QTL mapping in MAGIC populations

Part 1

Karl Broman

Biostatistics & Medical Informatics, UW–Madison

kbroman.org
github.com/kbroman
@kwbrman

Slides: kbroman.org/Talk_MAGIC2021
Intercross

P₁ × P₂ → F₁

F₁ × F₁ → F₂
QTL mapping
Congenic line/NIL
Improving precision

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

\[ P \quad \begin{array}{c}
A \\
B \\
\end{array} \]

\[ F_2 \]

\[ F_3 \]

\[ F_4 \]

\[ F_7 \]

\[ F_{10} \]
Recombinant inbred lines

\[ \text{P}_1 \rightarrow \times \rightarrow \text{P}_2 \]

\[ \text{F}_1 \rightarrow \times \rightarrow \text{F}_1 \]

\[ \text{F}_2 \rightarrow \times \rightarrow \text{F}_2 \]

\[ \text{F}_3 \rightarrow \times \rightarrow \text{F}_3 \]

\[ \text{F}_4 \rightarrow \times \rightarrow \text{F}_4 \]

\[ \vdots \]

\[ \text{F}_\infty \]
Recombinant inbred lines

Diagram:

- **P₁**: Parent 1
- **P₂**: Parent 2
- **F₁**: First generation
- **F₂**, **F₃**, **F₄**, **…**: Subsequent generations
- **F∞**:无限世代
Collaborative Cross

G₀
A  B  C  D

G₁
A  B

G₂
ABCDE

G₃

G₄

⋮

Gₘₙ

G₁
E  F

G₂
EFGH

G₃

Gₘₙ

G₀
A  B  C  D  E  F  G  H

A  B  C  D  E  F  G  H

G₁
A  B

G₂
ABCDE

G₃

G₄

⋮

Gₘₙ

G₀
A  B  C  D  E  F  G  H

A  B  C  D  E  F  G  H

G₁
A  B

G₂
ABCDE

G₃

G₄

⋮

Gₘₙ

G₀
A  B  C  D  E  F  G  H

A  B  C  D  E  F  G  H

G₁
A  B

G₂
ABCDE

G₃

G₄

⋮

Gₘₙ

G₀
A  B  C  D  E  F  G  H

A  B  C  D  E  F  G  H

G₁
A  B

G₂
ABCDE

G₃

G₄

⋮

Gₘₙ

G₀
A  B  C  D  E  F  G  H

A  B  C  D  E  F  G  H

G₁
A  B

G₂
ABCDE

G₃

G₄

⋮

Gₘₙ

G₀
A  B  C  D  E  F  G  H

A  B  C  D  E  F  G  H

G₁
A  B

G₂
ABCDE

G₃

G₄

⋮

Gₘₙ

G₀
A  B  C  D  E  F  G  H

A  B  C  D  E  F  G  H

G₁
A  B

G₂
ABCDE

G₃

G₄

⋮

Gₘₙ

G₀
A  B  C  D  E  F  G  H

A  B  C  D  E  F  G  H

G₁
A  B

G₂
ABCDE

G₃

G₄

⋮

Gₘₙ
Heterogeneous stock
MAGIC is magic

- Genetic diversity
- High-precision mapping
- Predictable linkage disequilibrium
- No rare alleles
- Phenotype replicates to reduce individual variation
- Pool phenotypes from multiple labs, environments, treatments
- Genotype once
MAGIC is magic

- Genetic diversity
- High-precision mapping
- Predictable linkage disequilibrium
- No rare alleles
- Phenotype replicates to reduce individual variation
- Pool phenotypes from multiple labs, environments, treatments
- Genotype once
- Cool name
MAGIC lines

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

How many?

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

How many?
Which?

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

Valdar et al., Genetics 172:1783, 2006

How many?
Which?

How long?
MAGIC lines

Valdar et al., Genetics 172:1783, 2006
The goal

Identify QTL

▶ Power

▶ Mapping precision
The goal

Identify QTG

- Power
- Mapping precision
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 ⇒ more QTL getting in the way?
- More QTL alleles ⇒ 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 $\Rightarrow$ 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?
Selfing or DH?

- Inbreeding gives added recombination
- But not so much as at the mixing stage
- If doubled haploids are feasible, use them
Sharing is also key

- The greatest power of MAGIC comes from sharing
  Pooling data, exploring multiple environments/treatments

- Common software needs
  Analysis software, database infrastructure

- Many students need to learn the same stuff
  Joint training opportunities
Summary

▶ How many founders?
  – Tradeoff between diversity and information about particular alleles

▶ Which founders?
  – Diverse, interesting, no breeding problems, star phylogeny

▶ How long to mix?
  – Tradeoff between power and precision

▶ How to fix?
  – Doubled haploids are great if feasible

▶ Let’s share!
  – Lines, data, software, training
Slides: kbroman.org/Talk_MAGIC2021

kbroman.org

kbroman.org/qltl2

github.com/kbroman

@kwbroman