The breakpoint process on RI chromosomes

Karl W Broman
Department of Biostatistics
Johns Hopkins University

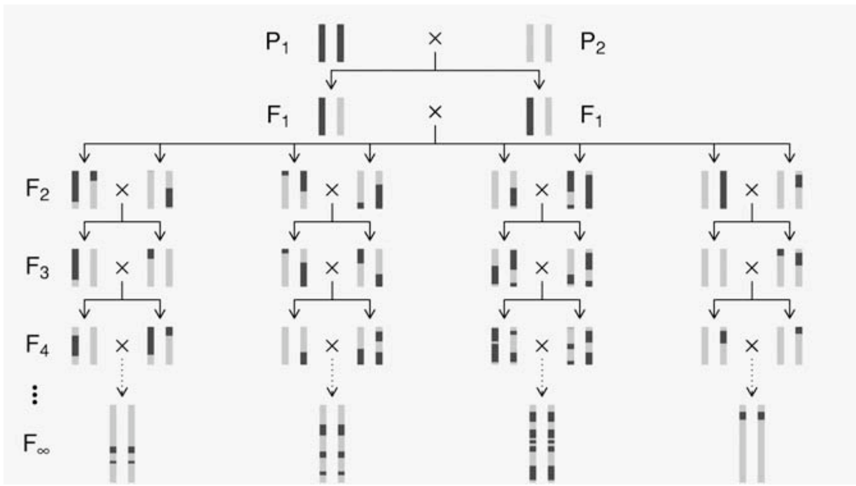
http://www.biostat.jhsph.edu/~kbroman

C57BL/6

2

Recombinant inbred lines

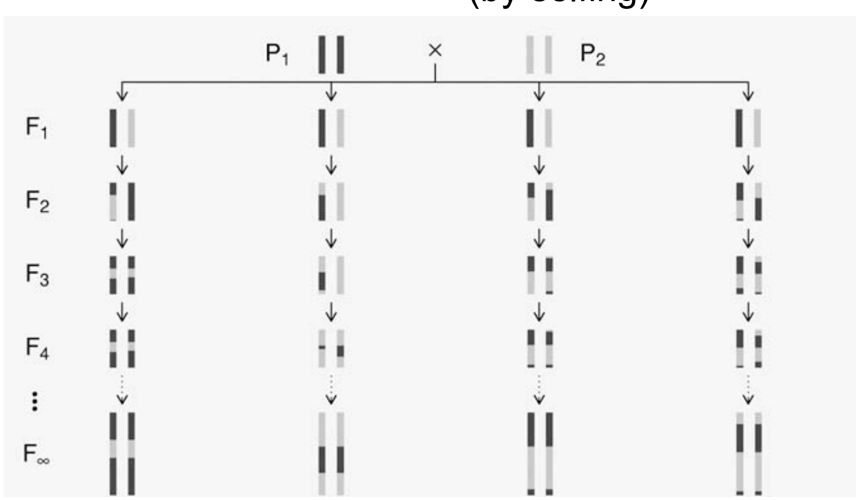
(by sibling mating)



3

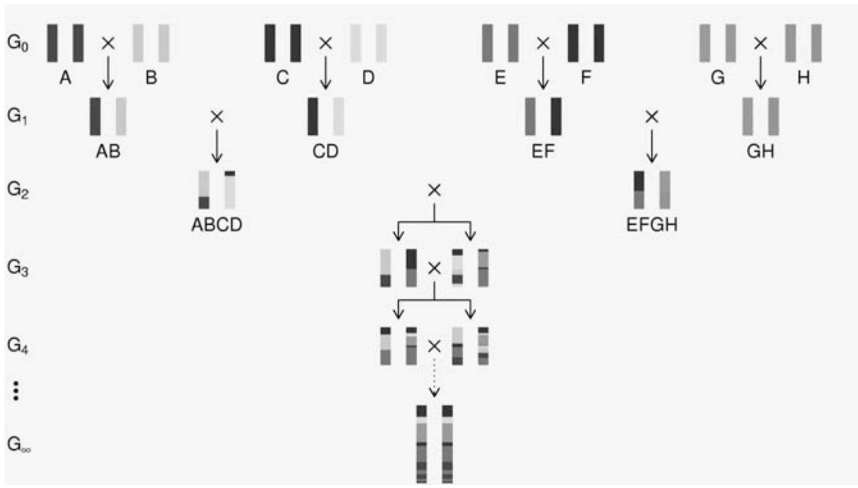
Recombinant inbred lines

(by selfing)



4

The “Collaborative Cross”



5

Genome of an 8-way RI

6

The goal

- Characterize the breakpoint process along a chromosome in 8-way RILs.
  - Understand the two-point transition matrix.
  - Study the clustering of the breakpoints, as a function of crossover interference in meiosis.
- Why?
  - It's interesting.
  - Statistical analyses of such 8-way RIL data will require a model for this breakpoint process.

7

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.

8

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

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.

9

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.
...- $2x)d_n$

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

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

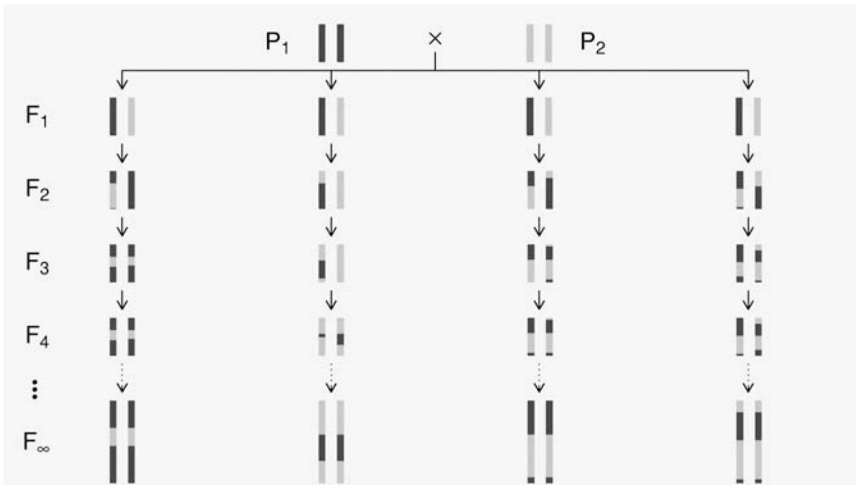
$$\therefore\ C_\infty + D_\infty = 1,\ C_\infty - D_\infty = c_\infty\ \therefore\ y = \tfrac{1}{2}(1-c_\infty).$$

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

10

Recombinant inbred lines

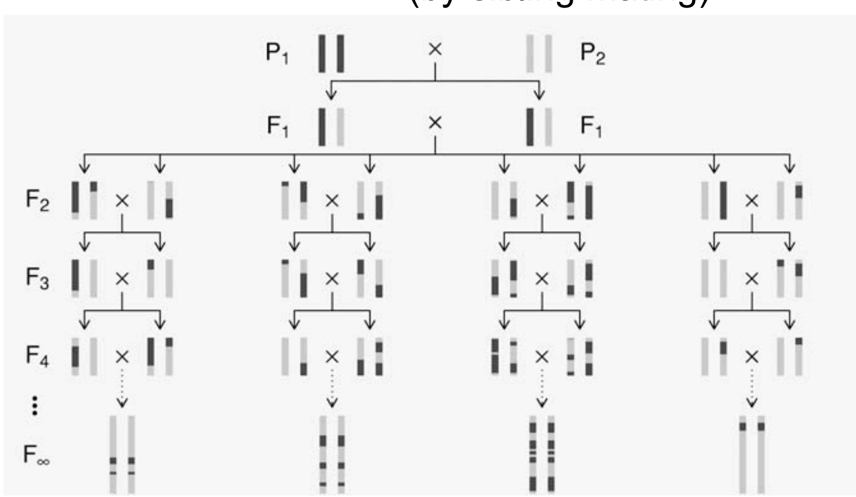
(by selfing)



11

Recombinant inbred lines

(by sibling mating)



12

Equations for sib-mating

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

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

13

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).

14

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

$Pr(G_1=A) = 2/3$ $\quad Pr(G_1=B) = 1/3$
$Pr(G_2=B \mid G_1=A) = 2r / (1+4r)$
$Pr(G_2=A \mid G_1=B) = 4r / (1+4r)$
$Pr(G_2 \neq G_1) = (8/3)\, r / (1+4r)$

15

8-way RILs

Autosomes

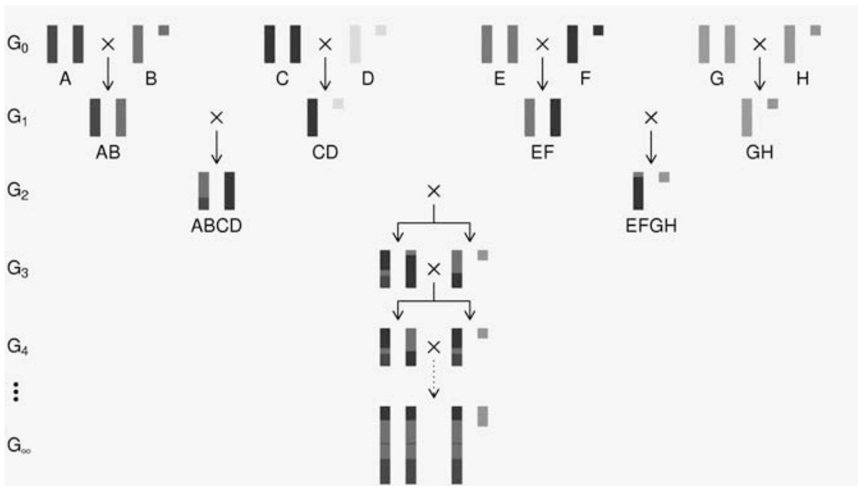
$Pr(G_1 = i) = 1/8$
$Pr(G_2 = j \mid G_1 = i) = r / (1+6r)$ for $i \neq j$
$Pr(G_2 \neq G_1) = 7r / (1+6r)$

X chromosome

$Pr(G_1=A) = Pr(G_1=B) = Pr(G_1=E) = Pr(G_1=F) = 1/6$
$Pr(G_1=C) = 1/3$
$Pr(G_2=B \mid G_1=A) = r / (1+4r)$
$Pr(G_2=C \mid G_1=A) = 2r / (1+4r)$
$Pr(G_2=A \mid G_1=C) = r / (1+4r)$
$Pr(G_2 \neq G_1) = (14/3)\, r / (1+4r)$

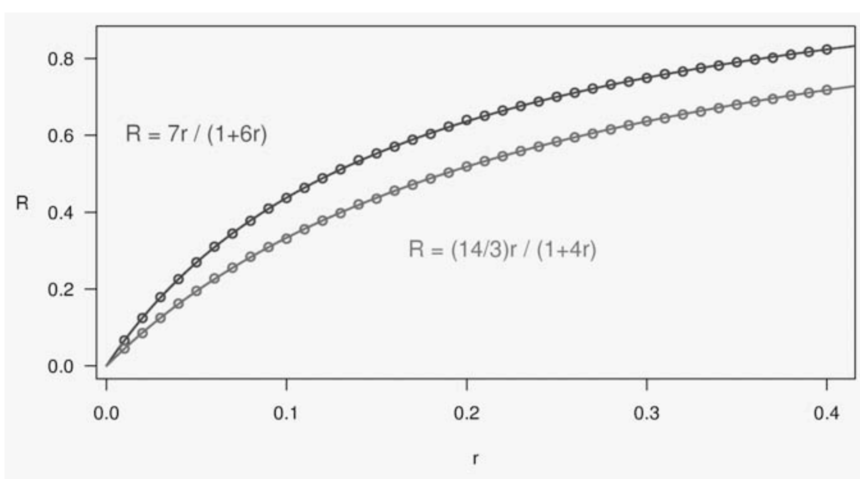
16

The X chromosome



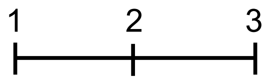
17

Computer simulations



18

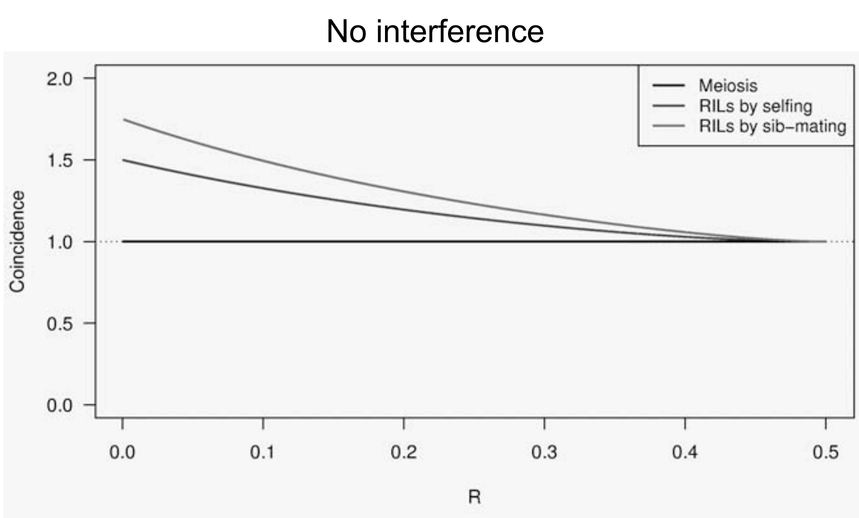
3 points in an RIL



- r_{ij} = recombination fraction for interval i,j
- Coincidence = Pr(rec'n in both intervals) / (r_{12} r_{23})
  = Pr(rec'n in 23 | rec'n in 12) / Pr(rec'n in 23)
- No interference → = 1
  Positive interference → < 1
  Negative interference → > 1
- H&W result + map function →
  coincidence on a 2-way RIL chromosome

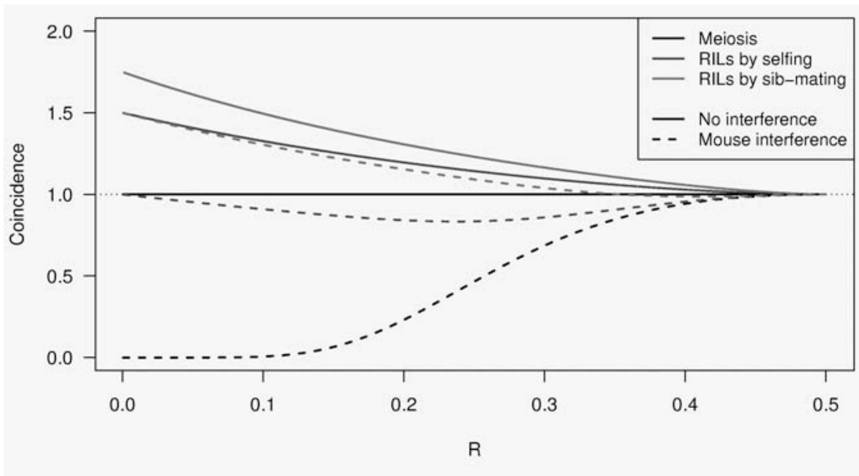
19

Coincidence



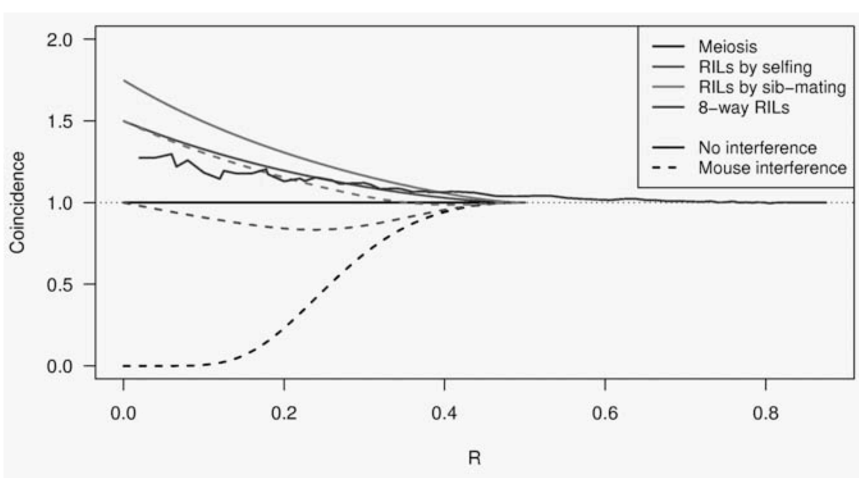
20

Coincidence



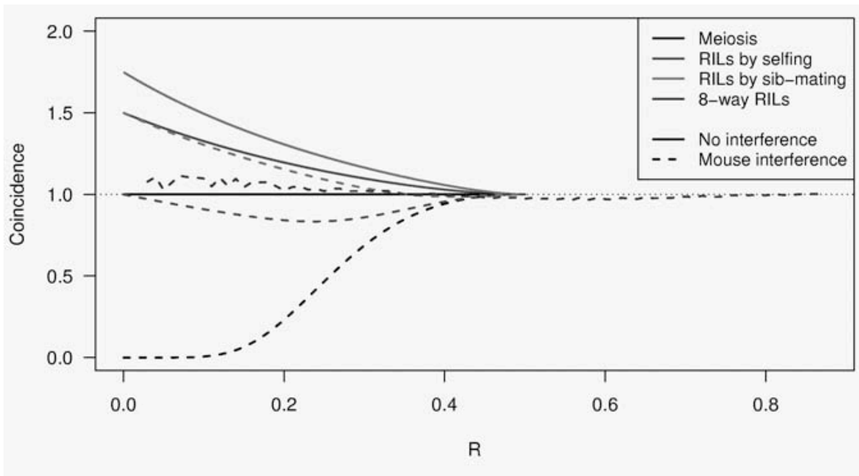
21

Coincidence



22

Coincidence



23

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.

24

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, but need to prove our results.
- 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’d like to:
  - Fully characterize the breakpoint process.
  - Study data (e.g., on 2-way RILs via selfing in Arabidopsis).

25