The genomes of recombinant inbred lines

Karl W Broman
Department of Biostatistics
Johns Hopkins University

http://www.biostat.jhsph.edu/~kbroman

Recombinant inbred lines
(by sibling mating)
Recombinant inbred lines
(by selfing)

The “Collaborative Cross”
The goal

- Characterize the breakpoint process along a chromosome in 8-way RILs.
  - Understand the two-point haplotype probabilities.
  - Study the clustering of the breakpoints, as a function of crossover interference in meiosis.
Why?

• It’s interesting.
• Later statistical analyses will require:
  – The two-point probabilities.
  – A model for the whole process.

Actually, we’ll probably just assume that:
  – The breakpoints follow a Poisson process.
  – The genotypes follow a Markov chain.

2 points in an RIL

\[ \begin{array}{c}
  1 \quad 2 \\
\end{array} \]

• \( r \) = recombination fraction = probability of a recombination in the interval in a random meiotic product.

• \( R \) = analogous thing for the RIL = probability of different alleles at the two loci on a random RIL chromosome.
Equations for selfing

\[ C_n, A_{BB} \text{ and } a_{bb}, \]
\[ D_n, A_{BB} \text{ and } a_{BB}, \]
\[ E_n, A_{BB}, A_{BB}, A_{BB}, \text{ and } a_{BB}, \]
\[ F_n, A_{BB} \text{ and } a_{BB}. \]

We assume \( 2C_n + 2D_n + 4E_n + 2F_n + G_n = 2 \), so that \( C_n = D_n = E_n = G_n = 0 \), and \( F_n = 2 \). Clearly \( E_n = F_n = G_n = 0 \), and \( D_n \) is the final proportion of crossover zygotes. Then considering the results of selfing each generation, we have:

\[
\begin{align*}
C_{n+1} &= C_n + \frac{1}{2} F_n + \frac{1}{2} (1 - \beta - \delta + \beta \delta) F_n + \frac{1}{2} (\delta G_n), \\
D_{n+1} &= D_n + \frac{1}{2} F_n + \frac{1}{2} (1 - \beta + \delta - \beta \delta) G_n, \\
E_{n+1} &= \frac{1}{2} F_n + \frac{1}{2} (1 - \beta - \delta + \beta \delta) G_n, \\
F_{n+1} &= \frac{1}{2} [1 - \beta - \delta + \beta \delta] F_n + \frac{1}{2} (1- \beta - \delta + \beta \delta) G_n, \\
G_{n+1} &= \frac{1}{2} [1 - \beta - \delta + \beta \delta] F_n + \frac{1}{2} (1- \beta - \delta + \beta \delta) G_n.
\end{align*}
\]

(1.1)

Put \( y = D_n \) (the final proportion of crossover zygotes)

\[ \therefore C_n + D_n = 1, \quad C_n - D_n = c_n \quad \therefore y = \frac{1}{2} (1 - c_n). \]

\[ \therefore y = \frac{2x}{1 + 2x} \]

(1.3)
Recombinant inbred lines
(by selfing)

Recombinant inbred lines
(by sibling mating)
Equations for sib-mating

\[ \begin{align*}
AABB \times AABB & \quad 2 \quad C_{\alpha} = C + \frac{1}{2}(a + 2b) + \frac{1}{6}(1 - 2x)(d + 2f + 2) + \frac{3}{2}k_a \\
AABB \times AABb & \quad 2 \quad D_{\alpha} = D + \frac{1}{2}(a + 2b) + \frac{1}{6}(1 - 2x)(d + 2f + 2) + \frac{3}{2}k_a \\
AABB \times AbB & \quad 2 \quad E_{\alpha} = E + \frac{1}{2}(a + 2b) + \frac{1}{6}(1 - 2x)(d + 2f + 2) + \frac{3}{2}k_a \\
AABB \times Abb & \quad 2 \quad F_{\alpha} = F + \frac{1}{2}(a + 2b) + \frac{1}{6}(1 - 2x)(d + 2f + 2) + \frac{3}{2}k_a \\
AABB \times AAb & \quad 8 \quad G_{\alpha} = G + \frac{1}{2}(a + 2b) + \frac{1}{6}(1 - 2x)(d + 2f + 2) + \frac{3}{2}k_a \\
AABB \times Aab & \quad 8 \quad H_{\alpha} = H + \frac{1}{2}(a + 2b) + \frac{1}{6}(1 - 2x)(d + 2f + 2) + \frac{3}{2}k_a \\
AABB \times AAB & \quad 14 \quad I_{\alpha} = I + \frac{1}{2}(a + 2b) + \frac{1}{6}(1 - 2x)(d + 2f + 2) + \frac{3}{2}k_a \\
AABB \times AaB & \quad 14 \quad J_{\alpha} = J + \frac{1}{2}(a + 2b) + \frac{1}{6}(1 - 2x)(d + 2f + 2) + \frac{3}{2}k_a \\
AABB \times AaB & \quad 14 \quad K_{\alpha} = K + \frac{1}{2}(a + 2b) + \frac{1}{6}(1 - 2x)(d + 2f + 2) + \frac{3}{2}k_a \\
AABB \times AaB & \quad 14 \quad L_{\alpha} = L + \frac{1}{2}(a + 2b) + \frac{1}{6}(1 - 2x)(d + 2f + 2) + \frac{3}{2}k_a \\
AABB \times AaB & \quad 14 \quad M_{\alpha} = M + \frac{1}{2}(a + 2b) + \frac{1}{6}(1 - 2x)(d + 2f + 2) + \frac{3}{2}k_a \\
\end{align*} \]

Result for sib-mating

Omitting some rather tedious algebra, the solution of these equations is:

\[ \begin{align*}
t &= \frac{q}{2 - 3q}, \quad \theta &= \frac{2q}{2 - 3q}, \quad \kappa &= \frac{1}{2 - 3q}, \\
\lambda &= \frac{1 - 2q}{2 - 3q}, \quad \mu &= \frac{1 - 2q}{2 - 3q}, \quad \rho &= \frac{2q}{2 - 3q}.
\end{align*} \]

as may easily be verified.

\[ \therefore \quad c_{\alpha} = c + 2c_a + \frac{1}{2}[1 - 2x](d + 2f + 2)x + \frac{1}{6}k_a \]

\[ + 2cx + 4x(b_a + i_a) \]  

(3.4)

and \( y = \frac{1}{2}(1 - c_{\alpha}) \).

In the case considered, \( d_a = 1 \), \( c_{\alpha} = c = d_a = 1 - 2x/1 + 6x \). Hence the proportion of crossover zygotes \( y = 4x/1 + 6x \) (3.5). \( \square \)
The “Collaborative Cross”

8-way RILs

Autosomes

Pr(G₁ = i) = 1/8
Pr(G₂ = j | G₁ = i) = r / (1+6r) for i ≠ j
Pr(G₂ ≠ G₁) = 7r / (1+6r)

X chromosome

Pr(G₁=A) = Pr(G₁=B) = Pr(G₁=E) = Pr(G₁=F) = 1/6
Pr(G₁=C) = 1/3
Pr(G₂=B | G₁=A) = r / (1+4r)
Pr(G₂=C | G₁=A) = 2r / (1+4r)
Pr(G₂=A | G₁=C) = r / (1+4r)
Pr(G₂ ≠ G₁) = (14/3) r / (1+4r)
Computer simulations

The X chromosome
3-point coincidence

1 2 3

- $r_{ij}$ = recombination fraction for interval i,j; assume $r_{12} = r_{23} = r$
- Coincidence = $c = \frac{\text{Pr(double recombinant) / } r^2}{\text{Pr(rec'n in 23 | rec'n in 12) / Pr(rec'n in 23)}}$
- No interference $\rightarrow = 1$
  Positive interference $\rightarrow < 1$
  Negative interference $\rightarrow > 1$
- Generally $c$ is a function of $r$.

3-points in 2-way RILs

1 2 3

- $r_{13} = 2r(1 - cr)$
- $R = f(r); \quad R_{13} = f(r_{13})$
- $\text{Pr(double recombinant in RIL)} = \{ R + R - R_{13} \} / 2$
- Coincidence (in 2-way RIL) = $\{ 2R - R_{13} \} / \{ 2R^2 \}$
Coincidence

No interference

Coincidence

R

Coincidence

R

Meiosis
RILs by selfing
RILs by sib-mating

Meiosis
RILs by selfing
RILs by sib-mating
No interference
Mouse interference
Why the clustering of breakpoints?

- The really close breakpoints occur in different generations.
- Breakpoints in later generations can occur only in regions that are not yet fixed.
- The regions of heterozygosity are, of course, surrounded by breakpoints.

Coincidence in 8-way RILs

- The trick that allowed us to get the coincidence for 2-way RILs doesn’t work for 8-way RILs.
- It’s sufficient to consider 4-way RILs.
- Calculations for 3 points in 4-way RILs is still astoundingly complex.
  - 2 points in 2-way RILs by sib-mating:
    55 parental types → 22 states by symmetry
  - 3 points in 4-way RILs by sib-mating:
    2,164,240 parental types → 137,488 states
- Even counting the states was difficult.
**Coincidence**

![Graph showing coincidence across different mating types and genome types.]

**Whole genome simulations**

- 2-way selfing, 2-way sib-mating, 8-way sib-mating
- Mouse-like genome, 1665 cM
- Strong positive crossover interference
- Inbreed to complete fixation
- 10,000 simulation replicates
No. generations to fixation

- 2-way selfing
- 2-way sib-mating
- 8-way sib-mating

mean = 10.5
mean = 35.6
mean = 38.9

No. gen’s to 99% fixation

- 2-way selfing
- 2-way sib-mating
- 8-way sib-mating

mean = 8.0
mean = 23.5
mean = 26.7
Percent genome not fixed

Number of segments
Segment lengths

- Median = 8.5 cM (2-way selfing)
- Median = 12.9 cM (2-way sib-mating)
- Median = 23.7 cM (8-way sib-mating)

Two chromosomes
X chromosome

Probability a segment is inherited intact

- Median = 8.5 cM (2-way selfing)
- Median = 12.9 cM (2-way sib-mating)
- Median = 23.7 cM (8-way sib-mating)
Length of smallest segment

No. segments < 1 cM
Summary

- RILs are useful.
- The Collaborative Cross could provide “one-stop shopping” for gene mapping in the mouse.
- Use of such 8-way RILs requires an understanding of the breakpoint process.
- We’ve extended Haldane & Waddington’s results to the case of 8-way RILs.
- We’ve shown clustering of breakpoints in RILs by sib-mating, even in the presence of strong crossover interference.
- Formulae for the 3-point problem in 8-way RILs elude us, but we can obtain numerical results.
- We used simulations to study other features of RILs.

The key points

- \[ R = \frac{7r}{1 + 6r} \]
- 2-point probabilities, for the autosomes of 8-way RILs, have all off-diagonal elements identical.
- 3-point coincidence on 8-way RIL is near 1.