact pattern seen ~100/3000 samples
  - Nested pattern seen ~2000/3000 samples
- estimate 95% confidence interval using quantiles

```
> best <- qb.best(qbHyper)
> summary(best)$best

    chrom locus locus.LCL locus.UCL     n.qtl
247     1  69.9  24.44875   95.7985 0.8026667
245     4  29.5  14.20000   74.3000 0.8800000
248     6  59.0  13.83333   66.7000 0.7096667
246    15  19.5  13.10000   55.7000 0.8450000

> plot(best)
```


what patterns are “near” the best?



- size & shade ~ posterior
- distance between patterns
  - sum of squared attenuation
  - match loci between patterns
  - squared attenuation = $(1-2r)^2$
  - sq.atten in scale of LOD & LPD
- multidimensional scaling
  - MDS projects distance onto 2-D
  - think mileage between cities

7. Software for Bayesian QTLs R/qtlbim: www.qtlbim.org

- Properties
  - cross-compatible with R/qtl
  - new MCMC algorithms
    - Gibbs with loci indicators; no reversible jump
  - epistasis, fixed & random covariates, GxE
  - extensive graphics
- Software history
  - initially designed (Satagopan, Yandell 1996)
  - major revision and extension (Gaffney 2001)
  - R/bim to CRAN (Wu, Gaffney, Jin, Yandell 2003)
  - R/qtlbim to CRAN (Yi, Yandell et al. 2006)
- Publications
  - Yi et al. (2005); Yandell et al. (2007); …

R/qtlbim: software history

- Bayesian module within WinQTLCart
  - WinQTLCart output can be processed using R/bim
- Software history
  - initially designed (Satagopan Yandell 1996)
  - major revision and extension (Gaffney 2001)
  - R/bim to CRAN (Wu, Gaffney, Jin, Yandell 2003)
  - R/qtlbim total rewrite (Yandell et al. 2007)

other Bayesian software for QTLs

- R/bim*: Bayesian Interval Mapping
  - Satagopan Yandell (1996; Gaffney 2001) CRAN
  - no epistasis; reversible jump MCMC algorithm
  - version available within WinQTLCart (statgen.ncsu.edu/qtlcart)
- R/qtl*
  - Broman et al. (2003 Bioinformatics) CRAN
  - multiple imputation algorithm for 1, 2 QTL scans & limited mult-QTL fits
- Bayesian QTL / Multimapper
  - Sillanpää Arjas (1998 Genetics) www.rni.helsinki.fi/~mjs
  - no epistasis; introduced posterior intensity for QTLs
- (no released code)
  - Stephens & Fisch (1998 Biometrics)
  - no epistasis
- R/bqtl
  - C Berry (1998 TR) CRAN
  - no epistasis, Haley Knott approximation

* Jackson Labs (Hao Wu, Randy von Smith) provided crucial technical support

many thanks

Karl Broman
Jackson Labs
Gary Churchill
Hao Wu
Randy von Smith
U AL Birmingham
David Allison
Nengjun Yi
Tapan Mehta
Samprit Banerjee
Ram Venkataraman
Daniel Shriner
Tom Osborn
David Butruille
Marcio Ferrera
Josh Udahl
Pablo Quijada
Alan Attie
Jonathan Stoehr
Hong Lan
Susie Clee
Jessica Byers
Mark Keller
Michael Newton
Hyuna Yang
Daniel Sorensen
Daniel Gianola
Liang Li
my students
Jaya Satagopan
Fei Zou
Patrick Gaffney
Chunfang Jin
Elias Chaibub
W Whipple Neely
Jee Young Moon
USDA Hatch, NIH/NIDDK (Attie), NIH/R01 (Yi, Broman)