Genetic maps
past, present and future

Karl W Broman
Biostatistics & Medical Informatics
University of Wisconsin – Madison

www.biostat.wisc.edu/~kbroman
Genetic maps
from my past, present and future

Karl W Broman
Biostatistics & Medical Informatics
University of Wisconsin – Madison

www.biostat.wisc.edu/~kbroman
Eucalypt genetic map

Byrne et al., Theor Appl Genet 91:869–875, 1995
What is a genetic map?

A sequence-based map measures distance between chromosome locations in basepairs.

A genetic map measures distance between chromosome locations via the recombination rate at meiosis.

Two markers are \( d \) centiMorgans (cM) apart if there is an average of \( d \) crossovers in the intervening interval in every 100 meiotic products.
Meiosis
Genetic distance

- Genetic distance between two markers (in cM) =
  
  Average number of crossovers in the interval in 100 meiotic products.

- “Intensity” of the crossover point process

- Recombination rate varies by
  - Organism
  - Sex
  - Chromosome
  - Position on chromosome
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.

Recombination fraction =

\[ Pr(\text{recombination in interval}) = Pr(\text{odd no. XOs in interval}) \]
Map functions

- A map function relates the **genetic length** of an interval and the **recombination fraction**.

\[ r = M(d) \]

- Map functions are related to **crossover interference**, but a map function is not sufficient to define the crossover process.

- **Haldane map function**: no crossover interference

- **Kosambi**: similar to the level of interference in humans

- **Carter-Falconer**: similar to the level of interference in mice
Ordering markers

Marker orders: A–B–C  A–C–B  B–A–C
Ordering markers

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th></th>
<th>A</th>
<th>abc</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>b</td>
<td>→</td>
<td>ABc</td>
<td>abC</td>
</tr>
<tr>
<td>C</td>
<td>c</td>
<td></td>
<td>Abc</td>
<td>aBC</td>
</tr>
</tbody>
</table>

Marker orders: A–B–C A–C–B B–A–C
### Ordering markers

<table>
<thead>
<tr>
<th>A</th>
<th>a</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>b</td>
</tr>
<tr>
<td>C</td>
<td>c</td>
</tr>
</tbody>
</table>

$\rightarrow$

<table>
<thead>
<tr>
<th>ABC</th>
<th>abc</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABc</td>
<td>abC</td>
</tr>
<tr>
<td>Abc</td>
<td>aBC</td>
</tr>
<tr>
<td>AbC</td>
<td>aBc</td>
</tr>
</tbody>
</table>

**Marker orders:** A–B–C    A–C–B    B–A–C
### Ordering markers

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>a</td>
<td>ABC</td>
</tr>
<tr>
<td>B</td>
<td>b</td>
<td>ABc</td>
</tr>
<tr>
<td>C</td>
<td>c</td>
<td>Abc</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AbC</td>
</tr>
</tbody>
</table>

Marker orders: A–B–C  A–C–B  B–A–C
Ordering markers

Marker orders: A–B–C  A–C–B  B–A–C

With M markers, there are $M!/2$ possible orderings.

For $M = 100$, $M!/2 \approx 10^{157}$
Eucalypt pedigree

118 progeny
Inferring phase

\[
\begin{pmatrix}
A_1 & B_1 \\
A_2 & B_2
\end{pmatrix}
\text{or}
\begin{pmatrix}
A_1 & B_1 \\
B_2 & A_2
\end{pmatrix}
\times
\begin{pmatrix}
C_1 & D_1 \\
C_2 & D_2
\end{pmatrix}
\text{or}
\begin{pmatrix}
C_1 & D_1 \\
D_2 & C_2
\end{pmatrix}
\]
Eucalypt genetic map

Byrne et al., Theor Appl Genet 91:869–875, 1995
CEPH pedigrees

1331 15 14 13 12
  2  1
  3 4 5 6 7 8 9 10 11 16 17

1332 16 15 14 13
  2  1
  3 4 5 6 7 8 9 10 11 12 17

1347 15 14 13 12
  2  1
  3 4 5 6 7 8 9 10 11 16

1362 16 15 14 13
  2  1
  3 4 5 6 7 8 9 10 11 12 17

1413 19 21 18 20
  2  1
  3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

1416 14 13 12 11
  2  1
  3 4 5 6 7 8 9 10 11 15 16

884 15 17 16
  2  1
  3 4 5 6 7 8 9 10 11 12 13 14

102
  2  1
  3 4 5 6 7 8 9 10 11 12 13 14 15 16
Marshfield maps: Tasks

- Assemble data
- Understand marker names
  - AFM, UT, CHLC (GATA etc.), Mfd, D*S*
- Identify cryptic duplicates
- Order markers and identify genotyping errors
  - Removed 764 / 969,425 genotypes
## Top of chr 22

<table>
<thead>
<tr>
<th>Marker</th>
<th>Dnumber</th>
<th>sex-ave(cM)</th>
<th>female(cM)</th>
<th>male(cM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ATA2G02</td>
<td>Unknown</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>2 GATA198B05</td>
<td>Unknown</td>
<td>1.79</td>
<td>0.00</td>
<td>2.60</td>
</tr>
<tr>
<td>3 AFM217xf4</td>
<td>D22S420</td>
<td>2.27</td>
<td>3.32</td>
<td>0.00</td>
</tr>
<tr>
<td>4 AFM288we5</td>
<td>D22S427</td>
<td>4.26</td>
<td>4.51</td>
<td>5.42</td>
</tr>
<tr>
<td>5 265yf5</td>
<td>D22S425</td>
<td>5.25</td>
<td>7.52</td>
<td>3.00</td>
</tr>
<tr>
<td>6 GGAA10F06</td>
<td>D22S686</td>
<td>0.03</td>
<td>0.00</td>
<td>0.65</td>
</tr>
<tr>
<td>7 AFMa037zd1</td>
<td>D22S539</td>
<td>0.84</td>
<td>0.00</td>
<td>0.82</td>
</tr>
<tr>
<td>8 AFM292va9</td>
<td>D22S446</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>9 Mfd51</td>
<td>D22S257</td>
<td>3.27</td>
<td>5.91</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Marker search

Mammalian Genotyping Service

Search for Markers

Enter the markers to be searched in the space below. Either probe or locus name may be used. Separate marker names with tabs, spaces, and/or "newlines".

Submit Form  Reset Form

http://research.marshfieldclinic.org/genetics/MarkerSearch/searchMarkers.asp
Figure 1  Plots of the female:male genetic-distance ratio against sex-averaged genetic location (in cM) along six selected chromosomes. Approximate locations of the centromeres are indicated by the triangles. The dashed lines correspond to equal female and male distances.
Total no. crossovers

Crossover locations

## Autozygosity

### Homozygous Segments for Individual 884-02

<table>
<thead>
<tr>
<th>Chromosome (Markers)</th>
<th>Cytogenetic Band(s)</th>
<th>Length (cM)</th>
<th>Proportion Homozygous</th>
<th>LOD Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 (D3S1571–D3S1617)</td>
<td>q28</td>
<td>4.9</td>
<td>9/9</td>
<td>5.53</td>
</tr>
<tr>
<td>4 (GATA144E02–D4S189)</td>
<td>p11-q12</td>
<td>11.1</td>
<td>21/21</td>
<td>12.26</td>
</tr>
<tr>
<td>5 (D5S398–D5S401)</td>
<td>q11-q14</td>
<td>29.8</td>
<td>77/77</td>
<td>46.21</td>
</tr>
<tr>
<td>6 (D6S1711–D6S278)</td>
<td>q11-q22</td>
<td>35.3</td>
<td>109/113</td>
<td>48.12</td>
</tr>
<tr>
<td>8 (D8S506–D8S385)</td>
<td>q22-q23</td>
<td>8.0</td>
<td>28/30</td>
<td>12.35</td>
</tr>
<tr>
<td>9 (D9S1802–D9S250)</td>
<td>q33</td>
<td>6.5</td>
<td>18/18</td>
<td>9.53</td>
</tr>
<tr>
<td>12 (D12S103–D12S1680)</td>
<td>q13-q21</td>
<td>11.3</td>
<td>43/43</td>
<td>21.82</td>
</tr>
<tr>
<td>16 (D16S494–D16S3107)</td>
<td>q21-q22</td>
<td>8.8</td>
<td>26/26</td>
<td>17.23</td>
</tr>
<tr>
<td>16 (D18S450–GATA51E05)</td>
<td>q21-q22</td>
<td>40.3</td>
<td>84/84</td>
<td>49.79</td>
</tr>
<tr>
<td>22 (D22S1156–D22S179)</td>
<td>q13</td>
<td>3.9</td>
<td>21/21</td>
<td>15.81</td>
</tr>
</tbody>
</table>

Crossover interference

Maternal chr 8
### Apparent triple XOs


<table>
<thead>
<tr>
<th>Marker</th>
<th>1362–10</th>
<th>1362–11</th>
<th>1413–03</th>
<th>1413–09</th>
<th>102–06</th>
<th>1332–05</th>
<th>Female CM</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8S561</td>
<td>●</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>○</td>
<td>3</td>
</tr>
<tr>
<td>D8S277</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>○</td>
<td>3</td>
</tr>
<tr>
<td>D8S1819</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>3</td>
</tr>
<tr>
<td>D8S439</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>○</td>
<td>3</td>
</tr>
<tr>
<td>D8S1706</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>○</td>
<td>3</td>
</tr>
<tr>
<td>D8S351</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>12</td>
</tr>
<tr>
<td>D8S1825</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>12</td>
</tr>
<tr>
<td>D8S1469</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>12</td>
</tr>
<tr>
<td>D8S542</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>12</td>
</tr>
<tr>
<td>D8S516</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>12</td>
</tr>
<tr>
<td>D8S1721</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>12</td>
</tr>
<tr>
<td>D8S376</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>12</td>
</tr>
<tr>
<td>D8S550</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>12</td>
</tr>
<tr>
<td>D8S265</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>12</td>
</tr>
<tr>
<td>D8S1695</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>12</td>
</tr>
<tr>
<td>D8S1759</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>12</td>
</tr>
<tr>
<td>D8S1755</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>12</td>
</tr>
<tr>
<td>D8S1130</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>12</td>
</tr>
<tr>
<td>D8S552</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>12</td>
</tr>
<tr>
<td>D8S308</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>12</td>
</tr>
<tr>
<td>KW306</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>12</td>
</tr>
<tr>
<td>D8S1106</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>12</td>
</tr>
<tr>
<td>D8S1754</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>12</td>
</tr>
<tr>
<td>GATA151F02</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>12</td>
</tr>
<tr>
<td>D8S1790</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>12</td>
</tr>
</tbody>
</table>
Chr 8p inversion

Broman et al., In: Science and Statistics: A Festschrift for Terry Speed, 2003
Comparison to sequence

Marshfield GL Map (N=85)

D22S529
D22S531
D22S534
D22S1175

Comparison to sequence (revisited)

Thanks to UCSC and Ensembl!
Dog genetic map

Neff et al., Genetics 151:803–820, 1999
Mellersh et al., Genomics 46:326–336, 1997
The dog X

Mellersh et al., Genomics 46:326–336, 1997
The dog X

Mellersh et al., Genomics 46:326–336, 1997

Neff et al., Genetics 151:803–820, 1999
C. savignyi (sea squirt)

Aim: Build linkage map to improve long-range ordering of draft genome sequence (which is currently represented as a few hundred “reftigs”)
C. savignyi pedigree

Markers: PCR amplicon (primers in exons, spanning an intron), digested with a restriction enzyme

→ 2, 3 or 4 banding patterns

Tricky bits: Which marker “phenotype” corresponds to which genotype?
Using information on locations of markers within “reftigs”
Example of two markers

<table>
<thead>
<tr>
<th></th>
<th>AC</th>
<th>AD</th>
<th>BC</th>
<th>BC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AC</strong></td>
<td>8</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>AD</strong></td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>BC</strong></td>
<td>2</td>
<td>0</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td><strong>BD</strong></td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>9</td>
</tr>
</tbody>
</table>
Example of two markers

<table>
<thead>
<tr>
<th>114467-MboI</th>
<th>AC</th>
<th>AD</th>
<th>BC</th>
<th>BC</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>8</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>AD</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>BC</td>
<td>2</td>
<td>0</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>BD</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>9</td>
</tr>
</tbody>
</table>
Example of two markers

<table>
<thead>
<tr>
<th></th>
<th>AC</th>
<th>AD</th>
<th>BC</th>
<th>BC</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>8</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>AD</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>BC</td>
<td>2</td>
<td>0</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>BD</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>9</td>
</tr>
</tbody>
</table>
**Example of two markers**

<table>
<thead>
<tr>
<th>114467-MboI</th>
<th>AC</th>
<th>AD</th>
<th>BC</th>
<th>BC</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>8</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>AD</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>BC</td>
<td>2</td>
<td>0</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>BD</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>9</td>
</tr>
</tbody>
</table>

**99481-HaeIII**
C. savignyi map
Shifman et al. *(PLoS Biology, 4:e395, 2006)* constructed a high-resolution genetic map of the mouse genome.

- 10,202 SNPs
- 80 families from the latest generations in a heterogeneous stock (HS) of outbred mice
- 4,048 meioses

- Valuable resource for mouse geneticists
- Characterization of recombination rate variation
  - Particularly regarding the sex difference in recombination.
Concerns

- In order to accommodate the analysis of 8 complex pedigrees, Shifman et al. used a sliding window of 5–15 SNPs.

- The remaining 72 families were all nuclear, and many lacked parental genotype data or had genotype data on just one parent, and many were small (as few as 2 siblings).

- The software used (CRIMAP, last revised in 1990) makes some approximations that result in biased estimates of genetic distances (even the sex-averaged ones) in the case of small sibships with incomplete parental genotype data.

- The length of the genetic map (and that of particular chromosomes) was quite different from previous mouse maps.
A complex pedigree

The diamonds are sibships.
What we did

- Obtained the raw data.
- Omitted 13 individuals with clear pedigree errors.
- Switched the sex of 26 individuals from female to male.
- Omitted 176 genotypes due to Mendelian inconsistencies.
- Split the large pedigrees into sibships (plus parents and grandparents).
- Split the larger sibships.
- Omitted sibships with no parental genotypes.
- Omitted small sibships ($\leq 8$ sibs) with genotype data on just one parent.
- Omitted 538 genotypes leading to apparent tight double crossovers.
Substantial differences

The revised genetic maps...

● Are much smaller.
  – The autosomal genome is 11% smaller in the revised maps

● Show a much smaller sex difference.
  – Shifman et al.: female autosomal genome is 26% longer than the male.
  – Revised maps: female autosomal genome is 9% longer than the male.

● Show fewer regions of unusually high recombination rate.
  – “Torrid” regions disappear or have markedly attenuated rec’n rates.
Recombination rates (chr 1)

Cox et al., Genetics 182:1335–1344, 2009
Recombination rates (chr 5)

Cox et al., Genetics 182:1335–1344, 2009
Recombination rates
(chr 18)

Cox et al., Genetics 182:1335–1344, 2009
Morals

- Genetic maps continue to be useful
- Be careful about automated map construction
- Careful, tedious work is often necessary
- The simplest things can have the greatest impact
- Artifacts can be more interesting than anything else
- Don’t give a sex-averaged map of the X chromosome
- Use care in data cleaning
- Split large pedigrees into non-overlapping sibships rather than resort to the use of a sliding window of markers
- Use computer simulations to verify the appropriateness of the choices you make in a complex analysis
Acknowledgments

Terry Speed, University of California, Berkeley
Jim Weber, PreventionGenetics (formerly Marshfield Medical Research Foundation)
Mark Neff, University of California, Davis
Matt Hill and Arend Sidow, Stanford
Beth Dumont, University of Wisconsin–Madison
Gary Churchill, Bev Paigen, Ken Paigen, et al., The Jackson Lab
Jonathan Flint, The Wellcome Trust