By the end of this lecture you should be able to explain:

- The four main types of information you need to for QTL analysis
- Why understanding recombination & genetic linkage is important for localising genes that control traits
- What a marker-trait association is
Objectives of QTL analysis

- The statistical study of the alleles that occur at a locus and the phenotypes (traits) that they produce
- Methods developed in the 1980s
- Next-gen sequencing = enabling
- Rapid, genome-wide analysis possible
- Better statistical methods
What you need for QTL analysis:

- (i) A large population of individuals that you can score
  - variation in phenotype
  - population designed using parents with contrasting phenotypes

- (ii) Markers over the genome to pinpoint QTL location

- (iii) A way to identify which markers from each parent have been inherited by the progeny

- (iv) A map of the genome to find out where you are
(i) Mapping lines

F1 = Heterozygous at all loci

Parents = Homozygous

F1 = Heterozygous at some loci

F7 RILs = Homozygous at all loci & heterogeneous

Crossing-over (recombination)

Many different individuals are obtained & separately selfed to develop RILs

F1 microspore culture of male gametes

DH population (homozygous)
(ii) Genetic markers

- Most common: molecular markers (DNA sequence differences)
- What else could you use?
(iii) A way to distinguish molecular markers

- Restriction enzymes e.g. EcoRI cut DNA only at a specific recognition sequence
- Compare restriction patterns:

Parent A

```
--------GAATTC--------GAATTC--------GAATTC--------
--------GAATTC--------GAATTC--------GAATTC--------
```

Parent B

```
--------GAATTC--------GATTTC--------GAATTC--------
--------GAATTC--------GATTTC--------GAATTC--------
```

First generation (F1)

```

```

Second generation (F2) from selfing F1:

```

```
(iv) Genetic map

- Physical map: lays out the sequence information and annotates it: promoters, genes etc.

- Linkage map:
  order of genetic markers and relative distances from each other
    - based on meiotic recombination (crossing over)
  between chromosomes

- Link genetic map to physical map
Genetic linkage is related to recombination frequency

- **Rf = 0.5 (50%)**
  - = no linkage
  - **Rf = recombination frequency**

- **More recombination so Rf = high (<0.5)**
  - = weak linkage

- **Some recombination so Rf = medium**
  - = quantifiable linkage

- **Little recombination so Rf = small**
  - = tight linkage
**Map distances and genetic linkage**

- Recombination frequency of 0.01 (1%) = a genetic map unit of 1 cM
- Recombination events occur randomly, once or twice per chromosome
- Linkage map made by characterising the recombination events that have taken place in a cross between two parental genotypes
- Assumes that linkage is the only cause of non-independence between markers and that segregation is Mendelian
- Haldane mapping function adjusts map distance to account for double crossovers that go undetected
- Kosambi mapping function also adjusts for crossover interference
Linkage groups are the basis of genetic maps

These should theoretically correspond to chromosomes, but if...

• Chromosomes very long

• Recombination frequency very high

• Mapping populations are not large enough

  ...one chromosome can statistically “break” into several linkage groups

• Also, centromeres and heterochromatin have suppressed recombination
Making a map: Selecting markers

• Are the markers polymorphic between the parental lines?

Marker 1 parental genotypes

• P1 274 bp
• P2 283 bp
Making a map: scoring genotypes

• The number of SSRs is highly variable among individuals
Making a map: scoring genotypes

<table>
<thead>
<tr>
<th>Markers</th>
<th>DH01</th>
<th>DH02</th>
<th>DH03</th>
<th>DH04</th>
<th>DH05</th>
<th>DH06</th>
<th>DH07</th>
<th>DH08</th>
<th>DH09</th>
<th>DH10</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>B</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>C</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>D</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
</tr>
<tr>
<td>E</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
</tr>
<tr>
<td>F</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
</tr>
<tr>
<td>G</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>H</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>I</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>J</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>a</td>
</tr>
</tbody>
</table>
An example by hand

Chart 1
• Determine all pairwise recombination frequencies (each marker with every other marker)
Determining map order: example continued

Chart 2
• Identify closely linked markers

Charts 3 and 4
• Determine the order of markers for each linkage group
Determining map order: LOD scores

- LOD score: likelihood of the observed linkage
- Statistical analysis of +/- of 100s of markers in (F7) progeny population (parental genotypes)

\[
LOD = \log_{10} \left( \frac{(1 - \theta)^{NR} \times \theta^R}{0.5^{(NR+R)}} \right)
\]

- NR = number of non-recombinant offspring; R = number of recombinant offspring
- Theta = recombinant fraction = R / (NR + R)
- Mapping software e.g. Mapmaker, JoinMap
Create linkage groups -> Linkage (genetic) map

Increasing LOD score

map units

cM
Map sorted genotype data = graphical genotypes

- Format and check the data
- Calculate all pairwise recombination frequencies
- Assign markers to linkage groups then map markers within each linkage group
Quantitative trait loci analysis & association mapping

Statistical analysis reveals correlation of the parental genotype of a segment of the genome with the value of the trait.
• Results from marker A/a: suggests that the gene is very close to the marker
• Results from marker B/b: suggests that the gene is not linked to the marker
Have to consider multiple loci
Input data for QTL analysis:

- marker data matrix
- trait data matrix
To map a quantitative trait (QTL analysis):

1. Make a suitable population
2. Genotype individuals - generate linkage map

3. Collect phenotypic measurements
   - Evaluate in uniform environment,
   - Evaluate in multiple environments
   - Data transformation (approach normal distribution)
   Look at correlations between traits, transgressive segregation?

4. Look for trait-marker associations

5. Estimate the effects of the QTLs on the quantitative trait:
   - many genes with small effect each or few genes with large effect each?
   - their effects on the trait: is gene action additive or dominant?
   - their positions in the genome: linkage and association, epistasis
   - their interaction with the environment

6. Identify candidate genes underlying the QTL and thus the trait

Variation in a trait in a population is caused by Genetic and Environmental contributions
\[ V_P = V_G + V_E \]
Variation due to genetical causes is the heritability of the trait
\[ h^2 = \frac{V_P}{V_G} \]
To map a quantitative trait (QTL analysis):

1. Cross parental lines A × B.

From the cross, develop a population of 100-300 progeny families. Evaluate each progeny both for phenotypic traits and for molecular markers.

2. Phenotypic evaluation

3. Marker evaluation

4. QTL statistical analysis:
   Analyze combined phenotypic and marker data

Identify and characterize QTLs

Variation in a trait in a population is caused by Genetic and Environmental contributions

\[ V_P = V_G + V_E \]

Variation due to genetical causes is the heritability of the trait

\[ h^2 = \frac{V_P}{V_G} \]
Next time:
(3) QTL and GWAS methods

By the end of this lecture you should be able to explain:

- Some of the principles underlying the statistical analysis of QTLs
- Under what conditions particular methods are suitable
- The core differences between QTL analysis and GWAS