The genomes of recombinant inbred lines

Karl W Broman
Department of Biostatistics
Johns Hopkins University

http://www.biostat.jhsph.edu/~kbroman
C57BL/6

The intercross
Recombinant inbred lines
(by sibling mating)

The RIX design
The Collaborative Cross

Genome of an 8-way RI
The Collaborative Cross

The goal
(for the rest of this talk)

- 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.
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 alleles at the two loci on a random RIL chromosome.

Haldane & Waddington 1931

INBREEDING AND LINKAGE

J. B. S. HALDANE AND C. H. WADDINGTON
John Innes Horticultural Institution, London, England

Received August 9, 1931

Genetics 16:357-374

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. See-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. See-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 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.
Recombinant inbred lines
(by selfing)

Markov chain

- Sequence of random variables \( \{X_0, X_1, X_2, \ldots \} \) satisfying
  \[
  \Pr(X_{n+1} \mid X_0, X_1, \ldots, X_n) = \Pr(X_{n+1} \mid X_n)
  \]
- Transition probabilities \( P_{ij} = \Pr(X_{n+1} = j \mid X_n = i) \)
- Here, \( X_n \) = “parental type” at generation \( n \)
- We are interested in absorption probabilities
  \[
  \Pr(X_n \rightarrow j \mid X_0)
  \]
Equations for selfing

\[ \begin{align*} 
C_n &= AABB \text{ and } aabb, \\
D_n &= AA BB \text{ and } aa BB, \\
E_n &= A B B B, A a BB, A a B B, \text{ and } a a BB, \\
F_n &= A B a b, \\
G_n &= A B a B. 
\end{align*} \]

We assume \( 2C_n + 3D_n + 3E_n + F_n + G_n = 2 \), so that \( C_1 = D_1 = E_1 = G_1 = 0 \), and \( F_1 = 2 \). 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:

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

(1.1)

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

\[
\begin{align*}
\therefore C_n + D_n &= 1, \\
C_n - D_n &= c_n \\
\therefore y &= \frac{1}{1 + 2x} .
\end{align*}
\]

(1.3)

Absorption probabilities

Let \( P_{ij} = \Pr(X_{n+1} = j \mid X_n = i) \) where \( X_n \) = state at generation \( n \).

Consider the case of absorption into the state \( AA|AA \).

Let \( h_i \) = probability, starting at \( i \), eventually 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.
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:

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

as may easily be verified.

\[ \therefore c_m = c_n + \frac{1}{1 + 6x} \left[ (1 - 2x)(d_n + 2f_n + 2) + \frac{1}{2} k_n \right] \]
\[ + 2g_n + 4x(h_n + l_n) \]  
(3.4)

and \( y = \frac{1}{4}(1 - c_m) \).

In the case considered, \( d_n = 1 \) : \( c_m = d_n = 1 - 2x/1 + 6x \). Hence the proportion of crossover zygotes \( y = 4x / 1 + 6x \)  
(3.5).

The “Collaborative Cross”
8-way RILs

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

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

Computer simulations

\[
\begin{align*}
R &= 7r / (1+6r) \\
R &= (14/3)r / (1+4r)
\end{align*}
\]
The X chromosome

3-point coincidence

- $r_{ij} =$ recombination fraction for interval $i,j$; assume $r_{12} = r_{23} = r$
- Coincidence $= c = \Pr(\text{double recombinant}) / r^2$
  $= \Pr(\text{rec'n in 23} \mid \text{rec'n in 12}) / \Pr(\text{rec'n in 23})$
- 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

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

Coincidence

No interference
Coincidence

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.
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

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

![Graph showing log probability ratio against recombination fraction for different values of x (A, B, C, D, E).]

Markov property

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

![Graph showing log probability ratio against recombination fraction for different values of x (A, B, C, D, E).]
Markov property

\[ \log_2 \left( \frac{\Pr(M_3 = A | M_2 = E, M_1 = x)}{\Pr(M_3 = A | M_2 = E)} \right) \]

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 complete fixation
- 10,000 simulation replicates
No. generations to fixation

- Mean = 10.5
- Mean = 35.6
- Mean = 38.9

No. gen’s to 99% fixation

- Mean = 8.0
- Mean = 23.5
- Mean = 26.7
Percent genome not fixed

Number of breakpoints
Segment lengths

Probability a segment is inherited intact
Length of smallest segment

No. segments < 1 cM
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.