The genomes of recombinant inbred lines

Karl W Broman
Department of Biostatistics
Johns Hopkins University

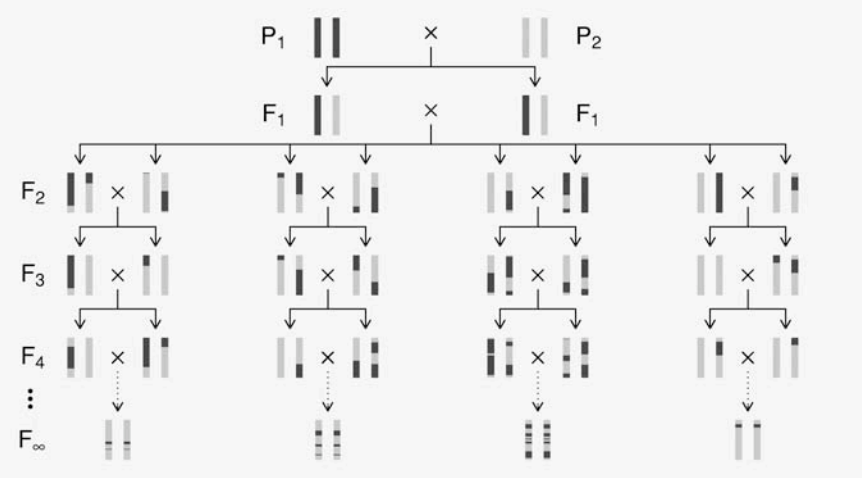
http://www.biostat.jhsph.edu/~kbroman

C57BL/6

2

Recombinant inbred lines

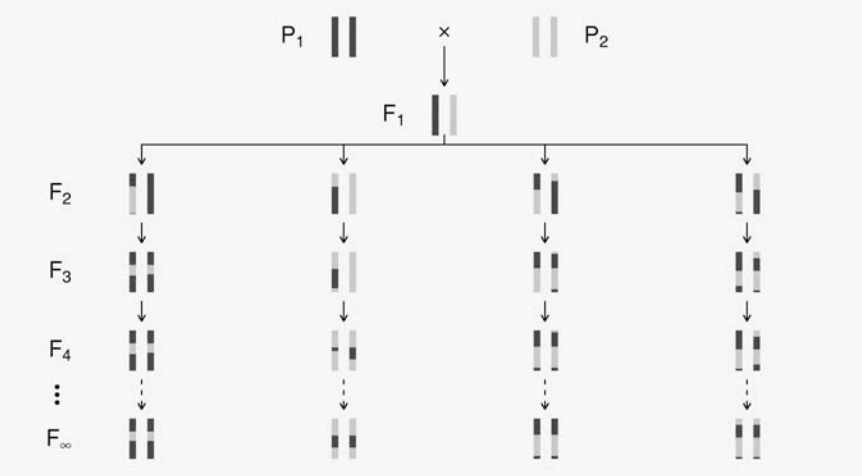
(by sibling mating)



3

Recombinant inbred lines

(by selfing)



4

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.

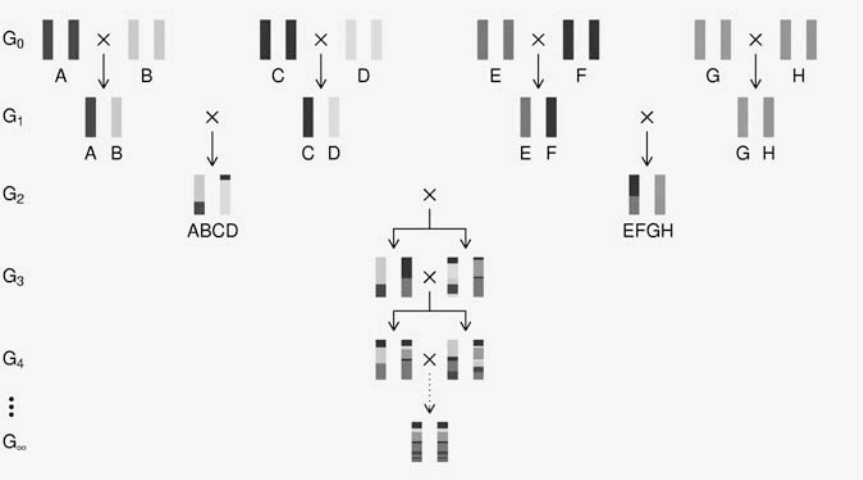
5

Disadvantages of RI lines

- Expensive and time consuming to create.
- The available panels are too small.
- Learn only about 2 alleles.

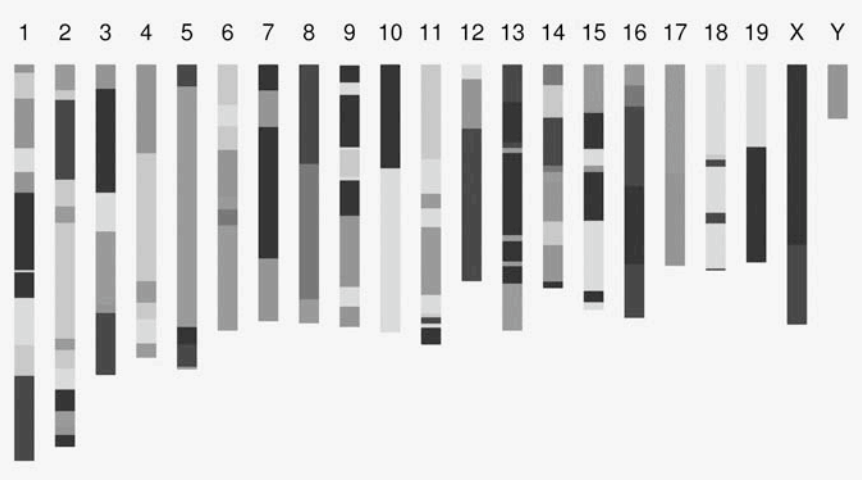
6

The “Collaborative Cross”



7

Genome of an 8-way RI



8

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.

9

Why?

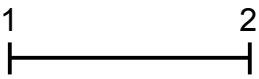
- It's interesting.
- Later statistical analyses will require:
  - The two-point probabilities.
  - A model for the whole process.

Actually, we'll probably just assume that:

- The breakpoints follow a Poisson process.
- The genotypes follow a Markov chain.

10

2 points in an RIL



- 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.

11

Haldane & Waddington 1931

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

	PAGE
Self-fertilization	358
Brother-sister mating. Sex-linked genes	360
Brother-sister mating. Autosomal genes	364
Parent and offspring mating. Sex-linked genes	367
Parent and offspring mating. Autosomal genes	368
Inbreeding with any initial population	370
Double crossing over	372
DISCUSSION	373
SUMMARY	374
LITERATURE CITED	374

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.

12

Equations for selfing

C_n *AABB* and *aabb*.
D_n *AAbb* and *aaBB*.
E_n *AABb*, *AaBB*, *Aabb*, and *aaBb*.
F_n *AB.ab*.
G_n *Ab.aB*.

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_∞ is the final proportion of crossover zygotes. Then considering the results of selfing each generation, we have:

$$\left.\begin{aligned}
C_{n+1} &= C_n + \tfrac{1}{2}E_n + \tfrac{1}{4}(1-\beta-\delta+\beta\delta)F_n + \tfrac{1}{4}\beta\delta G_n \\
D_{n+1} &= D_n + \tfrac{1}{2}E_n + \tfrac{1}{4}\beta\delta F_n + \tfrac{1}{4}(1-\beta-\delta+\beta\delta)G_n \\
E_{n+1} &= \tfrac{1}{2}E_n + \tfrac{1}{4}(\beta+\delta-2\beta\delta)(F_n+G_n) \\
F_{n+1} &= \tfrac{1}{2}(1-\beta-\delta+\beta\delta)F_n + \tfrac{1}{2}\beta\delta G_n \\
G_{n+1} &= \tfrac{1}{2}\beta\delta F_n + \tfrac{1}{2}(1-\beta-\delta+\beta\delta)G_n
\end{aligned}\right\} \quad (1.1)$$

or C_{n+1}, D_{n+1}, and F_{n+1}, G_{n+1},

d_n } (1.2)

r all values of n.
$- 2x)d_n$

$= \dfrac{1-2x}{1+2x}$.

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

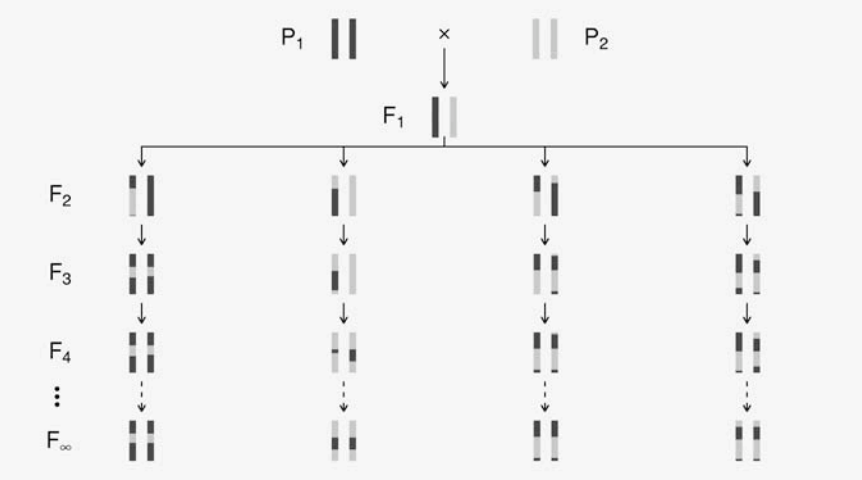
$\therefore\ C_\infty + D_\infty = 1,\ C_\infty - D_\infty = c_\infty\ \therefore\ y = \tfrac{1}{2}(1-c_\infty).$

$$\therefore\ y = \frac{2x}{1+2x}. \quad (1.3)$$

13

Recombinant inbred lines

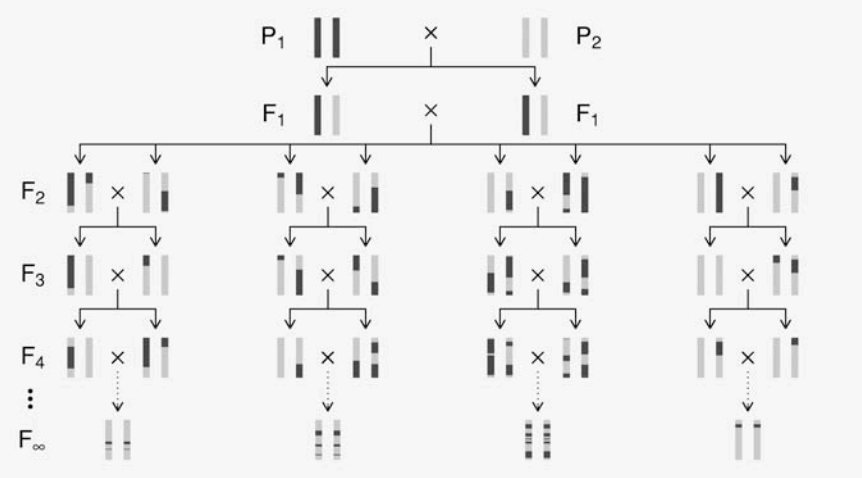
(by selfing)



14

Recombinant inbred lines

(by sibling mating)



15

Equations for sib-mating

Typical mating	*Number of types*	
AABB×*AABB*	2	$C_{n+1}=C_n+H+\frac{1}{4}(\alpha^2+\gamma^2)L+\frac{1}{4}(\beta^2+\delta^2)N+\frac{1}{8}Q+\frac{1}{8}R+\frac{1}{8}(\alpha^2+\gamma^2)$ $U+\frac{1}{8}(\beta^2+\delta^2)V+\frac{1}{16}\alpha^2\gamma^2W+\frac{1}{16}(\alpha^2\delta^2+\beta^2\gamma^2)X+\frac{1}{16}\beta^2\delta^2Y.$
AAbb×*AAbb*	2	$D_{n+1}=D+I+\frac{1}{4}(\alpha^2+\gamma^2)M+\frac{1}{4}(\beta^2+\delta^2)P+\frac{1}{8}Q+\frac{1}{8}S+\frac{1}{8}(\beta^2+\delta^2)$ $U+\frac{1}{8}(\alpha^2+\gamma^2)V+\frac{1}{16}\beta^2\delta^2W+\frac{1}{16}(\alpha^2\delta^2+\beta^2\gamma^2)X+\frac{1}{16}\alpha^2\gamma^2Y.$
AABB×*aabb*	2	$E_{n+1}=\frac{1}{16}\alpha^2\gamma^2W+\frac{1}{16}(\alpha^2\delta^2+\beta^2\gamma^2)X+\frac{1}{16}\beta^2\delta^2Y.$
AAbb×*aaBB*	2	$F_{n+1}=\frac{1}{16}\beta^2\delta^2W+\frac{1}{16}(\alpha^2\delta^2+\beta^2\gamma^2)X+\frac{1}{16}\alpha^2\gamma^2Y.$
AABB×*AAbb*	8	$G_{n+1}=\frac{1}{16}(\alpha\beta+\gamma\delta)(U+V)+\frac{1}{16}\alpha\beta\gamma\delta(W+2X+Y).$
AABB×*AABb*	8	$H_{n+1}=\frac{1}{2}H+$ $U+\frac{1}{16}($ $(\alpha\delta+\beta$
AAbb×*AABb*	8	$I_{n+1}=\frac{1}{2}I+$ $U+\frac{1}{16}($ $(\alpha\delta+\beta$
AABB×*Aabb*	8	$J_{n+1}=\frac{1}{16}($ $\beta\delta)(\alpha\delta$
AAbb×*AaBB*	8	$K_{n+1}=\frac{1}{16}$ $\beta\delta)(\alpha\delta$
AABB×*AB.ab*	4	$L_{n+1}=\frac{1}{4}(\alpha$ $\alpha^2\gamma^2W$
AAbb×*Ab.aB*	4	$M_{n+1}=\frac{1}{4}($ $\beta^2\delta^2W$

Typical mating	*Number of types*	
AABB×*Ab.aB*	4	$N_{n+1}=\frac{1}{2}R+\frac{1}{8}(\alpha\beta+\gamma\delta)(U+V)+\frac{1}{8}\alpha\beta\gamma\delta(W+2X+Y).$
AAbb×*AB.ab*	4	$P_{n+1}=\frac{1}{2}S+\frac{1}{8}(\alpha\beta+\gamma\delta)(U+V)+\frac{1}{8}\alpha\beta\gamma\delta(W+2X+Y).$
AABb×*AABB*	4	$Q_{n+1}=2G+\frac{1}{2}(H+I+J+K)+\frac{1}{4}(\alpha^2+\gamma^2)(L+M)+\frac{1}{4}(\beta^2+\delta^2)$ $(N+P)+\frac{1}{16}Q+\frac{1}{16}(R+S+T)+\frac{1}{16}(\alpha^2+\alpha\beta+\beta^2+\gamma^2+\gamma\delta+\delta^2)$ $(U+V)+\frac{1}{16}(\alpha\delta+\beta\gamma)^2(W+Y)+\frac{1}{16}(\alpha\gamma+\beta\delta)^2X.$
AABb×*AaBB*	4	$R_{n+1}=\frac{1}{16}(\beta^2+\delta^2)L+\frac{1}{16}(\alpha^2+\gamma^2)N+\frac{1}{16}R+\frac{1}{16}(\beta+\delta)U+\frac{1}{16}(\alpha+\gamma)V+$ $\frac{1}{16}(\alpha\delta+\beta\gamma)^2(W+Y)+\frac{1}{16}(\alpha\gamma+\beta\delta)^2X.$
AABb×*Aabb*	4	$S_{n+1}=\frac{1}{16}(\beta^2+\delta^2)M+\frac{1}{16}(\alpha^2+\gamma^2)P+\frac{1}{16}S+\frac{1}{16}(\alpha+\gamma)U+\frac{1}{16}(\beta+\delta)V+\frac{1}{16}$ $(\alpha\delta+\beta\gamma)^2(W+Y)+\frac{1}{16}(\alpha\gamma+\beta\delta)^2X.$
AABb×*aaBb*	4	$T_{n+1}=\frac{1}{16}(\alpha\beta+\gamma\delta)(U+V)+\frac{1}{16}(\alpha\delta+\beta\gamma)^2(W+V)+\frac{1}{16}(\alpha\gamma+\beta\delta)^2X.$
AABb×*AB.ab*	8	$U_{n+1}=\frac{1}{2}J+\frac{1}{16}(\alpha\beta+\gamma\delta)(L+N)+\frac{1}{16}(S+T)+\frac{1}{16}(\alpha+\gamma)U+\frac{1}{16}(\beta+\delta)$ $V+\frac{1}{16}\alpha\gamma(\beta\gamma+\alpha\delta)W+\frac{1}{16}(\alpha\gamma+\beta\delta)(\alpha\delta+\beta\gamma)X+\frac{1}{16}\beta\delta(\beta\gamma+\alpha\delta)Y.$
AABb×*Ab.aB*	8	$V_{n+1}=\frac{1}{2}K+\frac{1}{16}(\alpha\beta+\gamma\delta)(M+P)+\frac{1}{16}(R+T)+\frac{1}{16}(\beta+\delta)U+\frac{1}{16}(\alpha+\gamma)$ $V+\frac{1}{16}\beta\delta(\beta\gamma+\alpha\delta)W+\frac{1}{16}(\alpha\gamma+\beta\delta)(\alpha\delta+\beta\gamma)X+\frac{1}{16}\alpha\gamma(\beta\gamma+\alpha\delta)Y.$
AB.ab×*AB.ab*	1	$W_{n+1}=2(E+J)+\frac{1}{4}(\alpha^2+\gamma^2)L+\frac{1}{4}(\beta^2+\delta^2)N+\frac{1}{4}(S+T)+\frac{1}{4}(\alpha^2+\gamma^2)$ $U+\frac{1}{4}(\beta^2+\delta^2)V+\frac{1}{16}\alpha^2\gamma^2W+\frac{1}{16}(\alpha^2\delta^2+\beta^2\gamma^2)X+\frac{1}{16}\beta^2\delta^2Y.$
AB.ab×*Ab.aB*	2	$X_{n+1}=\frac{1}{2}T+\frac{1}{16}(\alpha\beta+\gamma\delta)(U+V)+\frac{1}{16}\alpha\beta\gamma\delta(W+2X+Y).$
Ab.aB×*Ab.aB*	1	$Y_{n+1}=2(F+K)+\frac{1}{4}(\alpha^2+\gamma^2)M+\frac{1}{4}(\beta^2+\delta^2)P+\frac{1}{4}(R+T)+\frac{1}{4}(\beta^2+$ $\delta^2)U+\frac{1}{16}(\alpha^2+\gamma^2)V+\frac{1}{16}\beta^2\delta^2W+\frac{1}{16}(\alpha^2\delta^2+\beta^2\gamma^2)X+\frac{1}{16}\alpha^2\gamma^2Y.$

16

Result for sib-mating

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

$$\zeta = \frac{q}{2-3q}, \quad \theta = \frac{2q}{2-3q}, \quad \kappa = \frac{1}{2-3q},$$

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

as may easily be verified.

$$\therefore \; c_\infty = c_n + 2e_n + \frac{1}{1+6x}[(1-2x)(d_n + 2f_n + 2j_n + \tfrac{1}{2}k_n) + 2g_n + 4x(h_n + i_n)] \qquad (3.4)$$

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

In the case considered, $d_0 = 1, \therefore c_\infty = \zeta d_0 = 1-2x/1+6x$. Hence the proportion of crossover zygotes, $y = 4x/1+6x$ (3.5).

17

Haldane & Waddington 1931

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

Autosomes

$Pr(G_1=A) = Pr(G_1=B) = 1/2$

$Pr(G_2=B \mid G_1=A) = Pr(G_2=A \mid G_1=B) = 4r / (1+6r)$

X chromosome

$Pr(G_1=A) = 2/3$ $\quad Pr(G_1=B) = 1/3$

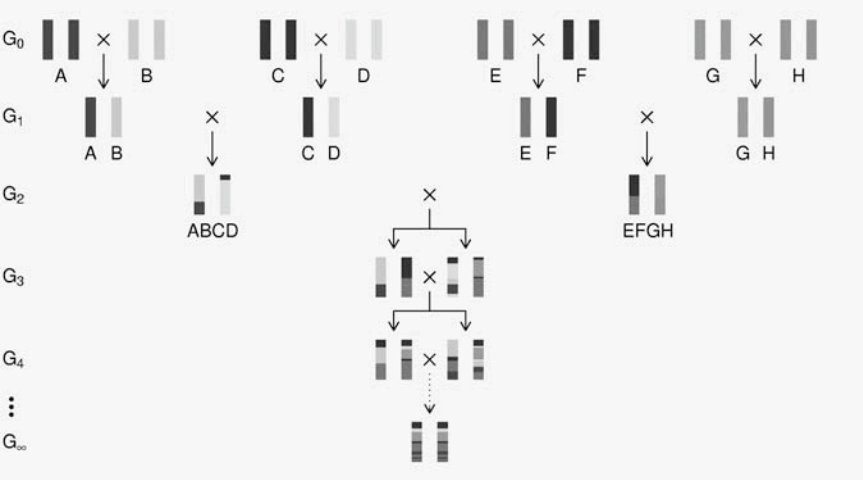
$Pr(G_2=B \mid G_1=A) = 2r / (1+4r)$

$Pr(G_2=A \mid G_1=B) = 4r / (1+4r)$

$Pr(G_2 \neq G_1) = (8/3)\, r / (1+4r)$

18

The “Collaborative Cross”



19

8-way RILs

Autosomes

$Pr(G_1 = i) = 1/8$

$Pr(G_2 = j \mid G_1 = i) = r / (1+6r)$ for $i \neq j$

$Pr(G_2 \neq G_1) = 7r / (1+6r)$

X chromosome

$Pr(G_1=A) = Pr(G_1=B) = Pr(G_1=E) = Pr(G_1=F) = 1/6$

$Pr(G_1=C) = 1/3$

$Pr(G_2=B \mid G_1=A) = r / (1+4r)$

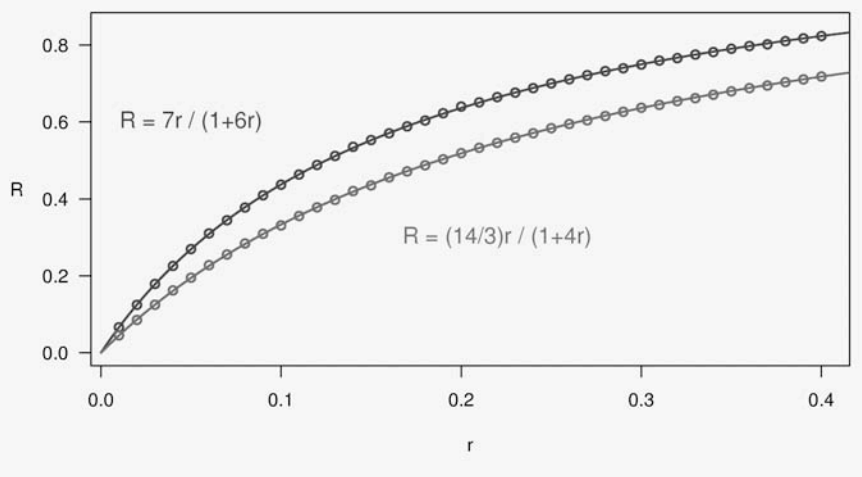
$Pr(G_2=C \mid G_1=A) = 2r / (1+4r)$

$Pr(G_2=A \mid G_1=C) = r / (1+4r)$

$Pr(G_2 \neq G_1) = (14/3)\, r / (1+4r)$

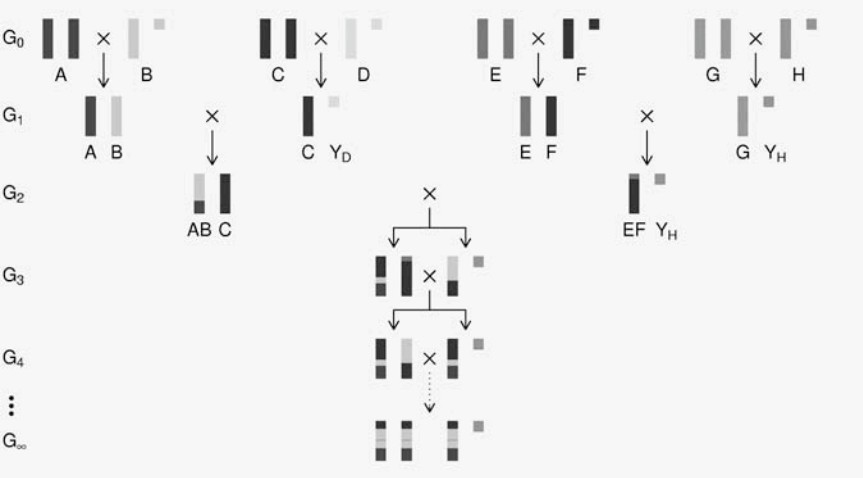
20

Computer simulations



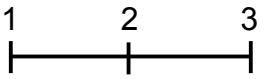
21

The X chromosome



22

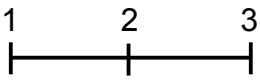
3-point coincidence



- r_{ij} = recombination fraction for interval i,j;
  assume $r_{12} = r_{23} = r$
- Coincidence = c = Pr(double recombinant) / r^2
  = 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.

23

3-points in 2-way RILs

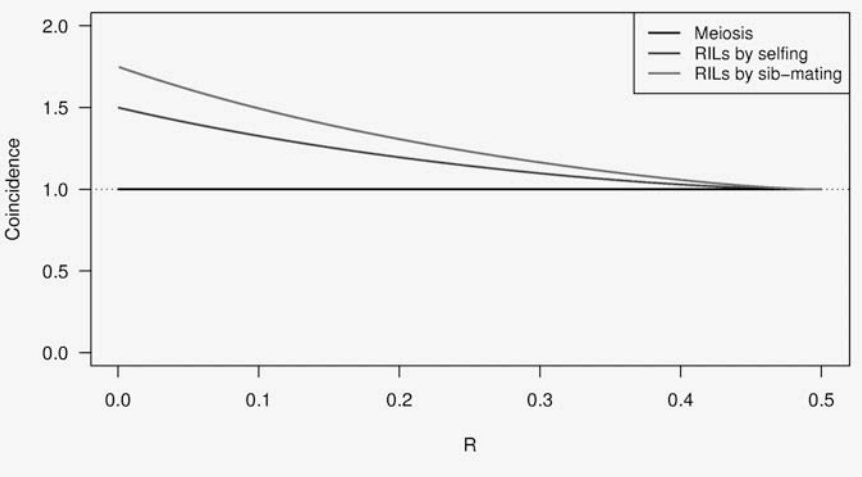


- $r_{13} = 2\ r\ (1 - c\ r)$
- $R = f(r)$; $R_{13} = f(r_{13})$
- Pr(double recombinant in RIL) = $\{ R + R - R_{13} \} / 2$
- Coincidence (in 2-way RIL) = $\{ 2\ R - R_{13} \} / \{ 2\ R^2 \}$

24

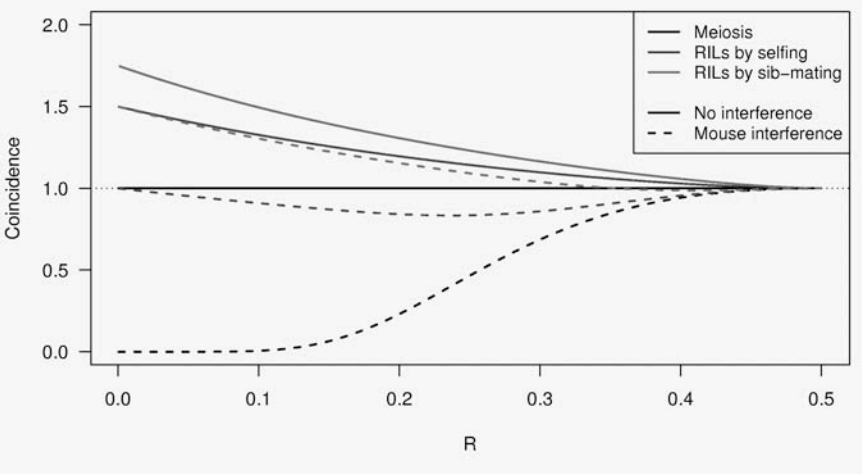
Coincidence

No interference



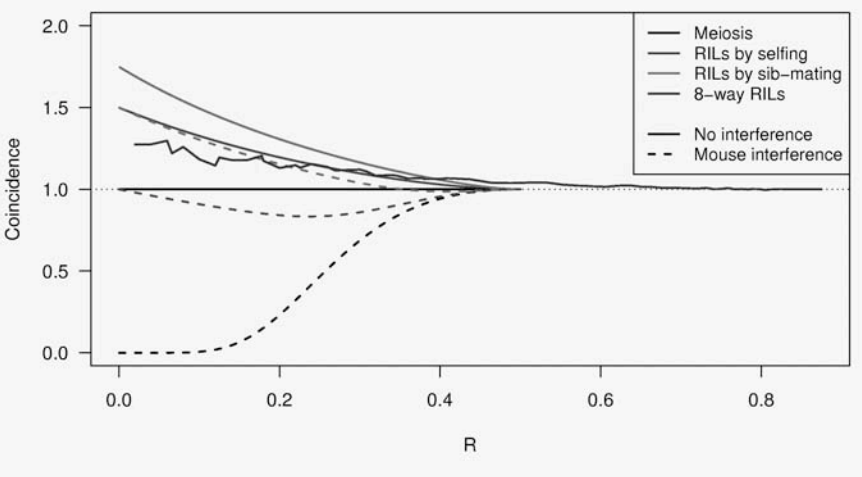
25

Coincidence



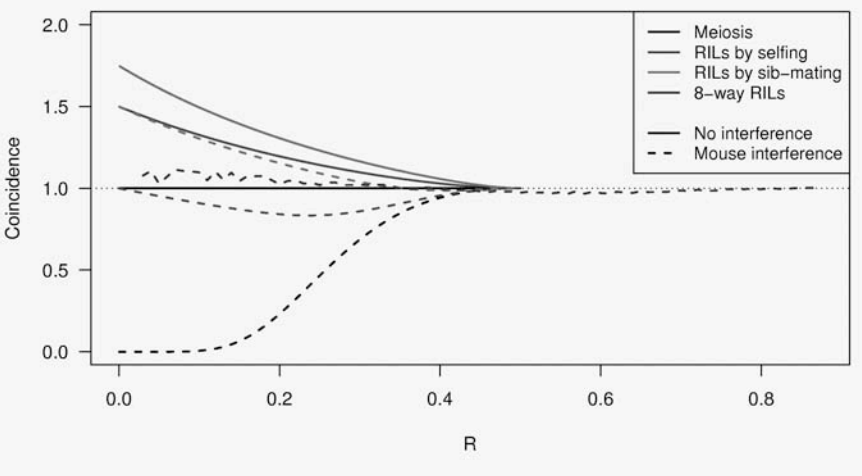
26

Coincidence



27

Coincidence



28

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.

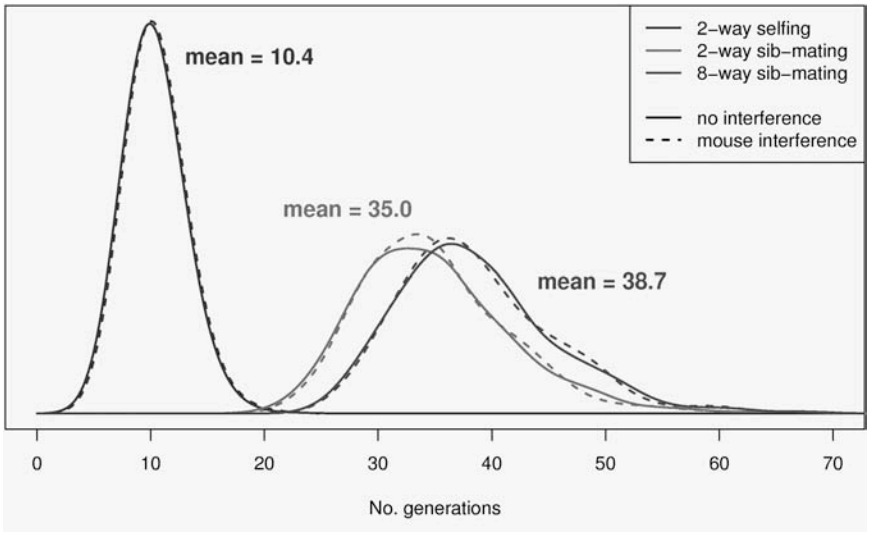
29

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

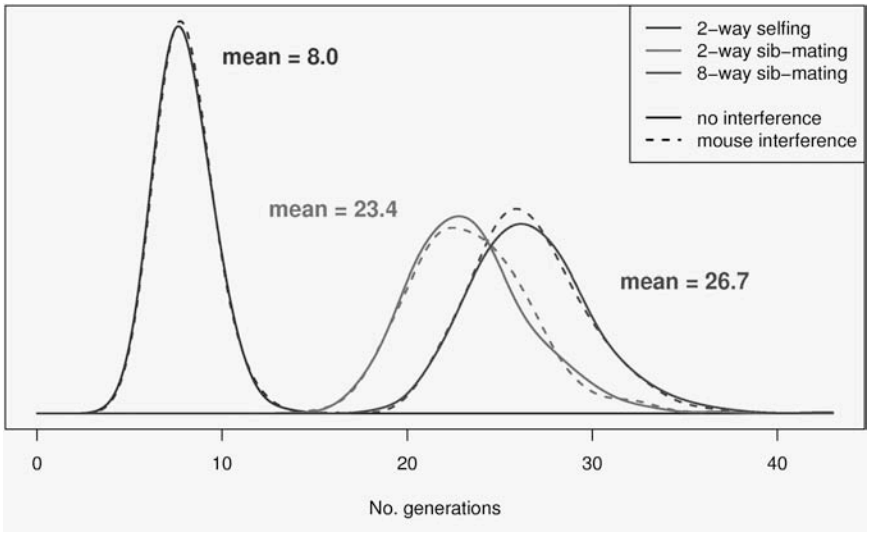
30

No. generations to fixation



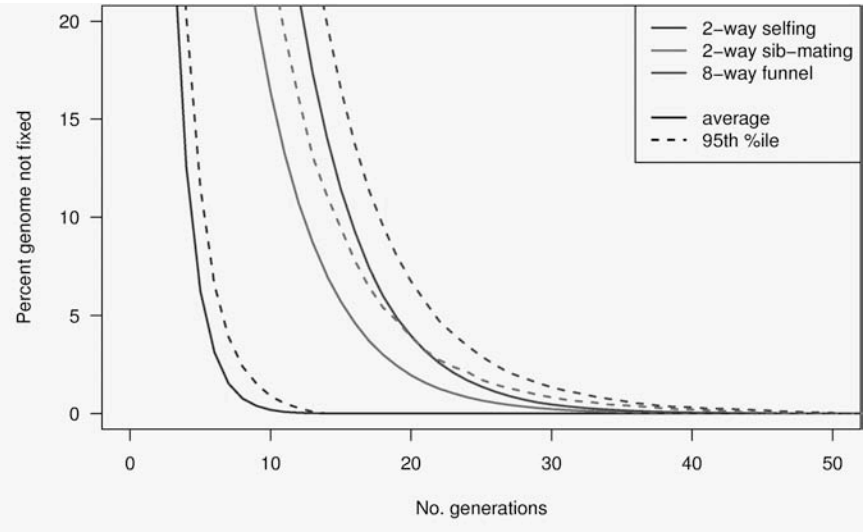
31

No. gen's to 99% fixation



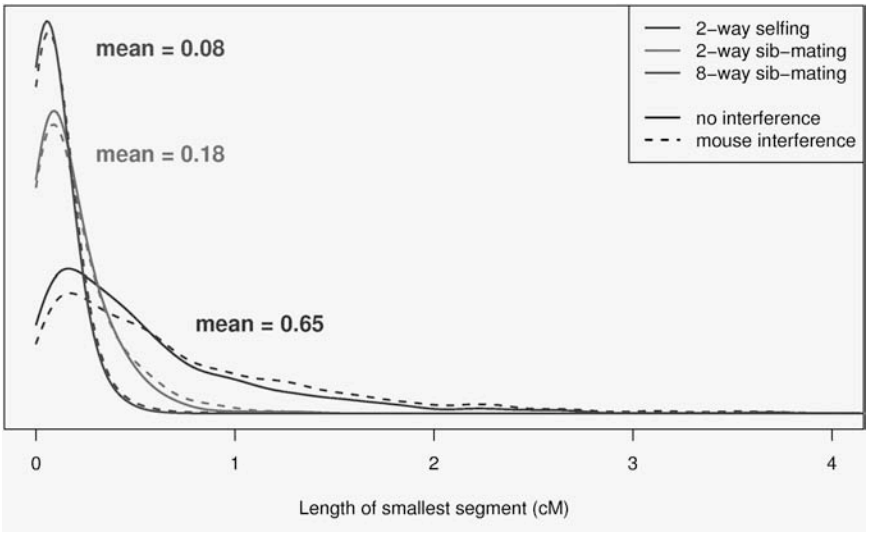
32

Percent genome not fixed



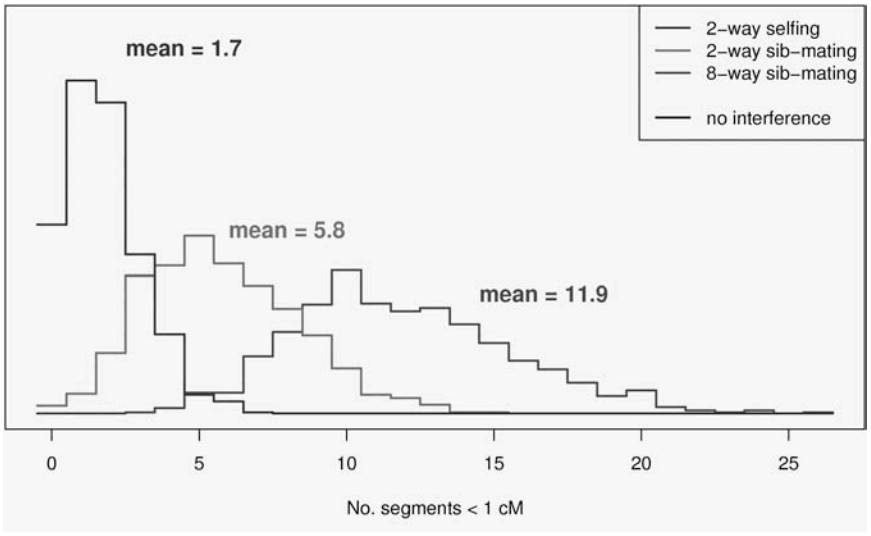
33

Length of smallest segment



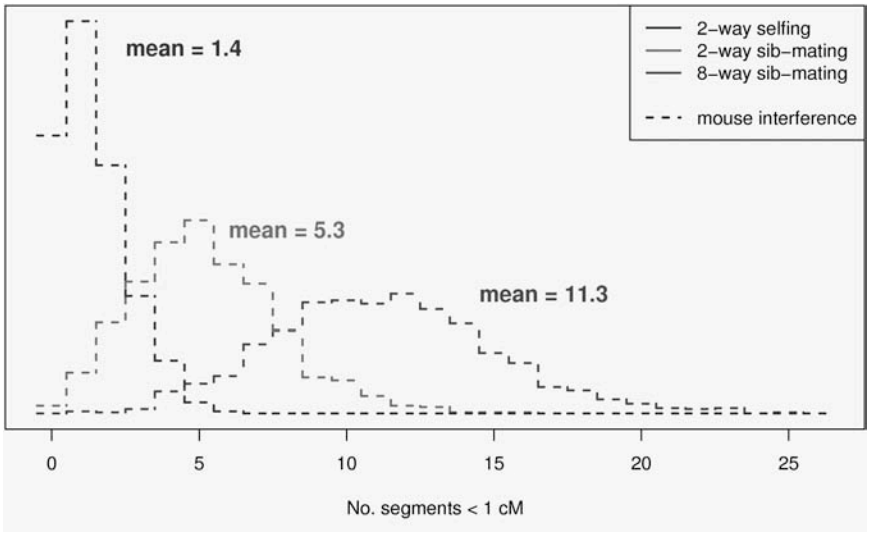
34

No. segments < 1 cM



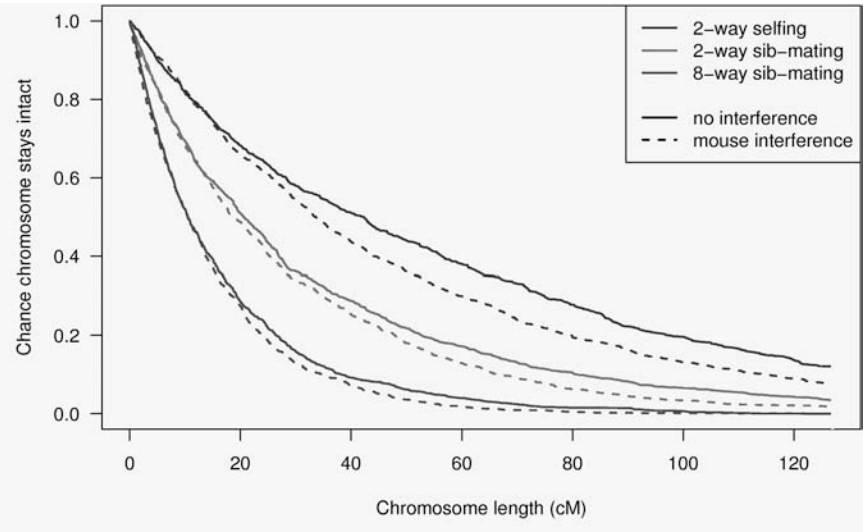
35

No. segments < 1 cM



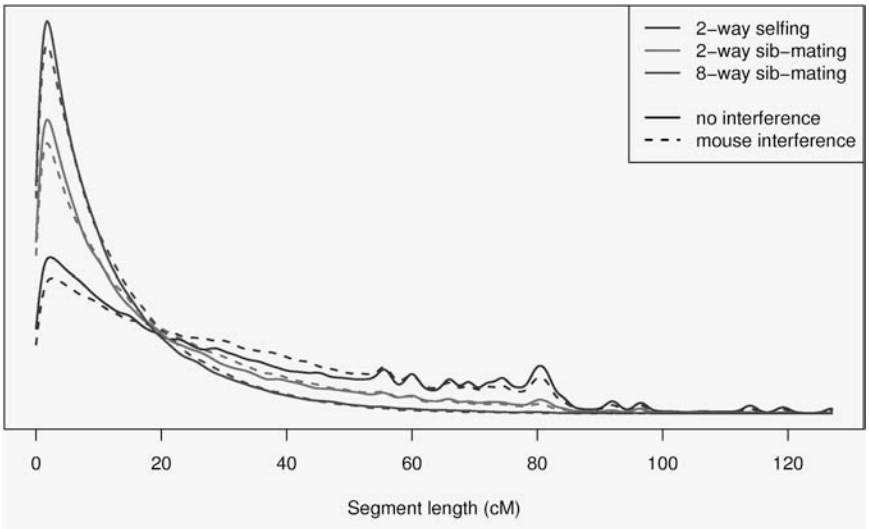
36

Probability that a chromosome is intact



37

Segment lengths



38

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.

39

The key points

- $R = 7\,r / (1 + 6\,r)$
- 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.

40