Basic Steps of a Pedigree Plant Breeding Program

1. Create F2 population and screening the population
   - a  b  c  n
   - a  b  c  n
   - a  b  c  n
   - Discard from program

2. Make selections and continue evaluation
   - b  b  b  n  n  n
   - b  b  b  n  n  n
   - b  b  b  n  n  n
   - Discard from program

3. Evaluate further and select best performing line for release
   - b
   - b
   - b

Challenge for Plant Breeding Programs

- Select lines with greatest performance for agronomic traits
  - Earlier selection the better
    - Reduces costs to produce an individual line for release
- If you reduce the cost, more lines can be evaluated
- But screening for quantitative traits can be expensive and difficult

Question for plant molecular genetics

- Can a marker system be developed that efficiently selects for important quantitative (agronomic) traits?
Plant Quantitative Trait Loci Analysis Is Not New

Sax (1923) Genetics 8:552-560

- **Species:** Common Bean
  - Parents are inbreds
    - Yellow Eye and Dot Eye
      - Pigmented, heavy weight
    - 1333
      - Non-pigmented, light weight
- **Specific results**
  - $PP \ 4.3 \pm 0.8$ centigrams heavier than $pp$
  - $Pp \ 1.9 \pm 0.6$ centigrams heavier than $pp$
- **General conclusion**
  - Offspring of multiple crosses
    - All heavy seeds were pigmented
    - All light seeds were non-pigmented
- **What gene is linked to seed weight**
  - Pigmentation gene
    - $P$ allele
      - Pigmentation in seed and flower
    - $p$ allele
      - No pigmentation in seed or flower
- **Conclusion**
  - *A factor linked to $P$ acts in an additive manner to control seed weight*

**Other examples**

- Lindstrom (1924) Science 60:182-183
  - Tomato
    - Fruit color is linked to fruit size
Applications of Molecular Markers to QTL Selection

1. Create a high yielding, high ear number corn population (FS10) by selection from an unselected population (UNS). The population was developed by ten cycles of selection for these traits.

2. Determine the allelic frequencies for eight isozymes in the original (UNS) and selected (FS10) population.

Table 1. Frequencies of alleles at eight allozyme loci in the FS10 population

<table>
<thead>
<tr>
<th>Allele</th>
<th>UNS frequency</th>
<th>FS10 frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acph1-c</td>
<td>0.198</td>
<td>0.528</td>
</tr>
<tr>
<td>β-glu1-k</td>
<td>0.571</td>
<td>0.903</td>
</tr>
<tr>
<td>Phi1-e</td>
<td>0.984</td>
<td>0.711</td>
</tr>
<tr>
<td>Pgm1-A9</td>
<td>0.735</td>
<td>1.000</td>
</tr>
<tr>
<td>Pgd1-A2</td>
<td>0.556</td>
<td>0.792</td>
</tr>
<tr>
<td>Pgd2-B5</td>
<td>0.667</td>
<td>1.000</td>
</tr>
<tr>
<td>Mdh1-A6</td>
<td>0.642</td>
<td>0.204</td>
</tr>
<tr>
<td>Mdh2-B3</td>
<td>0.255</td>
<td>0.447</td>
</tr>
</tbody>
</table>

3. Create a population (ALZ) which has essentially the same allelic frequencies as the selected population by selecting appropriate individuals from the base population.

4. Measure the yield and ear number/plant of the UNS, FS10, and ALZ populations in a replicated field trial over two years.

5. Results

- The ALZ population yield was equal to that found in the FS10 population after two rounds of selection.
- The ALZ population ear number was equal to that found in the FS10 population after 1½ cycles of selection.
Figure 9.1
Comparison of continuous variation (ear length in corn) with discontinuous variation (height in peas)

![Diagram with distributions and statistics]

Figure 21.2 Frequency distributions and descriptive statistics of time to maturity in four populations of wheat. P<sub>A</sub> and P<sub>B</sub> are inbred varieties that were crossed to produce F<sub>1</sub> hybrids. The F<sub>1</sub> plants were then intercrossed to produce an F<sub>2</sub>. Seed from all four populations was planted in the same season to determine the time to maturity. In each case, data were obtained from 40 plants. The mean maturation times (X) are indicated by the triangles; the sample variances (s<sup>2</sup>) and standard deviations (s) are also given.
Fig. 1. Recombinant inbred frequency distribution for temperature injury in maize. Arrows indicate mean values of the two parental lines: $P_1 = Pa33$, $P_2 = B37$.

Table 1. RFLP loci showing significant effect on membrane injury. $b$ and $R^2$ value are, respectively, the estimated effects and the proportion of between-lines variability. The denomination of the loci is according to Burr et al. (1988).

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Locus</th>
<th>$b$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.21</td>
<td>-0.096</td>
<td>0.099</td>
</tr>
<tr>
<td>2</td>
<td>5.21B</td>
<td>0.104</td>
<td>0.138</td>
</tr>
<tr>
<td></td>
<td>6.20</td>
<td>0.116</td>
<td>0.131</td>
</tr>
<tr>
<td></td>
<td>5.61B</td>
<td>0.118</td>
<td>0.154</td>
</tr>
<tr>
<td></td>
<td>NPl298</td>
<td>0.126</td>
<td>0.157</td>
</tr>
<tr>
<td>4</td>
<td>ZPL2A</td>
<td>-0.087</td>
<td>0.099</td>
</tr>
<tr>
<td>8</td>
<td>NPl220</td>
<td>0.136</td>
<td>0.229</td>
</tr>
<tr>
<td>9</td>
<td>3.06</td>
<td>-0.124</td>
<td>0.155</td>
</tr>
<tr>
<td></td>
<td>WX1</td>
<td>-0.133</td>
<td>0.170</td>
</tr>
<tr>
<td></td>
<td>5.04</td>
<td>-0.114</td>
<td>0.131</td>
</tr>
<tr>
<td></td>
<td>7.13</td>
<td>-0.109</td>
<td>0.124</td>
</tr>
<tr>
<td>10</td>
<td>NPl269B</td>
<td>-0.118</td>
<td>0.142</td>
</tr>
</tbody>
</table>

Fig. 2. RFLP analysis for temperature injury in recombinant inbreds from T32 x CM37 F$_1$ hybrid. Horizontal bars indicate degree ($R^2$) of correlation between RFLP loci and CMS. *: significant values ($p < 0.05$). A cluster, indicated as a circle, of significant values stands for a single QTL.
Table 3. Genomic location, genetic effects, and cumulative percentage of phenotypic variability for plant height accounted for by QTLs in four maize populations

<table>
<thead>
<tr>
<th>Population</th>
<th>Nearest RFLP locus</th>
<th>Chromosome</th>
<th>Distance* (cM)</th>
<th>Possible genetic loci</th>
<th>Estimated genetic effects</th>
<th>Cumulative % Var</th>
</tr>
</thead>
<tbody>
<tr>
<td>B73 x MO17</td>
<td>bn18.35</td>
<td>3</td>
<td>65</td>
<td>d1</td>
<td>-9.1</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>wx1</td>
<td>9</td>
<td>45</td>
<td>d3</td>
<td>-7.1</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>umc131</td>
<td>2</td>
<td>85</td>
<td>br2</td>
<td>7.2</td>
<td>-5.3</td>
</tr>
<tr>
<td></td>
<td>pio150033</td>
<td>10</td>
<td>65</td>
<td>st1</td>
<td>-6.9</td>
<td>-4.8</td>
</tr>
<tr>
<td></td>
<td>bnl12.06</td>
<td>1</td>
<td>105</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>umc42</td>
<td>4</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B73 x G35</td>
<td>umc83</td>
<td>1</td>
<td>185</td>
<td>br1, an1</td>
<td>-8.2</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>umc61</td>
<td>2</td>
<td>55</td>
<td>d5</td>
<td>-7.8</td>
<td>-7.7</td>
</tr>
<tr>
<td></td>
<td>bn15.37</td>
<td>3</td>
<td>120</td>
<td>yd2, na1</td>
<td>-6.6</td>
<td>8.6</td>
</tr>
<tr>
<td></td>
<td>pio200006</td>
<td>3</td>
<td>65</td>
<td>cr1, d1</td>
<td>-6.4</td>
<td>NS</td>
</tr>
<tr>
<td>K05 x W65</td>
<td>bnl17.56</td>
<td>5</td>
<td>65</td>
<td>glt1</td>
<td>6.4</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>pio1000014</td>
<td>5</td>
<td>145</td>
<td>na2, td1, bvt</td>
<td>5.8</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>bnl10.39</td>
<td>8</td>
<td>50</td>
<td>ct1, Sdw1</td>
<td>5.8</td>
<td>NS</td>
</tr>
<tr>
<td>J40 x V94</td>
<td>pio200569</td>
<td>7</td>
<td>40</td>
<td>d3</td>
<td>6.6</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>umc81</td>
<td>9</td>
<td>45</td>
<td></td>
<td>7.6</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>pio20095</td>
<td>6</td>
<td>40</td>
<td></td>
<td>7.1</td>
<td>NS</td>
</tr>
</tbody>
</table>

* Distance is measured from the terminal marker on the short arm of the chromosome.

** All effects within a population were estimated simultaneously using MAPMAKER/QTL (Lander and Lincoln, unpublished). The sign of the estimated additive effects is associated with the allele from the male parent. For example, the estimated additive effect for the first QTL listed (−9.1) indicates that the allele from MO17 is associated with families that are, on the average, 9.1 cm shorter. The sign associated with estimated dominance effects indicates the effect of the allele from the male parent for the heterozygous condition. For example, the first QTL has an estimated dominance deviation of 3.8 cm, which indicates that the heterozygotes are 3.8 cm taller than expected, based upon the estimated additive effects (−9.1 cm) of the MO17 allele.
Phenotypes of Genes that Mapped near Plant Height QTL

<table>
<thead>
<tr>
<th>Cross</th>
<th>Chromosome</th>
<th>Map location</th>
<th>Nearest gene</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>B73 x MO17</td>
<td>3</td>
<td>65</td>
<td>d1</td>
<td>dwarf plant</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>45</td>
<td>d3</td>
<td>dwarf plant</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>105</td>
<td>br2</td>
<td>small plant</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>100</td>
<td>st1</td>
<td>small plant</td>
</tr>
<tr>
<td>B73 x G35</td>
<td>1</td>
<td>185</td>
<td>br1</td>
<td>small plant, GA response</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>185</td>
<td>an1</td>
<td>dwarf height</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>55</td>
<td>d5</td>
<td>dwarf plant</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>120</td>
<td>yd2</td>
<td>yellow dwarf</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>120</td>
<td>na1</td>
<td>short, dwarf plant</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>65</td>
<td>cr1</td>
<td>short plant</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>65</td>
<td>d1</td>
<td>dwarf plant</td>
</tr>
<tr>
<td>K05 x W65</td>
<td>5</td>
<td>65</td>
<td>gl17</td>
<td>glossy, crossbend leaf</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>145</td>
<td>na2</td>
<td>short plant</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>145</td>
<td>td1</td>
<td>dwarf plant</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>145</td>
<td>bv1</td>
<td>short internodes, short plant</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>50</td>
<td>ct1</td>
<td>compact plant</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>50</td>
<td>Sdw1</td>
<td>semi-dwarf plant</td>
</tr>
<tr>
<td>J40 x V94</td>
<td>6</td>
<td>45</td>
<td>d3</td>
<td>dwarf plant</td>
</tr>
</tbody>
</table>

Important Points

1. Different QTL affect a quantitative trait in different crosses

2. Multiple QTL are found throughout the genome that affect the quantitative trait

3. QTL reside near genes that are known to control the phenotype of a quantitative trait

4. In general, QTL are marker loci that are located near a functional locus that affects a quantitative trait.
Statistical Approaches to Quantitative Trait Analysis in Plants

Quantitative traits
- Distributed as a continuum of phenotypes
- Often scored numerically.
- Term quantitative used to describe these traits

Distribution of a segregating population (such as $F_2$)
- Normal distribution.
  - $F_2$ population from cross of high and low yielding cultivars
  - Distribution will contain plants
    - Transgressive segregation
    - But majority of the plants
      - Yield value is between the parent phenotype
      - Most near the mean value

Genetic control of quantitative genetics
- Multiple genes affecting the trait are segregating in the population
- Goal of quantitative trait analysis
  - Locate the position in the genome where these genes reside

These genes are called quantitative trait loci (QTL) and any one locus is called a quantitative trait locus (QTL).

Note about nomenclature: By convention, QTL is both singular and plural. It is not acceptable to use QTLs to designate multiple genes affecting a quantitative trait.

Single Marker QTL Analysis

Single marker analysis
- First technique used to associate a specific marker with a quantitative trait.
- Basic principles are clearly described
  - Edwards et al. (1987; Genetics 116:113-125)
- Null hypothesis
  - Phenotypic trait value and marker genotype are independent.
- Alternative hypothesis
  - Phenotypic value and marker genotypes are not independent
    - Implies a gene affecting the quantitative trait is linked to the marker.
- Goal of single marker analysis
  - Discover those markers to which a QTL is linked
Steps in single marker QTL analysis

- Develop a segregating population from parents with contrasting phenotypes
  - Most common populations
    - F2 and recombinant inbred populations
- Analysis of population
  - Collect phenotypic data
    - Normally uses replicated trials
  - Collect molecular genotypic data
- Analysis
  - Data used to discover association between the marker and the quantitative trait
  - Procedure
    - F2 population
      - For each marker, individuals assigned to
        - Homozygous classes
          - M1M1 or m1m1
        - Heterozygous class
          - M1m1
    - Recombinant inbred population
      - For each marker, individuals assigned to
        - Homozygous classes
          - M1M1 or m1m1
    - Dominant markers
      - Individuals assigned to just two classes
        - M1_ or m1m1
- One-way analysis of variance
- Analysis performed for each marker
  - Does the mean phenotypic value of each class differ?
    - Yes
      - A QTL is linked to that marker.
    - No
      - The marker is not located near a QTL for the trait
Example of Phenotypic Marker Data and Single Factor Analysis

<table>
<thead>
<tr>
<th>Line</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phenotype</td>
<td>29</td>
<td>16</td>
<td>32</td>
<td>44</td>
<td>46</td>
<td>18</td>
<td>14</td>
<td>45</td>
<td>35</td>
<td>15</td>
<td>47</td>
<td>30</td>
</tr>
<tr>
<td>Marker M1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Marker M2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

1=mnmn, 2=Mnmn, 3=MnMn

**Marker M1**

<table>
<thead>
<tr>
<th>45</th>
<th>30</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1m1</td>
<td>M1m1</td>
<td>M1M1</td>
</tr>
</tbody>
</table>

**Marker M2**

<table>
<thead>
<tr>
<th>45</th>
<th>30</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1m1</td>
<td>M1m1</td>
<td>M1M1</td>
</tr>
</tbody>
</table>
Figure 1. Graphical representation of the relationship between the three marker classes observed in an F2 and the phenotypic mean of all individuals that possess that marker genotype. The two markers are M1 and M2, and these have alleles M1, m1, and M2, m2. (A) This distribution shows a positive association between the phenotypic value and marker M1. (B) This distribution shows no relationship between M2 marker genotype and phenotypic value.

Assessing the effect of the QTL

- $R^2$
  - Regress the phenotypic value onto the genotypes of the marker classes
    - Equivalent to the square of the correlation coefficient
      - Estimates the proportion of the phenotypic variation due to the different markers
      - What does (and does not) this value represent.

- Fig. 2
  - Phenotypic distribution about the three marker classes for two different markers
    - Each accounts for different amount of the phenotypic variation
    - Note the difference in the distributions
      - Means for marker M3 are broader (with a larger variance) than for M1
      - Results in a lower correlation
      - And a lower $R^2$ value for M3.

This does not mean that the effect of the QTL (= gene) linked to M3 is necessarily less than the effect of the QTL linked to M1.

- Statistic actually tells us nothing about the effect of the QTL.
- So why does it not give an indication of the size of the effect of the QTL?
- Distance between marker and QTL has an effect on the $R^2$ value
Figure 2. Phenotypic distributions about the marker classes for two markers determined by single-factor analysis of variance to be significantly associated with levels of expression of a quantitative trait. Marker M1 (A) has an $R^2$ value larger than that observed for marker M3 (B). Notice the differences in the breadth of the distributions about each marker class for each marker.

Understanding this concept

- Must consider the distribution of the phenotypes relative to
  - The marker
  - The QTL that is linked to it
- Fig. 3A
  - Shows the marker (M1) and a QTL (Q1) relationship
  - Parents have
    - Contrasting marker (M1 vs. m1) and QTL (Q1 vs. q1) genotypes
    - F1 will generate four different gamete types.
  - Linkage theory predicts
    - The frequency of each of these parental and recombinant gametes
      - Depends upon the linkage distance between M1 and Q1.
    - Closer the marker and the QTL
      - Fewer recombinant gametes are created.
    - Further apart the marker and the QTL
      - Larger the frequency of the recombinant gametes.
How linkage distance affects the phenotypic distribution

- For each marker class
  - Consider the different genotypes that constitute each marker class (Fig. 4.)
    - Q1 allele = 10 phenotypic units
    - q1 allele = 2 phenotypic units.
- Case 1: Marker and QTL are unlinked
  - Each marker class has an equal ratio of Q1 and q1 alleles.
  - Phenotypic mean phenotype of each marker class would be equal
- Case 2: Marker and QTL are linked
  - Fewer recombinant gametes and more parental gametes
    - M1M1 marker class
      - Contains more Q1 alleles than q1 alleles
  - Why
    - Q1 is linked to M1
    - Gamete distribution skewed toward parental gametes
  - Result
    - Higher mean yield for M1M1 class
**Figure 4.** Distribution of possible marker/QTL genotypes within each marker class for an F₂ generation. For each individual, the chromosome contributed by the female is listed on top and the chromosome contributed by the male is on the bottom. P refers to a parental chromosome and R a recombinant chromosome. It is important to note that lacking linkage, each of the marker classes contains equal frequency of parental and recombinant gametes.
Table 1. How linkage distance affects means of marker classes. Three cases are shown for the M1M1 and m1m1 marker classes; A. Marker M1 and QTL Q1 unlinked; B. Marker M1 and QTL Q1 linked at 10 cM; C. Marker M2 and QTL Q1 linked at 5 cM.

A. Unlinked genes

<table>
<thead>
<tr>
<th>Marker class</th>
<th>Frequency</th>
<th>Phenotypic value</th>
<th>Class mean contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1Q1/M1Q1</td>
<td>0.25</td>
<td>10 + 10</td>
<td>5</td>
</tr>
<tr>
<td>M1Q1/M1q1</td>
<td>0.25</td>
<td>10 + 2</td>
<td>3</td>
</tr>
<tr>
<td>M1q1/M1Q1</td>
<td>0.25</td>
<td>2 + 10</td>
<td>3</td>
</tr>
<tr>
<td>M1q1/M1q1</td>
<td>0.25</td>
<td>2 + 2</td>
<td>1</td>
</tr>
</tbody>
</table>

M1M1 class mean 10

<table>
<thead>
<tr>
<th>Marker class</th>
<th>Frequency</th>
<th>Phenotypic value</th>
<th>Class mean contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1Q1/m1Q1</td>
<td>0.25</td>
<td>10 + 10</td>
<td>5</td>
</tr>
<tr>
<td>m1Q1/m1q1</td>
<td>0.25</td>
<td>10 + 2</td>
<td>3</td>
</tr>
<tr>
<td>m1q1/m1Q1</td>
<td>0.25</td>
<td>2 + 10</td>
<td>3</td>
</tr>
<tr>
<td>m1q1/m1q1</td>
<td>0.25</td>
<td>2 + 2</td>
<td>1</td>
</tr>
</tbody>
</table>

m1m1 class mean 10

B. M1 and Q1 linked at 10 cM

<table>
<thead>
<tr>
<th>Marker class</th>
<th>Frequency</th>
<th>Phenotypic value</th>
<th>Class mean contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1Q1/M1Q1</td>
<td>0.64</td>
<td>10 + 10</td>
<td>12.80</td>
</tr>
<tr>
<td>M1Q1/M1q1</td>
<td>0.16</td>
<td>10 + 2</td>
<td>1.92</td>
</tr>
<tr>
<td>M1q1/M1Q1</td>
<td>0.16</td>
<td>2 + 10</td>
<td>1.92</td>
</tr>
<tr>
<td>M1q1/M1q1</td>
<td>0.04</td>
<td>2 + 2</td>
<td>0.16</td>
</tr>
</tbody>
</table>

M1M1 class mean 16.80

<table>
<thead>
<tr>
<th>Marker class</th>
<th>Frequency</th>
<th>Phenotypic value</th>
<th>Class mean contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1Q1/m1Q1</td>
<td>0.04</td>
<td>10 + 10</td>
<td>0.80</td>
</tr>
<tr>
<td>m1Q1/m1q1</td>
<td>0.16</td>
<td>10 + 2</td>
<td>1.92</td>
</tr>
<tr>
<td>m1q1/m1Q1</td>
<td>0.16</td>
<td>2 + 10</td>
<td>1.92</td>
</tr>
<tr>
<td>m1q1/m1q1</td>
<td>0.64</td>
<td>2 + 2</td>
<td>2.56</td>
</tr>
</tbody>
</table>

m1m1 class mean 7.20

C. M2 and Q2 linked at 5 cM

<table>
<thead>
<tr>
<th>Marker class</th>
<th>Frequency</th>
<th>Phenotypic value</th>
<th>Class mean contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>M2Q2/M2Q2</td>
<td>0.81</td>
<td>10 + 10</td>
<td>16.20</td>
</tr>
<tr>
<td>M2Q2/M2q2</td>
<td>0.09</td>
<td>10 + 2</td>
<td>1.08</td>
</tr>
<tr>
<td>M2q2/M2Q2</td>
<td>0.09</td>
<td>2 + 10</td>
<td>1.08</td>
</tr>
<tr>
<td>M2q2/M2q2</td>
<td>0.01</td>
<td>2 + 2</td>
<td>0.04</td>
</tr>
</tbody>
</table>

M2M2 class mean 18.40

<table>
<thead>
<tr>
<th>Marker class</th>
<th>Frequency</th>
<th>Phenotypic value</th>
<th>Class mean contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>m2Q2/m2Q2</td>
<td>0.01</td>
<td>10 + 10</td>
<td>0.20</td>
</tr>
<tr>
<td>m2Q2/m2q2</td>
<td>0.09</td>
<td>10 + 2</td>
<td>1.08</td>
</tr>
<tr>
<td>m2q2/m2Q2</td>
<td>0.09</td>
<td>2 + 10</td>
<td>1.08</td>
</tr>
<tr>
<td>m2q2/m2q2</td>
<td>0.81</td>
<td>2 + 2</td>
<td>3.24</td>
</tr>
</tbody>
</table>

m2m2 class mean 5.60
Confounding of linkage distance and strength of the QTL effect

- Another example:
  - M2 linked to QTL Q2
    - 5 cM. apart
  - Q2 = 10 phenotypic units
  - q2 = 2 phenotypic unit
- Because of a closer linkage distance
  - M2M2 marker class mean greater than the M1M1 marker class
- Why
  - M2M2 contains more parental gametes
- Conversely, the m2m2 class mean will be less than the m1m1 class.
- Assumption:
  - Strength of the Q2 is greater than the strength of Q1
- But the two QTL have equal strength
  - Closer linkage leads to a perceived greater strength

This is the confounding effect: there is a relationship between the marker/QTL linkage distance and the mean phenotypic expression level within each marker class.

Early QTL mapping experiments in plants

- Single marker analysis used
- Limitations
  - Location of the QTL could not be determined
  - Size of the QTL effect could not be calculated.
  - To overcome these limitations
    - New statistical approaches were developed

Interval Mapping QTL Analysis

Usefulness of molecular marker genetic linkage maps

- Enabled researchers
  - To estimate the location of a QTL
  - To calculate the magnitude of the effect of the QTL.

![Figure 5](image-url)  

Figure 5. The chromosomal distribution of two markers (M1 and M2) and a QTL (Q1). The map distance between the markers is $R$, and the distance between M1 and Q1 is $r_1$, and the distance between Q1 and M2 is $r_2$. 
How the physical arrangement of linked markers relates to a linked QTL.

- Fig. 5.
- Important parameters in this figure
  - Distance between two markers
    - $R$
  - Distance from each marker and QTL
    - $r_1$ and $r_2$
  - Therefore:
    - $R = r_1 + r_2$
- Interval mapping
  - Measures the effect of a QTL at intervals between the two markers
- If an effect exceeding a specific significance threshold is discovered
  - QTL is said to exist at that location.
- Quantitative genetic theory necessary to understand this procedure

Table 2

Theoretical effects for each genotype of a QTL

- Three parameters
  - Must be calculated to determine the effect of the QTL
    - Midparent value ($m$)
    - Additive effect ($a$)
    - Dominance effect ($d$)
  - How can we calculate these effects?

- Calculate the expected value for each of the nine marker classes.
  - M1M1M2M2  M1m1M2M2  m1m1M2M2
  - M1M1M2m2  M1m1M2m2  m1m1M2m2
  - M1M1m2m2  M1m1m2m2  m1m1m2m2

- Calculations of the expected values
  - Beyond the scope of this class
  - Values presented in Table 3

Table 2. The genotypic effects for each F$_2$ genotype of a QTL. $m =$ midparent value; $a =$ additive effect of Q1 allele; $d =$ dominance effect.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Genotypic effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1Q1</td>
<td>$m + a$</td>
</tr>
<tr>
<td>Q1q1</td>
<td>$m + d$</td>
</tr>
<tr>
<td>q1q1</td>
<td>$m - a$</td>
</tr>
</tbody>
</table>
Table 3. Additive (a) and dominance (d) coefficients for calculating the expected genotypic effects for each of the nine marker classes segregating in a F\textsubscript{2} population.

<table>
<thead>
<tr>
<th>Marker genotypes</th>
<th>Coefficients of expected genotypic effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(a = \text{additive genetic effect} )</td>
</tr>
<tr>
<td>M1M1M2M2</td>
<td>(<a href="1-r2">1-(r1)^2</a>^2 - r1^2 r2^2]/(1-R)^2 )</td>
</tr>
<tr>
<td>M1M1M2m2</td>
<td>([(1-r1)^2r2(1-r2) - r1^2 r2(1-r2)]/R(1-R) )</td>
</tr>
<tr>
<td>M1m1M2m2</td>
<td>([(1-r1)^2r2^2 - r2^2(1-r2)^2)]/R^2 )</td>
</tr>
<tr>
<td>M1M1M2M2</td>
<td>([r1(1-r1)(1-r2)^2 - r1(1-r1)r2^2)]/R(1-R) )</td>
</tr>
<tr>
<td>M1m1M2m2</td>
<td>([r1(1-r1)r2^2 - r1(1-r1)r2^2)]/R(1-R) )</td>
</tr>
<tr>
<td>m1M1M2M2</td>
<td>([r1^2(1-r2)^2 - (1-r1)^2r2^2)]/R(1-R) )</td>
</tr>
<tr>
<td>m1M1M2m2</td>
<td>([r1^2r2(1-r2) - (1-r1)^2r2(1-r2)]/R(1-R) )</td>
</tr>
<tr>
<td>m1m1m2m2</td>
<td>([r1^2r2^2 - (1-r1)^2r2^2)]/(1-R)^2 )</td>
</tr>
</tbody>
</table>

How to use the table

- For each class
  - Midparent (m) is added to the additive effect (a) times the additive coefficient plus the dominance effect (d) times the dominance coefficient.
- Two examples:
  - \(E(M1M1M2M2) = m + a(1-(r1)^2)(1-r2)^2 - r1^2 r2^2]/(1-R)^2 + d[2r1(1-r1)r2(1-r2)]/(1-R)^2 \)
  - \(E(M1M1M2m2) = m + a[(1-r1)^2r2(1-r2) - r1^2 r2(1-r2)]/R(1-R) + d[r1(1-r1)(1-r2)^2 + r1(1-r1)r2^2]/R(1-R) \)
- The table shows
  - Key values are \(R, r1\), and \(r2\)
- Interval mapping
  - Utilizes the predetermined \(R\) values
- If a QTL is assigned to a location in the interval
  - We know \(r1\) and \(r2\).
- With these three recombination values
  - The equations can be solved
  - Genotypic effects can be determine the genotypic effects.

Solving the nine (marker class) different equations

- Need to estimate the three variables: \(m, a, \) and \(d\).
- Methods
  - Regression
    - Procedure: Simple Interval QTL Mapping
  - Maximum likelihood
    - Procedure: Interval QTL Mapping
Basic approach is the same for each method

- Fix a QTL at M1
  - At this position
    - \( r1=0 \) and \( r2=R \)
- Solve for the three variables
  - Significance of the least squares (regression) method
    - Tested by a likelihood ratio test
  - Significance of the least squares (regression) method
    - Tested by maximum likelihood ratio
- Procedure repeated at a predetermined interval
  - Typically every 2 cM
  - Next interval would be
    - \( r1=2 \) and \( r2=R-2 \)
  - Calculate the new significance level
  - Significance values are plotted versus map position.

![Graphs showing LOD scores for different QTL scenarios](image)

**Figure 6.** Examples of interval mapping graphs depicting different types patterns of significant.

**Fig. 6 Examples**

- **Fig. 6A.**
  - Single peak
  - Range of locations where the significance level is above 3.0
    - Peak position considered a single QTL at this location
- **Fig. 6B**
  - Two significant QTL along the linkage group
- **Fig. 6C.**
  - Broad peak.
  - Does not imply multiple QTL.
    - Observing are *ghost QTL*
      - Loci that are significant because of linkage to neighboring QTL
• Fig. 6D
  o Difficulty with interval mapping at ends of linkage groups
    ▪ First location along the linkage group is the peak position
    ▪ Cannot conclude a QTL is at that location
    ▪ Why??
      • You only observe one side of a graphic peak
      • A QTL probably exists but
        ▪ The exact location cannot be determined

**Composite Interval Mapping**

**Dealing with Ghost QTL**

- *Composite Interval Mapping* procedure
- Goal:
  o Discovering the significant QTL in the interval
- Accounts for Ghost QTL
- Basic steps
  o Combines regression and maximum likelihood procedures
  o First step
    ▪ Single marker analysis
    ▪ Discovers the major QTL.
  o Multiple regression model built with these QTL
    ▪ These loci are removed
    ▪ Single marker analysis is again performed
    ▪ Discovers other potential QTL
    ▪ These were masked by the major QTL in the model
  o Newly discovered QTL are considered possible cofactors.
    ▪ Then
      ▪ Interval mapping performed
      ▪ Considers the cofactors and their positions along the linkage group.
  o Null hypothesis
    ▪ A QTL exists near the cofactor
  o Alternative hypothesis
    ▪ QTL exists in the interval.
  o Cofactor is dropped from the model
    ▪ QTL is discovered
  o Last step performed multiple times
    ▪ Determines which QTL are still significant

**Examples**

- Fig. 7A is the same graph as Fig. 6C.
- Fig. 7B is a hypothetical reanalysis of the same data in Fig. 7A
  o Result from composite interval mapping
    ▪ Ghost QTL are eliminated
      ▪ Two significant QTL discovered
Statistical Significance of QTL Experiments

Setting of the significance level
- Critical aspects of any statistical analysis
  - Important for any analysis in which the same data set is analyzed repeatedly
- Want to protect against are Type I errors
  - Declaring a QTL significant when indeed it is not significant
    - 5% of the time (5 out of 100 incremental positions with interval mapping
      - Declare a position significant by random chance alone
        - These are not necessarily real QTL

How to address this issue
- Single marker analysis
  - Declare a locus significant if
    - $P<0.01$ or even
    - $P<0.001$
- Interval mapping approach
  - Set the critical LOD level at 3.0 or higher.
- These approaches will generally reduce Type I error
  - But do not fully protect the researcher from these errors
  - Why??
    - Experiment-wide error rate may be higher than that value.
- If critical value set above the experiment-wide error rate
  - Risk of missing truly significant QTL

Figure 7. A comparison of hypothetical interval mapping and composite interval of the same data.
Permutation Tests and QTL Analysis

- QTL/marker analysis consists of
  - Population
    - Phenotype data for members of the population
    - Genotype data for each member
  - QTL analysis performed
    - Uses one of the three standard techniques.
      - Single factor analysis
        - F statistic is recorded for each comparison
      - Interval mapping
        - LOD value is recorded
  - Table 4: A simple case
    - Five individuals from an RI population
    - Two markers
    - Phenotype
      - Mineral deficiency tolerance
      - Rating
        - 1=tolerant
        - 5=susceptible.

Table 4. Original phenotypic and marker data.

<table>
<thead>
<tr>
<th>Line rating</th>
<th>Marker 1 Genotype</th>
<th>Marker 2 Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>A1A1</td>
<td>a2a2</td>
</tr>
<tr>
<td>1.5</td>
<td>A1A1</td>
<td>A2A2</td>
</tr>
<tr>
<td>2.5</td>
<td>a1a1</td>
<td>A2A2</td>
</tr>
<tr>
<td>3.7</td>
<td>a1a1</td>
<td>a2a2</td>
</tr>
<tr>
<td>4.9</td>
<td>a1a1</td>
<td>a2a2</td>
</tr>
</tbody>
</table>
The permutation

- Reshuffle the column with the phenotypic ratings
  - Keep the marker columns fixed
  - Table 5: An example of the first reshuffling
- Perform QTL/marker association analyses on reshuffled data set
- Reshuffling is performed many times
  - 1000 reshuffling is sufficient for the standard test when $\alpha=0.05$
    - Churchill and Doerge (1994) suggestion
- Develop a distribution
  - Use the maximum test statistics from all of “reshuffled”
    - Single factor analysis
    - Maximum F statistic
    - Interval mapping
    - LOD value
- Critical test statistic value
  - The value corresponding to the $\alpha$ value of the experiment.
  - If $\alpha=0.05$
  - Then go to the 95% $[100(1-\alpha)]$ point in the distribution
    - Record the statistic value at that point
    - This is your critical experiment-wide statistic value

Any marker whose test statistic is equal to or greater than the critical value determined from the distribution of permutation test statistics can be said to be significantly associated with the trait.

Table 5. Reshuffled phenotypic and marker data.

<table>
<thead>
<tr>
<th>Line rating</th>
<th>Marker 1 genotype</th>
<th>Marker 2 genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.7</td>
<td>A1A1</td>
<td>a2a2</td>
</tr>
<tr>
<td>1.2</td>
<td>A1A1</td>
<td>A2A2</td>
</tr>
<tr>
<td>4.9</td>
<td>a1a1</td>
<td>A2A2</td>
</tr>
<tr>
<td>2.5</td>
<td>a1a1</td>
<td>a2a2</td>
</tr>
<tr>
<td>1.5</td>
<td>a1a1</td>
<td>a2a2</td>
</tr>
</tbody>
</table>

Table 6. Significance determine with and without permutation test (critical value = Lod 3.5).

<table>
<thead>
<tr>
<th>Marker</th>
<th>LOD value</th>
<th>Significant No Permutation test</th>
<th>Significant Permutation test</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>3.7</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>M2</td>
<td>1.2</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>M3</td>
<td>3.1</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>M4</td>
<td>2.5</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>M5</td>
<td>4.8</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Table 6: An example

- Table shows LOD values
  - Without permutation test
    - Historical “rule-of-thumb”
    - LOD 3.0 is significant
      - Markers M1, M3, M5 considered significant
  - With permutation test, critical value
    - LOD 3.5
      - Markers M1, and M5 significant
        - Test statistic greater than critical value
      - Marker M3 not significant
        - Test statistic less than critical value
  - Without permutation test
    - This would be a Type I error

Concerns With QTL Analyses: More Markers or More Lines?

- Most significant factor limiting QTL discovery
  - Size of the confidence interval for the QTL
    - Often in the range of 30 cM
      - Effectively limits the number of QTL that can be mapped
        - About three per linkage group
  - How to address this problem
    - Increasing the number of markers?
      - Will not improve the resolution
        - Not enough recombination events
    - Increase the population size
      - Calculations suggest to uncover QTL of varying effects
        - Major and minor
          - 400 lines needed
  - But remember
    - Cloned QTL have started from analysis of small populations