Crossover interference in the mouse

Karl W Broman
Dept of Biostatistics
Johns Hopkins University
Meiosis
Interference

• Strand choice
  → Chromatid interference

• Spacing
  → Chiasma (crossover) interference
Why study interference?

• Obtain a model of meiosis for simulation and analysis
• It's interesting
• Determine shortest possible distance between crossovers

Goals

• Compare stochastic models for meiosis
• Characterize the level of interference in the mouse
• Compare level of interference between chromosomes
Recombination

Crossovers on 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 distance

distance (cM) = average # crossovers
in 100 meiotic products

per Morgan \{ 
2 chiasmata on 4-strand bundle
1 crossover on meiotic product

Map function

recombination fraction as a function of genetic distance

Haldane \quad r(d) = \frac{1}{2} \left[ 1 - \exp(-2d) \right]
Kosambi \quad r(d) = \frac{1}{2} \tanh(2d)
Carter-Falconer \quad d(r) = \left[ \tanh^{-1}2r + \tan^{-1}2r \right] / 4
The usual data

- Lots of meioses
- A few linked markers
- Look at frequency of rare multiple recombination events

*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>
Our data

C57BL/6J  

Mus spretus  

F₁  

C57BL/6J

94 BSB progeny
Typed at 1372 markers

C57BL/6J  

SPRET/Ei  

F₁  

SPRET/Ei

94 BSS progeny
Typed at 4913 markers
Genetic markers: STRPs or microsatellites

GATAGATA \ldots \ GATA
CTATCTAT \ldots \ CTAT
Basic methods

• Form integrated genetic map for the two crosses
• Identify intervals showing a recombination event
• Assume that recombination events indicate single crossovers, and that no double crossovers occurred
• Assume crossovers occurred at the center of the relevant interval (i.e., ignore interval censoring)
• Assume genetic distances known exactly (i.e., ignore sampling error)
Models

- **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 fitting

• Count-location model

\[ m_i = \# \text{ crossovers} \]
\[ n_i = \text{underlying} \# \text{ chiasmata} \]

\[ n_i \sim (p_0, p_1, p_2, \ldots) \]
\[ m_i \mid n_i \sim \text{binomial}(n_i, 1/2) \]

MLEs via a version of the EM algorithm
Model fitting

• Gamma model

\[ x_1, x_2, \ldots \sim f(u, 2u) \]
\[ x_0 \sim g = 2[1-F(u, 2u)] \]
\[ x_i \text{'s independent} \]

\[ y_1, y_2, \ldots \sim \sum (\frac{1}{2})^k f(ku, 2u) \]
\[ y_0 \sim \frac{1}{2} g + \sum (\frac{1}{2})^{(k+1)} g * f(ku, 2u) \]
\[ y_i \text{'s independent} \]

• MLE of \( u \) using \( y_i \text{'s} \)
• \( g \) calculated numerically
• Convolutions calculated numerically
• Maximization performed using a quasi-Newton method
### Distributions of # XOs / chr

<table>
<thead>
<tr>
<th>Source of Data</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chr. 1:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>53</td>
<td>87</td>
<td>46</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Poisson</td>
<td>70.3</td>
<td>69.2</td>
<td>34.0</td>
<td>14.5</td>
<td>23.9</td>
</tr>
<tr>
<td>Truncated Poisson</td>
<td>60.9</td>
<td>86.0</td>
<td>31.7</td>
<td>9.5</td>
<td>13.4</td>
</tr>
<tr>
<td>CL</td>
<td>52.7</td>
<td>88.6</td>
<td>44.6</td>
<td>2.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Obligate-chiasma CL</td>
<td>50.6</td>
<td>91.9</td>
<td>43.4</td>
<td>2.2</td>
<td>0.5</td>
</tr>
<tr>
<td>Gamma</td>
<td>49.3</td>
<td>89.1</td>
<td>44.7</td>
<td>5.0</td>
<td>2.1</td>
</tr>
<tr>
<td><strong>Chr. 14:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>76</td>
<td>108</td>
<td>4</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Poisson</td>
<td>101.4</td>
<td>62.6</td>
<td>19.3</td>
<td>4.7</td>
<td>56.1</td>
</tr>
<tr>
<td>Truncated Poisson</td>
<td>89.7</td>
<td>93.9</td>
<td>4.3</td>
<td>0.1</td>
<td>4.4</td>
</tr>
<tr>
<td>CL</td>
<td>89.5</td>
<td>93.8</td>
<td>4.6</td>
<td>0.1</td>
<td>4.4</td>
</tr>
<tr>
<td>Obligate-chiasma CL</td>
<td>89.5</td>
<td>93.9</td>
<td>4.5</td>
<td>0.1</td>
<td>4.3</td>
</tr>
<tr>
<td>Gamma</td>
<td>82.5</td>
<td>90.9</td>
<td>14.5</td>
<td>0.1</td>
<td>11.4</td>
</tr>
</tbody>
</table>
Discussion

• Approximations
  – Correct marker order
  – Correct genetic distances
  – All crossovers observed
  – Interval censoring unimportant
  – Interference constant across chromosome

• Conclusions
  – Gamma model fits well
  – Count-location model fits poorly
  – Gamma parameter, $u \approx 11$
    (stronger than Carter-Falconer, $u \approx 7.2$)
  – Apparent variation between chromosomes, with stronger interference in smaller chromosomes