Statistical Methods for Mapping Quantitative Trait Loci in Experimental Crosses

Karl W. Broman, Biostatistics, Johns Hopkins U
Gary A. Churchill, Statistical Genetics, Jackson Lab
Brian S. Yandell, Biometry, Statistic&Horticulture, UW-Madison
Zhao-Bang Zeng, Statistical Genetics \& Statistics, NCSU

www.stat.wisc.edu/~yandell/statgen

Rough Weekly/Chapter Schedule

1. General Biology Relevant to QTL
2. Key Statistical Ideas for QTL
3. Marker Regression Analysis
4. Interval Mapping for Single QTL
5. Sample Size, Power & Thresholds
6. Model Selection for Multiple QTL
7. Multiple QTL Algorithms
8. QTL for Multiple Crosses & Fine Mapping
9. Multiple Traits, Multiple Environments & Microarrays
10. Extensions of the Phenotype Model

chapters overview

1. General Biology Relevant to QTL
   - what is a QTL?
   - DNA dogma
   - crossover and recombination
   - experimental crosses
2. Key Statistical Ideas for QTL
   - recombination model and map distance
   - phenotype model and model likelihood
   - Bayesian posterior
   - missing data concepts

chapters overview

3. Marker Regression Analysis
   - Backcross & F2 intercross
   - LOD scores and thresholds
   - advantages & disadvantages
4. Interval Mapping for Single QTL
   - basic idea
   - maximum likelihood
   - Bayesian interval mapping
   - Haley-Knott regression
   - advantages & shortcomings

chapters overview

5. Sample Size, Power & Thresholds
   - review of sample size for t-test (known QTL)
   - sample size, marker spacing & power
   - thresholds & similar tests
   - positive false detection rates and multiple testing
6. Model Selection for Multiple QTL
   - reality of multiple QTL
   - comparing QTL models
   - QTL model selection criteria
   - issues on detecting epistasis

chapters overview

7. Multiple QTL Algorithms
   - Haley-Knott regression & composite IM
   - multiple interval mapping
   - multiple imputation
   - transdimensional Markov chain Monte Carlo
8. QTL for Multiple Crosses & Fine Mapping
   - four-way cross, multiple crosses of inbred parents
   - IM with pedigrees, association mapping
   - RIL, congenics, consomics
   - mutation, transgenics & confirmation studies

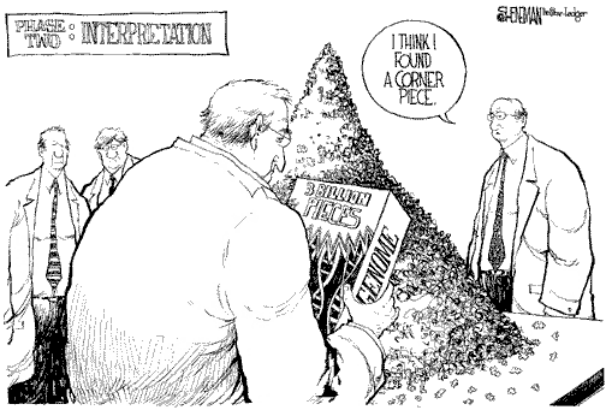
chapters overview

9. Multiple Traits, Multiple Environments & Microarrays
   - experimental design: blocks, covariates, other factors
   - two traits or environments on same cross
   - traits on independent crosses
   - multivariate dimension reduction (PC, DA, biclustering)
   - cis- and trans-action on expression in biochemical networks
10. Extensions of the Phenotype Model
   - binomial, Poisson, exponential, negative binomial
   - semi-parametric models: exponential families
   - non-parametric models
   - power and sensitivity issues

Genetics or Genomics?

- genetics: study single genes or a few genes
  - first identify mutant organism with change of interest
  - characterize effects of mutation
  - but only a fraction of 30k human genes directly studied!
- genomics: genes as dynamics system
  - over space (chromosomes) & time (evolution)
  - gene interactions, biological networks
- gene ontology (www.geneontology.org)
  - molecular function: what gene does
  - biological process: objective via assemblies of molecular functions
  - cellular component: of anatomical structure or gene product group
- (www.genomicglossaries.com)

interpretation of the genome



Drew Sheneman, The Newark Star Ledger, New Jersey
http://cagle.slate.msn.com/news/gene/gene14.asp

How to study the genome?

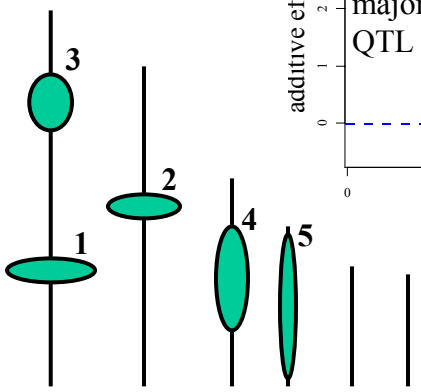
- Comparative genomics
  - chromosomal homeology among related taxa
- Functional genomics (www.functionalgenomics.org.uk)
  - mutagenesis & gene disruption
  - DNA arrays, proteomics, structural genomics
  - in silico: data prediction of cellular systems
- Reverse genetics
  - mutate gene, observe phenotypic effects
  - large screening panel or entire genome
  - knock-out (delete) or knock-in (replace)
- QTL: phenotype to genotype
  - initial screen (10cM regions)
  - recombinant inbred/congenic/consomic lines, fine mapping

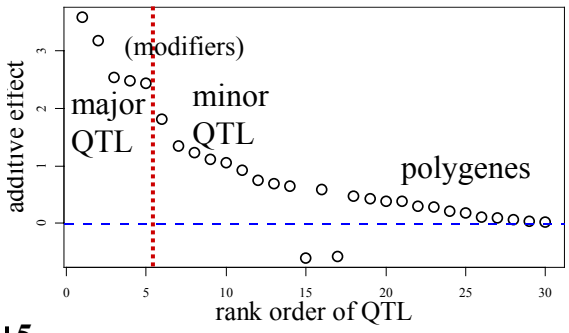
What is a QTL?

- QTL = quantitative trait locus (or loci)
  - trait = phenotype = characteristic of interest
  - quantitative = measured somehow
    - qualitative traits can often be directly mapped
    - quantitative traits not readily mapped
  - locus = location in genome affecting trait
    - gene or collection of tightly linked genes
    - some physical feature of genome

Pareto diagram of QTL effects

major QTL on linkage map



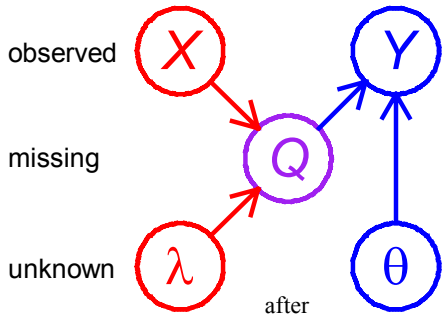


how many (detectable) QTL?

- many, many QTL may affect most any trait
  - how many QTL are detectable with these data?
    - limits to useful detection (Bernardo 2000)
    - depends on sample size, heritability, environmental variation
  - consider probability that a QTL is in the model
    - avoid sharp in/out dichotomy
    - major QTL usually selected, minor QTL sampled infrequently
- build m = number of QTL detected into QTL model
  - directly allow uncertainty in genetic architecture
  - model selection over number of QTL, architecture
  - use Bayes factors and model averaging
    - to identify "better" models

interval mapping basics

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



after Sen Churchill (2001)

Model Selection for QTL

- what is the genetic architecture?
  - M = model = (λ,θ,m)
  - λ = QT locus (or loci)
  - θ = genetic action (additive, dominance, epistasis)
  - m = number of QTL
- how to assess models?
  - MIM: various flavors of AIC, BIC (Broman Speed 2002)
  - BIM: Bayes factors
- how to search model space?
  - MIM: sequential forward selection/backward elimination
    - scan loci systematically across genome
  - BIM: sample forward/backward: transdimensional MCMC
    - sample loci at random across genome

B. napus 8-week vernalization whole genome study

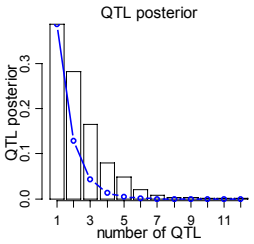
- 108 plants from double haploid
  - similar genetics to backcross: follow 1 gamete
  - parents are Major (biennial) and Stellar (annual)
- 300 markers across genome
  - 19 chromosomes
  - average 6cM between markers
    - median 3.8cM, max 34cM
  - 83% markers genotyped
- phenotype is days to flowering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

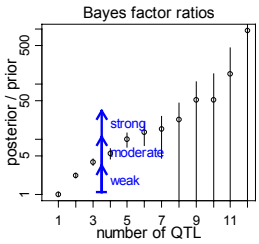
Bayesian model assessment

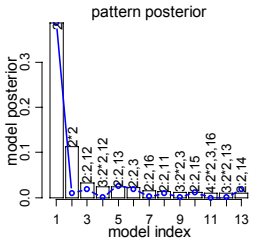
row 1: # QTL
row 2: pattern

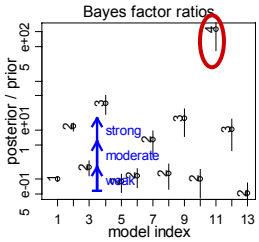
col 1: posterior
col 2: Bayes factor
note error bars on bf

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



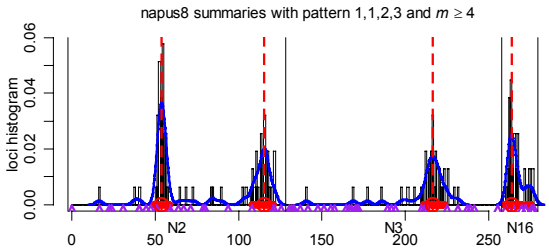




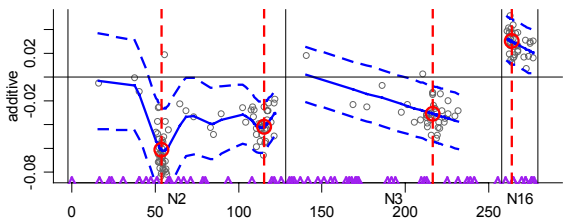


Bayesian estimates of loci & effects

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

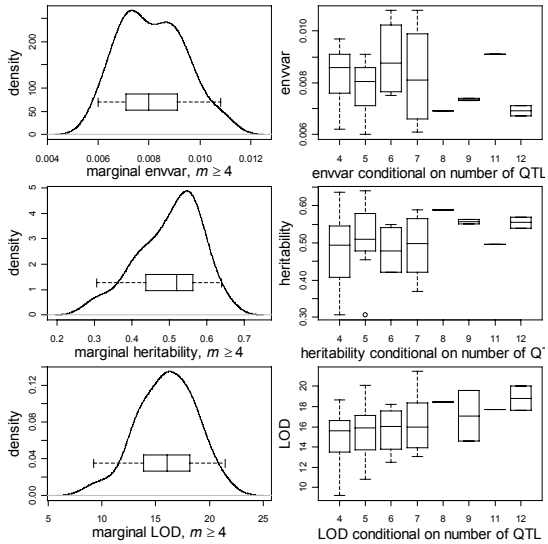


Bayesian model diagnostics

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

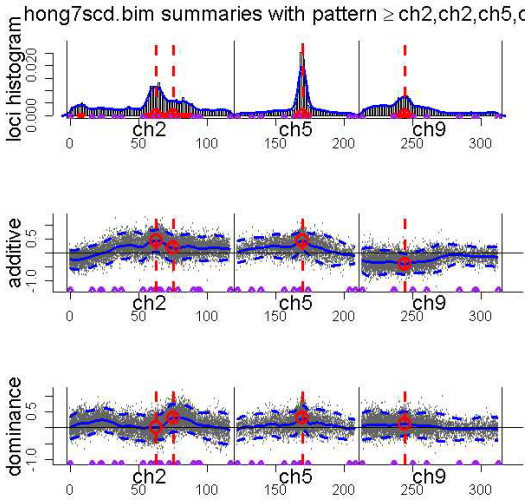
environmental variance
$\sigma^2 = .008$, $\sigma = .09$
heritability
$h^2 = 52\%$
LOD = 16
(highly significant)

but note change with *m*

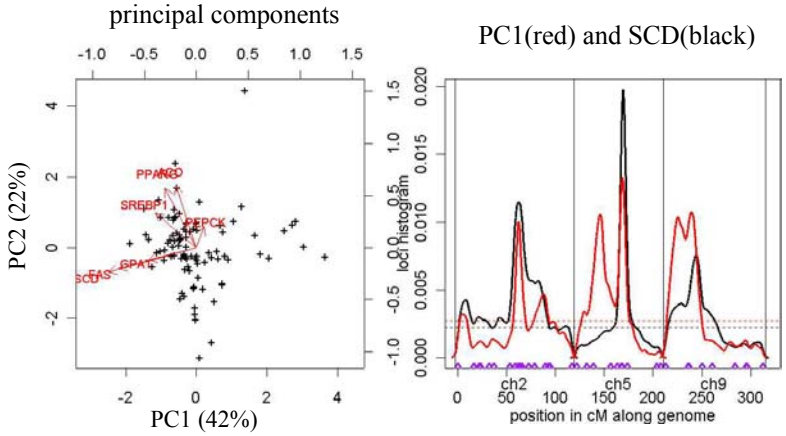


mapping gene expression

- 108 F2 mice
- microarray data
  prototype now
  mRNA via RT-PCR
- multivariate screen
  - clustering
  - PC analysis
- highlight key mRNA
  SCD
- Lan et al. (2003)
- note ch2 dominance



multivariate screen for gene expressing mapping



Some QTL software

- Regression-based (Haley-Knott) mapping
  - PLABQTL (www.uni-hohenheim.de/~ipspwww/soft.html)
  - MapManager (mapmgr.roswellpark.org/mapmgr.html)
- Classical interval mapping
  - QTL Cartographer (statgen.ncsu.edu/qtlcart/cartographer.html)
  - R/qtl (www.biostat.jhsph.edu/~kbroman)
- Bayesian software for QTLs
  - BIM/Bmapqtl (www.stat.wisc.edu/~yandell/qtl/software/Bmapqtl)
  - Bayesian QTL / Multimapper (www.rni.helsinki.fi/~mjs)
  - Yi & Xu (2002); Stephens & Fisch (1998)
- linkage analysis software lists
  - www.stat.wisc.edu/biosci/linkage.html
  - linkage.rockefeller.edu/soft/list.html

some QTL references

- Bernardo RN (2000) What if we knew all the genes for a quantitative trait in hybrid crops? *Crop Sci. 41*: 1-4.
- Broman KW (2001) Review of statistical mentods for QTL mapping in experimental crosses. *Lab Animal 30*: 44-52.
- Mackay TFC (2001) Quantitative trait loci in *Drosophila. Nature Reviews Genetics 2*: 11-20.

- Bernardo RN (2002) *Breeding for Quantitative Traits in Plants*. Stemma Press. ISBN 0972072403.
- Lynch M and Walsh M (1997) *Genetics and Analysis of Quantitative Traits.* Sinauer Assoc. ISBN 0878934812.