Human crossover interference

Karl W Broman
Dept of Biostatistics
Johns Hopkins University

Joint work with James L. Weber,
Marshfield Medical Research Foundation

http://biosun01.biostat.jhsph.edu/~kbroman
Things I’ve learned

• Don’t be afraid of approximations.
• Study data aberrations.
• Data analysis is often a muddle.
• Consider making plots on the log scale.
Meiosis
Interference

- Strand choice
  → Chromatid interference
- Spacing
  → Crossover interference
Why study interference?

• Estimate the probability of a double crossover in a small interval.

• Obtain a model of meiosis for simulation and analysis.

• Compare human meiosis to that of other organisms.

Goals

• Demonstrate the presence of interference in human meiosis.

• Find a good model.

• Estimate the level of interference.
Recombination

Crossovers on a random meiotic product

Typical data: recombination information

We generally do not observe the locations of crossovers; rather, we observe the grandparental origin of DNA at a set of genetic markers.

Recombination across an interval indicates an odd number of crossovers.
Genetic markers

Short tandem repeat polymorphisms (STRPs)

- Also known as microsatellites.
- Individuals differ in the number of repeats.
- Use PCR w/ tagged primers to amplify the segment of DNA.
- Use gel electrophoresis to determine the fragment lengths.
Genetic distance

Distance (cM) =

average no. crossovers in 100 meiotic products

Per 100 cM:

2 chiasmata on 4-strand bundle
1 crossover on meiotic product

Map functions

Recombination fraction as a function of genetic distance (Here d is in Morgans.)

Haldane:

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

Kosambi:

\[ r = \frac{\tanh(2d)}{2} \]
Sex-specific recombination rate
Model organisms

\[\text{Drosophila data (Morgan et al. 1935)}\]

<table>
<thead>
<tr>
<th>Event</th>
<th>Count</th>
<th>Event</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000</td>
<td>10,431</td>
<td>1001</td>
<td>46</td>
</tr>
<tr>
<td>1000</td>
<td>771</td>
<td>0101</td>
<td>53</td>
</tr>
<tr>
<td>0100</td>
<td>1,579</td>
<td>0011</td>
<td>25</td>
</tr>
<tr>
<td>0010</td>
<td>1,221</td>
<td>1110</td>
<td>1</td>
</tr>
<tr>
<td>0001</td>
<td>1,994</td>
<td>1101</td>
<td>1</td>
</tr>
<tr>
<td>1100</td>
<td>4</td>
<td>1011</td>
<td>1</td>
</tr>
<tr>
<td>1010</td>
<td>7</td>
<td>0111</td>
<td>1</td>
</tr>
<tr>
<td>0110</td>
<td>4</td>
<td>1111</td>
<td>1</td>
</tr>
</tbody>
</table>

- Many meioses.
- A few linked markers.
- Consider rare multiple recombination events.
Human data

- **http://research.marshfieldclinic.org/genetics**
- **8 CEPH families**
  
  - 3 generations; 11 – 15 progeny; 92 meioses
- **~ 8,000 STRPs**
  
  - 90% typed; 0.5 cM spacing
- **Data cleaning**
  
  Removed 764 / 964,425 (~8/10,000) genotypes resulting in tight double recombinants.
Data cleaning

CRI-MAP *chrompic*
CEPH individual 1331–11; maternal chr 10

1111111--- 11–111–11--11i 1–11---11--
111111i--i-- 11–11111–11--1111111 1111–111111
11--1111111 111–111i–i 11111111111 10000–0–00
0--000–000 0000–00000 0000--0000 0000--0000
0000–00000 000–0--0-- --0–11–11– 11i1i11i–1
---1–i–1–i 1111–i--11 11111–111i1 11i–11111
–1----i1111 1i1111–111 11i1–11111 11–1111111i
111–i111i– 1111111–i– 1111111–1i 1i–111i11–
i1i–1–11–1 11i–1i–1–1 1–1----1–1 1i–1i1i1i11
1i--1--1i– 1ii11–111 11--1i111i 1i1i–11111
i–0---0000 00000–0000 00–000
Autozygosity

Chromosome 6
Family 884
The data
**Models**

- **4-strand bundle**
  - $x_0, x_1, x_2, x_3, x_4, x_n$

- **Random meiotic product**
  - $y_0, y_1, y_2, y_m$

**Count-location model**

- $n \sim (p_0, p_1, p_2, \ldots)$
- Locations $| n \sim \text{iid uniform}$

**Gamma model**

- $x_i's \sim \text{stationary gamma renewal process (shape } = u, \text{ rate } = 2u)$
- $y_i's \sim \text{mixtures of gammas}$
Model fit: C-L model

\[ m_i = \# \text{ crossovers} \]
\[ n_i = \text{ underlying } \# \text{ chiasmata} \]
\[ n_i \sim (p_0, p_1, p_2, \ldots) \]
\[ m_i | n_i \sim \text{ binomial}(n_i, 1/2) \]

MLEs by a version of the EM algorithm
Model fit: Gamma model

\[ x_1, x_2, \ldots \sim f(u, 2u) \]
\[ y_1, y_2, \ldots \sim \sum (\frac{1}{2})^k f(ku, 2u) \]
\[ x_0 \sim g = 2[1-F(u, 2u)] \]
\[ y_0 \sim \frac{1}{2} g + \sum (\frac{1}{2})^{(k+1)} g \ast f(ku, 2u) \]
\[ x_i \text{'s independent} \]
\[ y_i \text{'s independent} \]

- MLE of \( u \) using \( y_i \)’s.
- \( g \) calculated numerically.
- Convolutions calculated numerically.
- Maximization performed using a quasi-Newton algorithm.
Estimated level of interference
The data
Maternal chr 1

Inter-crossover distance (cM)
Paternal chr 1
Maternal chr 2

Inter-crossover distance (cM)
Paternal chr 2

Inter-crossover distance (cM)
Maternal chr 8

inter-XO distance (cM)
Chr 8p: maternal chromosomes
<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>David H. Ledbetter</td>
<td>U Chicago</td>
</tr>
<tr>
<td>Naomichi Matsumoto</td>
<td>U Chicago, Nagasaki U</td>
</tr>
<tr>
<td>Sabrina Giglio</td>
<td>U Chicago, U di Pavia</td>
</tr>
<tr>
<td>Christa M. Lese</td>
<td>U Chicago</td>
</tr>
<tr>
<td>Jessica A. Roseberry</td>
<td>U Chicago</td>
</tr>
</tbody>
</table>
Chr 8p: Interphase FISH

Homozygous inversion
Homozygous normal
Heterozygous
Chr 8p: Metaphase FISH
Approximations

• Correct marker order.
• Correct inter-marker distances.
• All crossovers observed.
• Interval censoring unimportant.
• No individual variation in recombination.
• Interference constant along chromosomes.
Summary

• Gamma model fits well.
• Count-location model fits poorly.
• Gamma parameter, $u \approx 4.3$
  (stronger than Kosambi, $u \approx 2.6$)
• No significant variation in interference between chromosomes.
• Possible individual variation in interference among mothers.
• Other findings: autozygosity; inversion on 8p
Pr(double XO | no recombination)

Probability of double crossover given no recombination

Interval length (cM)