The genomes of recombinant inbred lines

Karl W Broman
Department of Biostatistics
Johns Hopkins University

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

C57BL/6
Recombinant inbred lines
(by sibling mating)

$P_1 \times P_2$

$F_1 \times F_1$

$F_2$

$F_3$

$F_4$

$\vdots$

$F_{\infty}$

Recombinant inbred lines
(by selfing)

$P_1 \times P_2$

$F_1$

$F_2$

$F_3$

$F_4$

$\vdots$

$F_{\infty}$
Advantages of RI lines

• Each strain is an eternal resource.
  – Only need to genotype once.
  – Reduce individual variation by phenotyping multiple individuals from each strain.
  – Study multiple phenotypes on the same genotype.

• Greater mapping precision.
  – More dense breakpoints on the RI chromosomes.

Disadvantages of RI lines

• Expensive and time consuming to create.

• The available panels are too small.

• Learn only about 2 alleles.
The “Collaborative Cross”

Genome of an 8-way RI
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

1 — 2

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

• \( R \) = analogous thing for the RIL = probability of different genotypes at the two loci in a random RIL.

INBREEDING AND LINKAGE
J. B. S. Haldane and C. H. Waddington
John Innes Horticultural Institution, London, England
Received August 9, 1930

Genetics 16:357-374

TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-fertilization</td>
<td>358</td>
</tr>
<tr>
<td>Brother-sister mating.</td>
<td>360</td>
</tr>
<tr>
<td>Autosomeal genes.</td>
<td>364</td>
</tr>
<tr>
<td>Parent and offspring mating.</td>
<td>367</td>
</tr>
<tr>
<td>Inbreeding with any initial population.</td>
<td>370</td>
</tr>
<tr>
<td>Double crossing over</td>
<td>372</td>
</tr>
<tr>
<td>Discussion</td>
<td>373</td>
</tr>
<tr>
<td>Summary</td>
<td>374</td>
</tr>
<tr>
<td>Literature cited</td>
<td>374</td>
</tr>
</tbody>
</table>

When a heterozygous population is self-fertilized or inbred the ultimate result (apart from effects of mutation) is complete homozygosis. The final proportions of the various genotypes are usually independent of the system of inbreeding adopted, although, as Jenkins (1916) and others have shown, the speed at which equilibrium is approached is greater in the case of self-fertilization than of brother-sister mating, and so on.
Equations for selfing

C<sub>n</sub>, AABB and aabb.
D<sub>n</sub>, AAbb and aabb.
E<sub>n</sub>, AAbb, aAbb, aabb, and aAbb.
F<sub>n</sub>, AAbb.
G<sub>n</sub>, AAbb.

We assume 2C<sub>n</sub> + 2D<sub>n</sub> + 4E<sub>n</sub> + F<sub>n</sub> + G<sub>n</sub> = 2, so that C<sub>1</sub> = D<sub>1</sub> = E<sub>1</sub> = G<sub>1</sub> = 0, and F<sub>1</sub> = 2. Clearly E<sub>n</sub> = F<sub>n</sub> = G<sub>n</sub> = 0, and D<sub>n</sub> is the final proportion of crossover zygotes. Then considering the results of selfing each generation, we have:

\[ C_{n+1} = C_n + \frac{1}{2}E_n + \frac{1}{2}(1 - \beta - \delta + \beta\delta)F_n + \beta\delta G_n \]
\[ D_{n+1} = D_n + \frac{1}{2}E_n + \frac{1}{2}(1 - \beta - \delta + \beta\delta)G_n \]
\[ E_{n+1} = \frac{1}{2}E_n + \frac{1}{2}(1 - \beta - \delta + \beta\delta)F_n + \beta G_n \]
\[ F_{n+1} = \frac{1}{2}(1 - \beta - \delta + \beta\delta)F_n + \beta\delta G_n \]
\[ G_{n+1} = \frac{1}{2}(1 - \beta - \delta + \beta\delta)G_n \]

(1.1)

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

\[ \therefore C_n + D_n = 1, \quad C_n - D_n = \epsilon \quad \therefore y = \frac{1}{1 + 2x} \]

(1.3)

Recombinant inbred lines
(by selfing)

[Diagram of successive generations of selfing showing increasing recombination and inbreeding.]
Recombinant inbred lines
(by sibling mating)

Equations for sib-mating
Result for sib-mating

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

\[ \iota = \frac{q}{2 - 3q}, \quad \theta = -\frac{2q}{2 - 3q}, \quad \kappa = \frac{1}{2 - 3q}, \]
\[ \lambda = \frac{1 - 2q}{2 - 3q}, \quad \mu = \frac{1 - 2q}{2 - 3q}, \quad \nu = \frac{2q}{2 - 3q} \]

as may easily be verified.

\[ \therefore \quad c_x = c_y + 2c_a + \frac{1}{1 + 6a} \left[ (1 - 2x)(d_a + 2f_a + 2) + \frac{1}{2}k_a \right] \]
\[ + 2g_a + 4x(b_a + b_s) \] (3.4)

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

In the case considered, \( d_a = 1 \), \( c_w = c_b = 1 - \frac{1}{2}x + 6x \). Hence the proportion of crossover zygotes \( y = \frac{4x}{1 + 6x} \) (3.5). \( \square \)

Haldane & Waddington 1931

\( r \) = recombination fraction per meiosis between two loci
\( G_i \) = allele at marker \( i \) in an RIL by sib-matings.

Autosomes
\[ \Pr(G_1 = A) = \Pr(G_1 = B) = 1/2 \]
\[ \Pr(G_2 = B \mid G_1 = A) = \Pr(G_2 = A \mid G_1 = B) = \frac{4r}{(1 + 6r)} \]

X chromosome
\[ \Pr(G_1 = A) = 2/3 \quad \Pr(G_1 = B) = 1/3 \]
\[ \Pr(G_2 = B \mid G_1 = A) = \frac{2r}{(1 + 4r)} \]
\[ \Pr(G_2 = A \mid G_1 = B) = \frac{4r}{(1 + 4r)} \]
\[ \Pr(G_2 \neq G_1) = \frac{8/3}{r} \frac{r}{(1 + 4r)} \]
The “Collaborative Cross”

8-way RILs

Autosomes
\[ \Pr(G_1 = i) = 1/8 \]
\[ \Pr(G_2 = j \mid G_1 = i) = r / (1+6r) \text{ for } i \neq j \]
\[ \Pr(G_2 \neq G_1) = 7r / (1+6r) \]

X chromosome
\[ \Pr(G_1 = A) = \Pr(G_1 = B) = \Pr(G_1 = E) = \Pr(G_1 = F) = 1/6 \]
\[ \Pr(G_1 = C) = 1/3 \]
\[ \Pr(G_2 = B \mid G_1 = A) = r / (1+4r) \]
\[ \Pr(G_2 = C \mid G_1 = A) = 2r / (1+4r) \]
\[ \Pr(G_2 = A \mid G_1 = C) = r / (1+4r) \]
\[ \Pr(G_2 \neq G_1) = (14/3) r / (1+4r) \]
Computer simulations

\[ R = \frac{7r}{1+6r} \]

\[ R = \frac{(14/3)r}{1+4r} \]

The X chromosome
3-point coincidence

1 2 3
• $r_{ij}$ = recombination fraction for interval $i,j$;
  assume $r_{12} = r_{23} = r$
• Coincidence = $c = Pr(\text{double recombinant}) / r^2$
  $= Pr(\text{rec'n in 23} | \text{rec'n in 12}) / Pr(\text{rec'n in 23})$
• 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
• $r_{13} = 2r (1 - cr)$
• $R = f(r); \quad R_{13} = f(r_{13})$
• $Pr(\text{double recombinant in RIL}) = \{ R + R - R_{13} \} / 2$
• Coincidence (in 2-way RIL) = $\{ 2R - R_{13} \} / \{ 2R^2 \}$
Coincidence

![Graph showing coincidence as a function of R.](image)

Coincidence

![Graph showing coincidence as a function of R.](image)
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.

Whole genome simulations

• 2-way selfing, 2-way sib-mating, 8-way sib-mating

• Mouse-like genome, 1665 cM

• No interference or strong positive interference

• Inbreed to complete fixation

• 1000 simulation replicates
No. generations to fixation

![Graph showing the number of generations to fixation with various mating types and interference levels.](image)

No. gen’s to 99% fixation

![Graph showing the number of generations to reach 99% fixation with various mating types and interference levels.](image)
Percent genome not fixed

![Graph showing the percent genome not fixed over the number of generations.]

Length of smallest segment

![Graph showing the length of the smallest segment over the number of generations.]

mean = 0.08
mean = 0.18
mean = 0.65
No. segments < 1 cM

- 2-way sib-mating
- 8-way sib-mating
- No interference

mean = 1.7

mean = 5.8

mean = 11.9

No. segments < 1 cM

No. segments < 1 cM

- 2-way sib-mating
- 8-way sib-mating
- Mouse interference

mean = 1.4

mean = 5.3

mean = 11.3

No. segments < 1 cM
Probability that a chromosome is intact

Segment lengths
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 still elude us.
• We used simulations to study other features of RILs.

The key points

• $R = \frac{7r}{(1 + 6r)}$
• 2-point prob’s, for the autosomes of 8-way RILs, have all off-diagonal elements identical.
• 3-point coincidence on 8-way RIL is near 1.