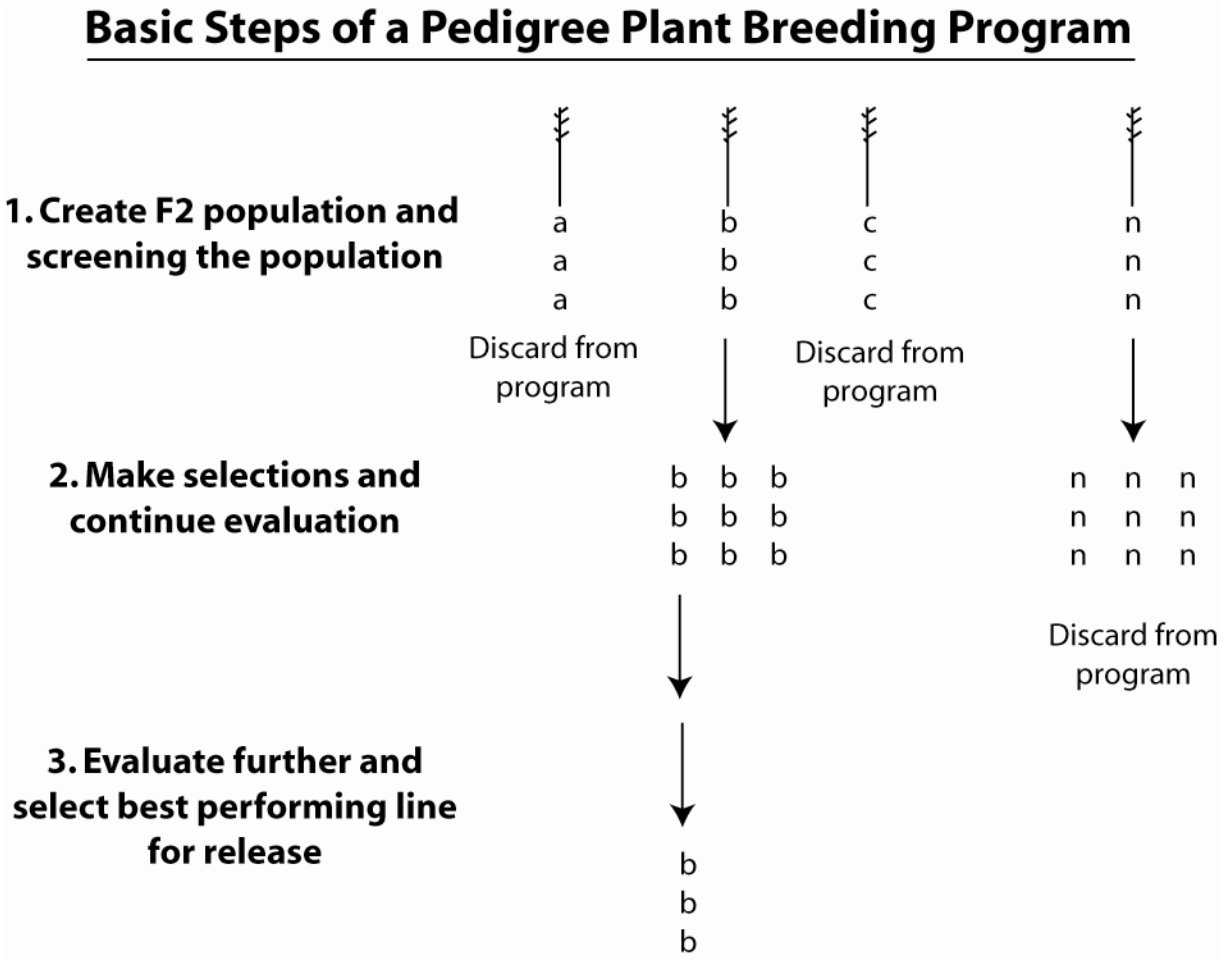


Challenge for Plant Breeding Programs

- 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

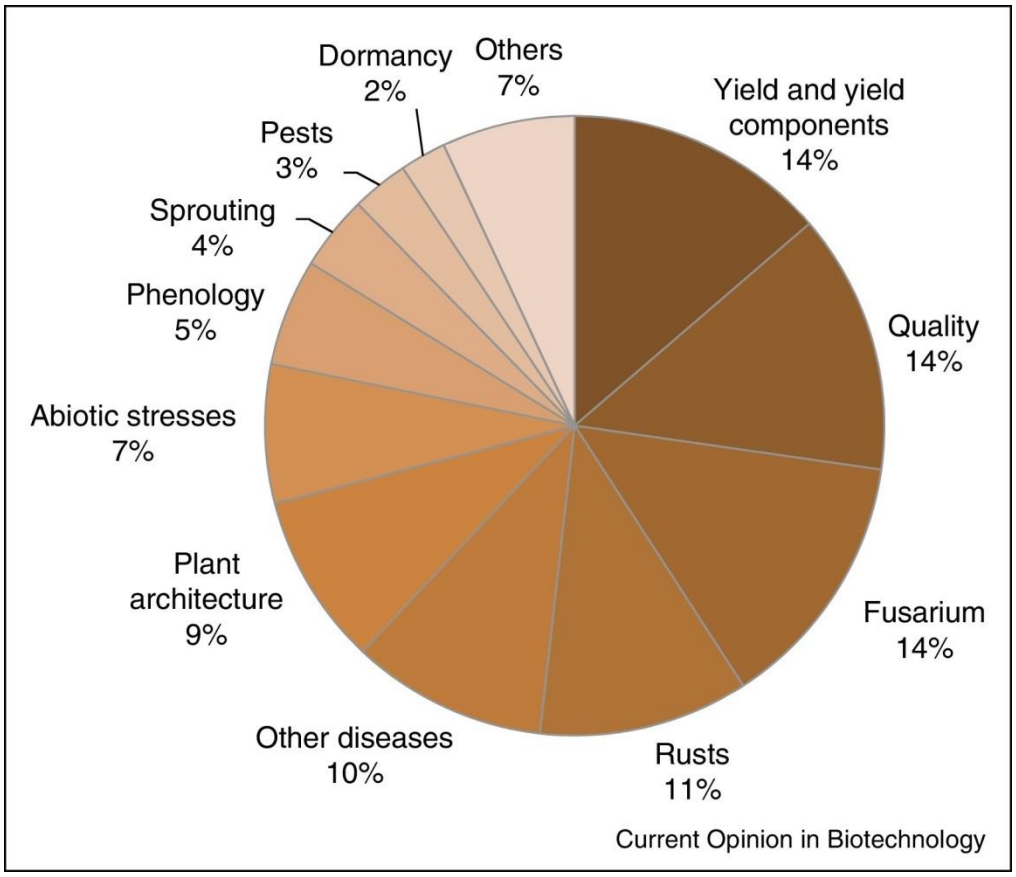
- Can a marker system be developed that efficiently selects for important (agronomic) quantitative trait loci (QTL)?

What is a Quantitative Trait Loci (QTL)?

- Genetic position in the crop genome
    - Accounts for a portion of the variance in expression of a quantitative trait
- A functional gene exists at that position that is primarily or partially responsible for the expression of the trait
    - The gene can have a major or minor effect on trait expression

How Extensive is the QTL Research World?

- Distribution of wheat QTL studies



From: Salvi and Tuberosa (2015) The crop QTLome comes of age. Current Opinion in Biotechnology 32: 179.

How many genes control a quantitative trait?

- Model organism studies
    - 1000 to 6000 genes
    - 4% to 20% of a genomes gene
        - This is a functional perspective that does not speak to effect size!!!

Crop breeding perspective: how many QTL?

- A few to several in any one cross
    - But across many biparental crosses,
- MANY QTL
    - Wheat (1992 – 2014 studies)
        - Yield: 133 QTL
        - Rust disease: 361 QTL
            - Some QTL overlap between studies

Species-wide studies using GWAS

- How many high heritability QTL
    - Tens of flowering time and plant height QTL in maize and rice

Remember

- A functional gene underlies each QTL

CLONED PLANT QTL GENES

Species	**Trait**	**QTL**	**Gene**	**Mutant effect**
Maize	Architecture	*tb1*	Transcription factor	Reduced expression
Tomato	Fruit weight	*fw2.2*	Regulate fruit size	Altered timing of expression
Tomato	Sugar content	*Brix9-2-5*	Invertase	Protein function altered
Rice	Flowering time	*Hd6*	Kinase	Loss of function

Plant Quantitative Trait Loci Analysis Is Not New

Sax (1923) Genetics 8:552-560

- Species: Common Bean
    - Parents mated; segregating generations evaluated
    - Parents are inbreds
        - Parent 1: Yellow Eye and Dot Eye
            - ***Pigmented, heavy weight***
        - Parent 2: 1333
            - ***Non-pigmented, light weight***
- Genetic factor controlling pigmentation
    - ***P*: Pigmentation gene**
        - *P* allele
            - Pigmentation in seed and flower
        - *p* allele
            - No pigmentation in seed or flower
- Specific results from segregating population
    - *PP* 4.3 ± 0.8 centigrams heavier than *pp*
    - *Pp* 1.9 ± 0.6 centigrams heavier than *pp*
- General conclusion
    - Offspring of multiple crosses
        - All heavy seeds were pigmented
        - All light seeds were non-pigmented
- Conclusion
    - ***A factor linked to P acts in an additive manner to control seed weight***
    - ***The P gene is a marker for the quantitative trait seed weight***

Other examples

- Lindstrom (1924) Science 60:182-183
    - Tomato
        - Fruit color is linked to fruit size

Application of Molecular Markers to QTL Selection

Stuber et al. Crop Science (1982) 22:737.

1. **Base population**: UNS, unselected for yield and ear number.
2. **Ten cycles of selection on UNS population**: high yielding, high ear number corn population (FS10).
3. **Compare 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

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

3. **Artificially create a population** (ALZ) with essentially the same allelic frequencies as the selected population by selecting appropriate individuals from the base population.

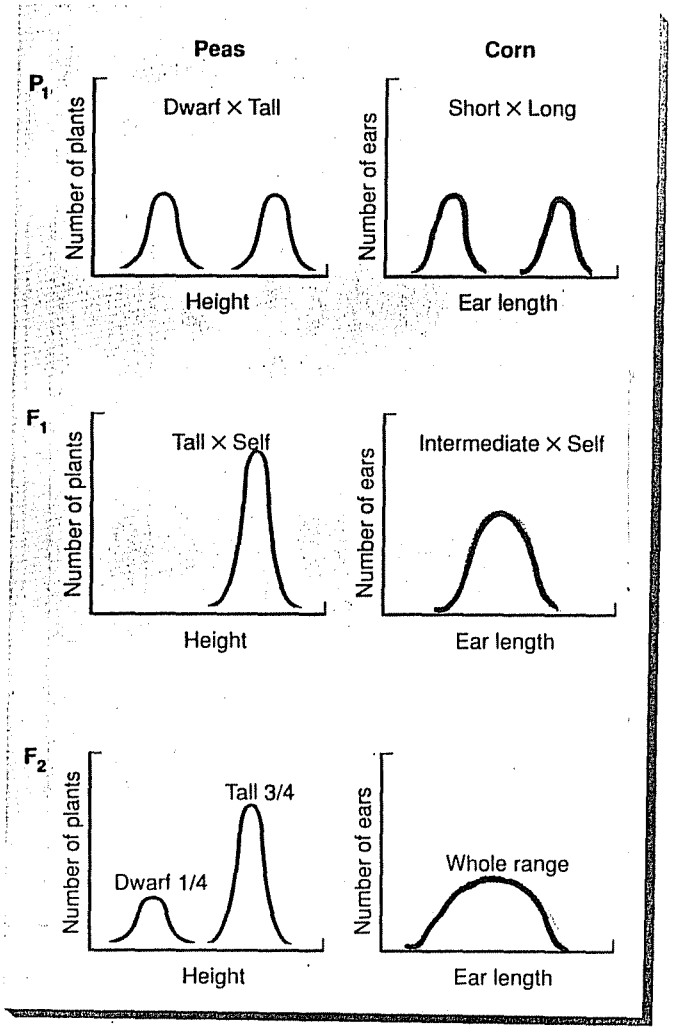
4. **Compare**: yield and ear number/plant of the UNS, FS10, and ALZ populations in a replicated field trial over two years.

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 1 ½ cycles of selection.

Results

- **Marker selection can change population mean values!!!**

Figure 9.1
Comparison of continuous variation (ear length in corn) with discontinuous variation (height in peas)



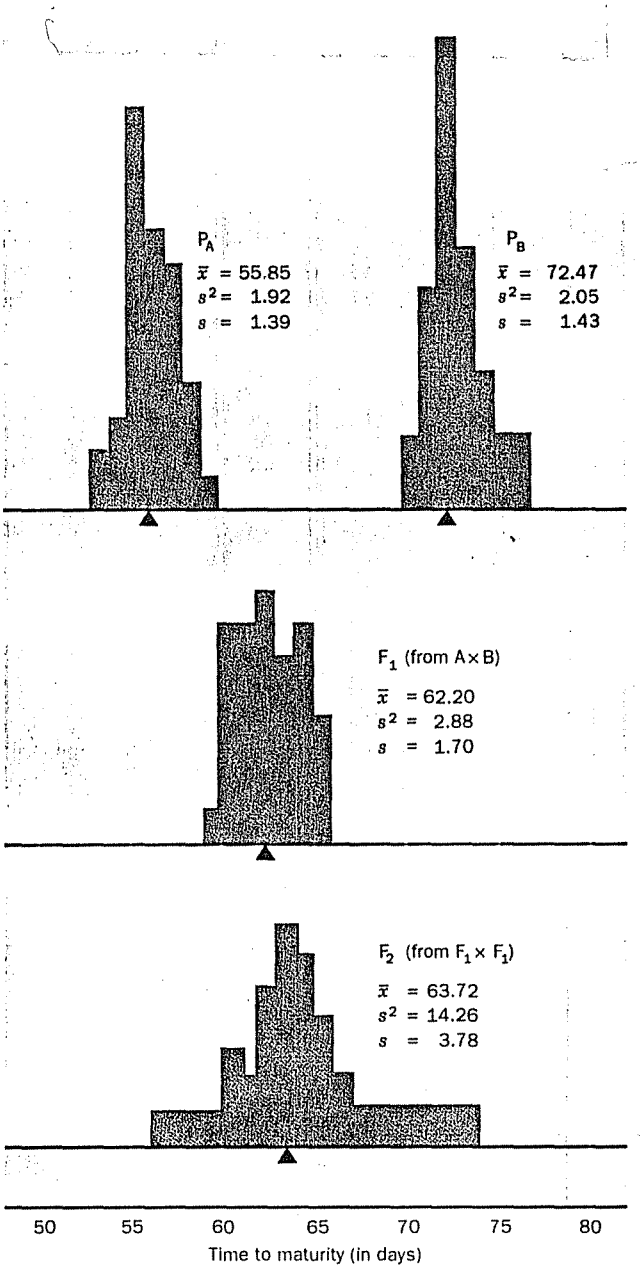


Figure 21.2 Frequency distributions and descriptive statistics of time to maturity in four populations of wheat. P_A and P_B are inbred varieties that were crossed to produce F_1 hybrids. The F_1 plants were then intercrossed to produce an F_2. Seed from all four populations was planted in the same season to determine the time to maturity. In each case, data were obtained from 40 plants. The mean maturation times ($\bar{X}$) are indicated by the triangles; the sample variances (s^2) and standard deviations (s) are also given.

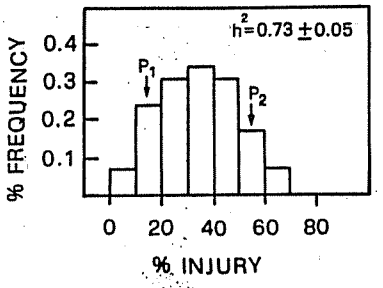


Fig. 1. Recombinant inbred frequency distribution for temperature injury in maize. *Arrows* indicate mean values of the two parental lines: P_1 = Pa33, P_2 = B37

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)

Chromosome	Locus	*b*	R^2
1	7.21	−0.096	0.099
2	5.21B	0.104	0.138
	6.20	0.116	0.131
	5.61B	0.118	0.154
	NPI298	0.126	0.157
4	ZPL2A	−0.087	0.099
8	NPI220	0.136	0.229
9	3.06	−0.124	0.155
	WX1	−0.133	0.170
	5.04	−0.114	0.131
	7.13	−0.109	0.124
10	NPI269B	−0.118	0.142

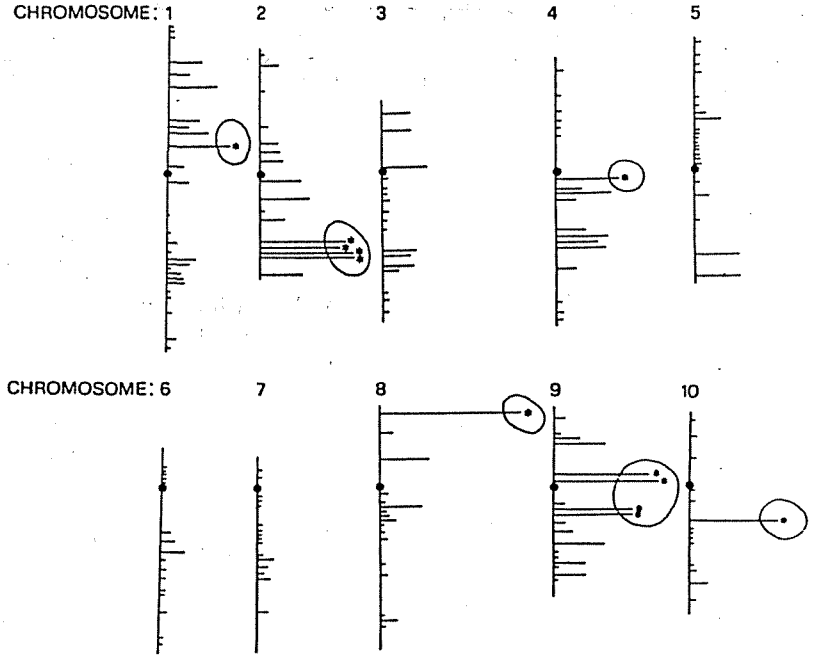


Fig. 2. RFLP analysis for temperature injury in recombinant inbreds from T32 × CM37 F_1 hybrid. *Horizontal bars* indicate degree (R^2) of correlation between RFLP loci and CMS. *: significant values ($p < 0.05$). A cluster, indicated as a circle, of significant values stands for a single QTL

Early Maize QTL Example: Dry Weight

Reiter, R. S., Coors, J. G., Sussman, M. R., & Gabelman, W. H. (1991). Genetic analysis of tolerance to low-phosphorus stress in maize using restriction fragment length polymorphisms. *Theoretical and Applied Genetics*, *82*(5), 561-568.

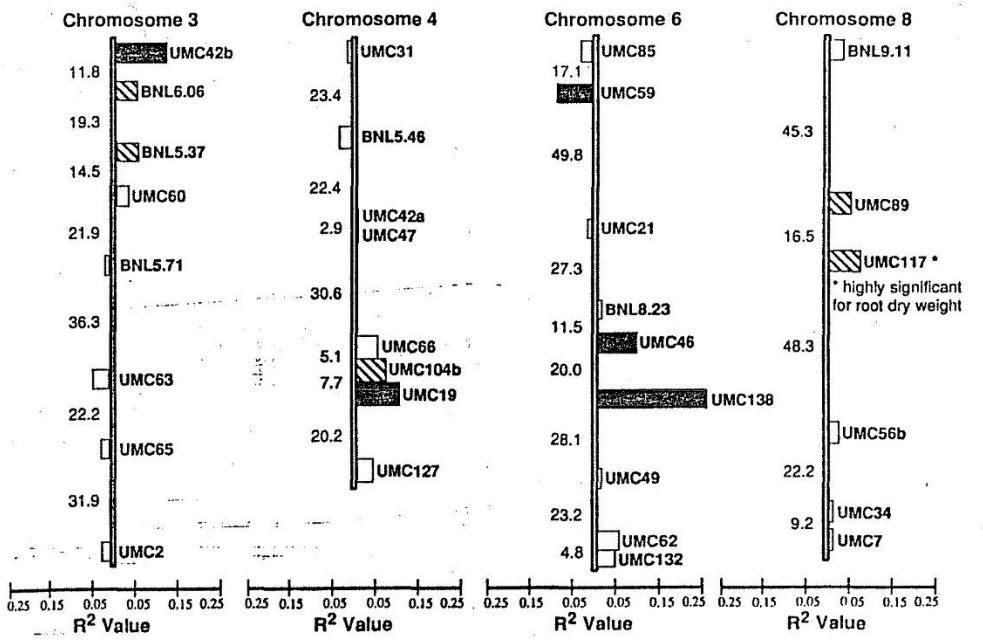


Fig. 2. The location and R^2 values for individual marker loci associated with total dry weight. Highly significant marker loci ($P<0.01$) are represented by *solid bars*, flanking marker loci ($P<0.05$) by *hatched bars*, and non-significant marker loci ($P\leq 0.05$) by *open bars*. Distances, in cM, between marker loci are indicated on the *left*. *Bars* lying to the *right* indicate the alleles with positive effects came from the tolerant parent, NY821. *Bars* lying to the *left* indicate the alleles with positive effects came from the intolerant parent, H99

Plant Height QTL Map Near Known Classical Plant Height Genes

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

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, a QTL is a genetic position, bounded by marker loci, that is located near a functional gene 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
    - Yields greater and less than the high and low yielding parents observed
      - ***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***

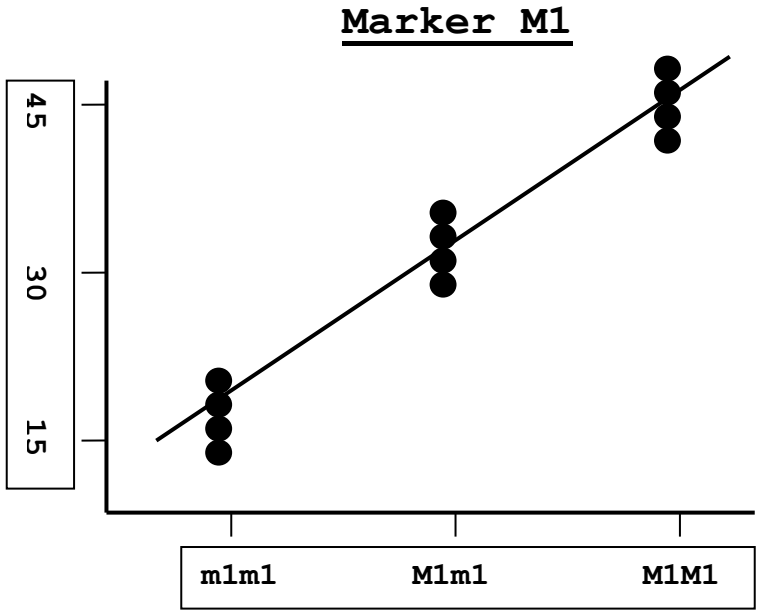
Steps in single marker QTL analysis

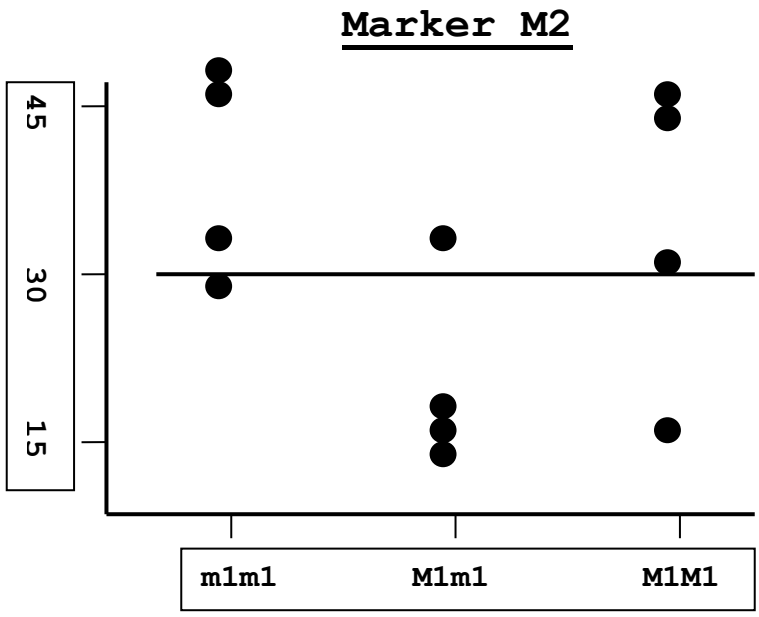
- Develop a segregating population from parents with contrasting phenotypes
  - Most common populations
    - F_2 and recombinant inbred populations
- Analysis of population
  - Collect phenotypic data
    - Normally uses replicated trials
  - Collect molecular genotypic data
- Analysis
  - Data used to discover association between the marker and the quantitative trait
  - Procedure
    - F_2 population
      - For each marker, individuals assigned to
        - Homozygous classes
          - M_1M_1 or m_1m_1
        - Heterozygous class
          - M_1m_1
    - Recombinant inbred population
      - For each marker, individuals assigned to
        - Homozygous classes
          - M_1M_1 or m_1m_1
    - Dominant markers
      - Individuals assigned to just two classes
        - $M_1\_$ or m_1m_1
- 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

	Line											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Phenotype	29	16	32	44	46	18	14	45	35	15	47	30
Marker M1	2	1	2	3	3	1	1	3	2	1	3	2
Marker M2	1	2	1	3	1	2	2	3	2	3	1	3

1=mnmn, 2=Mnmn, 3= MnMn





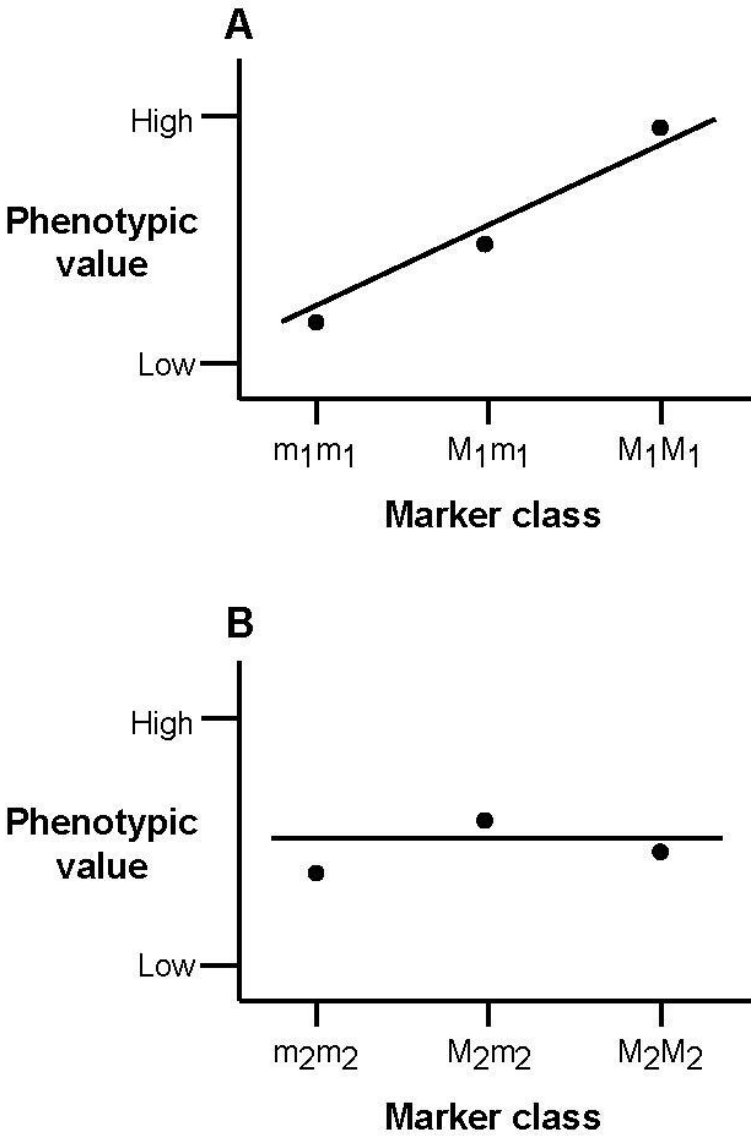


Figure 1. Graphical representation of the relationship between the three marker classes observed in an F_2 and the phenotypic mean of all individuals that possess that marker genotype. The two markers are M_1 and M2, and these have alleles M_1, m_1, and M_2, m_2. (**A**) This distribution shows a positive 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
    - Equal to the square of the correlation coefficient (R^2)
      - 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
      - Distributin 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

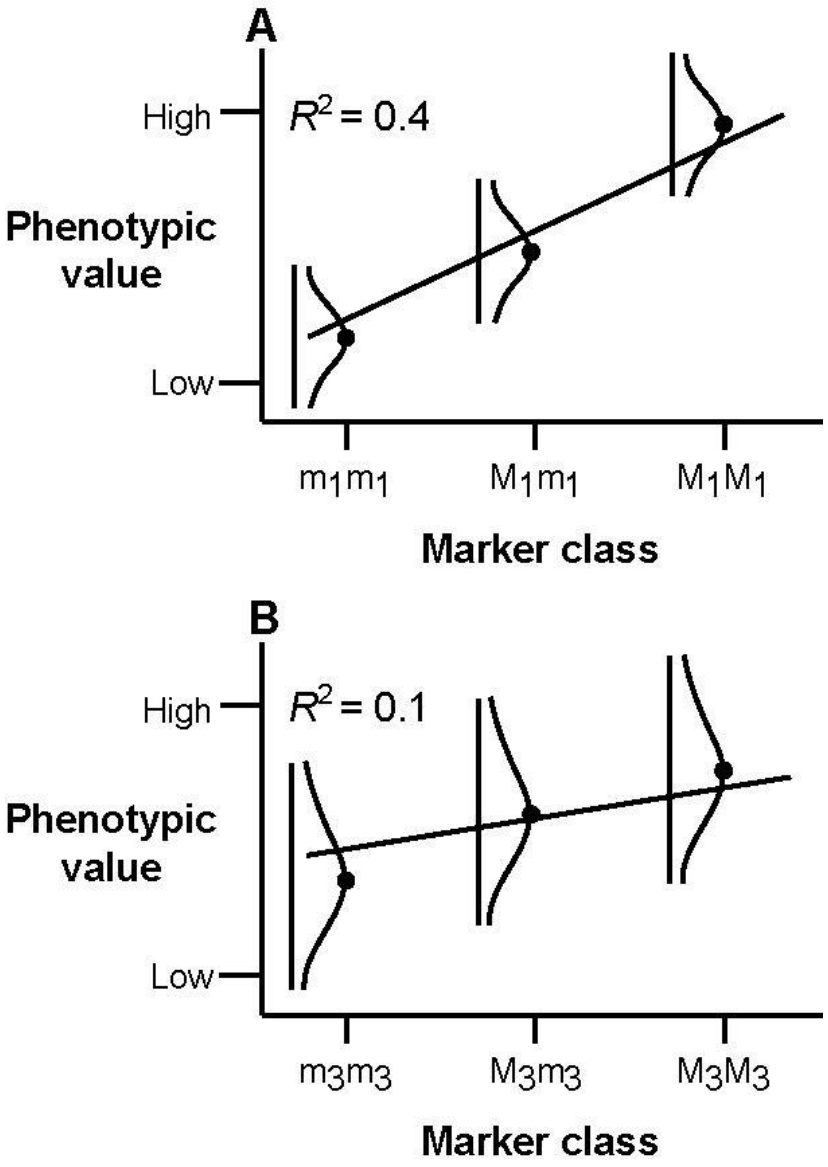
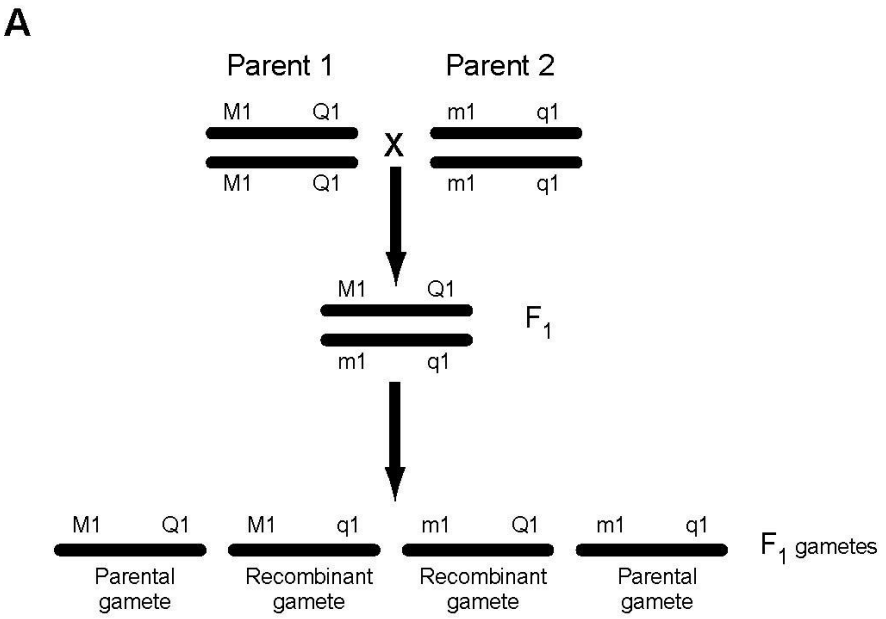


Figure 2. Phenotypic distributions about the marker classes for two markers determined by single-factor analysis of variance to be significantly associated with levels of expression of a quantitative trait. Marker M1 (**A**) has an R^2 value larger than that observed for marker M3 (**B**). Notice the 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
  - F_1 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 distance between the marker and the QTL
      - Fewer recombinant gametes are created.
    - Longer distance between the marker and the QTL
      - Larger the frequency of the recombinant gametes.



B

Male gametes	Female gametes: M1 Q1	Female gametes: M1 q1	Female gametes: m1 Q1	Female gametes: m1 q1
M1 Q1	M1 Q1 / M1 Q1	M1 Q1 / M1 Q1	M1 Q1 / M1 Q1	M1 Q1 / M1 Q1
M1 q1	M1 Q1 / M1 q1	M1 Q1 / M1 q1	M1 Q1 / M1 q1	M1 Q1 / M1 q1
m1 Q1	M1 Q1 / m1 Q1	M1 Q1 / m1 Q1	M1 Q1 / m1 Q1	M1 Q1 / m1 Q1
m1 q1	M1 Q1 / m1 q1	M1 Q1 / m1 q1	M1 Q1 / m1 q1	M1 Q1 / m1 q1

Figure 3. Parent and F_1 chromosomal arrangement (**A**) and F_2 gamete distribution for a marker (M1) and a linked QTL (Q1) (**B**).

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 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

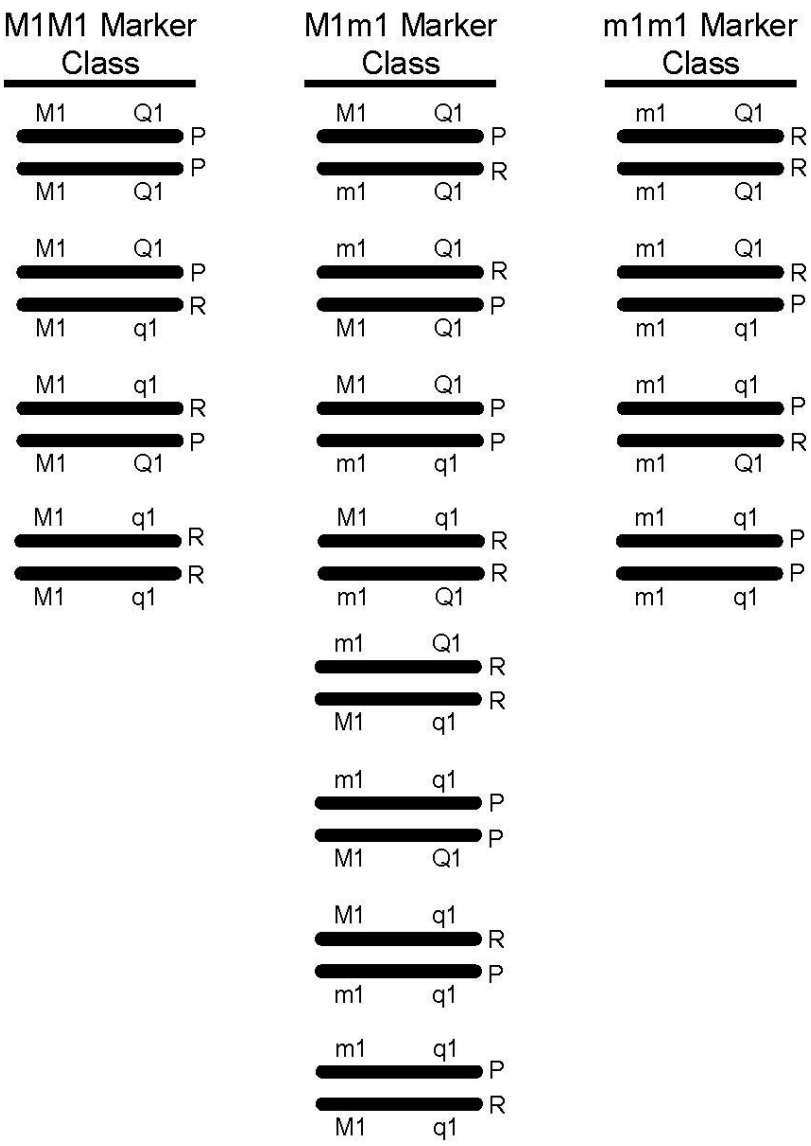


Figure 4. Distribution of possible marker/QTL genotypes within each marker class for an F_2 generation. For each individual, the chromosome contributed by the female is listed on top and the chromosome contributed by the male is on the bottom. P refers to a parental chromosome and R a recombinant chromosome. 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 Q2 linked at 5 cM.

A. Unlinked genes

Marker class	Frequency	Phenotypic value	Class mean contribution
M1Q1/M1Q1	0.25	10 + 10	5
M1Q1/M1q1	0.25	10 + 2	3
M1q1/M1Q1	0.25	2 + 10	3
M1q1/M1q1	0.25	2 + 2	1
		M1M1 class mean	**10**
m1Q1/m1Q1	0.25	10 + 10	5
m1Q1/m1q1	0.25	10 + 2	3
m1q1/m1Q1	0.25	2 + 10	3
m1q1/m1q1	0.25	2 + 2	1
		m1m1 class mean	**10**

B. M1 and Q1 linked at 10 cM

Marker class	Frequency	Phenotypic value	Class mean contribution
M1Q1/M1Q1	0.64	10 + 10	12.80
M1Q1/M1q1	0.16	10 + 2	1.92
M1q1/M1Q1	0.16	2 + 10	1.92
M1q1/M1q1	0.04	2 + 2	0.16
		M1M1 class mean	**16.80**
m1Q1/m1Q1	0.04	10 + 10	0.80
m1Q1/m1q1	0.16	10 + 2	1.92
m1q1/m1Q1	0.16	2 + 10	1.92
m1q1/m1q1	0.64	2 + 2	2.56
		m1m1 class mean	**7.20**

C. M2 and Q2 linked at 5 cM

Marker class	Frequency	Phenotypic value	Class mean contribution
M2Q2/M2Q2	0.81	10 + 10	16.20
M2Q2/M2q2	0.09	10 + 2	1.08
M2q2/M2Q2	0.09	2 + 10	1.08
M2q2/M2q2	0.01	2 + 2	0.04
		M2M2 class mean	**18.40**
m2Q2/m2Q2	0.01	10 + 10	0.20
m2Q2/m2q2	0.09	10 + 2	1.08
m2q2/m2Q2	0.09	2 + 10	1.08
m2q2/m2q2	0.81	2 + 2	3,24
		m2m2 class mean	**5.60**

Confounding of marker linkage distance and strength of the QTL effect

- Another example:
  - M2 linked to QTL Q2
    - 5 cM apart
  - Q2 = 10 phenotypic units (=Q1 effect)
  - q2 = 2 phenotypic unit (=q2 effect)
- 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.
- ***We conclude***:
  - 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***
    - ***<u>New statistical approaches were developed!!!</u>***

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

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
    - *r1* and *r2*
  - Therefore:
    - *R = r1 + r2*
- Interval mapping
  - Measures the effect of a QTL at intervals between the two markers
- If an effect exceeds a specific significance threshold:
  - QTL is said to exist at that location.
- Quantitative genetic theory necessary to understand this procedure

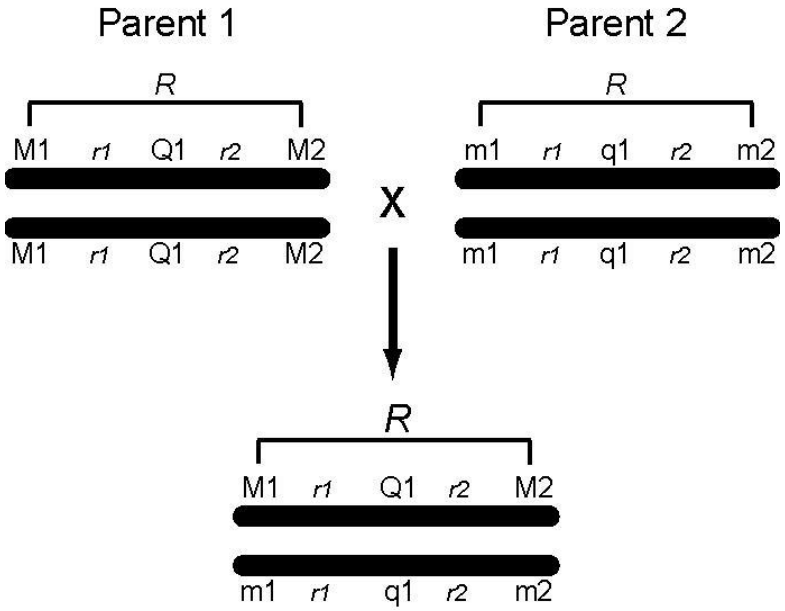


Figure 5. The chromosomal distribution of two markers (M1 and M2) and a QTL (Q1). The map distance between the markers is *R*, and the distance between M1 and Q1 is *r1*, and the distance between Q1 and M2 is *r2*.

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 value for each F_2 genotype of a QTL. *m* = midparent value; *a* = additive effect of Q1 allele; *d* = dominance effect.

Genotype	**Value**
Q1Q1	$m + a$
Q1q1	$m + d$
q1q1	$m - a$

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_2 population.

Marker genotypes	**Coefficients of expected genotypic effects**	
	***a* = additive genetic effect**	***d* = dominance genetic effect**
M1M1M2M2	$[(1-r1)^2(1-r2)^2 - r1^2r2^2]/(1-R)^2$	$[2r1(1-r1)r2(1-r2)]/(1-R)^2$
M1M1M2m2	$[(1-r1)^2r2(1-r2) - r1^2r2(1-r2)]/R(1-R)$	$[r1(1-r1)(1-r2)^2 + r1(1-r1)r2^2]/R(1-R)$
M1M1m2m2	$[(1-r1)^2r2^2 - r2^2(1-r2)^2]/R^2$	$[2r1(1-r1)r2(1-r2)]/R^2$
M1m1M2M2	$[r1(1-r1)(1-r2)^2 - r1(1-r1)r2^2]/R(1-R)$	$[(1-r1)^2r2(1-r2) - r1^2r2(1-r2)]/R(1-R)$
M1m1M2m2	0	$[r1^2r2^2 + r1^2(1-r2)^2 + (1-r1)2r2^2 + (1-r1)^2(1-r2)^2]/[R2 + (1-R)^2]$
M1m1m2m2	$[r1(1-r1)r2^2 - r1(1-r1)1-r2)^2]/R(1-R)$	$[(1-r1)^2r2(1-r2) + r1^2r2(1-r2)]/R(1-R)$
m1m1M2M2	$[r1^2(1-r2)^2 - (1-r1)^2r2^2]/R^2$	$[2r1(1-r1)r2(1-r2)]/R^2$
m1m1M2m2	$[r1^2r2(1-r2) - (1-r1)^2r2(1-r2)]/R(1-R)$	$[r1(1-r1)(1-r2)^2 + r1(1-r1)r2^2]/R(1-R)$
m1m1m2m2	$[r1^2r2^2 - (1-r1)^2(1-r2)^2]/(1-R)^2$	$[2r1(1-r1)r2(1-r2)]/(1-R)^2$

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:
  - $E(M1M1M2M2) = m + a[[(1-r1)^2(1-r2)^2 - r1^2r2^2]/(1-R)^2] + d[[2r1(1-r1)r2(1-r2)]/(1-R)^2]$
  - $E(M1M1M2m2) = m + a[[(1-r1)^2r2(1-r2) - r1^2r2(1-r2)]/R(1-R)] + d[[r1(1-r1)(1-r2)^2 + r1(1-r1)r2^2]/R(1-R)]$
- 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 determined.

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: m, a, d
  - 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.

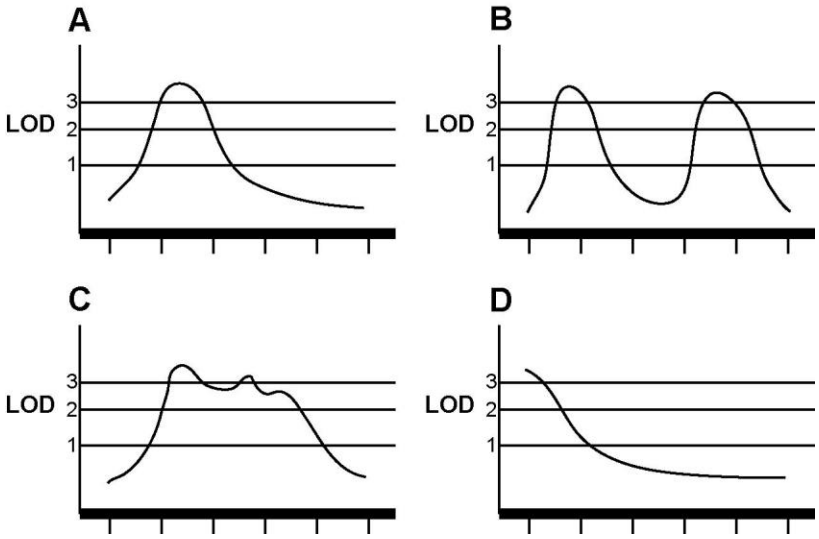
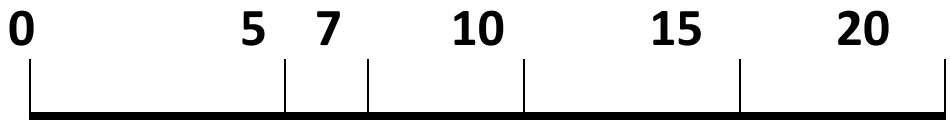


Figure 6. Examples of interval mapping graphs depicting different types patterns of significant.

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
  - Difficulty with interval mapping at ends of linkage groups
    - First location on 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**

Illustration of interval mapping using the molecular linkage map below



Step	Test QTL (Q) position (cM)	r1	r2	LOD value
1	0	0	5	
2	2	2	3	
3	4	4	1	
4	6	1	1	
5	8	1	2	
6	10	2	5	
…	…	…	…	
11	20	5	0	

Calculate the LOD value to test for the presence of a QTL at each of the test positions.

$$LOD = \frac{Probability\ of\ QTL\ at\ this\ position}{Probability\ of\ a\ QTL\ NOT\ at\ this\ position}$$

Ghost QTL

- Two linked QTL with the same phenotypic effect
  - Can be positive or negative
- Effect of ghost QTL during mapping
  - Broadens the QTL peak; or
  - Defines two linked peaks
- Solution to Ghost QTL?
  - Composite Interval Mapping

Composite Interval Mapping

Dealing with Ghost QTL

- *Composite Interval Mapping* procedure
- Goal:
  - Discovering the significant QTL in the interval
- Accounts for Ghost QTL
- Basic steps
  - Combines regression and maximum likelihood procedures
  - First step
    - Single marker analysis
    - Discovers the major QTL.
  - 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
  - Newly discovered QTL are considered possible ***cofactors***.
    - Then
      - Interval mapping performed
      - Considers the cofactors and their positions along the linkage group.
  - Null hypothesis
    - A QTL exists near the cofactor
  - Alternative hypothesis
    - QTL exists in the interval.
  - Cofactor is dropped from the model
    - QTL is discovered
  - 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
    - Result from composite interval mapping
        - Ghost QTL are eliminated
            - Two significant QTL discovered

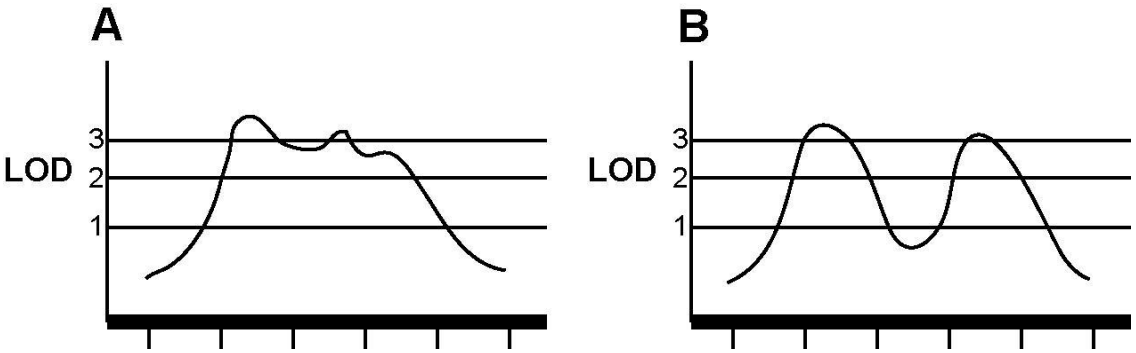


Figure 7. A comparison of hypothetical interval mapping and composite interval of the same data.

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

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.

Line rating	Marker 1 Genotype	Marker 2 Genotype
1.2	A1A1	a2a2
1.5	A1A1	A2A2
2.5	a1a1	A2A2
3.7	a1a1	a2a2
4.9	a1a1	a2a2

Table 5. Reshuffled phenotypic and marker data.

Line rating	Marker 1 genotype	Marker 2 genotype
3.7	A1A1	a2a2
1.2	A1A1	A2A2
4.9	a1a1	A2A2
2.5	a1a1	a2a2
1.5	a1a1	a2a2

The Permutation Test

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

Marker	LOD value	Significant No Permutation test	Significant Permutation test
M1	3.7	Yes	Yes
M2	1.2	No	No
M3	3.1	Yes	No
M4	2.5	No	No
M5	4.8	Yes	Yes

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