The genomes of recombinant inbred lines

Karl W Broman
Department of Biostatistics
Johns Hopkins University

www.biostat.jhsph.edu/~kbroman
Human vs mouse
Intercross
Recombinant inbred lines
(by sibling mating)
The RIX design

RI strains

RIX
The “Collaborative Cross”

Genome of an 8-way RIL
Genome of an 8-way RIL
Genome of an 8-way RIL
• Characterize the breakpoint process along a chromosome in 8-way RILs.
  – Understand the two-point haplotype probabilities.
  – Study the clustering of breakpoints, as a function of crossover interference in meiosis.
2 points in an RIL

- \( r = \text{recombination fraction} = \text{probability of a recombination in the interval in a random meiotic product.} \)

- \( R = \text{analogous thing for the RIL} = \text{probability of different alleles at the two loci on a random RIL chromosome.} \)
INBREEDING AND LINKAGE*

J. B. S. HALDANE AND C. H. WADDINGTON

John Innes Horticultural Institution, London, England

Received August 9, 1930

TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>TABLE OF CONTENTS</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-fertilization</td>
<td>358</td>
</tr>
<tr>
<td>Brother-sister mating. Sex-linked genes</td>
<td>360</td>
</tr>
<tr>
<td>Brother-sister mating. Autosomal genes</td>
<td>364</td>
</tr>
<tr>
<td>Parent and offspring mating. Sex-linked genes</td>
<td>367</td>
</tr>
<tr>
<td>Parent and offspring mating. Autosomal genes</td>
<td>368</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 JENNINGS (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.
Recombinant inbred lines
(by selfing)
• Sequence of random variables \( \{X_0, X_1, X_2, \ldots \} \) satisfying

\[
\Pr(X_{n+1} | X_0, X_1, \ldots, X_n) = \Pr(X_{n+1} | X_n)
\]

• Transition probabilities \( P_{ij} = \Pr(X_{n+1} | X_n) \)

• Here, \( X_n = \) “parental type” at generation \( n \).

• We are interested in absorption probabilities

\[
\Pr(X_n \rightarrow j | X_0)
\]
Absorption probabilities

Consider the case of absorption into the state \( \begin{array}{c|c} A \\ \hline A \\ A \end{array} \) (write this AA|AA)

Let \( h_i \) = probability, starting at \( i \), of being absorbed into AA|AA.

Then \( h_{AA|AA} = 1 \) and \( h_{AB|AB} = 0 \).

Condition on the first step: \( h_i = \sum_k P_{ik} h_k \)

For selfing, this gives a system of 3 linear equations.
Equations for selfing

We assume $2C_n + 2D_n + 4E_n + F_n + G_n = 2$, so that $C_1 = D_1 = E_1 = G_1 = 0$, and $F_1 = 2$. Clearly $E_{\infty} = F_{\infty} = G_{\infty} = 0$, and $D_{\infty}$ 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}E_n + \frac{1}{4}(1 - \beta - \delta + \beta\delta)F_n + \frac{1}{4}\beta\delta G_n \\
D_{n+1} &= D_n + \frac{1}{2}E_n + \frac{1}{4}\beta\delta F_n + \frac{1}{4}(1 - \beta - \delta + \beta\delta)G_n \\
E_{n+1} &= \frac{1}{2}E_n + \frac{1}{4}(\beta + \delta - 2\beta\delta)(F_n + G_n) \\
F_{n+1} &= \frac{1}{2}(1 - \beta - \delta + \beta\delta)F_n + \frac{1}{2}\beta\delta G_n \\
G_{n+1} &= \frac{1}{2}\beta\delta F_n + \frac{1}{2}(1 - \beta - \delta + \beta\delta)G_n
\end{align}

(1.1)

Put $y = D_{\infty}$ (the final proportion of crossover zygotes)

\[ \therefore C_{\infty} + D_{\infty} = 1, \quad C_{\infty} - D_{\infty} = c_{\infty} \therefore y = \frac{1}{2}(1 - c_{\infty}) \]

(1.3)

\[ \therefore y = \frac{2x}{1 + 2x} \]
Recombinant inbred lines (by sibling mating)
Equations for sib-mating

<table>
<thead>
<tr>
<th>Typical mating</th>
<th>Number of types</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AABB x AABB</strong></td>
<td>2</td>
</tr>
<tr>
<td><strong>AAbb x AAbb</strong></td>
<td>2</td>
</tr>
<tr>
<td><strong>AABB x aabb</strong></td>
<td>2</td>
</tr>
<tr>
<td><strong>AAbb x AaBB</strong></td>
<td>2</td>
</tr>
<tr>
<td><strong>AABB x AAbb</strong></td>
<td>8</td>
</tr>
<tr>
<td><strong>AAbb x AABb</strong></td>
<td>8</td>
</tr>
</tbody>
</table>

\[ C_{n+1} = C_0 + \frac{1}{2} \left( a^2 + \gamma^2 \right) L + \frac{1}{2} \left( \beta^2 + \delta^2 \right) N + \frac{1}{2} \left( \alpha^2 + \gamma^2 \right) R + \frac{1}{2} \left( a^2 + \gamma^2 \right) \]
\[ U + \frac{1}{2} \left( \beta^2 + \delta^2 \right) V + \frac{1}{2} \alpha^2 W + \frac{1}{2} \left( \alpha^2 + \beta^2 \gamma \right) X + \frac{1}{2} \alpha^2 \beta^2 Y. \]

\[ D_{n+1} = D_0 + \frac{1}{2} \left( a^2 + \gamma^2 \right) M + \frac{1}{2} \left( \beta^2 + \delta^2 \right) P + \frac{1}{2} \left( \alpha^2 + \gamma^2 \right) Q + \frac{1}{2} \left( \beta^2 + \delta^2 \right) R + \frac{1}{2} \left( a^2 + \gamma^2 \right) S + \frac{1}{2} \left( \beta^2 + \delta^2 \right) T + \frac{1}{2} \left( a^2 + \gamma^2 \right) U + \frac{1}{2} \left( \beta^2 + \delta^2 \right) V + \frac{1}{2} \alpha^2 W + \frac{1}{2} \left( \alpha^2 + \beta^2 \gamma \right) X + \frac{1}{2} \alpha^2 \beta^2 Y. \]

\[ E_{n+1} = E_0 + \frac{1}{2} \alpha^2 \gamma^2 W + \frac{1}{2} \alpha^2 \beta^2 Y + \frac{1}{2} \alpha^2 \beta^2 Y. \]
\[ F_{n+1} = F_0 + \frac{1}{2} \alpha^2 \gamma^2 W + \frac{1}{2} \alpha^2 \beta^2 Y + \frac{1}{2} \alpha^2 \beta^2 Y. \]
\[ G_{n+1} = G_0 + \frac{1}{2} \alpha^2 \gamma^2 (U + V) + \frac{1}{2} \alpha^2 \gamma^2 (W + 2X + Y). \]
\[ H_{n+1} = H_0 + \frac{1}{2} \alpha^2 \gamma^2 (U + V) + \frac{1}{2} \alpha^2 \gamma^2 (W + 2X + Y). \]

<table>
<thead>
<tr>
<th>Typical mating</th>
<th>Number of types</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AABB x AAbb</strong></td>
<td>8</td>
</tr>
<tr>
<td><strong>AAbb x AABb</strong></td>
<td>8</td>
</tr>
<tr>
<td><strong>AABb x AAbb</strong></td>
<td>8</td>
</tr>
<tr>
<td><strong>AABb x AAbb</strong></td>
<td>4</td>
</tr>
<tr>
<td><strong>AAbb x AAbb</strong></td>
<td>4</td>
</tr>
<tr>
<td><strong>AABb x AAbb</strong></td>
<td>4</td>
</tr>
<tr>
<td><strong>AAbb x AAbb</strong></td>
<td>4</td>
</tr>
</tbody>
</table>

\[ I_{n+1} = I_0 + \frac{1}{2} \left( a^2 + \gamma^2 \right) L + \frac{1}{2} \left( \beta^2 + \delta^2 \right) N + \frac{1}{2} \left( \alpha^2 + \gamma^2 \right) R + \frac{1}{2} \left( a^2 + \gamma^2 \right) \]
\[ U + \frac{1}{2} \left( \beta^2 + \delta^2 \right) V + \frac{1}{2} \alpha^2 W + \frac{1}{2} \left( \alpha^2 + \beta^2 \gamma \right) X + \frac{1}{2} \alpha^2 \beta^2 Y. \]

\[ J_{n+1} = J_0 + \frac{1}{2} \left( a^2 + \gamma^2 \right) L + \frac{1}{2} \left( \beta^2 + \delta^2 \right) N + \frac{1}{2} \left( \alpha^2 + \gamma^2 \right) R + \frac{1}{2} \left( a^2 + \gamma^2 \right) \]
\[ (U + V) + \frac{1}{2} \left( \alpha^2 \beta^2 \gamma \right) X + \frac{1}{2} \alpha^2 \beta^2 Y. \]

\[ K_{n+1} = K_0 + \frac{1}{2} \left( a^2 + \gamma^2 \right) L + \frac{1}{2} \left( \beta^2 + \delta^2 \right) N + \frac{1}{2} \left( \alpha^2 + \gamma^2 \right) R + \frac{1}{2} \left( a^2 + \gamma^2 \right) \]
\[ (U + V) + \frac{1}{2} \left( \alpha^2 \beta^2 \gamma \right) X + \frac{1}{2} \alpha^2 \beta^2 Y. \]

\[ L_{n+1} = L_0 + \frac{1}{2} \left( a^2 + \gamma^2 \right) L + \frac{1}{2} \left( \beta^2 + \delta^2 \right) N + \frac{1}{2} \left( \alpha^2 + \gamma^2 \right) R + \frac{1}{2} \left( a^2 + \gamma^2 \right) \]
\[ (U + V) + \frac{1}{2} \left( \alpha^2 \beta^2 \gamma \right) X + \frac{1}{2} \alpha^2 \beta^2 Y. \]

\[ M_{n+1} = M_0 + \frac{1}{2} \left( a^2 + \gamma^2 \right) L + \frac{1}{2} \left( \beta^2 + \delta^2 \right) N + \frac{1}{2} \left( \alpha^2 + \gamma^2 \right) R + \frac{1}{2} \left( a^2 + \gamma^2 \right) \]
\[ (U + V) + \frac{1}{2} \left( \alpha^2 \beta^2 \gamma \right) X + \frac{1}{2} \alpha^2 \beta^2 Y. \]
Omitting some rather tedious algebra, the solution of these equations is:

\[ \xi = \frac{q}{2 - 3q}, \quad \theta = \frac{2q}{2 - 3q}, \quad \kappa = \frac{1}{2 - 3q}, \]

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

as may easily be verified.

\[ \therefore c_\infty = c_n + 2e_n + \frac{1}{1 + 6x} [(1 - 2x)(d_n + 2f_n + 2j_n + \frac{1}{2}k_n) + 2g_n + 4x(h_n + i_n)] \]

(3.4)

and \( y = \frac{1}{3}(1 - c_\infty) \).

In the case considered, \( d_0 = 1 \). \( \therefore c_\infty = \xi d_0 = 1 - 2x / 1 + 6x \). Hence the proportion of crossover zygotes, \( y = \frac{4x}{1 + 6x} \) (3.5).
The “Collaborative Cross”

8-way RILs

Autosomes

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

X chromosome

\[
\begin{align*}
\Pr(G_1 = A) &= \Pr(G_1 = B) = \Pr(G_1 = E) = \Pr(G_1 = F) = 1/6 \\
\Pr(G_1 = C) &= 1/3 \\
\Pr(G_2 = B \mid G_1 = A) &= r/(1 + 4r) \\
\Pr(G_2 = C \mid G_1 = A) &= 2r/(1 + 4r) \\
\Pr(G_2 = A \mid G_1 = C) &= r/(1 + 4r) \\
\Pr(G_2 \neq G_1) &= (14/3)r/(1 + 6r)
\end{align*}
\]
X chromosome

G₀

G₁

G₂

G₃

G₄

G∞
Computer simulations

\[ R = \frac{7r}{1+6r} \]
\[ R = \frac{14/3}{1+4r} \]
3-point coincidence

1 2 3

- \( r_{ij} \) = recombination fraction for interval \((i, j)\)
  Assume \( r_{12} = r_{23} = r \).

- Coincidence = \( c = \frac{\Pr(\text{double recombinant})}{r^2} \)
  = \( \frac{\Pr(\text{rec’n in 23} \mid \text{rec’n in 12})}{\Pr(\text{rec’n in 12})} \)

- No interference = 1
  Positive interference < 1
  Negative interference > 1

- Generally \( c \) is a function of \( r \)
3 points in 2-way RILs

1  2  3

- \( r_{13} = 2r(1 - cr) \)

- \( R = f(r); \quad R_{13} = f(r_{13}) \)

- \( \Pr(\text{double “recombinant” in RIL}) = \{R + R - R_{13}\}/2 \)

- \( \text{Coincidence (in 2-way RIL)} = \{2R - R_{13}\}/\{2R^2\} \)
Why the clustering of breakpoints?

- The really close breakpoint occur in different generations.
- Breakpoints in later generations can occur only in regions that are not yet fixed.
- These regions of heterozygosity are surrounding 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 $\rightarrow$ 22 states by symmetry
  - 3 points in 4-way RILs by sib mating: 2,164,240 parental types $\rightarrow$ 137,488 states by symmetry
- Even counting the states was difficult.
Coincidence

- Meiosis
- 2-way RILs by selfing
- 2-way RILs by sib-mating
- 8-way RILs by sib-mating
- No interference
- Mouse interference

The graph shows the coincidence values for different mating patterns, with the x-axis representing the recombination fraction (r) and the y-axis representing the coincidence values.
But there is an easier way...
### Equations for sib-mating

<table>
<thead>
<tr>
<th>Typical mating</th>
<th>Number of types</th>
</tr>
</thead>
<tbody>
<tr>
<td>AABB × AABB</td>
<td>2</td>
</tr>
<tr>
<td>AAbb × AAbb</td>
<td>2</td>
</tr>
<tr>
<td>AABXaabb</td>
<td>2</td>
</tr>
<tr>
<td>AABB × AAbb</td>
<td>8</td>
</tr>
<tr>
<td>ABB×XaAb</td>
<td>8</td>
</tr>
<tr>
<td>AAB×XAb, ab</td>
<td>8</td>
</tr>
<tr>
<td>AAbb×XAb, ab</td>
<td>8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Typical mating</th>
<th>Number of types</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABB×XaAb</td>
<td>4</td>
</tr>
<tr>
<td>ABB×XAb, ab</td>
<td>4</td>
</tr>
<tr>
<td>ABB×XaAb</td>
<td>4</td>
</tr>
<tr>
<td>ABB×XAb, ab</td>
<td>4</td>
</tr>
</tbody>
</table>

\[
C_{n+1} = C_n + H + \frac{1}{2}(a^2 + \gamma^2)L + \frac{1}{4}(b^2 + \delta^2)N + \frac{1}{2}Q + \frac{1}{2}R + \frac{1}{2}(a^2 + \gamma^2)U + \frac{1}{4}(b^2 + \delta^2)V + \frac{1}{8}a^2 \gamma^2 W + \frac{1}{8}(a^2 \beta^2 + \beta^2 \gamma^2)X + \frac{1}{8}a^2 \beta^2 \gamma^2 Y.
\]

\[
D_{n+1} = D_n + \frac{1}{4}(a^2 + \gamma^2)M + \frac{1}{4}(b^2 + \delta^2)P + \frac{1}{4}Q + \frac{1}{4}S + \frac{1}{4}(b^2 + \delta^2)U + \frac{1}{4}(a^2 + \gamma^2)X + \frac{1}{4}a^2 \beta^2 \gamma^2 Y.
\]

\[
E_{n+1} = \frac{1}{2}a^2 \beta^2 W + \frac{1}{4}(a^2 \beta^2 + \beta^2 \gamma^2)X + \frac{1}{2}a^2 \beta^2 \gamma^2 Y.
\]

\[
F_{n+1} = \frac{1}{2}a^2 \beta^2 W + \frac{1}{2}a^2 \beta^2 \gamma^2 Y.
\]

\[
G_{n+1} = \frac{1}{2}(a^2 + \gamma^2)(U + V) + \frac{1}{2}a^2 \gamma^2 (W + 2X + Y).
\]

\[
H_{n+1} = \frac{1}{2}a^2 \beta^2 (U + V) + \frac{1}{2}a^2 \beta^2 \gamma^2 (W + 2X + Y).
\]

\[
I_{n+1} = \frac{1}{2}a^2 \gamma^2 X + \frac{1}{2}(a^2 + \gamma^2)Y + \frac{1}{2}a^2 \gamma^2 (W + 2X + Y).
\]
Consider the cross $W_1 W_2 | X_1 X_2 \times Y_1 Y_2 | Z_1 Z_2$

Let 
\begin{align*}
q_1 &= \Pr(W_1 W_2 \text{ fixed}) \\
q_2 &= \Pr(W_1 X_2 \text{ fixed}) \\
q_3 &= \Pr(W_1 Y_2 \text{ fixed})
\end{align*}

First generation: $W_i \equiv X_i \equiv A, Y_i \equiv Z_i \equiv B$

Then \(\Pr(\text{AA fixed}) = 2(q_1 + q_2)\)

\(\Pr(\text{AB fixed}) = 4q_3\)
The simpler method

\[ W_1 W_2 | X_1 X_2 \times Y_1 Y_2 | Z_1 Z_2 \]

\[ q_1 = \Pr(W_1 W_2 \text{ fixed}) \quad q_2 = \Pr(W_1 X_2 \text{ fixed}) \quad q_3 = \Pr(W_1 Y_2 \text{ fixed}) \]

**Second generation:** \( W_i \equiv Y_i \equiv A, \quad X_i \equiv Z_i \equiv B \)

Then \( \Pr(AA \text{ fixed}) = 2(q_1 + q_3) \)

Thus \( q_2 = q_3 \)
The simpler method

\[ W_1 W_2 | X_1 X_2 \times Y_1 Y_2 | Z_1 Z_2 \]

\[ q_1 = \Pr(W_1 W_2 \text{ fixed}) \quad q_2 = \Pr(W_1 X_2 \text{ fixed}) \quad q_3 = \Pr(W_1 Y_2 \text{ fixed}) \]

Now we use the usual trick, condition on the first step:

\[ q_1 = \left( \frac{1 - r}{2} \times q_1 \times 4 \right) + \left[ \frac{1}{2} \times \frac{1}{2} \times q_2 \times 12 \right] \]

Combined with the previous results, we obtain

\[ q_2 = \frac{r}{2(1 + 6r)} \]

And so \( \Pr(AB \text{ fixed}) = 4q_3 = \frac{4r}{1 + 6r} \)
The formula

\[ C = \frac{(1 + 6r)[280 + 1208r - 848r^2 + 5c(7 - 28 - 368r^2 + 344r^3) - 2c^2(49 - 324r + 452r^2)r^2 - 16c^3(1 - 2r)r^4]}{49(1 + 12r - 12cr^2)(5 + 10r - 4(2 + c)r^2 + 8cr^3)} \]
3-point symmetry

$$\Pr(M_2 = x \mid M_1 = A, M_2 \neq A, M_3 = A)$$
Markov property

\[ \log_2 \left\{ \frac{\Pr(M_3 = A \mid M_2 = A, M_1 = x)}{\Pr(M_3 = A \mid M_2 = A)} \right\} \]
Markov property

\[ \log_2 \left\{ \frac{\Pr(M_3 = A \mid M_2 = B, M_1 = x)}{\Pr(M_3 = A \mid M_2 = B)} \right\} \]
Markov property

$$\log_2 \left\{ \frac{\Pr(M_3 = A \mid M_2 = C, M_1 = x)}{\Pr(M_3 = A \mid M_2 = C)} \right\}$$
Markov property

\[
\log_2 \left\{ \frac{\Pr(M_3 = A \mid M_2 = E, M_1 = x)}{\Pr(M_3 = A \mid M_2 = E)} \right\}
\]
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 complex fixation
• 10,000 simulation replicates
No. generations to fixation

- 2-way selfing: mean = 10.5
- 2-way sib-mating: mean = 35.6
- 8-way sib-mating: mean = 38.9
No. generations
to 99% fixation

- 2-way selfing: mean = 8.0
- 2-way sib-mating: mean = 23.5
- 8-way sib-mating: mean = 26.7
Percent genome not fixed

![Graph showing percent genome not yet fixed over no. generations for different mating systems.]

- 2-way selfing
- 2-way sib-mating
- 8-way sib-mating
- Average
- 95th percentile
Segment lengths

- Median for 2-way selfing: 23.7 cM
- Median for 2-way sib-mating: 12.9 cM
- Median for 8-way sib-mating: 8.5 cM

Graph showing distribution of segment lengths for different mating systems.
Segment lengths

- 2-way selfing: median = 23.7 cM
- 2-way sib-mating: median = 12.9 cM
- 8-way sib-mating: median = 8.5 cM

Two chromosomes

X chromosome

Segment lengths (cM)
Probability a segment is inherited intact

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

Length of segment (cM) vs. Probability segment inherited intact.
Length of smallest segment

95th %ile = 0.26 cM
95th %ile = 0.58 cM
95th %ile = 2.2 cM
No. segments < 1 cM

mean = 1.4
mean = 5.2
mean = 11.2

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

- 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: $R = \frac{7r}{1 + 6r}$

- We’ve shown clustering of breakpoints in RILs by sib-mating, even in the presence of strong crossover interference.


  Teuscher F, Broman KW. Haplotype probabilities for multiple-strain recombinant inbred lines. *Genetics*, to appear
Friedrich Teuscher
Research Institute for the Biology of Farm Animals
Dummerstorf, Germany