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 $\chi^2$ model
- Data: humans and mice
Mitosis: ordinary cell division
Mitosis: ordinary cell division

Chromosomes duplicate
Mitosis: ordinary cell division

- Chromosomes duplicate
- Chromosomes line up
Mitosis: ordinary cell division

1. Chromosomes duplicate
2. Chromosomes line up
3. Chromosome's pull apart and cell divides
Meiosis: production of sex cells
Meiosis: production of sex cells

Chromosomes duplicate
Meiosis: production of sex cells

Chromosomes duplicate

Chromosomes pair up
Meiosis: production of sex cells

- Chromosomes duplicate
- Chromosomes pair up
- Chr’s exchange material and cell divides
Meiosis: production of sex cells

- Chromosomes duplicate
- Chromosomes pair up
- Chr’s exchange material and cell divides
- Chr’s pull apart and cells divide
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
But we don’t observe crossovers

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

odd no. crossovers = recombination event

even no. crossovers = no recombination

Recombination fraction = Pr(recombination event in interval)
Microsatellite markers

aka Short Tandem Repeat Polymorphisms (STRPs)

- Tandem repeat of something like GATA at a specific position in the genome.
- Number of repeats varies
- Use PCR to “amplify” region
- Use gel electrophoresis to determine length of region
Map functions

Connect **genetic distance** (average no. crossovers) to **recombination fraction** (chance of an odd no. crossovers).

\[ r = M(d) \quad d = M^{-1}(r) \]

We require a model for the crossover process.
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, \( m \mid n \sim \text{Binomial}(n, 1/2) \)

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 $n \sim \text{Poisson}(2 \ d)$
where $d$ is the genetic length of the interval (in Morgans)

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

Thus

$$r = \frac{1}{2} \left\{ 1 - \exp(-2 \ d) \right\}$$
Models for recombination

- Assuming NCI, thin $\chi$-process by 1/2, independently, to get the XO-process.
- Models:
  - Count-location model
  - Gamma model, $\chi^2$ model
Count-location (CL) model

Let \( n \) = no. chiasmata on 4-strand bundle

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

locations \( n \sim \text{iid uniform}(0,L) \)

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

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

Let \( m \) = no. crossovers on random meiotic product

Then \( m \mid n \sim \text{Binomial}(n, 1/2) \) and

\[
\Pr(m = i) = \sum_{n=0}^{\infty} p_n(i) (\frac{1}{2})^n
\]
The CL model stinks

**Advantage:** Can easily incorporate obligate chiasma

**Disadvantage:** Fits data poorly!

→ Allows crossovers to be too close together.
**Gamma model**

Locations of chiasmata according to a stationary gamma renewal process.

Increments are iid gamma\((\text{shape } = \nu, \text{ rate } = 2\nu)\)
(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
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 $\chi^2$ 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](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: $0.6 \pm 1.2$ cM
  - Male: $0.4 \pm 1.0$ cM
- Data cleaning
  - Removed 764/954,425 (~0.08%) genotypes resulting in tight double recombinants
<table>
<thead>
<tr>
<th>CEPH individual 1331–11</th>
<th>maternal chr 10 haplotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>11111111--- 11-111-11-- 11-11---1111 1-11---11-11</td>
<td></td>
</tr>
<tr>
<td>11111111--1 11-11111-11 1111111111 1111-111111</td>
<td></td>
</tr>
<tr>
<td>11--111111 111-11111-1 1111111111 10000-0-00</td>
<td></td>
</tr>
<tr>
<td>0---000-000 0000-000000 0000--0000 0000--0000</td>
<td></td>
</tr>
<tr>
<td>0000-000000 000-0--0-- 0---0-11-11-11111111-1</td>
<td></td>
</tr>
<tr>
<td>---1-1-1-1-1 1111-1--11 111111-1111 -111-111111</td>
<td></td>
</tr>
<tr>
<td>-1-----1111 111111-111 -1111-111-- 11-11111111</td>
<td></td>
</tr>
<tr>
<td>111-111111-11111111-1 11111111-11 11-1111111-11</td>
<td></td>
</tr>
<tr>
<td>11--1-11-1 111-11-1-1 1-1-----1-1 11-11111111</td>
<td></td>
</tr>
<tr>
<td>11--1-11-11-111111--111 11--1111111 1111-111111</td>
<td></td>
</tr>
<tr>
<td>1-0---0000 00000-0000 0-0-000</td>
<td></td>
</tr>
</tbody>
</table>
The data

Progeny

Location on female genetic map (cM)
MLEs of interference parameter

![Graph showing MLEs of interference parameter across different chromosomes for female and male subjects.]

- Female (open circles)
- Male (filled circles)

Chromosome axis ranges from 1 to 22, and the y-axis represents the interference parameter, ranging from 0 to 15.
Goodness of fit?
Maternal chromosome 1

Inter-crossover distance (cM)
Paternal chromosome 1

Inter-crossover distance (cM)
Maternal chromosome 2

Inter-crossover distance (cM)
Paternal chromosome 2
Mouse data


- **Two interspecific backcrosses with common F$_1$ 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
Mouse data

Backcross individuals

Chromosome position (cM)
MLEs: mouse
Crossover locations: Chr 1
Crossover locations: Chr 4
Inter-XO distances: Chr 1

Chr. 1

(n = 46)

Count

0 20 60 100

Inter–crossover distance
Inter-XO distances: Chr 4

Chr. 4
(n = 33)

Count

Inter–crossover distance
Acknowledgements

Terry Speed,  University of California, Berkeley, and WEHI

Hongyu Zhao,  Yale University

Mary Sara McPeek,  University of Chicago

Jim Weber,  Marshfield Medical Research Foundation
References


