

The genomes of recombinant inbred lines

Karl W Broman

Department of Biostatistics
Johns Hopkins University

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

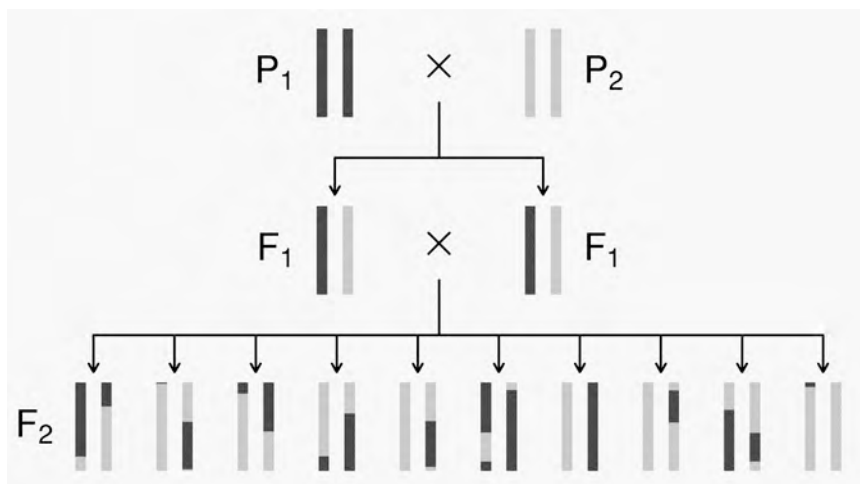


C57BL/6



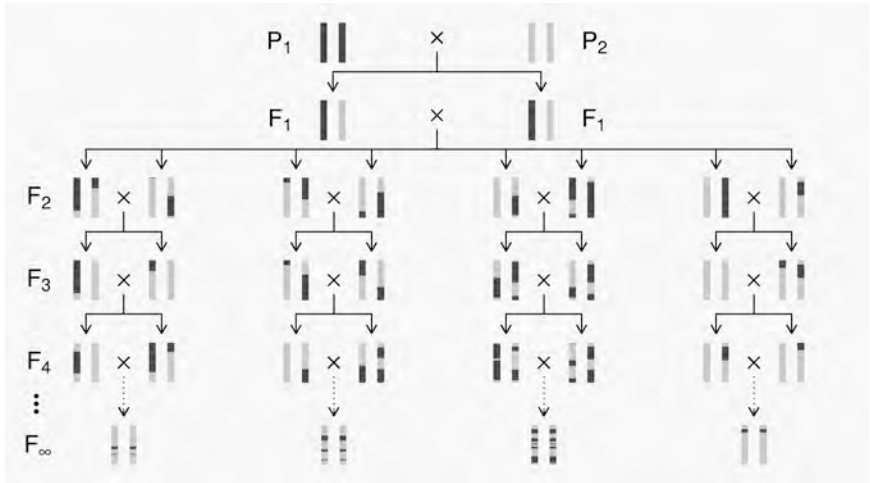
3

The intercross



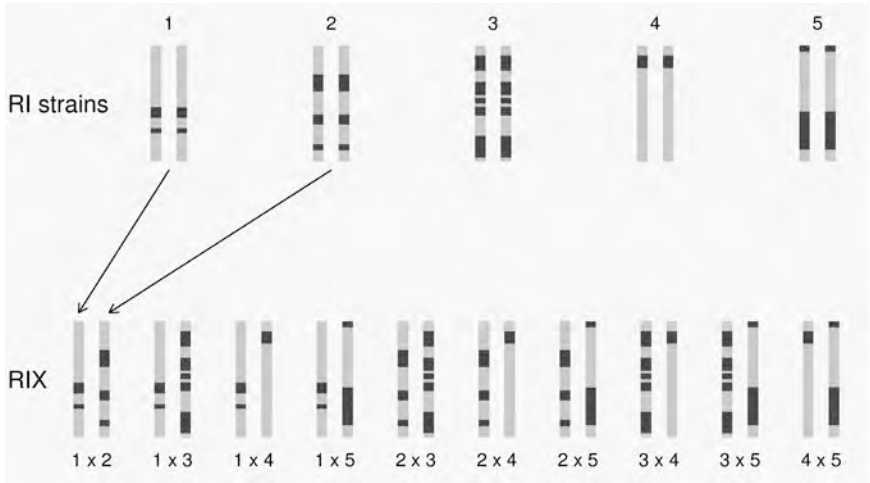
4

Recombinant inbred lines (by sibling mating)



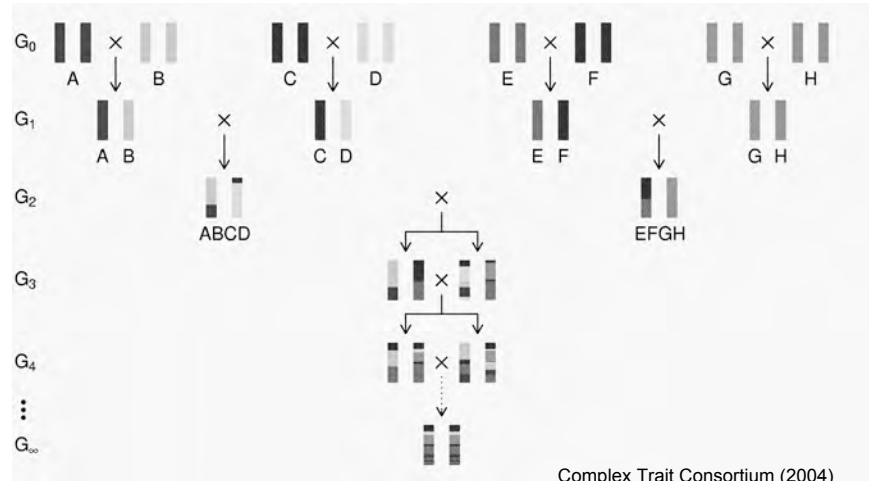
5

The RIX design



6

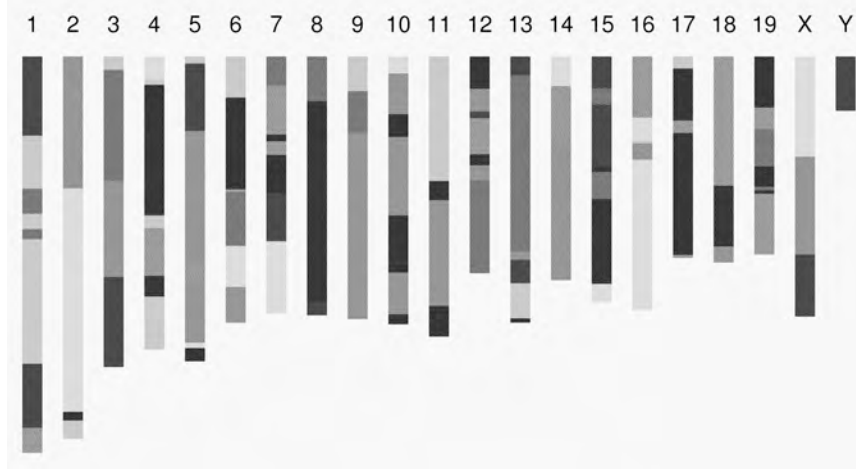
The Collaborative Cross



Complex Trait Consortium (2004)
Nat Genet 36:1133-1137

7

Genome of an 8-way RI



8

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.

9

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 alleles at the two loci on a random RIL chromosome.

10

Haldane & Waddington 1931

INBREEDING AND LINKAGE*

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

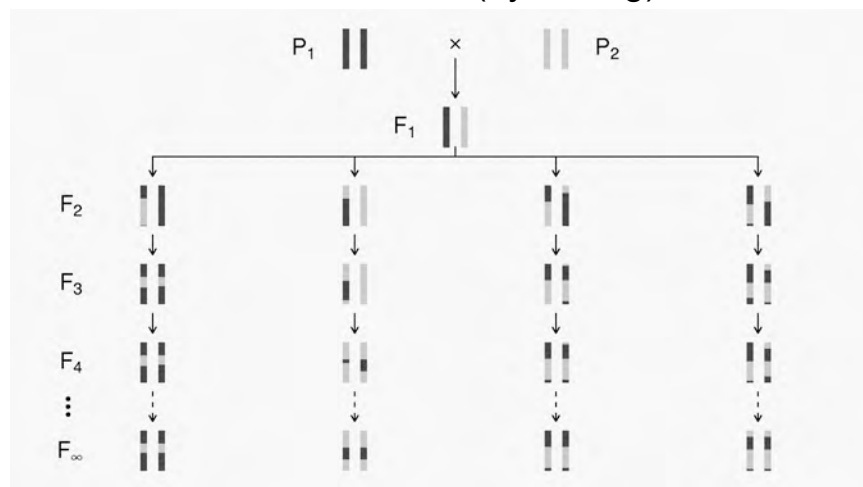
Received August 9, 1930 Genetics 16:357-374

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.

Recombinant inbred lines (by selfing)



Markov chain

- Sequence of random variables $\{X_0, X_1, X_2, \dots\}$ satisfying $\Pr(X_{n+1} | X_0, X_1, \dots, X_n) = \Pr(X_{n+1} | X_n)$
- Transition probabilities $P_{ij} = \Pr(X_{n+1}=j | X_n=i)$
- Here, $X_n =$ "parental type" at generation n
- We are interested in absorption probabilities

$$\Pr(X_n \rightarrow j | X_0)$$

13

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:

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

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

$$d_n \} \quad (1.2)$$

for all values of n .

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

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

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

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

14

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.

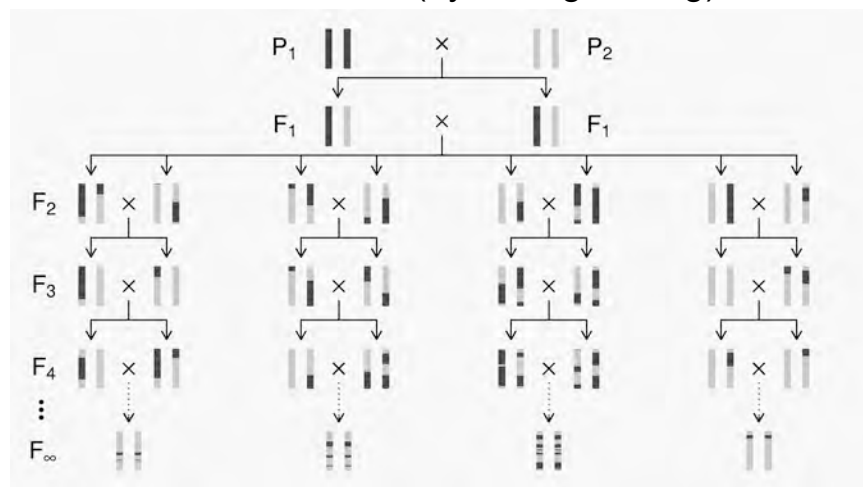
Then $h_{AA|AA} = 1$ and $h_{AB|AB} = 0$.

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

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

15

Recombinant inbred lines (by sibling mating)



16

Equations for sib-mating

Typical mating	Number of types	Equation
$AABB \times AABB$	2	$C_{n+1} = C_n + H + J(\alpha^2 + \gamma^2)L + \frac{1}{2}(\beta^2 + \delta^2)N + \frac{1}{2}Q + \frac{1}{2}R + \frac{1}{2}(\alpha^2 + \gamma^2)U + \frac{1}{2}(\beta^2 + \delta^2)V + \frac{1}{2}\alpha\alpha^2\gamma^2W + \frac{1}{2}(\alpha^2\beta^2 + \beta^2\gamma^2)X + \frac{1}{2}\alpha\beta^2\delta^2Y.$
$AAbb \times AAbb$	2	$D_{n+1} = D + I + \frac{1}{2}(\alpha^2 + \gamma^2)M + \frac{1}{2}(\beta^2 + \delta^2)P + \frac{1}{2}Q + \frac{1}{2}S + \frac{1}{2}(\beta^2 + \delta^2)U + \frac{1}{2}(\alpha^2 + \gamma^2)V + \frac{1}{2}\alpha\beta^2\delta^2W + \frac{1}{2}(\alpha^2\beta^2 + \beta^2\gamma^2)X + \frac{1}{2}\alpha\alpha^2\gamma^2Y.$
$AABB \times aabb$	2	$E_{n+1} = \frac{1}{2}\alpha\alpha^2\gamma^2W + \frac{1}{2}(\alpha^2\beta^2 + \beta^2\gamma^2)X + \frac{1}{2}\alpha\beta^2\delta^2Y.$
$AAbb \times aaBB$	2	$F_{n+1} = \frac{1}{2}\alpha\beta^2\delta^2W + \frac{1}{2}(\alpha^2\beta^2 + \beta^2\gamma^2)X + \frac{1}{2}\alpha\alpha^2\gamma^2Y.$
$AABB \times AAbb$	8	$G_{n+1} = \frac{1}{2}(\alpha\beta + \gamma\delta)(U + V) + \frac{1}{2}\alpha\alpha^2\beta\delta(W + 2X + Y).$
$AAbb \times AABB$	8	$H_{n+1} = \frac{1}{2}(\alpha\delta + \beta\gamma)(U + V) + \frac{1}{2}\alpha\alpha^2\beta\delta(W + 2X + Y).$
$AAbb \times ABab$	4	$I_{n+1} = \frac{1}{2}(\alpha\delta + \beta\gamma)(U + V) + \frac{1}{2}\alpha\alpha^2\beta\delta(W + 2X + Y).$
$AABB \times Abab$	4	$J_{n+1} = \frac{1}{2}(\alpha\delta + \beta\gamma)(U + V) + \frac{1}{2}\alpha\alpha^2\beta\delta(W + 2X + Y).$
$AAbb \times AbBB$	4	$K_{n+1} = \frac{1}{2}(\alpha\delta + \beta\gamma)(U + V) + \frac{1}{2}\alpha\alpha^2\beta\delta(W + 2X + Y).$
$AABB \times ABab$	4	$L_{n+1} = \frac{1}{2}(\alpha\delta + \beta\gamma)(U + V) + \frac{1}{2}\alpha\alpha^2\beta\delta(W + 2X + Y).$
$AAbb \times AbaB$	4	$M_{n+1} = \frac{1}{2}(\alpha\delta + \beta\gamma)(U + V) + \frac{1}{2}\alpha\alpha^2\beta\delta(W + 2X + Y).$
$AABB \times AbaB$	4	$N_{n+1} = \frac{1}{2}(\alpha\delta + \beta\gamma)(U + V) + \frac{1}{2}\alpha\alpha^2\beta\delta(W + 2X + Y).$
$AAbb \times ABab$	4	$P_{n+1} = \frac{1}{2}(\alpha\delta + \beta\gamma)(U + V) + \frac{1}{2}\alpha\alpha^2\beta\delta(W + 2X + Y).$
$AAbb \times AAbb$	4	$Q_{n+1} = 2G + \frac{1}{2}(H + I + J + K) + \frac{1}{2}(\alpha^2 + \gamma^2)(L + M) + \frac{1}{2}(\beta^2 + \delta^2)(N + P) + \frac{1}{2}Q + \frac{1}{2}(R + S + T) + \frac{1}{2}(\alpha^2 + \alpha\beta + \beta^2 + \gamma^2 + \gamma\delta + \delta^2)(U + V) + \frac{1}{2}\alpha(\alpha\delta + \beta\gamma)(W + Y) + \frac{1}{2}(\alpha\gamma + \beta\delta)X.$
$AAbb \times AbBB$	4	$R_{n+1} = \frac{1}{2}(\beta^2 + \delta^2)L + \frac{1}{2}(\alpha^2 + \gamma^2)N + \frac{1}{2}(\beta^2 + \delta^2)U + \frac{1}{2}(\alpha + \gamma)V + \frac{1}{2}\alpha(\alpha\delta + \beta\gamma)(W + Y) + \frac{1}{2}(\alpha\gamma + \beta\delta)X.$
$AAbb \times Abbb$	4	$S_{n+1} = \frac{1}{2}(\beta^2 + \delta^2)M + \frac{1}{2}(\alpha^2 + \gamma^2)P + \frac{1}{2}(\alpha + \gamma)U + \frac{1}{2}(\beta + \delta)V + \frac{1}{2}\alpha(\alpha\delta + \beta\gamma)(W + Y) + \frac{1}{2}(\alpha\gamma + \beta\delta)X.$
$AAbb \times aaBb$	4	$T_{n+1} = \frac{1}{2}(\alpha\delta + \beta\gamma)(U + V) + \frac{1}{2}\alpha\alpha^2\beta\delta(W + Y) + \frac{1}{2}(\alpha\gamma + \beta\delta)X.$
$AAbb \times ABab$	8	$U_{n+1} = \frac{1}{2}(\alpha\delta + \beta\gamma)(U + V) + \frac{1}{2}\alpha\alpha^2\beta\delta(W + Y) + \frac{1}{2}(\alpha\gamma + \beta\delta)X.$
$AAbb \times AbaB$	8	$V_{n+1} = \frac{1}{2}K + \frac{1}{2}(\alpha\delta + \beta\gamma)(M + P) + \frac{1}{2}(R + T) + \frac{1}{2}(\beta + \delta)U + \frac{1}{2}(\alpha + \gamma)V + \frac{1}{2}\alpha\beta(\beta\gamma + \alpha\delta)W + \frac{1}{2}(\alpha\gamma + \beta\delta)(\alpha\delta + \beta\gamma)X + \frac{1}{2}\alpha\alpha^2\beta\delta(W + Y) + \frac{1}{2}(\alpha\gamma + \beta\delta)X.$
$ABab \times ABab$	1	$W_{n+1} = 2(E + J) + \frac{1}{2}(\alpha^2 + \gamma^2)L + \frac{1}{2}(\beta^2 + \delta^2)N + \frac{1}{2}(S + T) + \frac{1}{2}(\alpha^2 + \gamma^2)U + \frac{1}{2}(\beta^2 + \delta^2)V + \frac{1}{2}\alpha\alpha^2\gamma^2W + \frac{1}{2}(\alpha^2\beta^2 + \beta^2\gamma^2)X + \frac{1}{2}\alpha\beta^2\delta^2Y.$
$ABab \times AbaB$	2	$X_{n+1} = \frac{1}{2}T + \frac{1}{2}(\alpha\delta + \gamma\delta)(U + V) + \frac{1}{2}\alpha\alpha^2\beta\delta(W + 2X + Y).$
$Abab \times AbaB$	1	$Y_{n+1} = 2(F + K) + \frac{1}{2}(\alpha^2 + \gamma^2)M + \frac{1}{2}(\beta^2 + \delta^2)P + \frac{1}{2}(R + T) + \frac{1}{2}(\beta^2 + \delta^2)U + \frac{1}{2}(\alpha\delta + \beta\gamma)V + \frac{1}{2}\alpha\beta^2\delta^2W + \frac{1}{2}(\alpha^2\beta^2 + \beta^2\gamma^2)X + \frac{1}{2}\alpha\alpha^2\gamma^2Y.$

17

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.

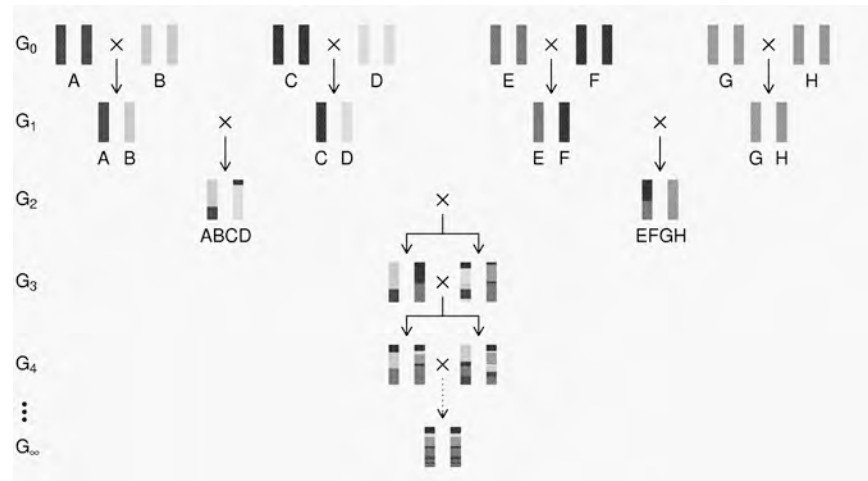
$$\therefore c_{\infty} = c_n + 2e_n + \frac{1}{1 + 6\kappa} [(1 - 2\kappa)(d_n + 2f_n + 2j_n + \frac{1}{2}k_n) + 2g_n + 4x(h_n + i_n)] \quad (3.4)$$

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

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

18

The "Collaborative Cross"



19

8-way RILs

Autosomes

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

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

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

X chromosome

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

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

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

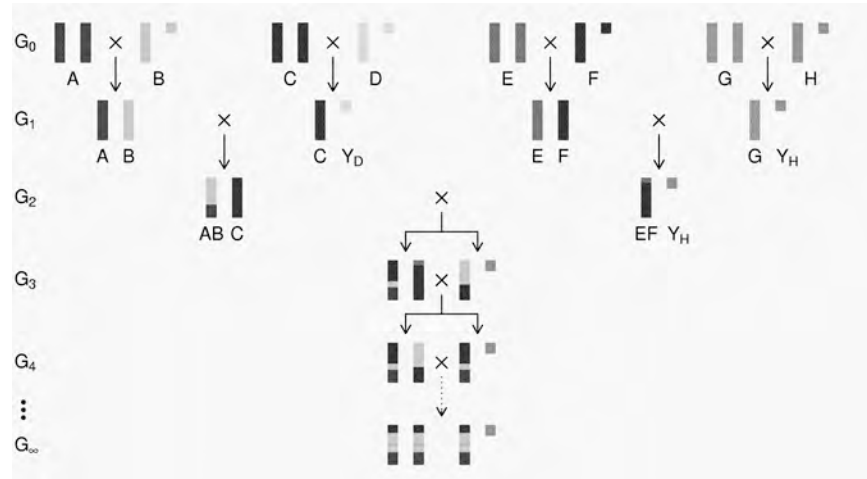
$$\Pr(G_2=C \mid G_1=A) = 2r / (1+4r)$$

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

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

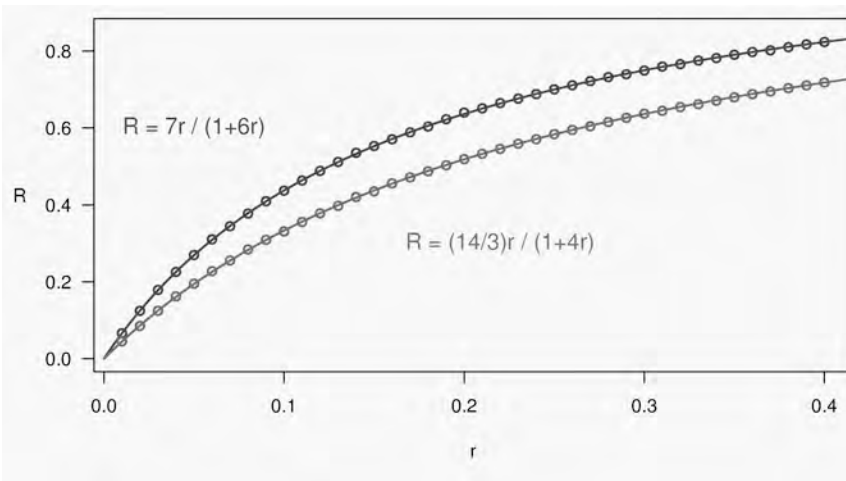
20

The X chromosome



21

Computer simulations



22

$r \rightarrow R$, autosomes

	Selfing	Sib-mating
2-way	$2r / (1+2r)$	$4r / (1+6r)$
4-way	$3r / (1+2r)$	$6r / (1+6r)$
8-way	$r(4 - r) / (1+2r)$	$7r / (1+6r)$
16-way	$r(5 - 3r + r^2) / (1+2r)$	$r(8 - r) / (1+6r)$
2^n -way	$1 - (1 - r)^{n-1} / (1+2r)$	$1 - (1 - r)^{n-2} / (1+6r)$

23

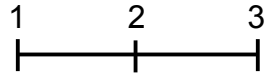
3-point coincidence



- r_{ij} = recombination fraction for interval i,j ;
assume $r_{12} = r_{23} = r$
- Coincidence = $c = \text{Pr}(\text{double recombinant}) / r^2$
= $\text{Pr}(\text{rec'n in 23} \mid \text{rec'n in 12}) / \text{Pr}(\text{rec'n in 23})$
- No interference $\rightarrow = 1$
Positive interference $\rightarrow < 1$
Negative interference $\rightarrow > 1$
- Generally c is a function of r .

24

3-points in 2-way RILs

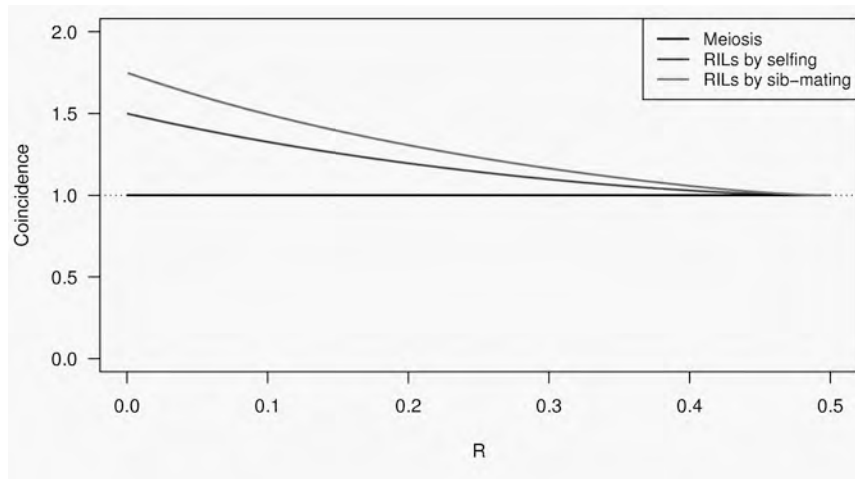


- $r_{13} = 2 r (1 - c r)$
- $R = f(r); \quad R_{13} = f(r_{13})$
- $\text{Pr}(\text{double recombinant in RIL}) = \{ R + R - R_{13} \} / 2$
- $\text{Coincidence (in 2-way RIL)} = \{ 2 R - R_{13} \} / \{ 2 R^2 \}$

25

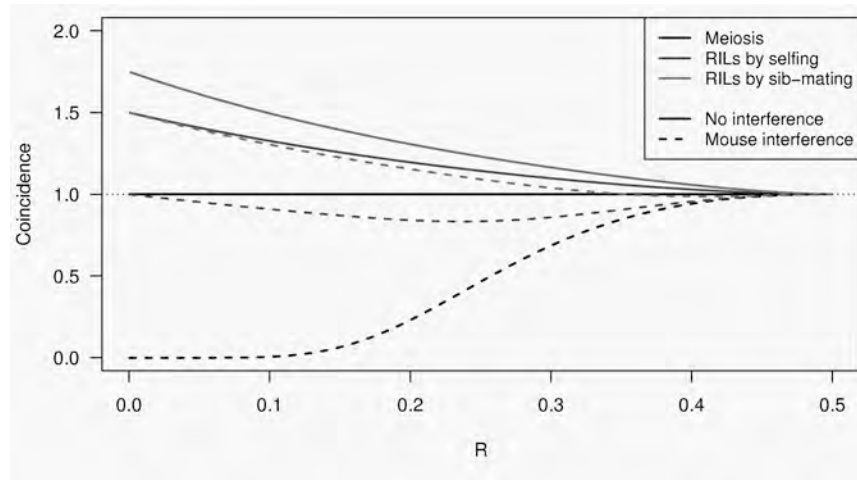
Coincidence

No interference



26

Coincidence



27

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.

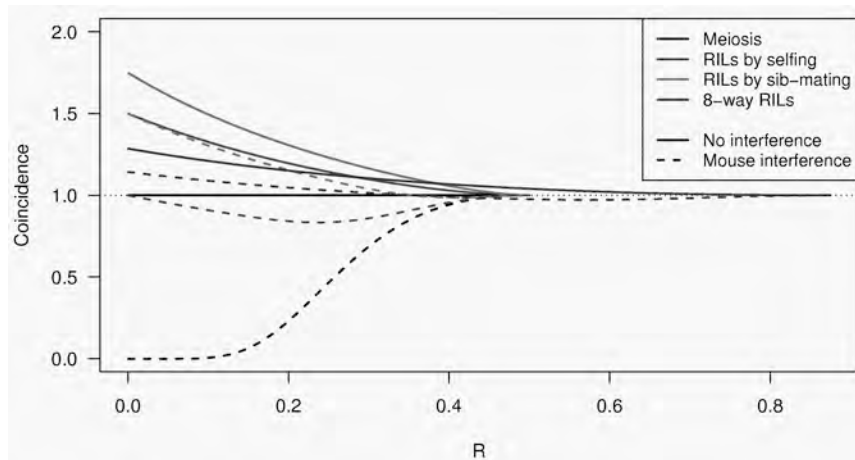
28

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
- Even counting the states was difficult.

29

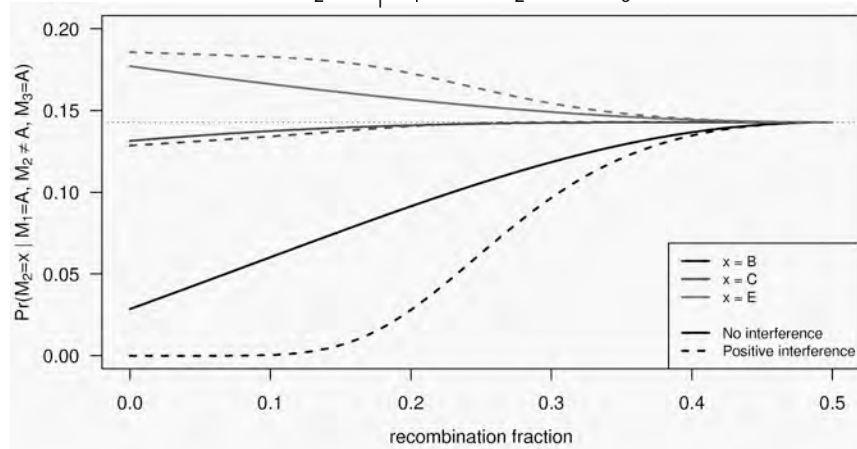
Coincidence



30

3-point symmetry

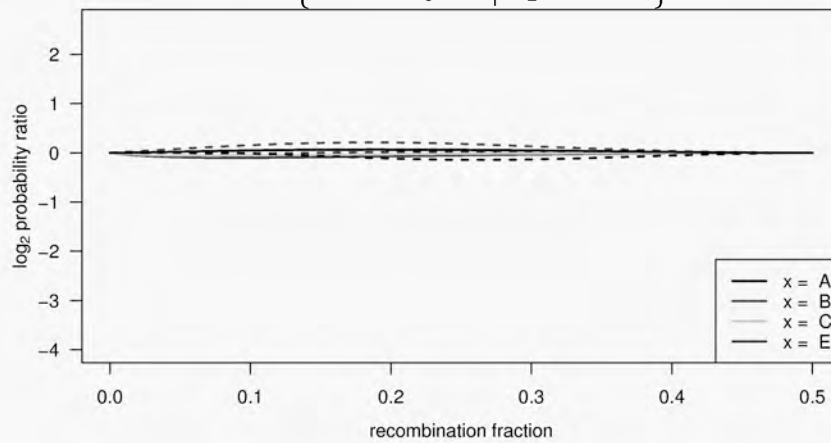
$$\Pr(M_2 = x \mid M_1 = A, M_2 \neq A, M_3 = A)$$



31

Markov property

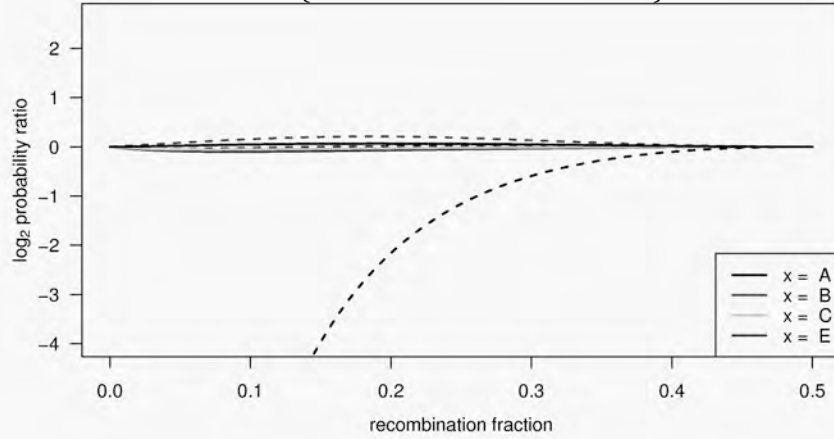
$$\log_2 \left\{ \frac{\Pr(M_3 = A \mid M_2 = A, M_1 = x)}{\Pr(M_3 = A \mid M_2 = A)} \right\}$$



32

Markov property

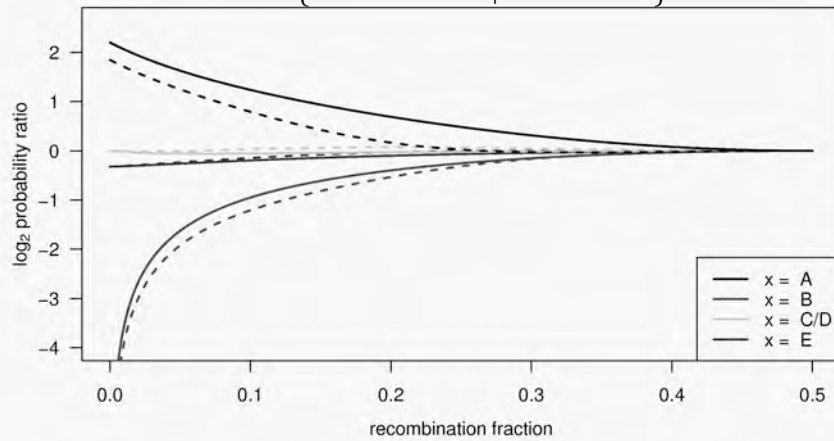
$$\log_2 \left\{ \frac{\Pr(M_3 = A \mid M_2 = B, M_1 = x)}{\Pr(M_3 = A \mid M_2 = B)} \right\}$$



33

Markov property

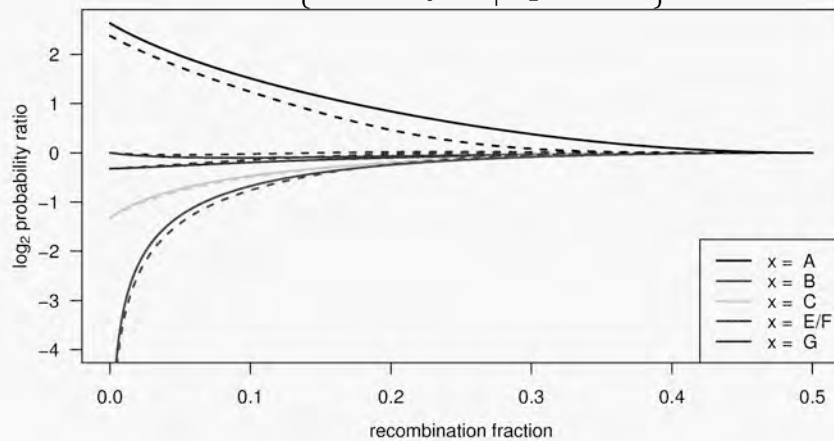
$$\log_2 \left\{ \frac{\Pr(M_3 = A \mid M_2 = C, M_1 = x)}{\Pr(M_3 = A \mid M_2 = C)} \right\}$$



34

Markov property

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



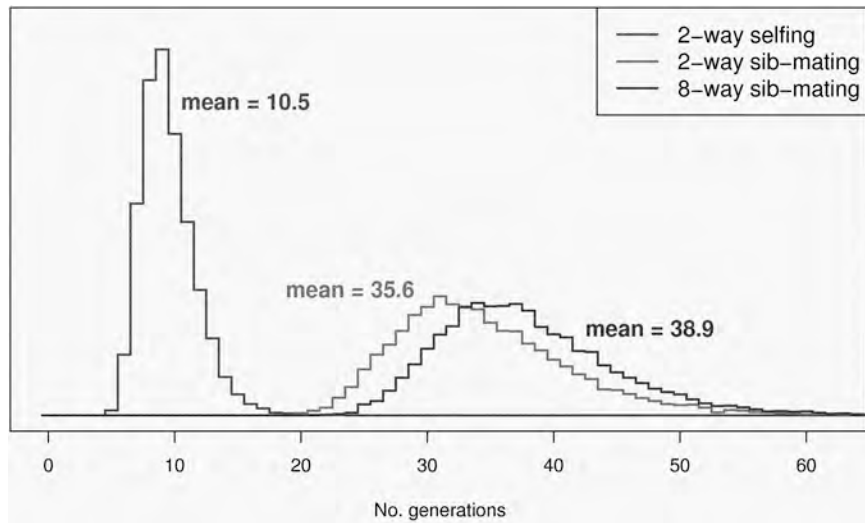
35

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

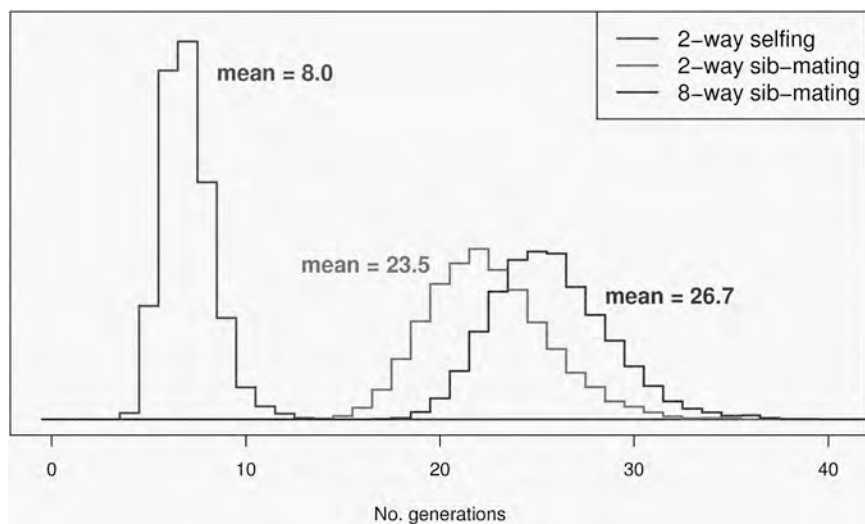
36

No. generations to fixation



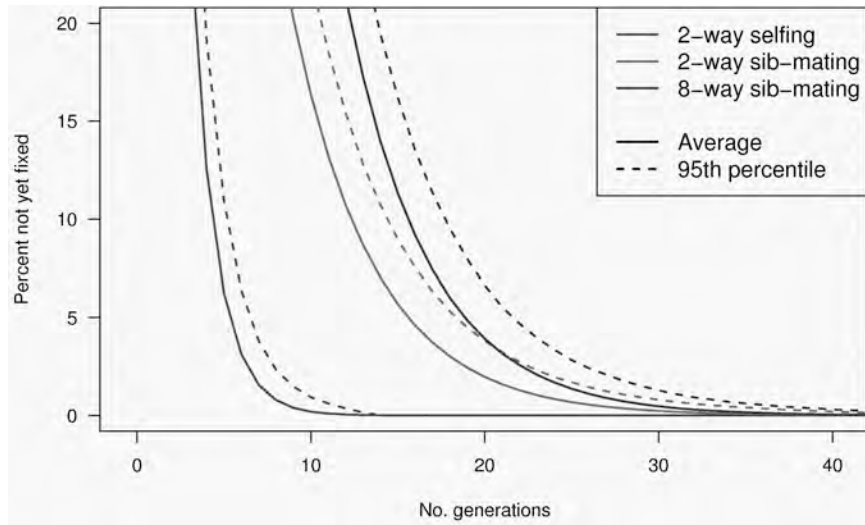
37

No. gen's to 99% fixation



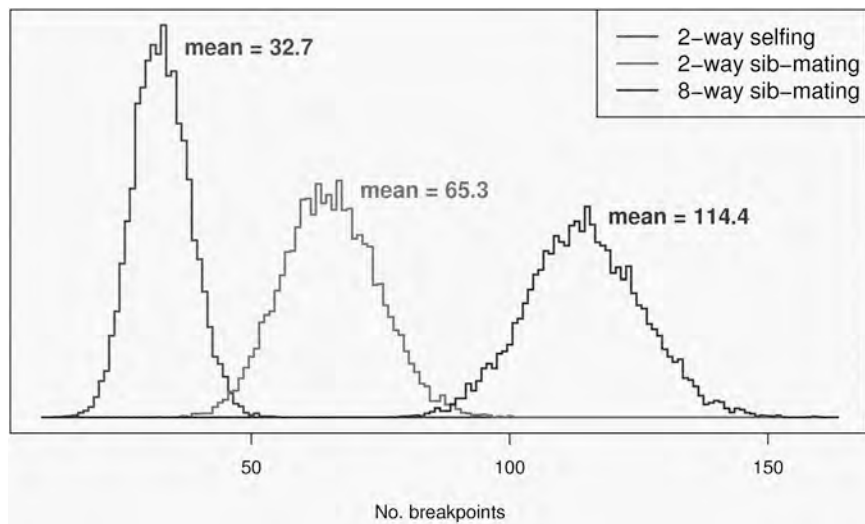
38

Percent genome not fixed



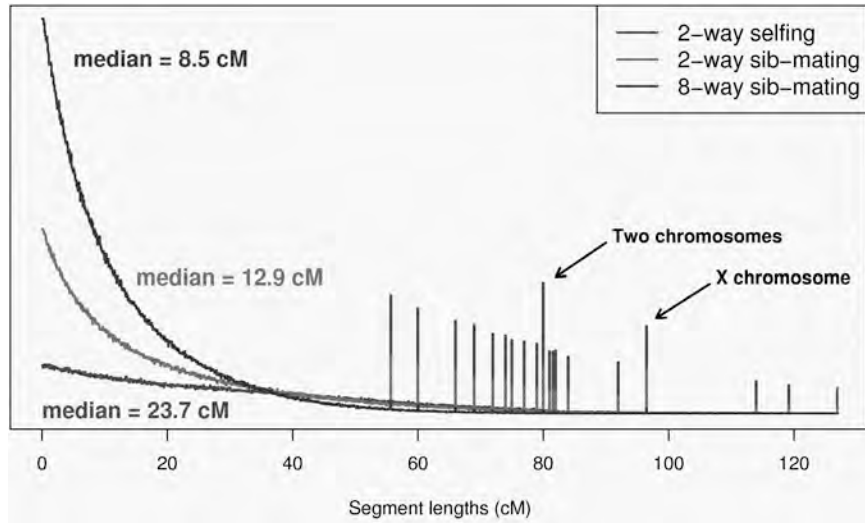
39

Number of breakpoints



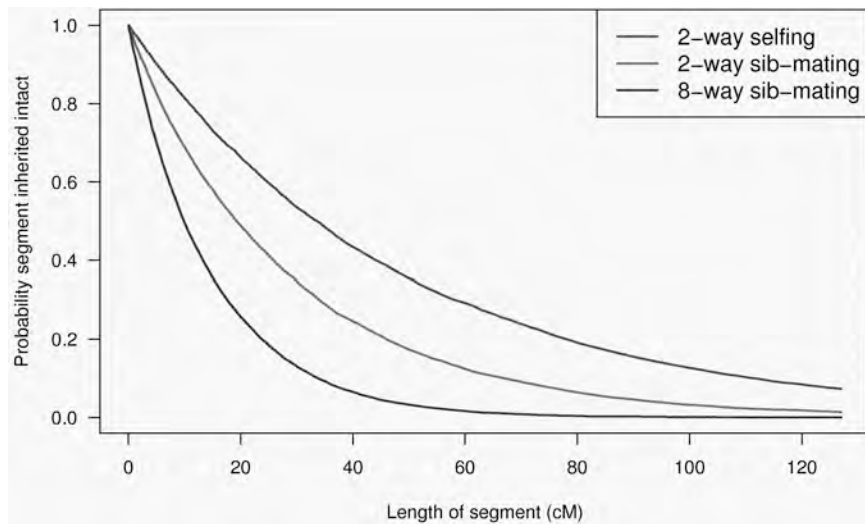
40

Segment lengths



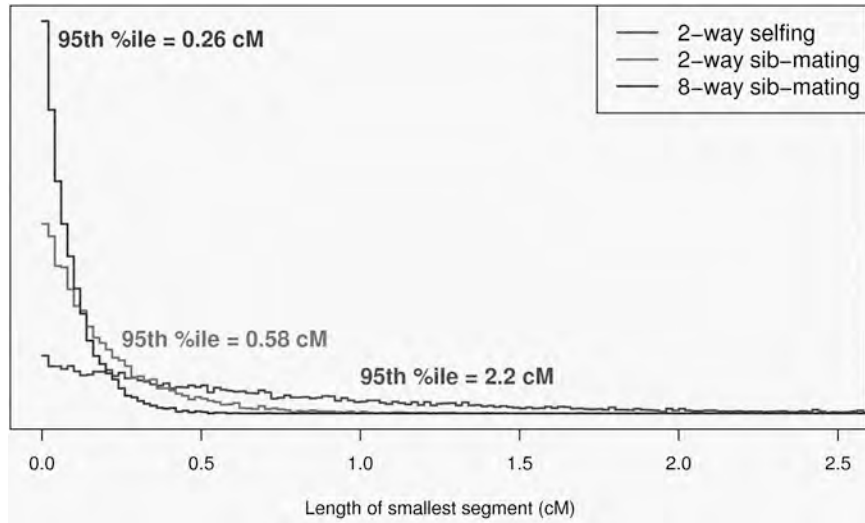
41

Probability a segment is inherited intact



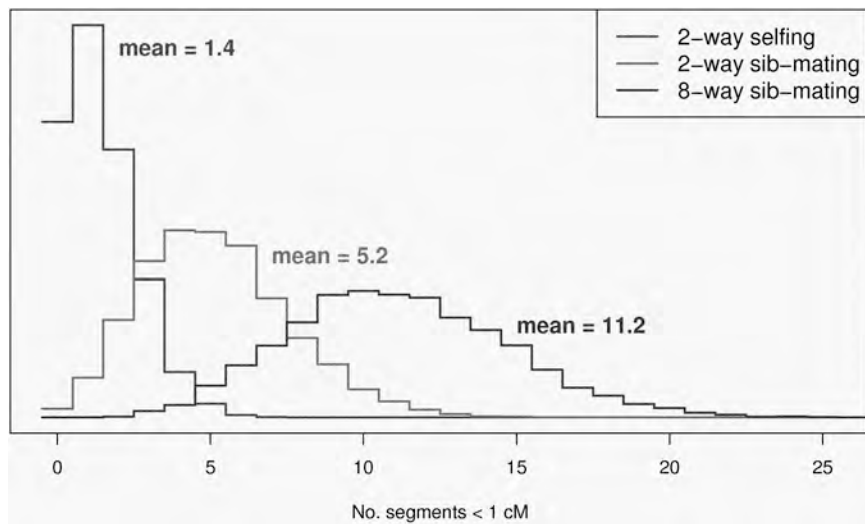
42

Length of smallest segment



43

No. segments < 1 cM



44

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: $R = 7r / (1 + 6r)$.
- We’ve shown clustering of breakpoints in RILs by sib-mating, even in the presence of strong crossover interference.
- Broman KW (2005) The genomes of recombinant inbred lines. *Genetics* 169:1133-1146