Lecture 5: Recombination, IBD distributions and linkage

Magnus Dehli Vigeland

Statistical methods in genetic relatedness and pedigree analysis
NORBIS course, 6th – 10th of January 2020, Oslo
Outline

• Review from yesterday:
  – Meiotic recombination
  – IBD segments

• Consequences of the discrete nature of recombination

• Measuring distance in the genome
  1) Physical distance
  2) Genetic map distance ( = crossover rate)
  3) Recombination rate

• Map functions: Translating between 2) and 3)

• Using linked markers for relatedness
Recombination and IBD
IBD segments

IBD status
Consequences of the discrete nature of recombination

- Distant relatives: Possible to have no IBD sharing!
- 100 % inbred?
- Indistinguishable relationships - or not?
  - Half siblings
  - Grandparent/grandchild
  - Aunt/nephew
Let's not forget – Prince Harry and Meghan Markle are actually (very distant) cousins
P(any IBD) ≈ 0

13th cousins once removed
Is 100 % inbreeding possible?

- Full sibs mating scheme
- Easy to show:
  - inbreeding coefficient $f \rightarrow 1$
- But never $f = 1!$ (in a finite pedigree)

Realised inbreeding
After ~30 generations, usually $f_{\text{real}} = 1$. 
Indistinguishable relationships?

Statistical methods in genetic relatedness and pedigree analysis
Distributions of IBD segments

Clear difference between GP and the other two!
Outline

• Review from yesterday:
  – Meiotic recombination
  – IBD segments

• Consequences of the discrete nature of recombination

• Measuring distance in the genome
  1) Physical distance
  2) Genetic map distance ( = crossover rate)
  3) Recombination rate

• Map functions: Translating between 2) and 3)

• Using linked markers for relatedness
Physical distance

- The **physical distance** between two loci
  \[ \text{number of base pairs between them} \]

  \[
  \begin{array}{c}
  \text{ATCGCGACCATAATG} \\
  \text{TAGGCGTGTTTTAC}
  \end{array}
  \] 4 bp

  \[
  \begin{array}{c}
  \text{GTATCGGCGTCCA} \\
  \text{CATAGCGCGAGGT}
  \end{array}
  \]

- Units:
  - 1 bp (base pair)
  - 1 kb = 1000 bp ("1 kilobase")
  - 1 Mb = 1 000 000 bp ("1 megabase")

- The physical distance/position is often the ultimate goal, but **rarely accessible** by experiments
Map distance

- Chromosomal crossovers:

  ![Crossover](image1.png)
  ![Double crossover](image2.png)

- The **genetic map distance** between two loci
  
  = average number of crossovers between them

- Units:
  - 1 Morgan (M) = average 1 crossover per meiosis
  - 1 centiMorgan (cM) = 0.01 M

- The entire human genome: Ca 30 Morgan
**Map distance**

- Rule of thumb:
  \[ 1 \text{ cM} \approx 1 \text{ Mb} \]

- But: crossover rates vary
  - across the genome
  - males vs. females
Females have a much longer genome!
Consequences
Napoleon Bonaparte (1769 - 1821)  

10 gen. paternal line  

you

\[ P(\text{any IBD sharing}) \approx 19\% \]

Jane Austen (1775 - 1817)  

10 gen. maternal line  

you

\[ P(\text{IBD sharing}) \approx 33\% \]
Can we separate these??

HS paternal

HS maternal
Yes!

Statistical methods in genetic relatedness and pedigree analysis
• So far:
  – crossover rates: meiotic process
  – IBD distributions
  – has nothing to do with markers

• Now: Change focus
  – marker genotypes
  – what can we observe?
Map distance

- Crossovers (and thus map distances) are not directly observable with markers
- Reason: multiple crossovers between markers

Remember: We don't observe the meioses themselves, only their outcome (the gametes)

What can we observe directly?
- Answer: Recombination rates between markers
Recombination rate

- The **recombination rate** between two loci
  
  \[ \text{recombination rate} = \text{average number of recombinant gametes} \]
Recombination rate

- Loci on different chromosomes: $\theta = 0.5$
- Loci far apart on the same chromosome: $\theta \approx 0.5$
- Loci right next to each other: $\theta = 0$

Definition: Two loci are linked $\theta < 0.5$ (In plain language: "On the same chromosome, not too far apart")
Crossover rate vs. recombination rate

Map distance: Crossover rate
\[ d = \mathbb{E}[\text{#crossovers}] \]

- Based on a fundamental property of the meiosis
- Very natural measure of distance
- But:
  - Hard to observe directly

Recombination rate:
\[ \theta = \mathbb{E}[\text{#recombinant gametes}] \]

- Perhaps not as intuitive
- Relative to markers
- But:
  - Easy to estimate using genotyping

Can we relate these in some way?
Haldane's map function

- Given observations on recombinations, we would like to compute the crossover rate.
- Requires a *statistical model* of the crossover process
- Haldane's model:
  - Crossover events occur completely at random, with fixed rate 1.
  - Events are independent (in reality not true: cannot be too close)
  - Poisson process!

- From this easy to compute the recombination rate, using that
  \[ \theta = P(\text{odd number of crossovers}) \]

Haldane's map function:

\[ \theta = \frac{1}{2} (1 - e^{-2d}) \]
Haldane's map function

Recombination rate vs. Morgan
Likelihood with linked markers

Unlinked markers:

\[ L = L_1 \times L_2 \]

This does not hold if the markers are linked!

Ignoring linkage can lead to serious errors
Advantages of linked markers

Cannot be distinguished with unlinked markers.

But they CAN with linked markers!
Summary

• Distributions of IBD segments

• Measures of genomic distance
  – physical
  – genetic distance (= crossover rate):
    ▪ centiMorgan
  – recombination rate

• Haldanes map function

• Marker linkage in relatedness analysis
  – bad (if ignored)
  – good (esp. for distinguished some relationships)