

Construction of a genetic map, mapping of major genes, and QTL analysis

A. Touré^a, B.I.G. Haussmann^b, N. Jones^c, H. Thomas^d, and H. Ougham^e

^a *IER, CRRA-Sotuba BP 261 Bamako, Mali, West Africa*

^b *University of Hohenheim, Institute of Plant Breeding, Seed Science, and Population Genetics, 70593 Stuttgart, Germany*

^c *University of Wales Aberystwyth, Institute of Biological Sciences, Aberystwyth SY23 3DD, UK*

^d *Institute of Grassland and Environmental Research, Plas Gogerddan, Aberystwyth SY23 3EB, UK*

Abstract

The mapping of chromosomal regions affecting qualitative or quantitative traits receives growing attention in plant breeding. In this contribution, we review the concepts and basic principles behind genetic map construction, and mapping of major genes and quantitative trait loci (QTL). The following topics are treated:

- definition of molecular markers and mapping;
- recombination and linkage;
- mapping populations;
- production of a molecular marker map and software for map construction;
- mapping of major genes;
- definition of QTL and principles of QTL detection;
- specific methods of QTL analysis and QTL mapping software.

Introduction

Traditional methods of plant breeding have made a significant contribution to crop improvement, but they have been slow in targeting complex traits like grain yield, grain quality, drought - or striga resistance. To meet the great increase in food production necessitated by population growth, and the higher standards of living expected by most of the developing countries, biotechnology brings new and powerful tools to plant breeders.

In: B.I.G. Haussmann, H.H. Geiger, D.E. Hess, C.T. Hash, and P. Bramel-Cox (eds.). 2000. Application of molecular markers in plant breeding. Training manual for a seminar held at IITA, Ibadan, Nigeria, from 16-17 August 1999. International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502 324, Andhra Pradesh, India.

One method receiving growing attention is the mapping of chromosomal regions affecting qualitative or quantitative traits. Polygenic characters which were very difficult to analyze using traditional plant breeding methods can now be tagged using molecular markers (DNA markers). Molecular markers allow geneticists and plant breeders to locate and follow the numerous interacting genes that determine a complex trait. Genetic linkage maps can provide a more direct method for selecting desirable genes *via* their linkage to easily detectable molecular markers (Tanksley et al., 1989). Combining marker-assisted selection methods with conventional breeding schemes can increase the overall selection gain and, therefore, the efficiency of a breeding program. With the use of molecular techniques it is possible to hasten the transfer of desirable genes between varieties and to introgress novel genes from wild species into crop plants. These new techniques also make it possible to establish genetic relationships between sexually incompatible crop plants.

The purpose of this paper is to review the concepts and basic principles behind genomic map construction, and to describe how major genes and quantitative trait loci (QTL) can be detected using genomic maps. The first part on map construction and mapping of major genes is largely based on a paper from Jones et al., 1997.

What are molecular markers?

Molecular markers reveal neutral sites of variation at the DNA sequence level. By 'neutral' is meant that, unlike morphological markers, these variations do not show themselves in the phenotype, and each might be nothing more than a single nucleotide difference in a gene or a piece of repetitive DNA. The most common DNA markers used in plant sciences today are restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNAs (RAPDs), amplified fragment length polymorphisms (AFLPs), and microsatellites or single sequence repeats (SSRs).

In the case of RFLP markers, used here to illustrate aspects of map construction, the DNA is cut by restriction enzymes. Each different restriction enzyme recognizes a specific and characteristic nucleotide sequence. Because even a single nucleotide alteration can create or destroy a restriction site, point mutations cause variation in the number of sites and, therefore, fragment lengths. In addition, insertions or deletions between two restriction sites can cause changes in the lengths of the fragments. Thus there is variation between individuals in the positions of cutting sites and the lengths of DNA between them, resulting in restriction fragment length polymorphism.

To make the polymorphism visible the DNA fragments are separated according to their size by electrophoresis, and then transferred from the gel to a nylon or nitrocellulose filter

membrane (Southern transfer) and denatured. A small piece of cloned genomic DNA, from the same or a related plant species, will match the whole or part of one of the fragments in the membrane. If it is labeled with a radioactive or chemical tag, this cloned bit will serve as probe in a Southern hybridization and will detect the single fragment with which it has sequence homology. A characteristic band pattern might result (Figure 1). A plant may show a single band if the two fragments from the diploid genome are homozygous, with restriction sites at identical places, so the probe detects both of them at the same place in the Southern blot. A second plant might give a variant of the same fragment that differs in length because it is homozygous for a mutation which has either destroyed one of the restriction sites or else created a new one within the original fragment. A third heterozygous plant – it could be a hybrid between plants 1 and 2 - will show two bands corresponding to the fragment sizes of plants 1 and 2. The two different sized fragments are alleles of one locus. The locus itself is identified by the probe used to detect it, and takes the name or number of that probe.

AA	aa	Aa

Figure 1. Southern hybridization pattern with a single probe using DNA from plants with three RFLP genotypes at one locus. Track AA is from the homozygote for one allele, aa from the genotype homozygous for the other allele, and Aa from the heterozygote. The codominance of RFLPs allows for all three genotypes at a single locus to be scored. (Modified from Jones et al., 1997).

What is mapping?

Mapping is putting markers (and genes or QTL) in order, indicating the relative distances among them, and assigning them to their linkage groups on the basis of their recombination values from all pairwise combinations. Knowledge about the genetic concepts of segregation and recombination is essential to the understanding of mapping.

Segregation and recombination

Genetic concepts of segregation and recombination will be illustrated with classical Mendelian traits. Each trait is determined by one gene locus with two alleles. The alleles are given upper and lower case letters, respectively. As a result of meiosis, the two alleles of a locus will segregate with equal frequencies to the gametes. If A and a are two such alleles, a

diploid heterozygous individual (genotype Aa) will give gametes half of which are A and half of which are a. Similarly, alleles B and b at a separate locus will segregate fifty-fifty into the gametes.

The situation becomes more complicated when two loci are considered simultaneously. The simplest way to follow such events is first to make a cross between two homozygous parents [P_1 (AABB) and P_2 (aabb)], and then to backcross the F_1 to the double recessive parent P_2 (Figure 2). There are four possible combinations in the gametes of the F_1 : AaBb AB, Ab, aB, ab. The gametes Ab and aB recombine the alleles of the homozygous parent lines. Their frequencies depend on the level of recombination between the two loci. With independent segregation (the two loci on different chromosomes), the recombinants will each comprise 25% of the gamete population. On the other hand, when the gene loci are linked on the same chromosome, the recombinants will only arise when crossover occurs between them, and then their frequency will be $\frac{1}{2}r$, with r being the recombination value between the two loci. The recombination value r can take values between 0 and 0.5, with 0 = complete linkage and 0.5 = free recombination. The maximal recombination value for linked genes is 0.5 because a crossover during meiosis involves only two of the four chromatids of a chromosome (Figure 3). A value of 0.5 for r will only occur when the loci are far apart, like at the opposite ends of the chromosomes. In the first backcross progeny (BC_1), expected genotype frequencies are equal to the gamete frequencies of the F_1 , because P_2 produces only one gamete type (Figure 2).

In an F_2 population, expected frequencies of the 16 possible genotypes (4×4 possible gamete combinations) are obtained by multiplying the respective individual gamete frequencies of two F_1 plants (not shown).

In summary, recombination is the process by which new combinations of parental genes or characters arise. It can occur by independent segregation of unlinked loci or by crossover between loci that are linked on a chromosome.

$P_1 \times P_2$	AABB	×	aabb		
Gametes	AB	×	ab		
Frequency	1	↓	1		
$F_1 \times P_2$		AaBb		×	aabb
Gametes	AB	Ab	aB	ab	×
Frequency					
- free recombination	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	1
- reduced recombination	$\frac{1}{2}(1-r)$	$\frac{1}{2}r$	$\frac{1}{2}r$	$\frac{1}{2}(1-r)$	1
- complete linkage	$\frac{1}{2}$	0	0	$\frac{1}{2}$	1
BC ₁ Genotypes	AaBb	Aabb	aaBb	aabb	↙
Frequency	$\frac{1}{2}(1-r)$	$\frac{1}{2}r$	$\frac{1}{2}r$	$\frac{1}{2}(1-r)$	

Figure 2. Expected frequencies of gametes and genotypes in a backcross breeding scheme with the parents in coupling phase. The recombination value “r” can take values between 0 and 0.5, with 0 = complete linkage and 0.5 = free recombination .

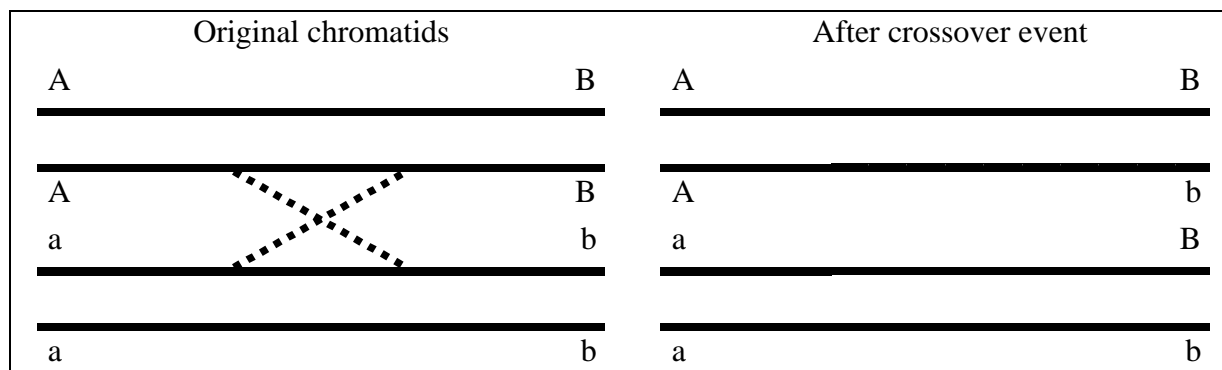
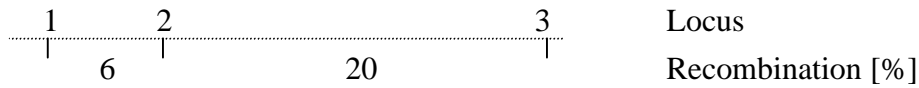


Figure 3. Diagram of a bivalent at the four-strand (diplotene) stage of meiosis, showing how a single chiasma involves only two of the four chromatids and can lead to a maximum of 50% recombination for genes at opposite ends of the chromosomes. When the two loci are closer together, chiasma formation will not always occur, and recombination will be < 50%. (Modified from Jones et al., 1997).

Recombination and linkage maps

The percentage of a segregating progeny that are recombinants for a pair of linked loci is the recombination frequency. The recombination frequency gives an estimate of the distance between two loci in a chromosome, on the assumption that the probability of crossover is

proportional to the distance between loci. Suppose the recombination between loci 1 and 2 = 6%, that between loci 2 and 3 = 20%, and that between loci 1 and 3 = 24%, then we can order the loci along the chromosome:



Recombination frequencies are not additive, as in the example given here: $6 + 20 = 26$ is the true distance between markers 1 and 3 (and not 24). The underestimated recombination frequency between loci 1 and 3 is due to double or multiple crossovers, which go undetected as recombinants (Figure 4).

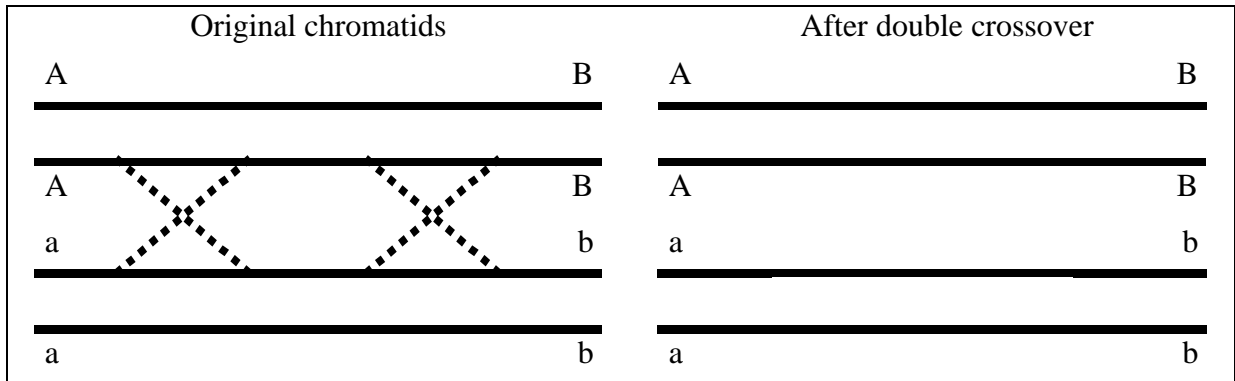


Figure 4. Diagram of a bivalent at the four-strand (diplotene) stage of meiosis, showing how double crossovers involving the same pair of chromatids go undetected as recombinants, and thus lead to an underestimation of the genetic distance on the basis of pure recombination frequencies. (Modified from Jones et al., 1997).

Production of a molecular marker map

To construct a molecular marker map, crosses are made between two homozygous, genetically different lines. Various types of mapping population may be produced from the heterozygous F_1 hybrids:

1. Double haploid lines (DHLs): plants are regenerated from pollen (which is haploid) of the F_1 plants and treated to restore diploid condition in which every locus is homozygous. Since the pollen population has been generated by meiosis, the DHLs represent a direct sample of the segregating gametes.
2. Backcross population: the F_1 plants are backcrossed to one of the parents.

3. F₂ population: F₁ plants are selfed.
4. F_{2:3} lines (F₃ plants tracing back to the same F₂ plant, also called F₂ families), further inbred generations or recombinant inbred lines (RILs) can be derived by selfing individual F₂ plants and further single seed descent. A population of RILs represents an ‘immortal’ or permanent mapping population.

The choice of the type of mapping population depends on the plant species (suitability for DHL production), type of marker system used, and the trait to be mapped later on. Information provided by codominant markers is best exploited by an F₂ population, while information obtained by dominant marker systems can be maximized by using RILs or DHLs. Double haploids, F₂ or F₃ families, or RILs are advantageous if the trait to be mapped cannot be accurately measured on a single plant basis but must be assessed in replicated field experiments.

Before the whole mapping population is scored with RFLP markers, a screening is performed to identify those probes which are polymorphic for the two parent lines. Only markers that differ between the parents will yield useful information for the map. Once a sufficient number of polymorphic probes has been identified, the parental lines, the F₁ (if still available), and all individuals of the mapping population are scored to determine their genotypes. Then recombination values for all pairs of markers are calculated. The example in Figure 5 is a highly simplified scheme with only 12 backcross progenies. It shows the outcome of two separate gels for probes 1 and 2. In the case of probe 1, the F₁ segregates its two alleles in equal numbers (6). These combine with the single allele of P₁ to give two kinds of backcross progeny in the sample. Probe 2 shows a differing banding pattern. The recombinants are all those which have three bands across the two panels; the other patterns are parental types. Four recombinants out of 12 backcross progenies make 33% recombination. In the same way, many other probes are used, and the data are analyzed making all possible pairwise combinations. The number of possible combinations is $N(N-1)/2$ with N being the number of polymorphic probes.

From the estimated recombination frequencies, the most probable order of markers along a chromosome is calculated. As recombination frequencies are not additive due to multiple crossover, maps are built up by adding small intervals. Two functions have been proposed to convert recombination frequencies into map distances. Haldane’s mapping function assumes absence of interference, i.e., no effect of one crossover on the occurrence of crossovers in neighboring regions. In contrast, Kosambi’s mapping function assumes presence of interference, i.e., a negative effect of one crossover on the occurrence of crossovers in the near neighborhood. In both cases, map distances are measured in centimorgans (cM). Two markers are said to be 1 cM apart if they are separated by recombination 1% of the time.

Markers that map together as one linkage group do so because they are all located on a single chromosome. The number of different linkage groups that we eventually find, given enough markers, will correspond to the basic chromosome number of the species.

Probe	P ₁	P ₂	F ₁	Backcross progeny: F ₁ × P ₂ (AaBb × aabb)												
				1	2	3	4	5	6	7	8	9	10	11	12	
1	AA	aa	Aa	aa	Aa	aa	Aa	aa	Aa	aa	Aa	aa	Aa	aa	Aa	
2	BB	bb	Bb	bb	Bb	Bb	Bb	bb	bb	bb	Bb	Bb	Bb	bb	bb	
Combined score				P	P	R	P	P	R	P	P	R	P	P	R	
Genotype				aabb	parental (like P ₂)				4		33% recombinants					
				aaBb	recombinant				2							
				Aabb	recombinant				2							
				AaBb	parental (like F ₁)				4							

Figure 5. Highly simplified procedure for RFLP mapping using a backcross. The mapping population consists of parents (P₁, P₂), the F₁ and the backcross progeny (12 individuals). RFLP alleles at two different loci are identified by probes 1 and 2, and the recombinants (R) are the genotypes which have three bands across both gels. The other individuals (P) correspond to one of the parental genotypes (i.e., F₁ or P₂). (modified from Jones et al., 1997).

It must be emphasized that the genetic map is based on recombination frequencies. Unlike other maps, the distance between points on a genetic map is not measured in any kind of physical unit, but is just a reflection on the recombination frequency between those two points. In cases where crossovers are clustered or suppressed in certain regions rather than being randomly distributed, the genetic map will be a distortion of the physical distances separating loci on the chromosomes.

Software for map construction

Several software packages are available on the internet for the construction of genetic maps:

- **MapMaker/Exp** (<ftp://genome.wi.mit.edu/pub/mapmaker3/>): freely distributed, analyzes F_2 and BC mapping populations.
- **GMendel**: <http://gnome.agrenv.mcgill.ca/info/gmendel.htm>): freely distributed, analyzes all types of mapping populations; can combine maps of different mapping populations provided there are common markers.
- **JoinMap** (<http://www.cpro.dlo.nl.cbw/>): analyzes all types of mapping populations; can combine maps of different mapping populations provided there are common markers. © CPRO-DLO; cost per license dependent on number of licenses.

Input files are usually simple ASCII files with all N markers in lines and G genotypes in columns, with the genotype codes "A" for parent 1, "B" for parent 2, and "H" for heterozygous:

```
;          Genotype sample
;          0000000000000000...G
;          0000000001111111...G
;          1234567890123456...G
Marker1   AHBHBABAHAHBBBA...B
Marker2   BBAAHHHBAHBBAAHH...A
...
MarkerN   AAHHBABHBAHHBAH...A
```

The program Joinmap (Stam and Van Ooijen, 1995), as an example, first assigns markers to linkage groups, using a chi-squared test for independence of segregation. Then recombination frequencies are estimated for all marker pairs within a linkage group. The procedure is a Maximum Likelihood Estimation of the recombination frequencies which maximizes the probability of obtaining the observed phenotype frequencies. A significance test uses user-defined thresholds for the minimal LOD values (likelihood odds ratio from the Maximum Likelihood Estimation) and for the maximal recombination frequencies. Finally, the genetic distances among the markers are estimated and markers aligned along each chromosome in the most probable order, using Haldane's or Kosambi's mapping function. As a result, a detailed map is obtained describing the relative position of large numbers of neutral DNA sequences (Figure 6). These neutral DNA sequences serve as signposts which can point to genes of interest. The procedures of "putting traits on the map" are described in the following.

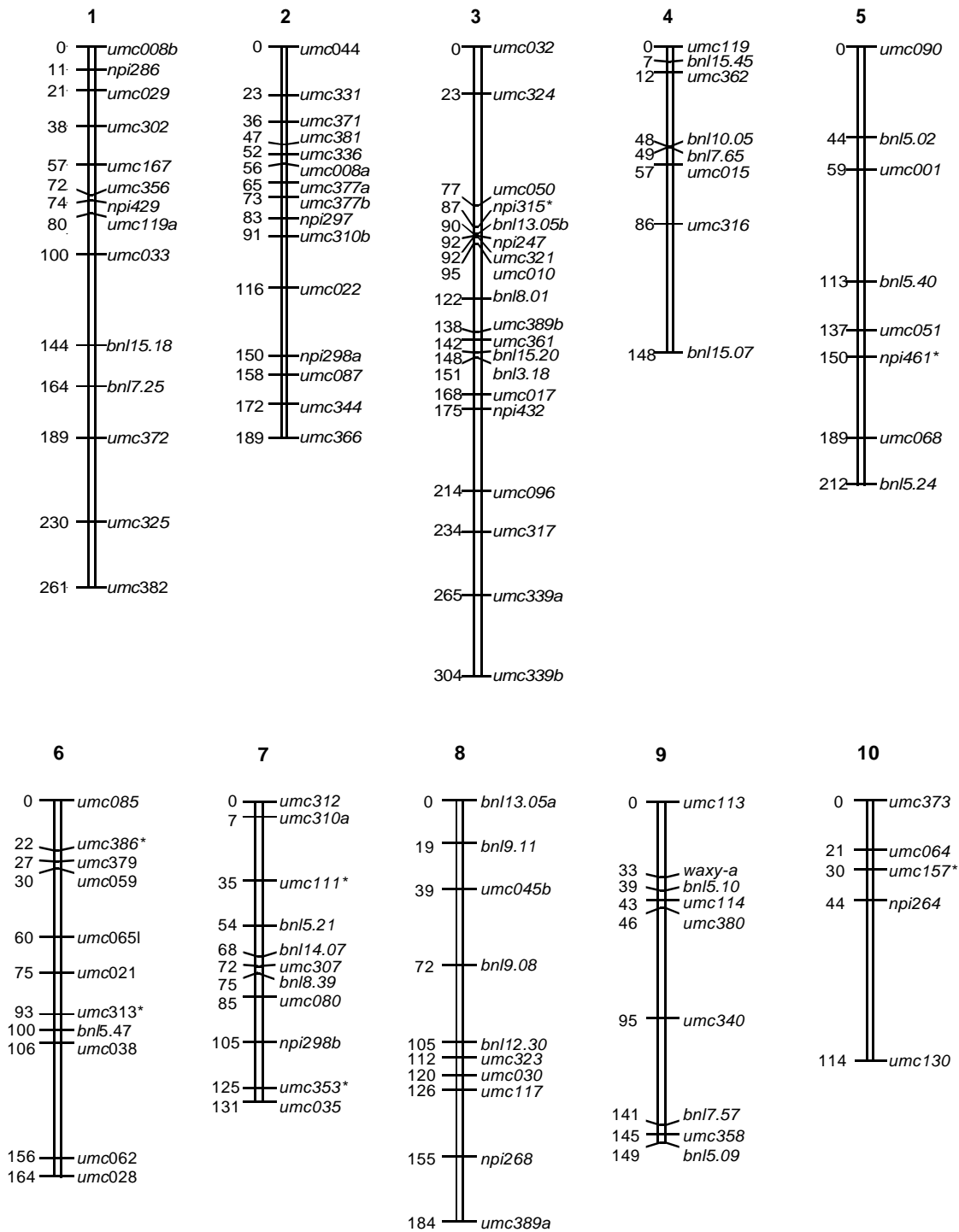


Figure 6. Example of a genome map: Linkage map of the maize population F_{2:3} (Lo951 × CML292) constructed with 110 polymorphic RFLP markers and 256 lines. Numbers on the left of the bars indicate relative distance in cM to the first locus on the short arm of the chromosome. To the right, the markers are specified. Asterisks indicate loci that mapped to different chromosomes compared to other populations (Maize Data Base 1997; Figure from Schechert et al., 1999).

Mapping of major genes

The identification of chromosomal regions carrying important genes requires

- a well saturated genetic map from the mapping population;
- a reliable screening method; and
- genetic variation within the mapping population for the trait of interest.

Major genes are inherited in a Mendelian manner and their allelic forms give qualitatively distinct phenotypes. Mapping of such genes is a relatively simple exercise. Provided there is a well saturated genetic map covering the whole genome, the alleles of the major gene will segregate together with a particular RFLP marker (Figure 7). By computing linkage values between the alleles encoding the trait and the RFLPs, the gene locus can be included in the map.

If a resistance gene, for instance, and an RFLP marker are so close that they map to the same location, then the RFLP marker becomes a very useful tight marker, or gene tag, for resistance. To find a truly coincident marker for a gene of interest is rare. Usually, marker and resistance gene may be nearby, about 5 cM apart. In this case there would be a 5% loss of resistant genotypes through recombination if selection relied on only such a neighboring RFLP marker.

Many physiological processes are modulated by major genes. Experimentally induced mutations in model species like *Arabidopsis* and near isogenic lines (NILs) have been valuable in mapping such genes.

$P_1 (AArr) \times P_2 (aaRR)$ $F_1 \times F_1:$ A, a: marker genotype R, r: alleles for susceptibility and resistance, respectively	$\frac{Ar}{aR} \times \frac{Ar}{aR}$		$\frac{Ar}{aR} \times \frac{Ar}{aR}$		F_2		Ar		aR			
							AArr		AaRr			
								AaRr		aaRR		
Plant sample (F_2 population)	1	2	3	4	5	6	7	8	9	10	11	12
Disease score	1	9	8	1	7	8	9	7	8	8	9	1
Genotype code for marker locus	A	B	H	A	H	H	B	H	H	B	H	A
1 : 3 for resistance (disease score 1) : susceptibility (scores 7 – 9) and 1 : 3 for marker genotypes A : (B + H)												

Figure 7. Highly simplified scheme showing how a tightly linked codominant marker cosegregates with a recessive resistance gene in a F_2 population. (Modified from Jones et al., 1997).

In many plants, identifying specific markers linked to a gene of interest is made more difficult by substantial genetic variation throughout the rest of the genome. A way round the problem is to use the approach called bulk segregant analysis (Michelmoore et al., 1991). This method is described in detail in the next paper (I. Kapran).

Most of the important agronomic characters like yield and yield components, plant height, days maturity etc. are controlled by several genes. The segregating progeny shows a continuous pattern of expression rather than the discrete classes characteristic for Mendelian, or qualitative, inheritance. These are known as "quantitative" traits. The number of genes and their interactive effects controlling the expression of quantitative traits are poorly understood.

What is a QTL?

A quantitative trait locus (QTL) is the location of a gene that affects a trait that is measured on a quantitative (linear) scale. QTL are identified via statistical procedures that integrate genotypic and phenotypic data. QTL are assigned to chromosome locations based on the positions of markers on a linkage map. QTL are located to regions of the genome at specified levels of statistical probability. Thus, mapping QTL is not as simple as mapping gene that affects a qualitative trait. Mapping QTL has become a reality in the past 10 years, primarily because of the availability of molecular markers (RFLPs, RAPDs, AFLPs, and SSRs). The markers segregate as single genes, and they are not affected by the environment. With polymorphic molecular markers and linkage maps as tools, mapping QTL is simply a matter of growing and evaluating large population of plants, and of applying the appropriate statistical tools.

A number of methods for mapping QTL and estimating their effects have been suggested and investigated (Edwards et al., 1987; Haley and Knott, 1992; Jiang and Zeng, 1995; Lander and Botstein, 1989; Jansen and Stam, 1994; Utz and Melchinger, 1994; Zeng, 1994). The strategy is to identify major levels of the total genetic variance that contribute to a trait's variation. In plants there is a great need to assess quantitative trait loci that contribute to important agronomic traits. The development of molecular genetic markers and the use of these markers in QTL analysis has become a powerful approach for studying the inheritance of complex traits (Edwards et al., 1987; Paterson et al., 1988). Once these traits are identified and mapped, marker-assisted selection (MAS) could be used to introduce them into a wide variety of populations. MAS can reduce breeding population sizes, continuous recurrent testing, and the time required to develop a superior line. Studies of genetic diversity are also appropriate and technically simple at this time (Lee, 1993). These studies will indicate what sources of genetic novelty can be used, for example, in sorghum improvement. QTL information can also be used as a basis for germplasm characterization and conservation.

Principle of QTL detection

The basic idea of QTL mapping has been known for over 30 years, since Thoday's seminal paper in 1961. The idea is simple: if genetic markers are scattered throughout the genome of an organism of interest, the segregation of these markers can be used to direct and estimate the effects of linked QTL, making possible the mapping and characterization of underlying QTL. The QTL method involves searching for associations between the segregating molecular markers and the character of interest in a segregating population, to identify the linkage of the marker to the QTL. To discover a marker/QTL linkage, one must have a segregating population. As mentioned above, experimental populations such as F_2 , backcross (BC), recombinant inbred lines (RILs), and double haploid lines (DHLs) are commonly used as mapping populations in plants. In the case of F_2 mapping populations, F_2 plants are usually used to genotype, and F_2 families to phenotype. Recombinant inbred lines are produced by single-seed descent. Near isogenic lines (NILs) are used for fine mapping and study of specific QTL effect. RI and DH populations are permanent populations and allow replicated evaluation for the phenotype. They are extremely useful for mapping traits which are difficult to measure.

Specific methods of QTL analysis

❶ **Single Marker Analysis (Point Analysis):** The traditional method to detect a QTL in the vicinity of a marker is studying single genetic markers one at a time. The phenotypic means for progeny of each marker class are compared (e.g., means of the marker classes AA, Aa, aa). The difference between two means provides an estimate of the phenotypic effect of substituting an A allele by an a allele at the QTL. To test whether the inferred phenotypic effect is significantly different from zero, a simple statistical test, such as t-test or F-test, is used. A significant value indicates that a QTL is located in the vicinity of the marker. Single point analysis does not require a complete molecular linkage map. The further a QTL is from the marker, the less likely it is to be detected statistically due to crossover events between the marker and the gene.

⇒ Anova, t-test or GLM approach

Follow these steps to determine the association between marker genotype and quantitative trait phenotype:

Classify progeny by marker genotype

Compare phenotypic mean between classes (t-test, GLM or ANOVA)

Significance = marker linked to QTL

Difference between means = estimate of QTL effect

⇒ **Regression approach**

Follow these steps to determine the association between marker genotype and quantitative trait phenotype:

Give numeric codes to marker genotypes (for example, aa = 0, AA = 1)

Regress phenotypes on codes

Significance = marker linked to QTL

Regression slope = estimate of QTL effect

Same result as t-test, GLM, ANOVA

New Strategies to map QTL: Because of some limitations in the basic QTL analysis approach, scientists have developed new strategies for QTL mapping.

⇒ **Interval mapping by maximum likelihood**

QTL interval mapping is probably the most common method of QTL analysis. A well known example is the Mapmaker/QTL developed by Lincoln et al. (1993). The principle behind interval mapping is to test a model for the presence of a QTL at many positions between two mapped marker loci. This model is a fit, and its goodness is tested using the method of maximum likelihood.

Maximum Likelihood approach: if it is assumed that a QTL is located between two markers, the 2-loci marker genotypes (i.e., AABB, AAbb, aaBB, aabb for DH progeny) each contain mixtures of QTL genotypes. Maximum likelihood involves searching for QTL parameters that give the best approximation for quantitative trait distributions that are observed for each marker class. Models are evaluated by computing the likelihood of the observed distributions with and without fitting a QTL effect. The map position of a QTL is determined as the maximum likelihood from the distribution of likelihood values (LOD scores: ratio of likelihood that the effect occurs by linkage : likelihood that the effect occurs by chance), calculated for each locus (Figure 8).

⇒ **Interval mapping by regression**

Interval mapping by regression (Haley and Knott, 1992) was developed primarily as a simplification of the maximum likelihood method. It is essentially the same as the method of basic QTL analysis (regression on coded marker genotypes) except that phenotypes are regressed on QTL genotypes. Since the QTL genotypes are unknown, they are replaced by probabilities estimated from the nearest flanking markers.

In most cases, regression mapping gives estimates of QTL position and effect that are almost identical to those given by the maximum likelihood method. The approximation deviates only at places where there are large gaps, or many missing genotypes.

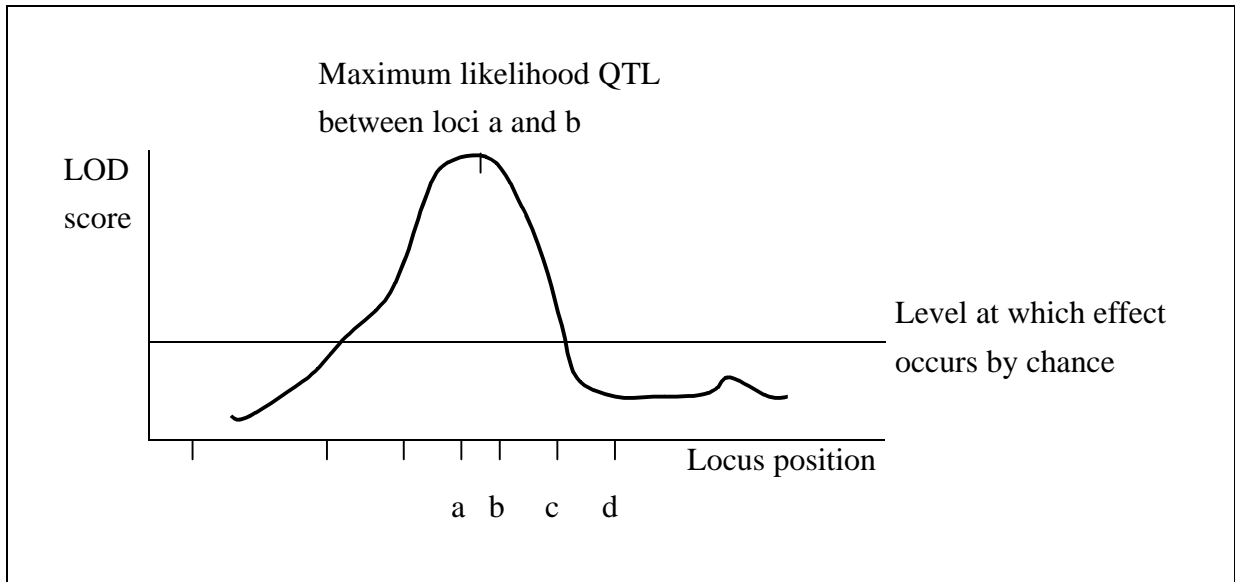


Figure 8. Simplified example for interval mapping by maximum likelihood: the map position of a QTL is determined as the maximum likelihood from the distribution of likelihood values (LOD scores: ratio of likelihood that the effect occurs by linkage : likelihood that the effect occurs by chance), calculated for each locus (modified from Jones et al., 1997).

⇒ Composite interval mapping

One of the factors that weakens interval mapping is fitting the model for a QTL at only one location. There are two problems with this approach:

- the effects of additional QTL will contribute to sampling variance;
- if two QTL are linked, their combined effects will cause biased estimates.

The method of composite interval mapping (CIM) was proposed as solution (Jansen and Stam, 1994; Utz and Melchinger, 1994; Zeng, 1994). CIM will perform the analysis in the usual way, except that the variance from other QTL is accounted for by including partial regression coefficients from markers (“cofactors”) in other regions of the genome. CIM gives more power and precision than simple interval mapping (SIM) because the effects of other QTL are not present as residual variance. CIM can remove the bias that can be caused by QTL that are linked to the position being tested.

QTL mapping software

There are over 100 genetic analysis software packages (linkage analysis and QTL mapping). Here, we list some features of the most commonly used software packages.

- **MapMaker/QTL** (<ftp://genome.wi.mit.edu/pub/mapmaker3/>) is the original QTL mapping software for IBM computer. It is user-friendly, freely distributed, and runs on almost all platforms. It will analyze F₂ or backcross data using standard interval mapping.
- **MQTL** is an IBM computer program for composite interval mapping in multiple environments. It can also perform simple interval mapping. Currently, MQTL is restricted to the analysis of data from homozygous progeny (double haploids, or recombinant inbred lines). Progeny types with more than two marker classes (e.g., F₂) are not handled.
- **PLABQTL** (<http://www.uni-hohenheim.de/~ipspwww/soft.html>) is a freely distributed IBM computer program for composite interval mapping and simple interval mapping of QTL. Its main purpose is to localize and characterize QTL in mapping populations derived from a biparental cross by selfing or production of double haploids. Currently, this program is the easiest software for composite interval mapping.
- **QTL Cartographer** (<http://statgen.mcsu.edu/qtlcart/cartographer.html>) is QTL software written for either UNIX, Macintosh, or Windows. It performs single-marker regression, interval mapping, and composite interval mapping. It permits analysis from F₂ or backcross populations. It displays map positions of QTL using the GNUPLOT software.
- **MapQTL** (<http://www.cpro.dlo.nl/cbw/>)
- **Qgene** is a QTL mapping and marker-aided breeding package written for Macintosh. It has a user-friendly graphical interface and produces graphical outputs. QTL mapping is conducted by either single-marker regression or interval regression.
- **SAS** is a general statistical analysis software. It can detect QTL by identifying associations between marker genotype and quantitative trait phenotype by single marker analysis approach such as ANOVA, t-test, GLM or REG.

Outlook

Once a genetic map has been constructed and QTL have been defined for a trait of interest, the potential use of the linked marker(s) in marker-assisted selection needs to be evaluated. For a realistic assessment of marker-assisted selection we need: high power of QTL detection; high accuracy and precision of QTL localization and estimated QTL effects; and validation of results across environmental samples, genotypic samples, generations, and across breeding populations.

Molecular markers tightly linked to economically important monogenic or oligogenic traits have potential for immediate utility in plant improvement. A major problem is when the linked marker used for selection is at a distance away from the gene of interest, leading to crossovers between the marker and the gene. In future, the success of marker assisted selection may depend on the possibility of tagging the favorable alleles themselves.

References

- Edwards, M.D., C.W. Stuber, and J.F. Wendel. 1987. Molecular-marker-facilitated investigations of quantitative trait loci in maize. I. Numbers, genomic distribution, and types of gene action. *Genetics* 116: 113 – 125.
- Haley, C.S., and S.A. Knott. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69: 315-324.
- Jiang, C., and Z.-B. Zeng. 1995. Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics* 140: 1111 – 1127.
- Jansen, R.C., and P. Stam. 1994. High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* 136: 1447 – 1455.
- Jones, N., H. Ougham, and H. Thomas. 1997. Markers and mapping: we are all geneticists now. *New Phytologist* 137: 165 – 177.
- Lander, E.S., and D. Botstein. 1989. Mapping Mendelian factors underlying quantitative traits by using RFLP linkage maps. *Genetics* 121: 185 – 199.
- Lee, M. 1993. DNA markers in plant breeding programs. *Advances in Agronomy* 55: 265 – 344.
- Lincoln, S.E., M.J. Daly, and E.S. Lander. 1993. Constructing genetic linkage maps with MAPMAKER/EXP version 3.0: A tutorial and reference manual. Whitehead Institute for Biomedical Research, Cambridge, MA.
- Michelmore R.W., I. Paran, and R.V. Kesseli. 1991. Identification of markers linked to disease resistance genes by bulk segregant analysis: a rapid method to detect markers in specific genomic regions using segregating populations. *Proceedings of the National*

- Academy of Science USA 88: 9828 - 9832.
- Paterson, A.H., E.S. Lander, J.D. Hewitt, S. Peterson, S.E. Lincoln, and S.D. Tanksley. 1988. Resolution of quantitative traits into Mendelian factors, using a complete linkage map of restriction fragment length polymorphisms. *Nature (London)* 335: 721 – 726.
- Schechert, A.W., H.G. Welz, and H.H. Geiger. 1999. QTL for resistance to *Setosphaeria turcica* in Tropical African Maize. *Crop Science* 39: 514-523.
- Stam, P., and J.W. Van Ooijen. 1995. JoinMap™ version 2.0: Software for the calculation of genetic linkage maps. CPRO-DLO, Wageningen.
- Tanksley, S.D., N.D. Young, A.H. Paterson, and M.W. Bonierbale. 1989. RFLP mapping in plant breeding: new tools for an old science. *Biotechnology* 7, 257-264.
- Thoday, J.M. 1961. Location of polygenes. *Nature (London)* 191: 368 – 370.
- Utz, H.F., and A.E. Melchinger. 1994. Comparison of different approaches to interval mapping of quantitative trait loci. Pp. 195 – 204 *in*: J.W. van Ooijen and J. Jansen (Eds.) *Biometrics in Plant Breeding: Applications of Molecular Markers*. Proceedings of the Ninth Meeting of the EUCARPIA Section Biometrics in Plant Breeding, 6 - 8 July 1994, Wageningen. CPRO-DLO, Wageningen, The Netherlands.
- Utz, H.F., and A.E. Melchinger. 1995. PLABQTL Version 1.0. A computer program to map QTL. Institut für Pflanzenzüchtung, Saatgutforschung und Populationsgenetik, Universität Hohenheim, 70593 Stuttgart, Germany.
- Zeng, Z.-B. 1994. Precision mapping of quantitative trait loci. *Genetics* 136: 1457 – 1468.