

Meiosis, recombination, and interference

Karl W Broman

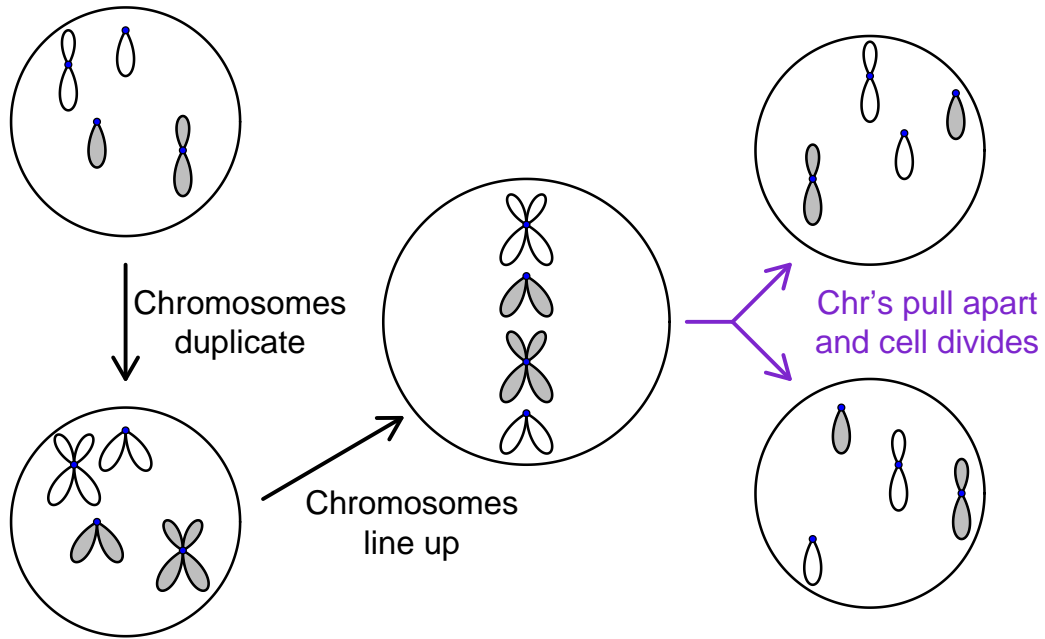
Department of Biostatistics
Johns Hopkins University
Baltimore, Maryland, USA

www.biostat.jhsph.edu/~kbroman

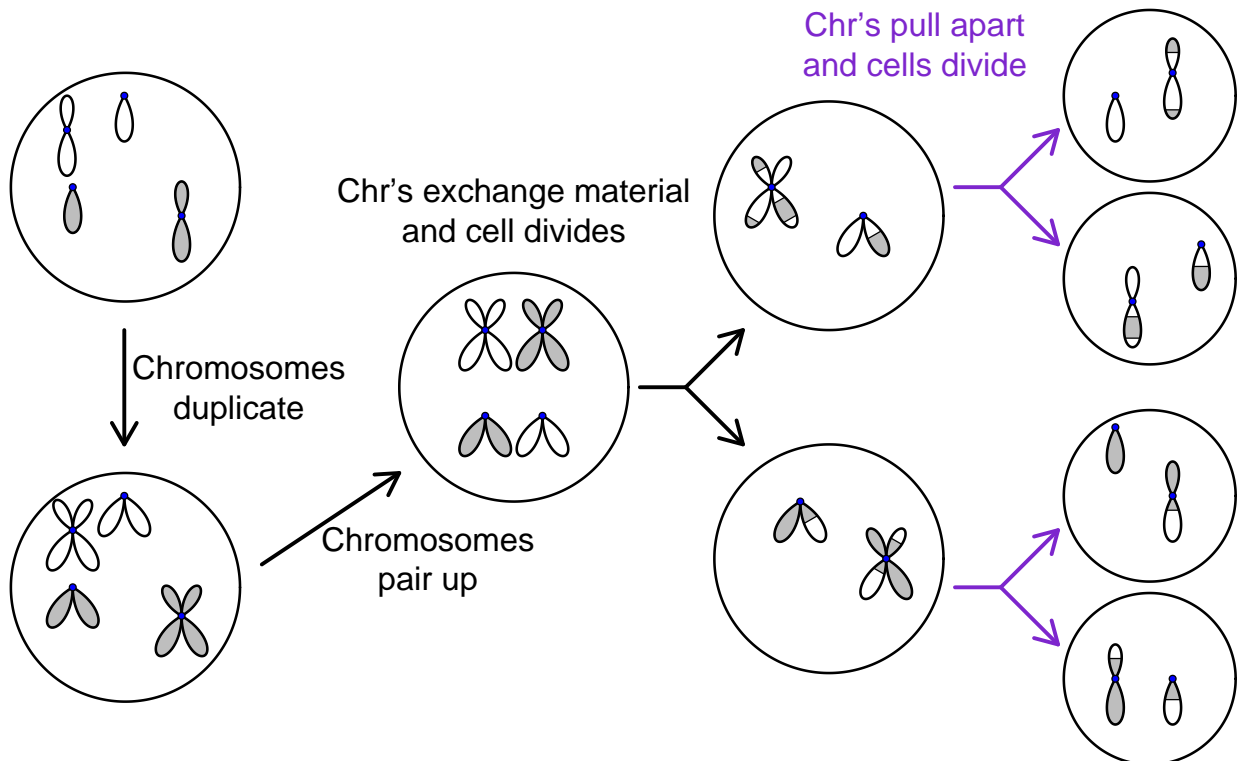
Outline

- Mitosis and meiosis
- Chiasmata, crossovers
- Genetic distance
- Genetic markers, recombination
- Chromatid and chiasma interference
- Mather's formula
- The count-location model
- The gamma model and the χ^2 model
- Data: humans and mice

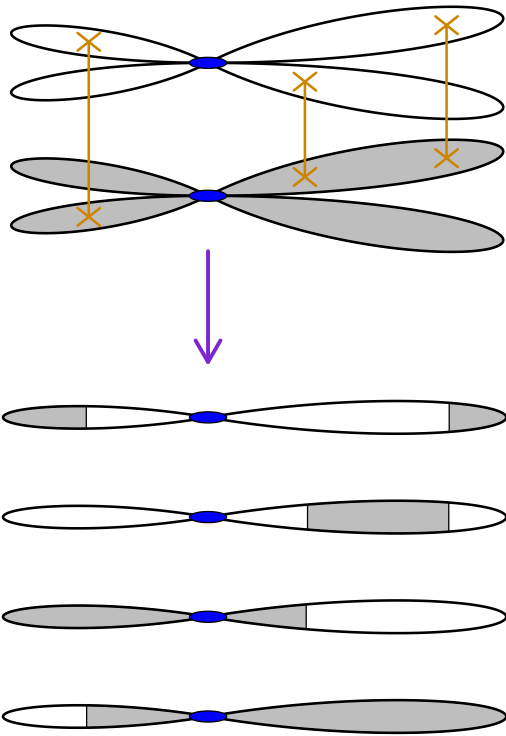
Mitosis: ordinary cell division



Meiosis: production of sex cells



The exchange process



Vocabulary

Four-strand bundle
Meiotic products

Sister chromatids
Non-sister chromatids

Chiasma, chiasmata
Crossovers

Obligate chiasma

Genetic distance

Two points are d Morgans apart if the average number of crossovers per meiotic product in the intervening interval is d .

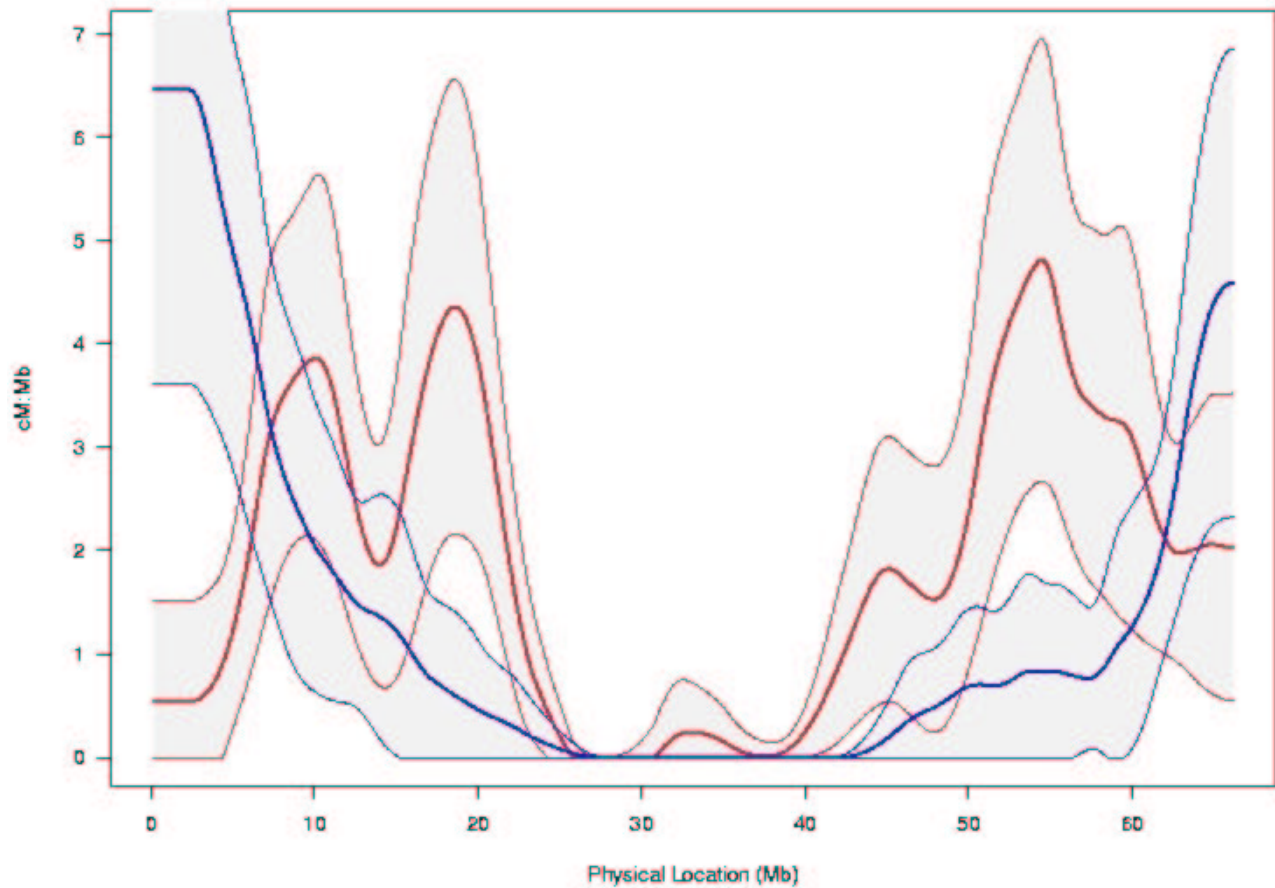
Usual units: centiMorgan (cM); 100 cM = 1 Morgan

Genetic distance \neq Physical distance

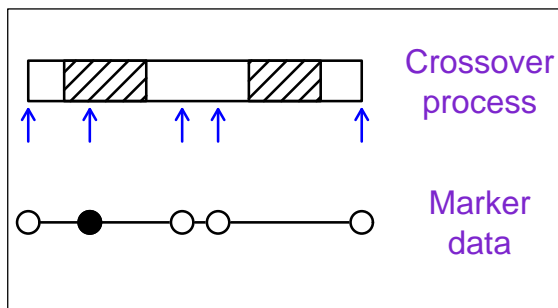
The intensity of the crossover process varies by

- Sex
- Individual
- Chromosome
- Position on chromosome
- Temperature

Chromosome 20



But we don't observe crossovers



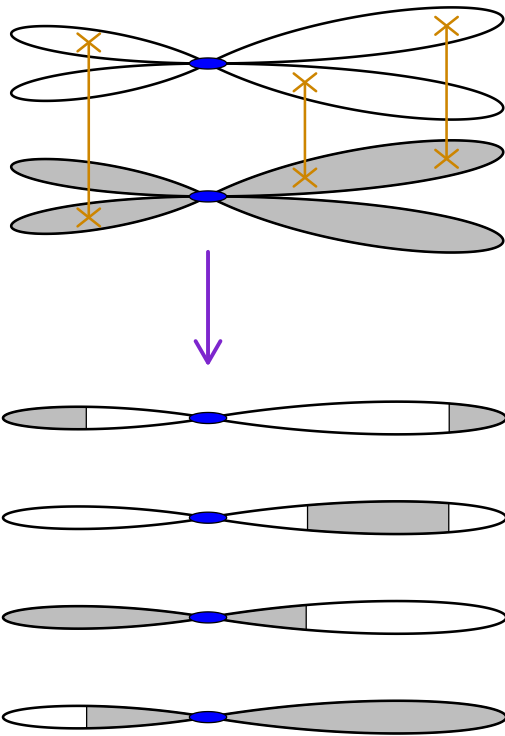
- Crossovers generally not observable
- We instead observe the origin of DNA at **marker loci**.

odd no. crossovers = recombination event

even no. crossovers = no recombination

Recombination fraction = $\Pr(\text{recombination event in interval})$

Interference



Chromatid interference:
strand choice

Chiasma interference:
positions of chiasmata

Mather's formula

Assuming no chromatid interference (NCI):

$$\Pr(\text{no rec'n in interval}) = \frac{1}{2} \{ 1 - \Pr(\text{no chiasma in interval}) \}$$

↑ ↑

[in random meiotic product] [on 4-strand bundle]

Let n = no. chiasmata in interval on 4-strand bundle
and m = no. crossovers in interval on random meiotic product

Under NCI, $m \mid n \sim \text{Binomial}(n, 1/2)$

$$\text{Thus } \Pr(m \text{ is odd} \mid n) = \begin{cases} 0 & \text{if } n = 0 \\ 1/2 & \text{if } n \geq 1 \end{cases}$$

Haldane map function

Under no interference, the locations of chiasmata on the 4-strand bundle are according to a Poisson process (rate: 2 per Morgan).

Thus $n \sim \text{Poisson}(2d)$

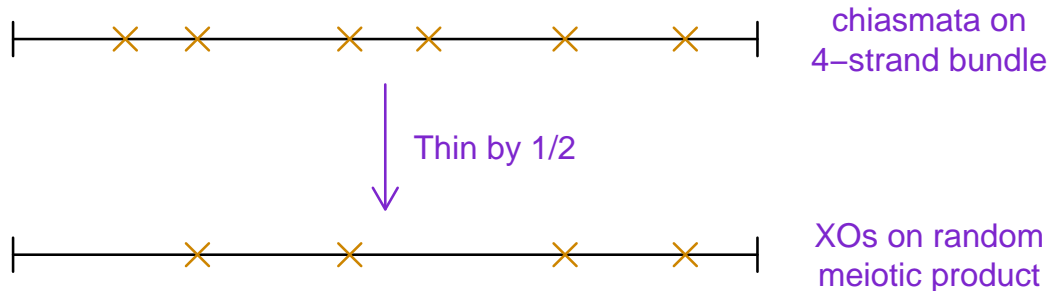
where d is the genetic length of the interval (in Morgans)

Thus $\Pr(n = 0) = \exp(-2d)$

Thus

$$r = \frac{1}{2} \{ 1 - \exp(-2d) \}$$

Models for recombination



- Assuming NCI, thin χ -process by 1/2, independently, to get the XO-process.
- Models:
 - Count-location model
 - Gamma model, χ^2 model

Count-location (CL) model

Let n = no. chiasmata on 4-strand bundle

Model: $n \sim \mathbf{p} = (p_0, p_1, p_2, \dots)$

locations | $n \sim$ iid uniform(0,L)

Note: $\mathbf{p} =$ Poisson(2L) \rightarrow no interference

Under NCI, crossovers on random meiotic product will also follow a count-location model.

Let m = no. crossovers on random meiotic product

Then $m | n \sim$ Binomial($n, 1/2$) and

$$\Pr(m = i) = \sum_{n=0}^{\infty} p_n \binom{n}{i} \left(\frac{1}{2}\right)^n$$

The CL model stinks

Advantage: Can easily incorporate **obligate chiasma**

Disadvantage: Fits data **poorly!**

\rightarrow Allows crossovers to be too close together.

Gamma model

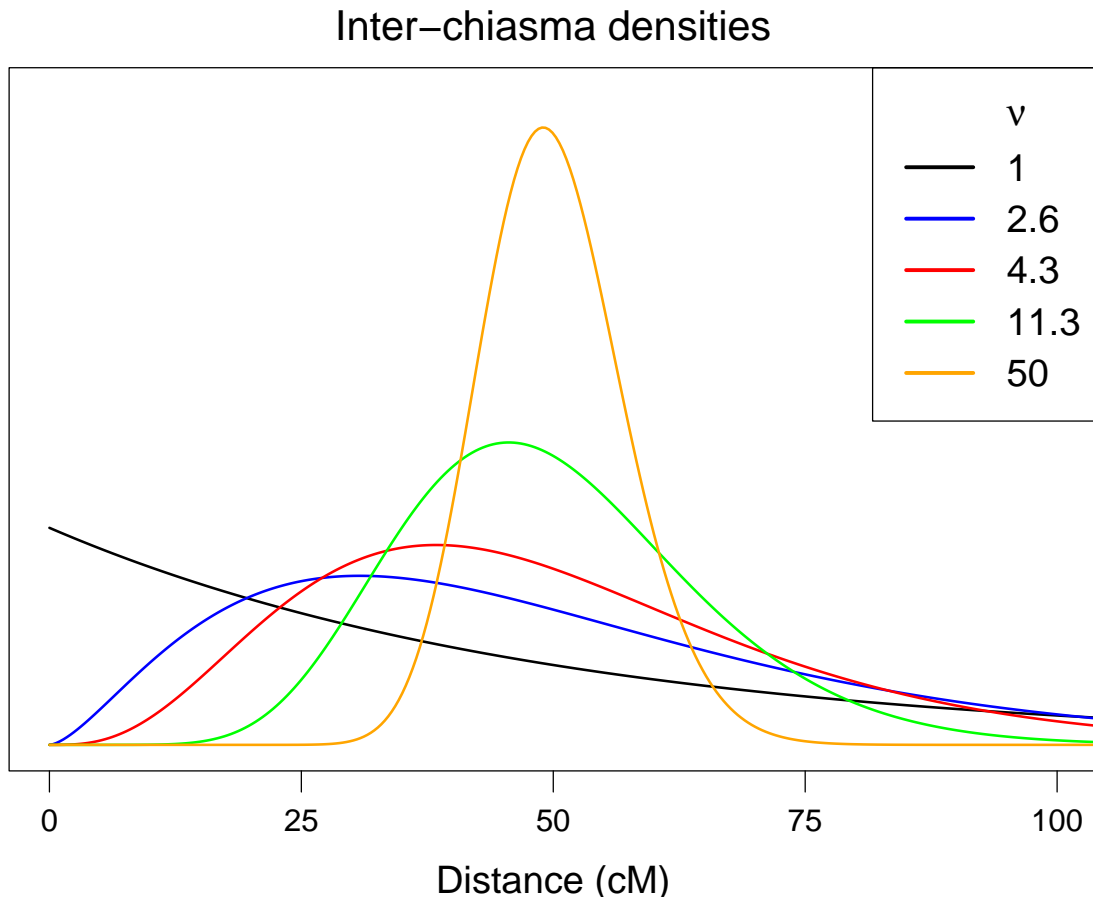
Locations of chiasmata according to a stationary gamma renewal process.

Increments are iid gamma(shape = ν , rate = 2ν)
(Constrained to have mean 1/2 Morgan.)

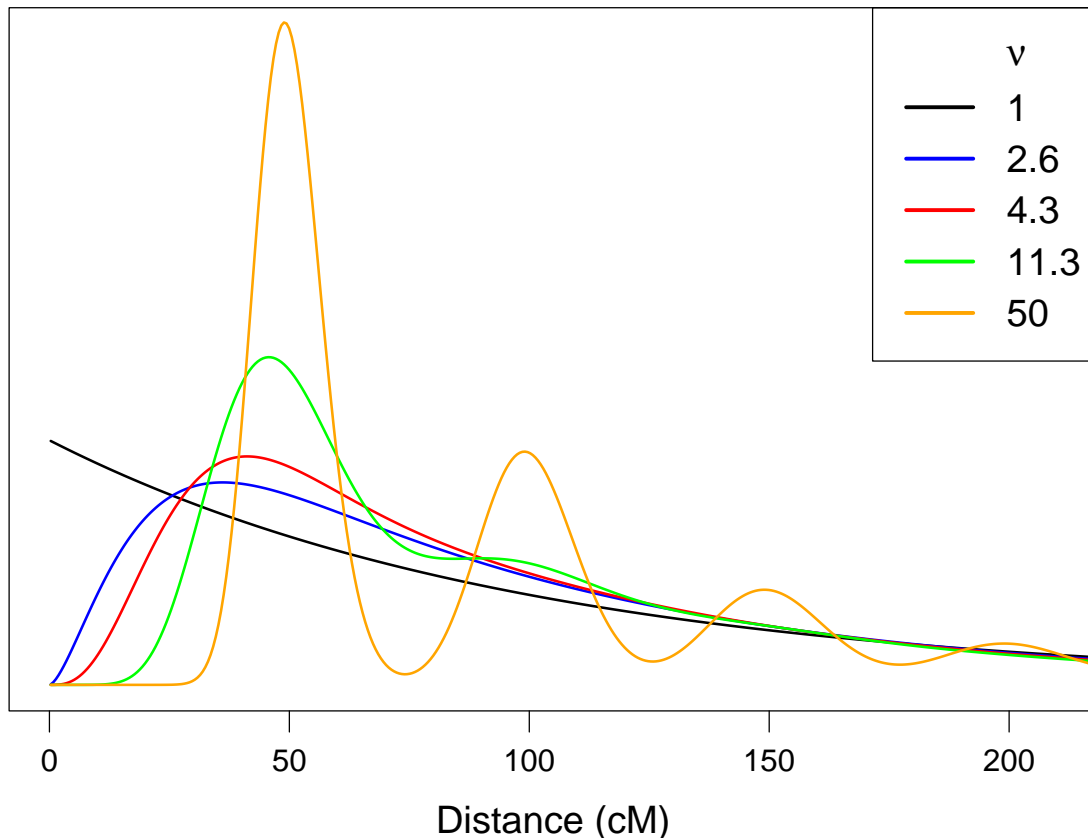
$$\nu \begin{cases} = 1 & \text{no interference} \\ > 1 & \text{positive chiasma interference} \end{cases}$$

Locations of crossovers on random meiotic product also a stationary renewal process.

Inter-arrival distribution is a mixture of gammas



Inter-crossover densities



Chi-square model

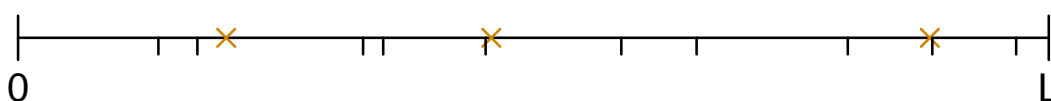
Special case of the gamma model when the parameter is a non-negative integer (take $m = \nu - 1$).

Computer simulations and many calculations are easier.

Chiasmata on the 4-strand bundle: take every m th point from a Poisson process with rate $2m$ per Morgan.

Inter-arrival distribution is a scaled version of a χ^2 distribution.

Example: ($m = 4$)



The gamma model

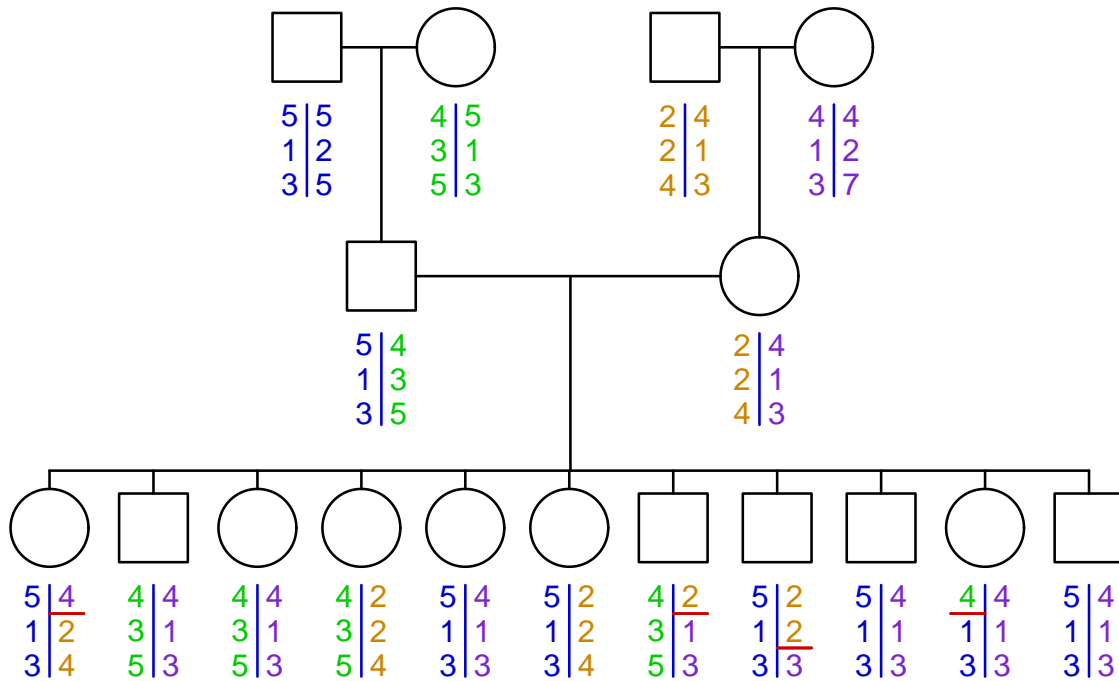
Advantage: Fits data reasonably well

Disadvantage: Doesn't account for obligate chiasma

Human data

- <http://research.marshfieldclinic.org/genetics>
- 8 CEPH families
 - three generations
 - 11 to 15 progeny
 - 92 meioses, total
- ~8000 STRP markers (~90% typed)
- Average spacing:
 - Female: 0.6 ± 1.2 cM
 - Male: 0.4 ± 1.0 cM
- Data cleaning
 - Removed 764/954,425 (~0.08%) genotypes resulting in tight double recombinants

CEPH pedigree (data on 3 markers)

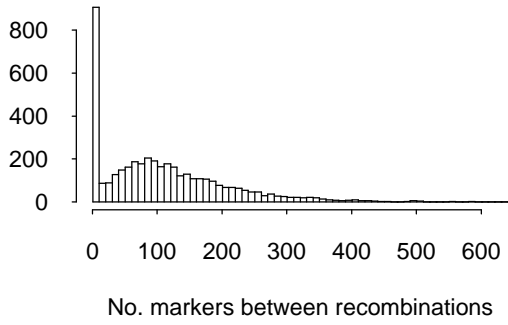


CEPH individual 1331-11 maternal chr 10 haplotype

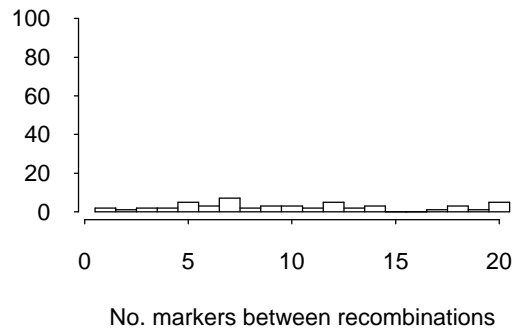
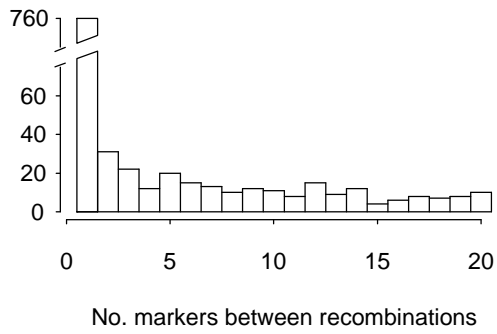
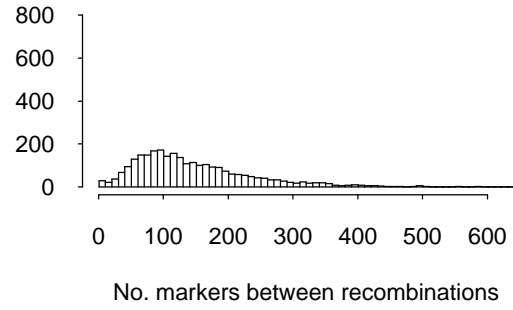
```

11111111--- 11-111-11-  --11--1111  1-11---11-
1111111--1-  11-1111-11  --11111-11  1111-11111
11--111111  111-1111-1  1111111111  10000-0-00
0--000-000  0000-00000  0000--0000  0000--0000
0000-00000  000-0--0--  --0-11-11-  -1111111-1
---1-1-1-1  1111-1--11  11111-1111  -111-11111
-1----1111  111111-111  -1111-111-  11-1111111
111-11111-  1111111-1-  1111111-11  11-111111-
11--1-11-1  111-11-1-1  1-1-----1-1  11-1111111
11--1--11-  11111--111  11--111111  1111-11111
1-0---0000  00000-0000  00-000
  
```

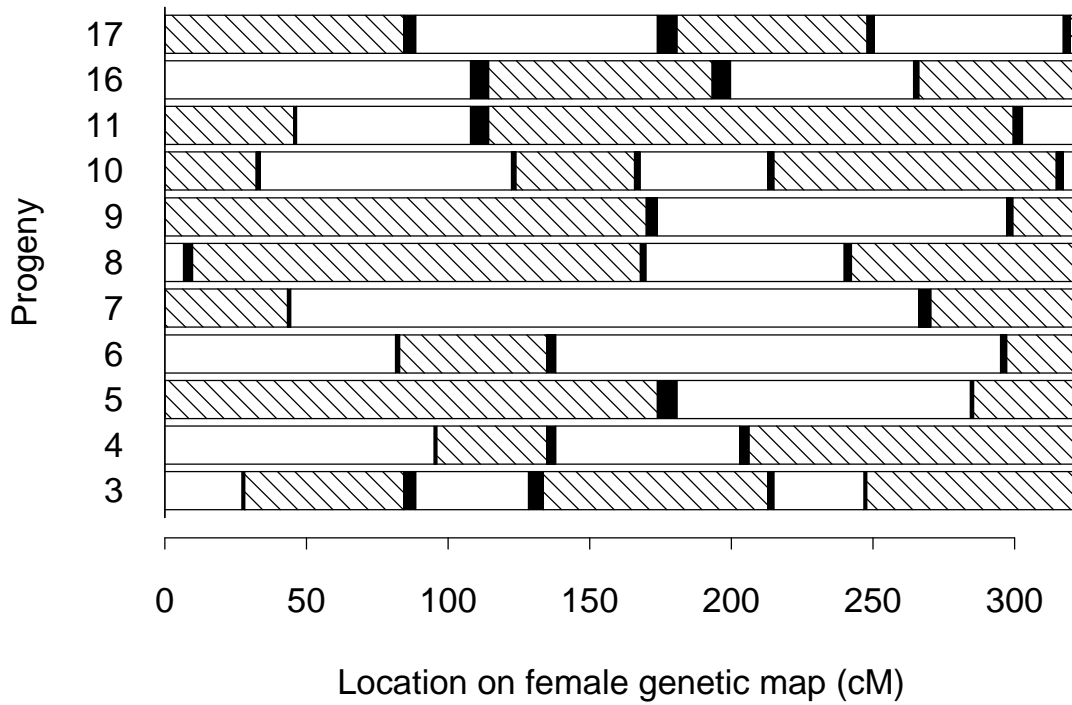
Raw Data



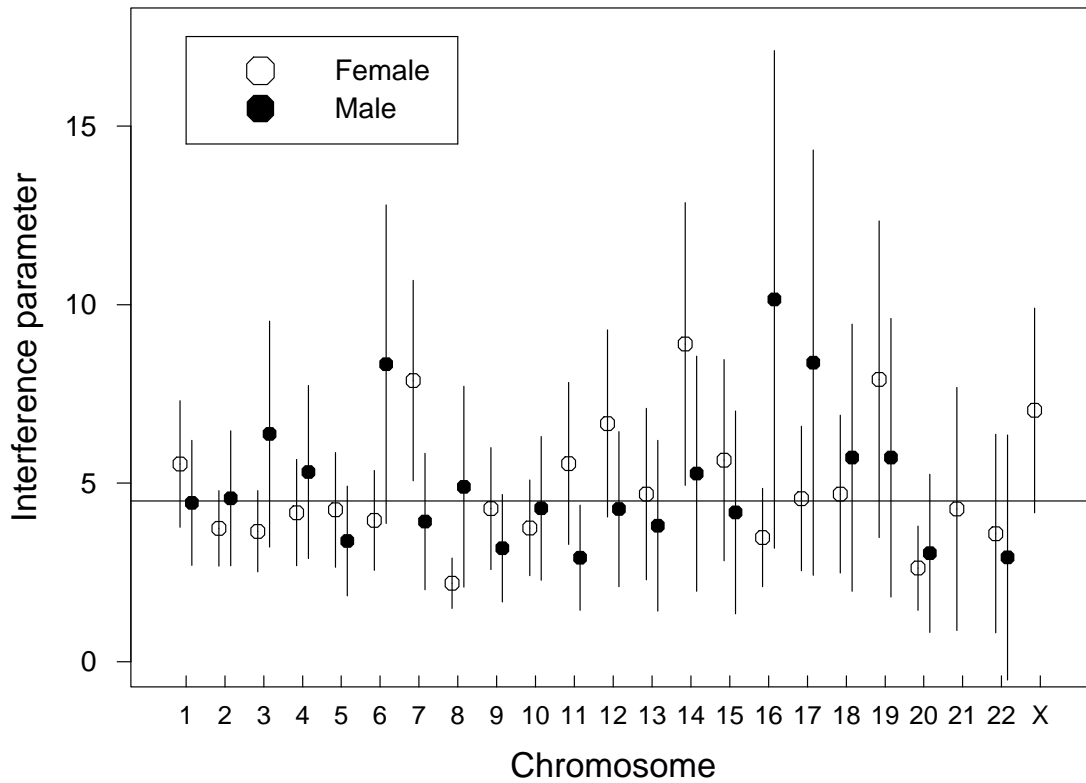
Clean Data



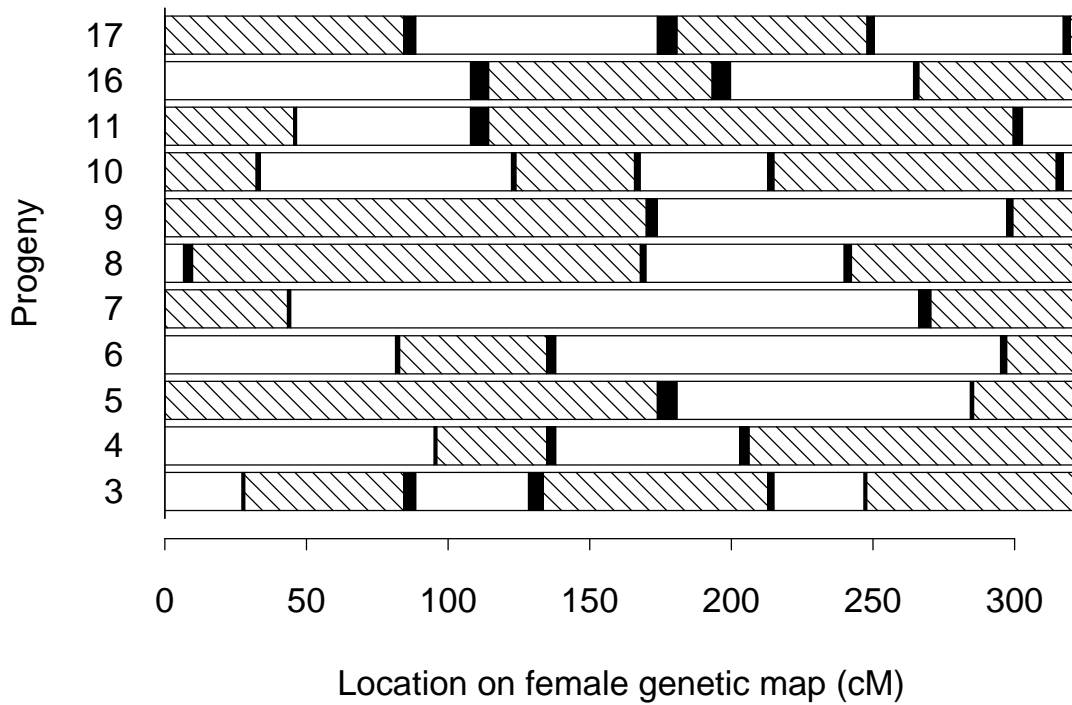
The data



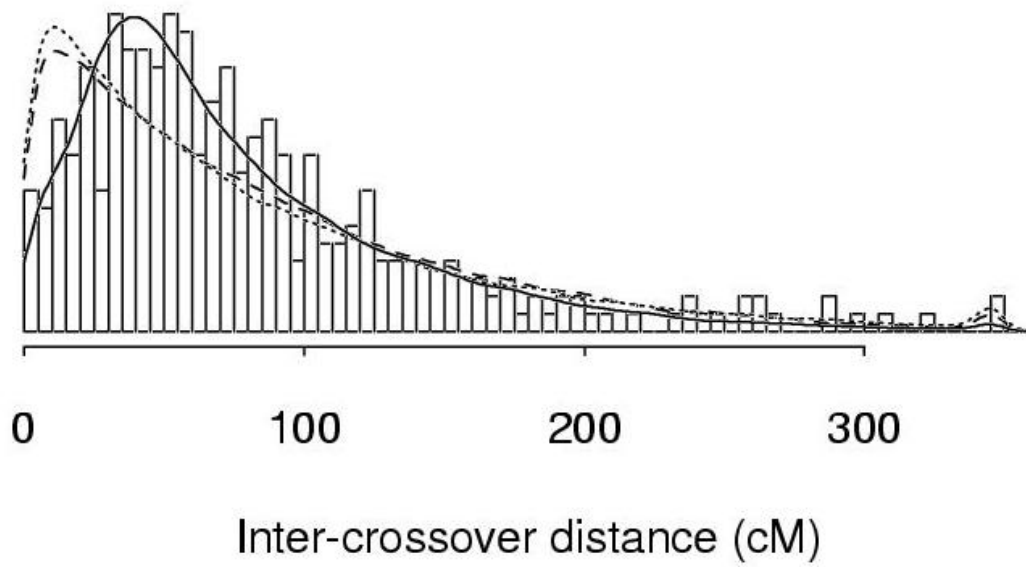
MLEs of interference parameter



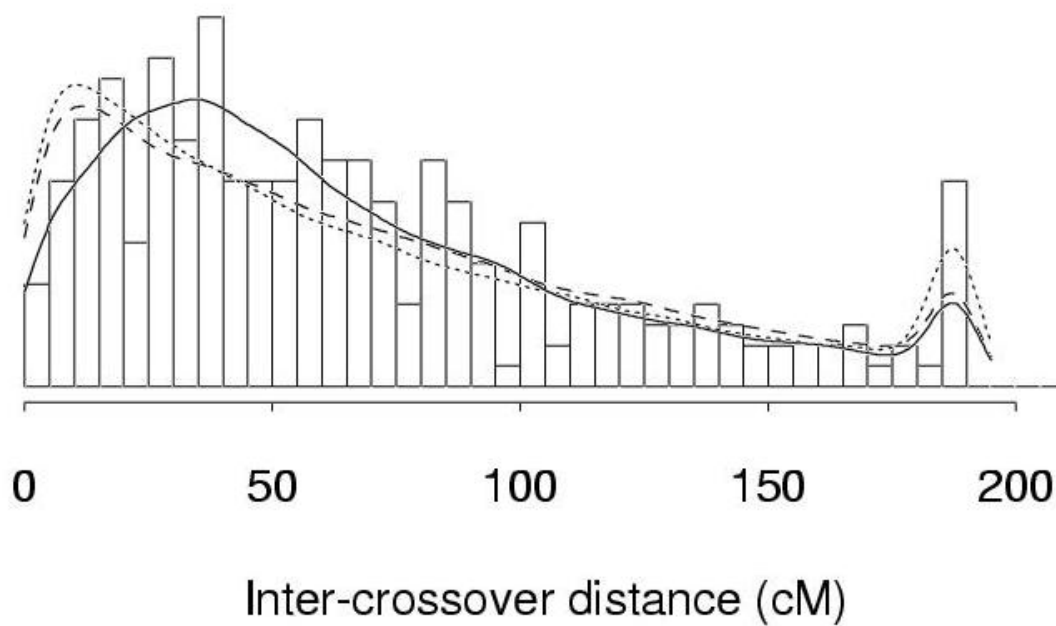
Goodness of fit?



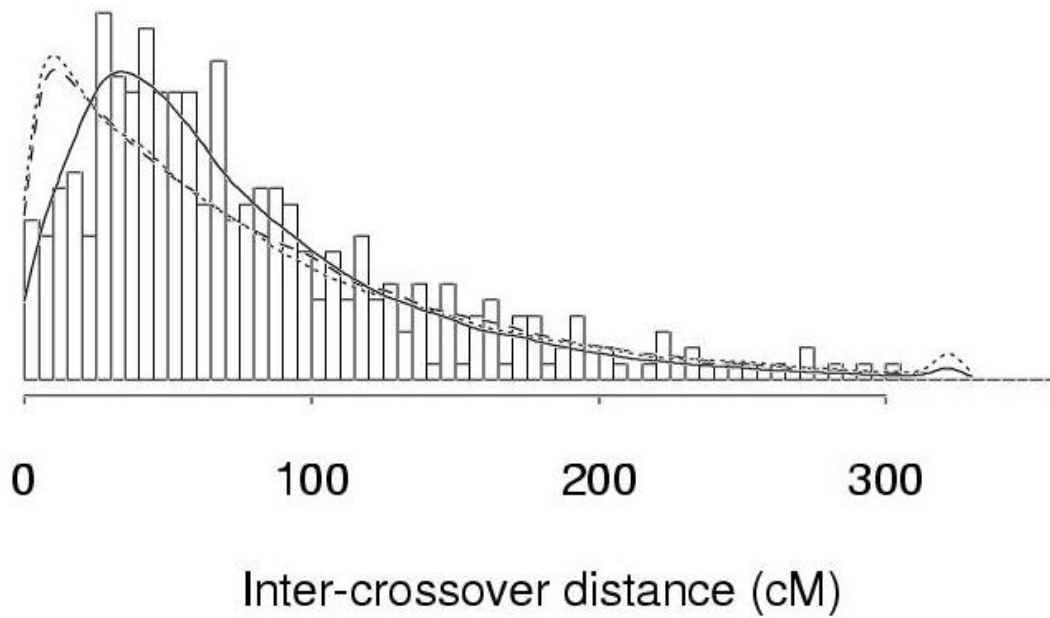
Maternal chromosome 1



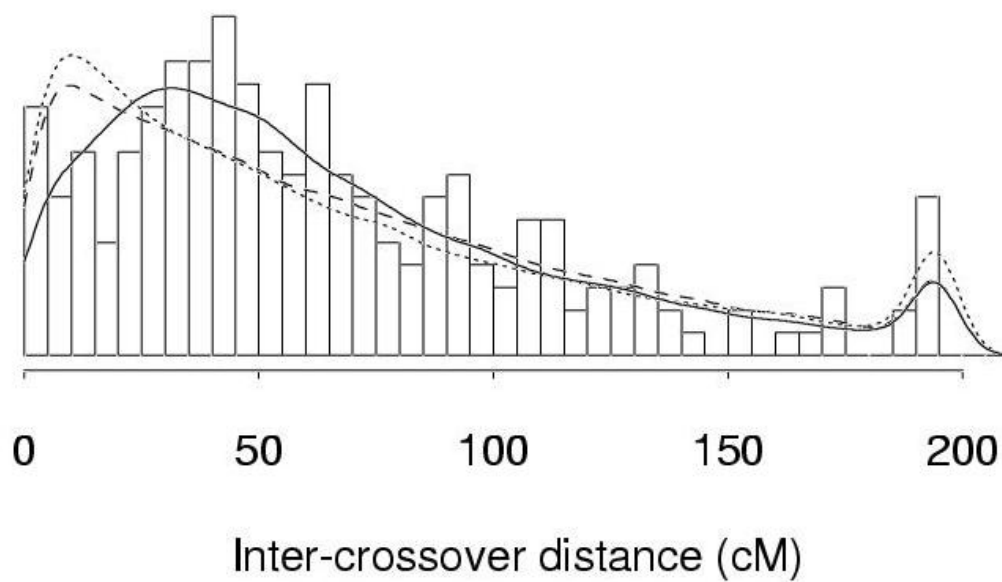
Paternal chromosome 1



Maternal chromosome 2



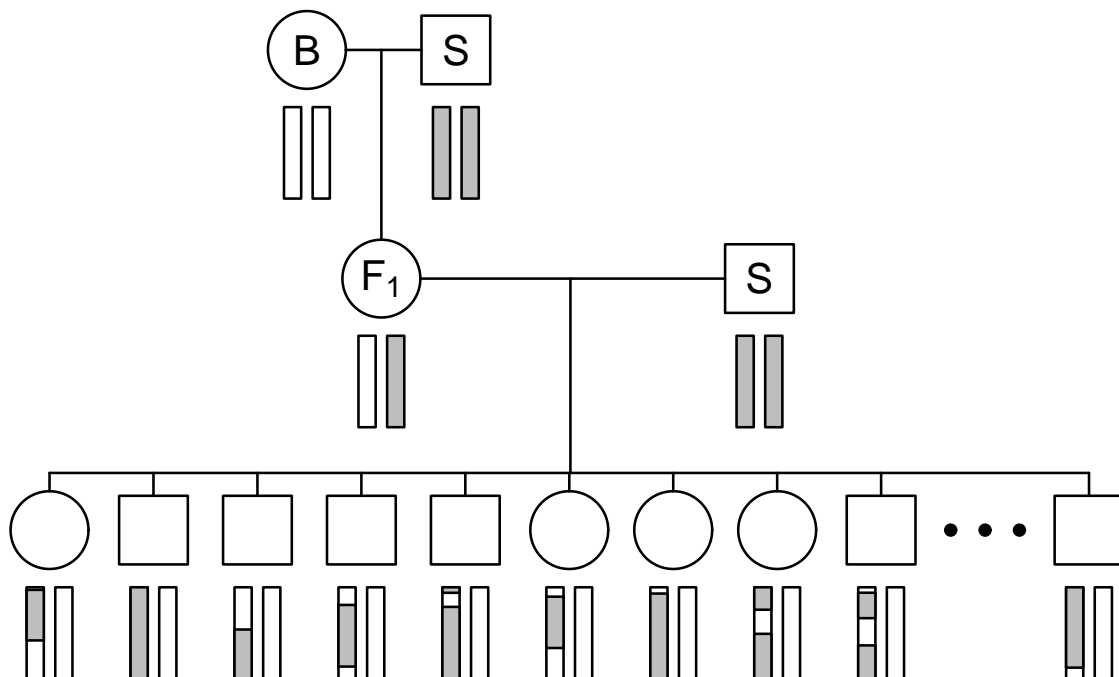
Paternal chromosome 2



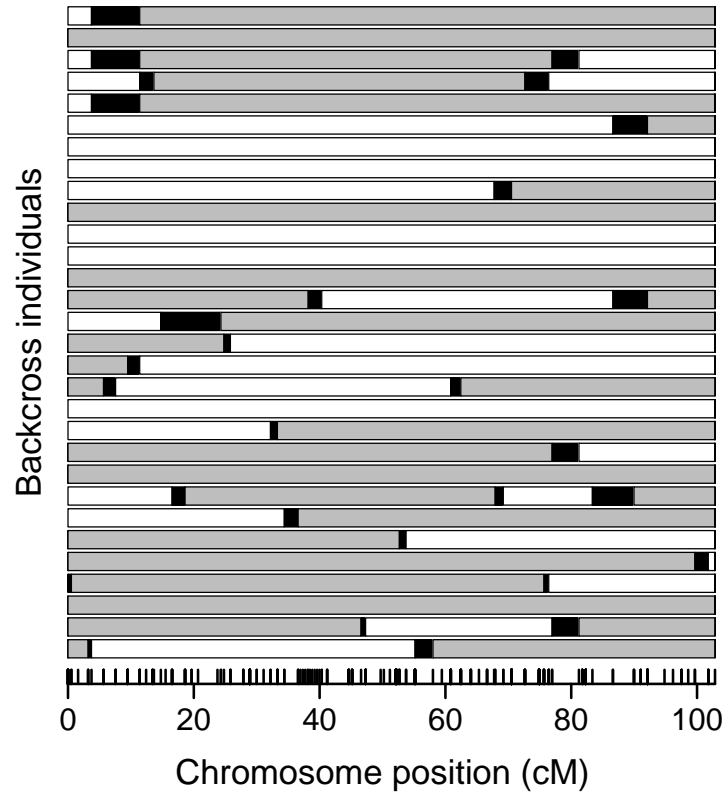
Mouse data

- <http://www.jax.org/resources/documents/cmdata/ftp.html>
- Two interspecific backcrosses with common F₁ parent
 - BSB: (C57BL/6J × M. spretus) × C57BL/6J
 - BSS: (C57BL/6J × SPRET/Ei) × SPRET/Ei
 - 94 individuals from each cross
- High-density STRP markers
 - BSB: 1372 markers
 - BSS: 4913 markers
- Average spacing:
 - BSB: 1.0 cM
 - BSS: 0.3 cM

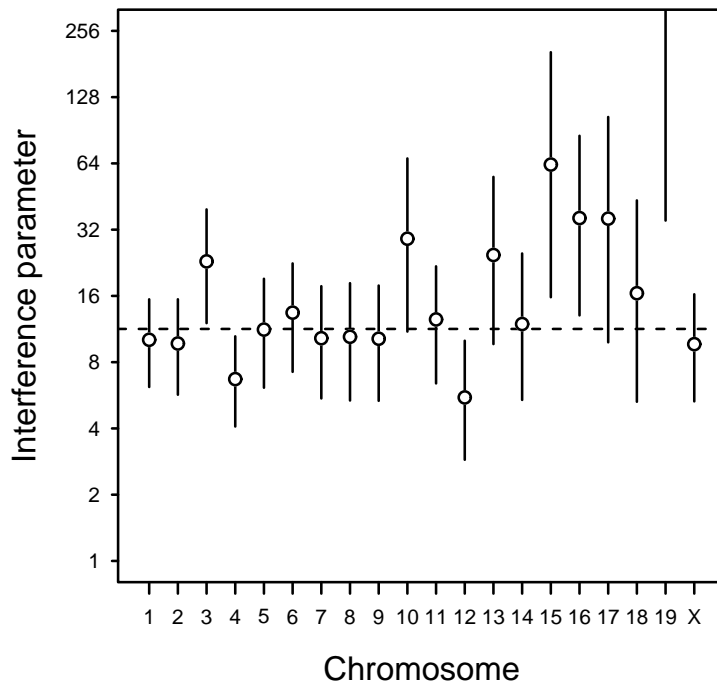
Backcross pedigree



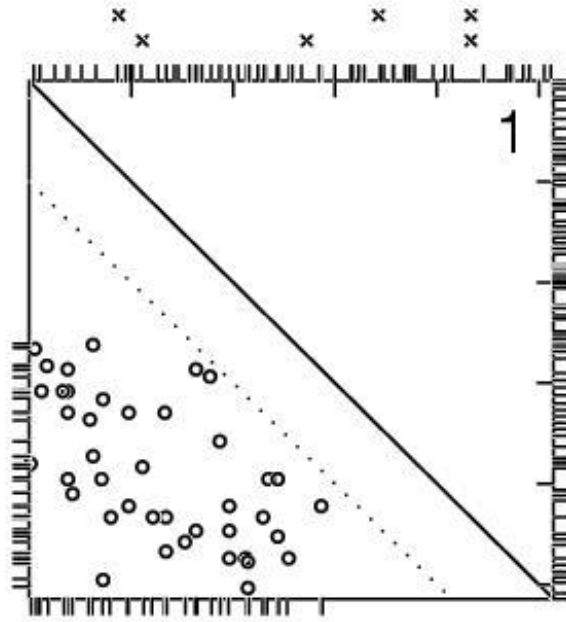
Mouse data



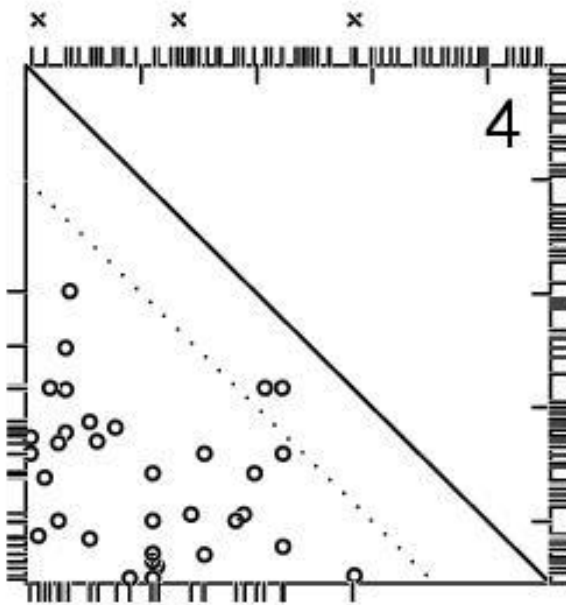
MLEs: mouse



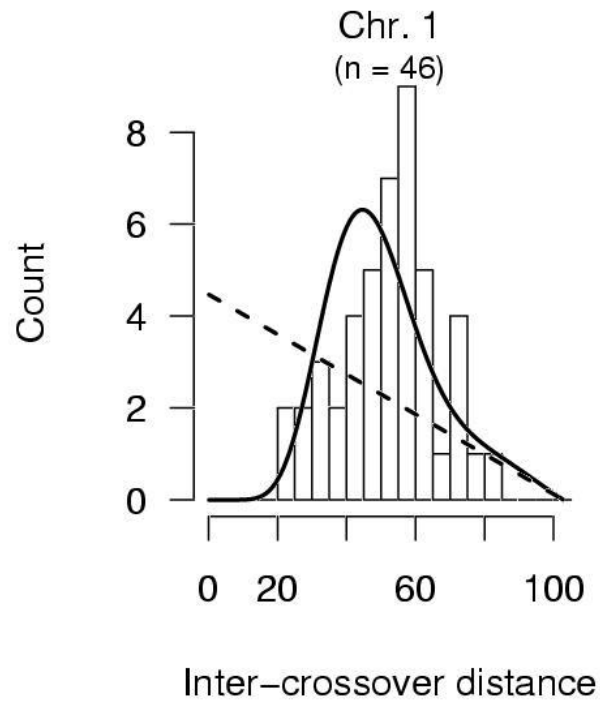
Crossover locations: Chr 1



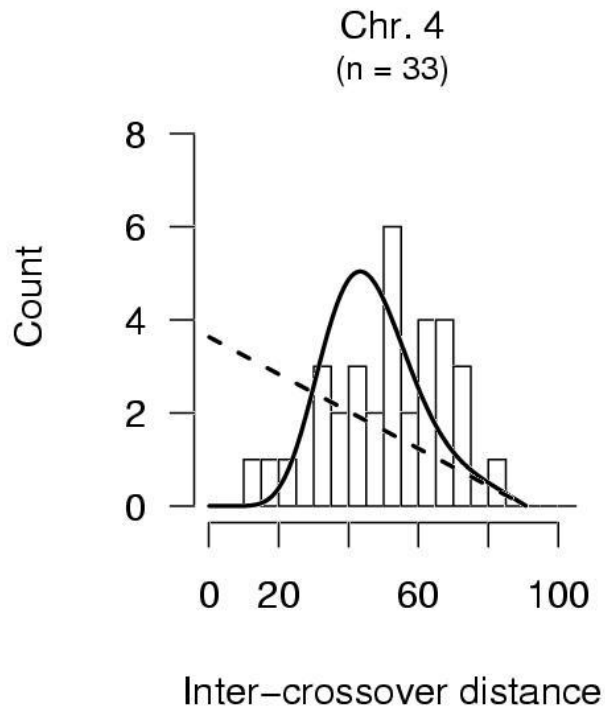
Crossover locations: Chr 4



Inter-XO distances: Chr 1



Inter-XO distances: Chr 4



Acknowledgements

Terry Speed, University of California, Berkeley, and WEHI

Hongyu Zhao, Yale University

Mary Sara McPeck, University of Chicago

Jim Weber, Marshfield Medical Research Foundation

References

- McPeck MS (1996) An introduction to recombination and linkage analysis. In: Speed TP, Waterman MS (eds) Genetic mapping and DNA sequencing. IMA volumes in mathematics and its applications. Vol 81. Springer-Verlag, New York, pp 1–14
- McPeck MS, Speed TP (1995) Modeling interference in genetic recombination. *Genetics* 139:1031–1044
- Zhao H, Speed TP, McPeck MS (1995) Statistical analysis of crossover interference using the chi-square model. *Genetics* 139:1045–1056
- Zhao H, McPeck MS, Speed TP (1995) Statistical analysis of chromatid interference. *Genetics* 139:1057–1065
- Speed TP (1996) What is a genetic map function? In: Speed TP, Waterman MS (eds) Genetic mapping and DNA sequencing. IMA volumes in mathematics and its applications. Vol 81. Springer-Verlag, New York, pp 65–88
- Zhao H, Speed TP (1996) On genetic map functions. *Genetics* 142:1367–1377

- Broman KW, Weber JL (2000) Characterization of human crossover interference. *Am J Hum Genet* 66:1911–1926
- Broman KW, Rowe LB, Churchill GA, Paigen K (2002) Crossover interference in the mouse. *Genetics* 160:1123–1131