Multiparent populations & R/qtl2

Karl Broman
Biostatistics and Medical Informatics
University of Wisconsin – Madison

kbroman.org/qtl2
kbroman.org
github.com/kbroman
@kwbroman
QTL mapping

The image shows a graph with LOD scores plotted against chromosomes. The horizontal line at a LOD score of 4.0 indicates the threshold for significant QTLs. The graph displays peaks at various chromosomal locations, suggesting regions of interest for QTL mapping.
Improving precision

- more recombinations
- more individuals
- more precise phenotypes
- lower-level phenotypes
  - transcripts, proteins, metabolites
Advanced intercross lines

P
A
B

F_2

F_3

F_4

F_7

F_{10}
Recombinant inbred lines
Recombinant inbred lines
Collaborative Cross/MAGIC
MAGIC lines

Valdar et al., Genetics 172:1783, 2006
MAGIC lines

Valdar et al., Genetics 172:1783, 2006

How many?
MAGIC lines

Valdar et al., Genetics 172:1783, 2006

How many?

Which?
MAGIC lines

![Diagram showing the process of combining, mixing, and fixing founders through foundation, outbred, and recombinant inbred stages.]

- How many?
- How long?
- Which?
MAGIC lines

combine mix fix

How many? How long? How?
Which?
MAGIC is magic

- Genetic diversity
- High-precision mapping
- Predictable linkage disequilibrium
- Phenotype replicates to reduce individual variation
- Pool phenotypes from multiple labs, environments, treatments
- Genotype once
The goal

Identify QTG

- Power
- Mapping precision
- Estimate QTL allele frequencies
Principles

- Avoid population structure

- Tradeoff between power for *de novo* discovery and mapping precision

- More QTL to find $\Rightarrow$ more QTL getting in the way?

- More QTL alleles $\Rightarrow$ less information about each

- Are QTL alleles common or rare?
How many founders?

More

- More general use
- More QTL
- Greater precision
- Estimate allele frequencies
- Haplotype analysis in founders

Fewer

- Lower residual variance
- Greater power for a particular QTL?
- Better power for epistasis
- Rare alleles are less rare
Which founders?

- Diverse
- Interesting
- No breeding problems
- Balanced: star phylogeny
How much mixing?

- More mixing ⇒ Greater mapping precision
- ...but lower power for *de novo* mapping
- Potential for population structure, missing alleles
- Random mating or curated mating?
- Start with many random cross directions?
Key analysis issues

How to deal with the multiple alleles?

- Full model (an effect for each allele)
- Diallelic QTL model
- Random effects model (like BLUP)

How to account for multiple QTL?

- Stepwise selection
- Bayesian model averaging
- Random effect for polygenes
Linear mixed models

\[ y_i = \mu + \sum_k \beta_{k,i} + \epsilon_i \]
\[ \epsilon_i \sim N(0, \sigma^2_e) \]

\[ = \mu + \eta_i + \epsilon_i \]
\[ \eta_i \sim N(0, \sigma^2_p) \]

\[ \text{COV}(\eta_i, \eta_j) = \sigma^2_p (2k_{ij}) \]
HS genotype reconstruction

Founders:
- A
- B
- C
- D
- E
- F
- G
- H

HS−1
HS genotype reconstruction

Founders

A
B
C
D
E
F
G
H

HS−1

HS−2
HS genotype reconstruction

Founders

A
B
C
D
E
F
G
H

HS−1

HS−2

HS−3

23
HS genome scans

Full model

Additive alleles

SNP associations
Why R/qtl2?

• High-dimensional data
  genotypes and phenotypes

• More diverse crosses
  especially multi-parent populations

• Linear mixed models
  especially in HS/AIL
Open problems

• Analysis of multi-parent populations
  — SNP vs haplotype vs random QTL effect
  — Identifying causal polymorphisms

• Joint analysis of high-dimensional phenotypes
  — gene expression, proteins, metabolites, etc.

• QTL × environment interactions