Annu. Rev. Genet. 1993. 27:205–233
Copyright © 1993 by Annual Reviews Inc. All rights reserved

MAPPING POLYGENES

Steven D. Tanksley
Department of Plant Breeding and Biometry, Cornell University, Ithaca, New York 14853

KEY WORDS: quantitative genetics, quantitative trait loci, gene action, heterosis

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org
by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

CONTENTS

INTRODUCTION 205
Tracking Polygenes with Single Gene Markers 206
Advent of Molecular Markers in Quantitative Genetics 208
Attributes of Molecular Markers with Respect to Quantitative Genetics 209
STRATEGIES FOR DETECTING POLYGENES 210
Controlled Matings 211
Random Matings 215
CHARACTERIZATION OF POLYGENIC TRAITS 216
Number of Polygenes 216
Magnitude of Effects 217
Gene Action 220
Epistasis 221
Environment by QTL Interaction 223
High Resolution Mapping of QTL 224
Relationship between Genes Controlling Polygenic and Monogenic Traits 227
Orthologous Polygenes 227
Heterosis 228
Transgressive Variation 229
PROGRESS AND PROSPECTS 230

INTRODUCTION

Macromutations (alleles with major phenotypic effects) have greatly simplified the lives of geneticists. Wingless flies, hairless mice, and dwarf plants can all be conditioned by macromutations at single loci. Loci for which macromutations exist are easy to study with the tools of Mendelian genetics since they allow the genotype of a particular locus to be unambiguously inferred from the phenotype of the individual. It is for this reason that, until recently, most of the loci displayed on genetic linkage maps corresponded to loci for which macromutations were available. While macro-

0066-4197/93/1215-0205$02.00

mutations are common in the laboratory, they are much rarer in nature. Most of the macromutations found in genetic stocks seldom occur in natural populations of a species. There is good reason since many, if not most, macromutations are deleterious to the individual and would be weeded out of the population by natural selection.

While the incidence of macromutations sustained in natural populations is relatively low, these populations are by no means deficient in phenotypic or genetic variation. However, the phenotypic variation is usually continuous instead of discrete and conditioned by allelic variation at several (and some times many) genetic loci, each with a relatively small effect. Characters whose phenotypic variation is continuous and determined by the segregation of multiple loci have often been referred to as *quantitative* traits and the inheritance as *polygenic*. The individual loci controlling a quantitative trait are referred to as polygenes or quantitative trait loci (QTL) (see later section for discussion). In nature, most genetic variation in readily observable traits is polygenic.

Lack of discrete phenotypic segregation has all but prevented the use of classical Mendelian techniques for studying polygenes. In the early part of this century a subspecialty of genetics, "quantitative genetics", began to emerge for dealing with quantitative traits. Instead of using discrete phenotypic segregation to study the inheritance of individual loci, quantitative genetics relied upon statistics to describe the characteristics of continuous phenotypic distributions. From these statistics several things could be estimated, including the approximate number of loci affecting the character in a particular mating, the average gene action (e.g. dominance, recessiveness), and the degree to which the various polygenes interact with each other and the environment in determining the phenotype (21, 36). What was not easy to decipher with this approach was the magnitude of effect, inheritance, and gene action of any *specific* locus that was affecting the character. While quantitative genetics has made many important contributions to basic genetics as well as to animal and plant breeding, the inability to describe, study, and ultimately clone individual genes affecting quantitative traits has seriously hampered the study of natural variation at loci for which macromutations do not exist.

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org
by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

Tracking Polygenes with Single Gene Markers

In 1923, Sax reported the association of seed size in beans (a quantitatively inherited character) with seed-coat pigmentation (a discrete monogenic trait) (48). He interpreted this finding as the linkage of the single gene controlling seed color with one or more of the polygenes controlling seed size. Subsequent reports have shown linkage of genes controlling quantitative variation with single gene markers (8, 50, 59, 60). The idea of using single gene markers to systematically characterize and map individual polygenes controlling quantita-

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org
by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

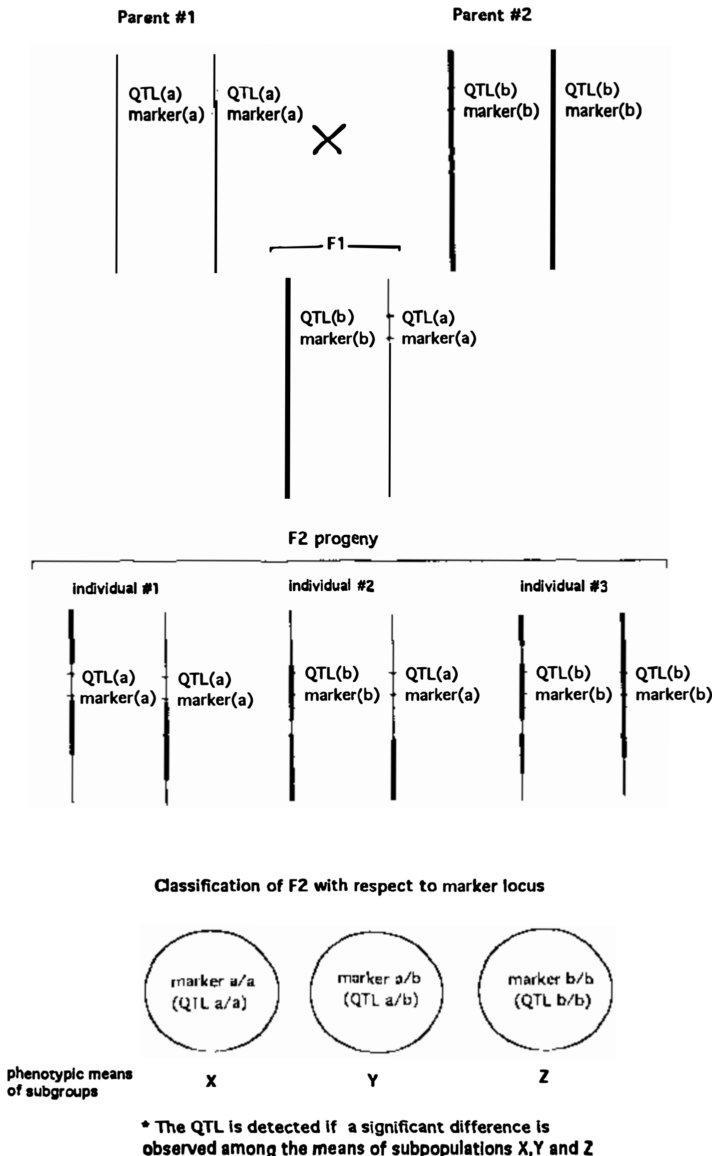


Figure 1 Marker-aided detection of a QTL (quantitative trait locus) in a theoretical F2 population segregating for quantitative trait. Top portion of figure depicts region of chromosomes in parent 1 (thin lines) and parent 2 (thick lines) that contains a marker gene and a linked QTL. F1 hybrid is heterozygous at both marker and QTL. F2 progeny segregate for recombinant chromosomes, but in most instances, genotype of linked marker accurately predicts the genotype for the QTL. Statistical comparisons of the phenotypic means of the subpopulations (comprised of each of the three possible marker genotypes —a/a, a/b, b/b) can permit statistical detection of the linked QTL.

tive traits was put forth by Thoday in 1961 (58). The idea was simple. If the segregation of a single gene marker could be used to detect and estimate the effect of a linked polygene and if single gene markers were scattered throughout the genome of an organism, it should be possible to map and characterize all of the polygenes affecting a character (Figure 1).

Putting Thoday's ideas into practice was difficult and fraught with problems. For most organisms only a few monogenic markers had been mapped, and even for those organisms where such markers were available, most were not suitable for studying quantitative traits. Often the marker genes had a larger effect on the quantitative character than did the linked polygene, thus preventing the detection of the polygene or causing a gross over/underestimate of the polygene's effect. For example, in plants a common monogenic marker is dwarfism. Individuals homozygous for the mutant allele are greatly reduced in stature compared to the wild type. It would be nearly impossible to use a dwarf marker to study the polygenic inheritance of plant height since the effect of the marker gene on height would be much greater than any individual polygene. Other problems with marker genes available at that time were dominance, epistasis, and lack of polymorphism in natural populations (see later section for discussion).

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org
by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

Advent of Molecular Markers in Quantitative Genetics

The discovery that the allelic forms of enzymes (often referred to as allozymes or isozymes) can be separated on electrophoretic gels and detected with histochemical activity stains heralded the era of molecular markers in genetics research (26, 51). No longer was it necessary for a gene to cause a discrete and visible change in the phenotype of an organism in order to study that gene. Enzyme coding genes could be screened for polymorphism in natural populations and mapped genetically using electrophoretic techniques independent of any phenotypic changes (33). By the early 1980s isozyme markers were being employed as a general tool for mapping polygenes and these studies met with considerably more success than previous studies using morphological markers (20, 55, 62, 65). The strategies for using isozyme markers for mapping polygenes were very similar to those proposed by Thoday (58). The difference in success was due to the nature of the markers themselves (see next section).

The next advance in molecular markers came with the introduction of DNA-based genetic markers, the first of which was restriction fragment length polymorphism (RFLP) (7). In the past few years a new generation of DNA-based genetic markers, based on the polymerase chain reaction (e.g. RAPDs and microsatellites), has been developed (63, 67). As with isozymes, allelic variation for DNA-based markers usually has no detectable phenotypic effect. But unlike isozymes, the genetic variation is surveyed directly at the DNA level and thus can reveal more polymorphism.

Attributes of Molecular Markers with Respect to Quantitative Genetics.

The advent of molecular markers has made it feasible to map and characterize the polygenes underlying quantitative traits in natural populations. The key properties that differentiate molecular markers from morphological markers and have permitted the rapid advance of polygene mapping are summarized below.

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

PHENOTYPIC NEUTRALITY The problem with the marker gene having a larger phenotypic effect than the linked polygene is largely overcome with molecular markers. Alternate alleles at molecular marker loci usually cause no obvious changes in the phenotype of the organism. For example, most allozymes differ by amino acid substitutions that change the migration rate of the protein on a gel, but do not significantly affect the function of the enzyme. For DNA-based markers, most of the allelic variation is in the noncoding portion of the genome. Phenotypically neutral molecular markers not only made it easier to detect linkage between the segregating marker and the polygene, but also provided an unbiased way to estimate the phenotypic effect of each polygene without interference by the marker locus.

POLYMORPHISM As discussed earlier, morphological markers identified by macromutant alleles are rare in natural populations. Without allelic variation there is no segregation, and without segregation no linkage tests can be performed to detect polygenes. The level of polymorphism maintained at any given locus in natural populations is determined by many factors, including population size, mating habit, selection, mutation rate, and migration. Two of these factors, relaxed selection pressure and higher mutation rates (in some types of DNA-based markers), cause allelic variation to be higher at molecular marker loci than at morphological marker loci. In addition, the laboratory techniques used to monitor molecular markers (e.g. gel electrophoresis, restriction enzyme analysis, and polymerase chain reaction amplification) are more sensitive in detecting existing variation (such as missense mutations and small deletions/insertions) at genetic loci than are the phenotypic screening methods used for scoring morphological marker loci. As a result, the proportion of informative (segregating) molecular marker loci can be high in crosses between individuals from the same or different populations.

ABUNDANCE If enough segregating marker genes are scattered throughout an entire genome, it is theoretically possible to detect and characterize all of the polygenes affecting a quantitatively inherited character. However, if some chromosomal regions are devoid of segregating markers, there is no

chance of detecting polygenes located in those areas. The number of useful morphological markers for quantitative genetics was so limited that in most studies only a few markers were used, representing only a small fraction of the genome. The situation improved with isozyme markers, but the number of markers was limited by the number of available enzyme activity stains and in no instances were there enough informative isozyme markers to cover an entire genome (20, 55, 62).

DNA-based markers solved the problem of limited marker abundance. By the late 1980s, complete RFLP linkage maps were available for several organisms and in 1988 the first study was published in which molecular markers, covering an entire genome, were used to map quantitative traits (43). The availability of complete genome maps also opened up the opportunity for new statistical approaches for detecting polygenes (e.g. interval analysis), which will be discussed later. Today, molecular linkage maps covering the entire genome are available for quantitative-trait studies in many organisms including humans (64), mice (12, 17), rats (49), and many plants (11, 44, 56, 61).

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org
by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

CODOMINANCE For loci with codominant alleles there is a one to one relationship between genotype and phenotype (i.e. all possible genotypes can be deduced directly from the phenotype in any generation). This is not true of loci with dominant-recessive alleles, where only homozygous recessive genotypes can be deduced unambiguously from the phenotype. Alleles of most molecular markers are codominant whereas most morphological marker loci segregate dominant-recessive alleles. Thus, the advent of molecular markers has allowed straightforward polygene mapping in virtually any segregating generation (e.g. F2, F3, backcross, recombinant inbreds).

EPISTASIS Epistasis is a form of interaction between nonallelic genes whereby one gene interferes with the phenotypic expression of another gene (4). With morphological marker loci, strong epistatic interactions among loci limit the number of segregating markers that can be unequivocally scored in the same generation. Because molecular marker loci do not normally exhibit epistatic or pleiotropic effects, a virtually limitless number of segregating markers can be used in a single population for mapping polygenes through an entire genome.

STRATEGIES FOR DETECTING POLYGENES

The underlying assumption of using marker loci to detect polygenes is that linkage disequilibrium exists between alleles at the marker locus and alleles of the linked polygene(s). Linkage disequilibrium can be defined as the

nonrandom association of alleles at different loci in a population and can be caused by a number of factors including selection and genetic drift. However, in primary segregating generations (e.g. F2, F3, or backcross populations) the predominant cause of linkage disequilibrium is physical linkage of loci and this has formed the basis for classical linkage mapping for the past century. Linkage disequilibrium due to physical linkage of loci is at its highest value in populations derived from controlled matings and as a consequence the ability to map and characterize polygenes using marker loci is also at its highest.

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org
by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

Controlled Matings

Backcross or F2/F3 populations have been used most commonly for detecting linkage between molecular markers and genes controlling quantitative traits (34, 43, 52, 55, 62). In species where severe inbreeding is tolerated, recombinant inbred populations (derived by inbreeding F2 progeny until they become virtually homozygous lines by selfing or sibbing) have also been used (9, 57). While the latter case has less linkage disequilibrium, due to more opportunity for meiotic recombination, it has the advantage of homozygous lines that can be replicated and retested for more accurate measurement of the quantitative trait (9).

There are several statistical procedures for determining whether a polygene is linked to a marker gene and they all share the same basic principle: to partition the population into different genotypic classes based on genotypes at the marker locus and then to use correlative statistics to determine whether the individuals of one genotype differ significantly compared with individuals of other genotype(s) with respect to the trait being measured. If the phenotypes differ significantly, it is interpreted that a gene(s) affecting the trait is linked to the marker locus used to subdivide the population (Figure 1). The procedure is then repeated for additional marker loci throughout the genome to detect as many QTL as possible. Normally, it is not possible to determine whether the effect detected with a marker locus is due to one or more linked genes affecting the trait. For this reason, the term quantitative trait locus (QTL) was coined to describe a region of a chromosome (usually defined by linkage to a marker gene) that has a significant effect on a quantitative trait. Determining whether a QTL is comprised of one or more genes is still one of the most difficult aspects of quantitative genetics (see later section).

SINGLE POINT ANALYSIS The simplest approach for detecting QTL is to analyze the data using one marker at a time (Figures 1, 2). This approach is often referred to as single point analysis or point analysis and does not require a complete molecular marker linkage map. It is for this reason that

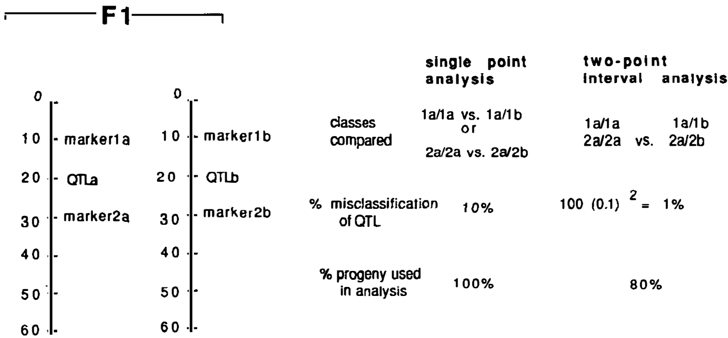


Figure 2 Point analysis versus interval analysis for detecting a QTL in a theoretical backcross population. Left side depicts F1 hybrid genotype for a QTL and two flanking marker loci (numbers on left are centimorgan distances). Note that F1 is heterozygous for both marker loci as well as QTL. For single point analysis (in backcross to parent contributing "a" alleles) statistical comparisons would be made on the *basis* of either the marker 1 genotype *or* marker 2 genotype in the backcross population. In this instance, the QTL genotype will be misclassified approximately 10% of the time since the map distance from the QTL to either marker locus is 10 cM. For interval analysis statistical comparisons are made on the basis of *both* marker loci simultaneously. Misclassification would occur only in the case of a simultaneous crossovers between the QTL and marker 1 and the QTL and marker 2. Misclassification is thus reduced to approximately 1%. Individuals recombinant between marker 1 and marker 2 are not used in the type of interval analysis shown here, therefore the number of individuals analyzed is less (80%) than with single point analysis (100%). In more sophisticated versions of interval analysis, recombinant individuals are also used in the detection and placement of QTL within interval between two markers (31).

point analyses were employed in the first molecular marker/quantitative genetic studies (20, 55, 65, 68). The disadvantages of point analysis are: (*a*) The further a QTL is from the marker gene, the less likely it is to be detected statistically due to crossover events between the marker and QTL that result in misclassification (Figure 2). (*b*) The magnitude of the effect of any detected QTL will normally be underestimated, due also to recombination between the marker locus and QTL. Both problems are minimized when a large number of segregating molecular markers are used, covering the entire genome (usually at intervals less than 15cM). Under these conditions, any potential QTL would be closely linked to at least one molecular marker.

INTERVAL ANALYSIS The availability of molecular linkage maps covering entire genomes has made it possible to overcome some problems with point analysis. To take the fullest advantage of linkage maps for quantitative studies, Lander & Botstein proposed a method called interval analysis (29–31). Instead of analyzing the population one marker at a time, sets of

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

linked markers are analyzed simultaneously with regard to their effects on quantitative traits (Figure 2). By using linked markers for analysis, it is possible to compensate for recombination between the markers and the QTL, increasing the probability of statistically detecting the QTL and also providing an unbiased estimate of the QTL effect on the character. Interval analysis, using a molecular linkage map of an entire genome, was first demonstrated on an interspecific backcross of tomato (43) and has subsequently been used successfully for several quantitative trait linkage studies (18, 27, 41, 53).

The maximum benefit of interval analysis versus point analysis is realized when linked markers are fairly far apart (> 20cM). Under these conditions there are likely to be many crossovers between the markers and QTL, which can be compensated for with interval analysis. Where the marker density is higher (markers < 15 cM apart) point and interval analysis give nearly identical results (53). When marker loci are very far apart (e.g. > 35 cM), even interval analysis is inefficient in detecting QTL in the interval between the marker loci.

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

DISTRIBUTIONAL EXTREMES Despite technological improvements in the speed and accuracy with which molecular markers can be assayed, it can still be time consuming and expensive to assay large populations. When the time and expense of assaying molecular markers is significantly greater than measuring the quantitative character of interest on each individual, it is possible to use a modified approach to detect QTL. The approach, also proposed by Lander & Botstein (31), starts with a large segregating population (e.g. F2 or backcross). A quantitative measure of the character of interest is taken on each individual in the population. Marker analysis is performed *only* on individuals in the extreme tails of the distribution (i.e. those with the lowest and highest values for the character) (Figure 3). If the allele frequency at any molecular marker locus differs significantly between the two extreme subpopulations, it is inferred that a QTL controlling the character of interest is located near the marker.

The benefit of distributional extreme analysis is in the savings of time and resources in assaying molecular markers. Given the same number of individuals assayed for molecular markers in total population analysis versus distributional extreme analysis, the statistical power of detecting QTL will be greater for the latter (31). The penalties for distributional extreme analysis are: (*a*) More segregating individuals must be analyzed for the quantitative phenotype to collect enough individuals in the distributional extremes. In many instances this is not a problem. However, in some situations the time and cost of characterizing a large population phenotypically outweighs the advantages; and for some organisms (especially animal species) it is difficult

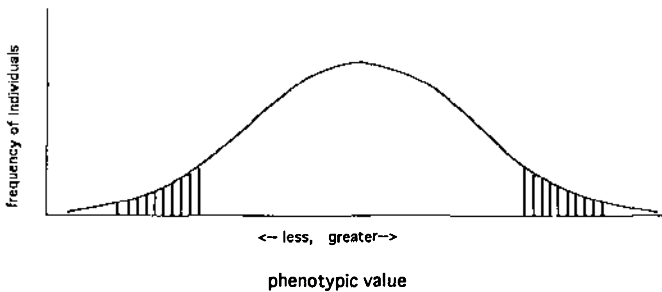


Figure 3 Distributional extreme analysis for detection of QTL. Quantitative characters often assume a continuous distribution (i.e. normal distribution shown in this figure). Individuals with extreme phenotypes (striped portions of distribution) are likely to have a large number of either "positive" or "negative" QTL alleles. Comparison of marker genotypes between individuals in the two extremes can significantly increase the probability of detecting QTL affecting the quantitative trait, but limits the ability to measure the individual effect of each QTL (see text for more detailed discussion).

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org
by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

to generate the large segregating populations required. (*b*) While distribution extreme analysis is more efficient at detecting linkage between marker loci and QTL, it is less efficient in determining individual QTL effects. Individuals in the extremes tend to have either a large number of positive or negative alleles at all QTL, depending on which extreme they represent. There is thus a deficiency of individuals with a mixture of positive and negative alleles in the subpopulations being analyzed, which confounds the ability to individually measure the effects of any specific QTL. (*c*) It is often impractical to use distributional extremes to map more than one quantitative character, since the individuals with extreme phenotypes for one character are not likely to represent the extremes for other characters.

LIMITATIONS ON PROGENY SIZE To detect QTL with molecular markers normally requires analysis of fairly large segregating populations (i.e. > 100 individuals). Although most plants and some animals readily produce offspring in such large numbers, not all species do. Where offspring numbers are limiting, alternative strategies (other than backcross or F2 analysis) must be employed. The types of modified populations employed for QTL mapping are largely a function of the reproductive characteristics of the species under study and the ingenuity of the investigator.

Half-sib analysis Half sibs are defined as the progeny derived from the mating of a single individual to random individuals in the population. If the original individual involved in the mating is heterozygous for both

molecular markers and QTL, then linkage between the molecular markers and the QTL can be detected by analyzing a large half-sib population. Both molecular markers and quantitative traits would be measured on the half sibs and either point or interval analysis would be employed to detect significant association between the alternate alleles donated by the original individual used to create the half sibs. This strategy can be used when the gametes of one sex (usually the female sex) are limiting. Prime candidates for this approach would be farm animals where a single heterozygous male could be mated to a large number of females to create large half-sib populations. Because the phenotype of half sibs is determined only in part (theoretically one half) by genes inherited from the original heterozygous individual, the statistical power of detecting linkage between segregating markers and QTL is decreased significantly, requiring even larger populations.

Random Matings

For most plants and many experimental organisms, it is possible to select any two individuals and make controlled matings between them. Controlled matings have the advantage of allowing the investigator to pick individuals that differ significantly for the character of interest. The greater the phenotypic difference between any two individuals, the more likely one is to detect significant QTL controlling that character in a derived, segregating population. Controlled matings are also advantageous in that they result in maximum linkage disequilibrium (due to physical linkage between loci) for detecting QTL with linked molecular markers. However, for some species, it is either impractical or unethical to make controlled matings. Humans are a prime example. Not only are natural family sizes too small for QTL analyses, but it is necessary to work with the matings that nature has given us. Unfortunately for the geneticists, individuals with the most extreme phenotypes do not always produce offspring. The only practical solution to this problem is to pool quantitative and molecular marker data from a multitude of families and attempt to analyze for QTL linkage using the entire dataset. This creates one of the most difficult problems in polygene mapping—trait heterogeneity.

HETEROGENEOUS TRAITS In random-mating populations, individuals with similar phenotypes do not necessarily have similar genotypes. The phenotype is the result of the combined action of all the genes affecting a character, and different combinations of alleles can create phenotypically similar individuals. The underlying assumption of pooling data from different segregating families is that the phenotypic variation segregating in the families is due to segregation of the same QTL. If this assumption is met,

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org
by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

then pooling data from families will increase the statistical power of detecting QTL. However, if different QTL are segregating in each family, pooling of data will not increase the statistical power and such QTL will likely go undetected. A comprehensive discussion of the problems and potential solutions for mapping polygenes controlling heterogeneous traits is presented by Lander & Botstein (29, 30).

CHARACTERIZATION OF POLYGENIC TRAITS

Number of Polygenes

Classical quantitative genetics theory has yielded several statistical approaches for estimating the number of genes segregating in a population that affect a given morphological character (21, 28, 32, 40). These estimates are based on characteristics of the phenotypic distribution in the segregating population(s) (especially on the phenotypic variance) and often depend on various underlying assumptions, including approximately equal effects of individual polygenes, independent assortment of polygenes, similar gene action (allelic interaction) of polygenes, and minimal epistasis among polygenes. In practice, one or more of these assumptions are normally violated, making the estimates suspect. Furthermore, until recently there was no practical way to independently verify or falsify one's estimates.

Estimating the number of polygenes using molecular marker approaches is straightforward, although not without limitations and biases. The approach is simply to add up the number of QTL detected in a particular study and to use that value as an estimate of the number of segregating polygenes affecting the character in that population. The limitations of this approach are : (*a*) Only those genes with a sufficiently large phenotypic effect to be detected statistically will be counted. Genes with lesser effects will fall below the threshold of detection, depending on the size of the segregating population (Figure 4). The larger the population, the more likely it is to statistically detect genes of lesser effects. Therefore, the molecular marker approach almost always underestimates the number of genes affecting a character and is biased towards the detection of genes of larger phenotypic effects. (*b*) Using typical population sizes (< 500 individuals) two or more polygenes closer together than approximately 20 cM will usually appear as a single QTL (i.e. they cannot be distinguished as separate genes). Again, this is a bias toward underestimating the number of genes.

The advantage of the molecular marker approach to estimate the number of polygenes segregating in a population is that results are quantifiable and testable. They are quantifiable in that the cumulative genetic and phenotypic variance attributable to the QTL detected can be estimated. For example,

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org
by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

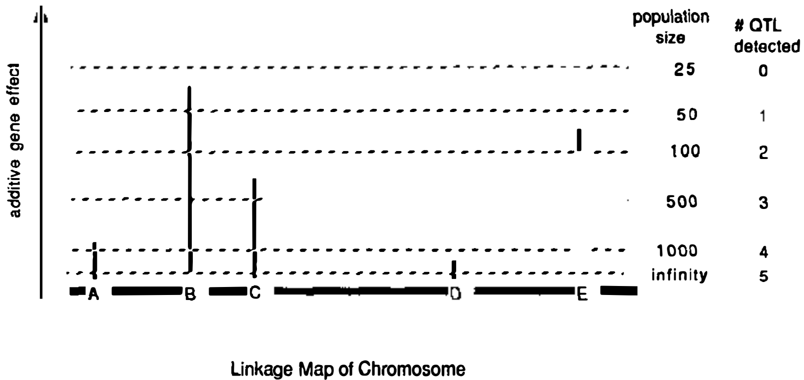


Figure 4 Relationship between the size of a segregating population and the probability of detecting QTL with varying effects. QTL (A,B,C, etc) are shown along X axis which represents a single chromosome. Y axis reflects magnitude of effect of individual QTL and is indicated by vertical height of bar above each QTL. Dotted lines correspond to different population sizes. If the dotted line intersects a vertical bar, the QTL will be detected statistically in that sized population. Note that with small population sizes (e.g. 25 or 50), only the major QTL are detected. To detect the QTL of smaller effects, very large population sizes are required.

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

if five QTL are detected and together account for 70% of the genetic variance, additional QTL (probably of smaller effects) must account for the remaining 30% of the variance. The results are testable since the experiment can be repeated (even in different environments) to see if the same number and genomic placement of QTL are observed (41, 53). Also, new populations or genetic stocks can be created to study any QTL in further detail, including high-resolution mapping to determine if any particular QTL is composed of more than one gene (see section on high-resolution mapping). Table 1 gives a sample of gene number estimates derived from molecular marker mapping in several organisms for a variety of characters.

Magnitude of Effects

It would be expected that different polygenes should be unequal in the magnitude of effect they exert on a character. However, in the past it was difficult to test this assumption. In the past 10 years enough data have accumulated from marker studies to establish definitively that polygenes do vary widely in their effects and that the assumption of many polygenes with equal effects (required for some of the previous statistical models) is not valid. More surprising has been the finding that in many instances a large proportion of quantitative variation can be explained by the segregation of a few major QTL. It is not uncommon to find individual QTL that can

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org
by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

Table 1 Summary of QTL detected for different characters in a variety of species

Organism	Trait	Population	# QTL	% phenotypic variance max	min	total	Ref.
Tomato	soluble solids	BC1 (237)	4			44	41
		F2/F3 (350	7	28	6.0	44	43
	fruit mass	BC1 (237)	6			58	41
		F2/F3 (350)	13	42	4.0	72	43
	fruit pH	BC1 (237)	5			48	41
		F2/F3 (350	9	28	4.2	34	43
	growth	F2 (432)	5	7	3.2	21	16
	leaflet shape	F2 (432)	9	30	3.1	60	16
	height	F2 (432)	9	8	3.1	42	16
Maize	height	F2 (112)	6			73	5
		F2 (112)	4			53	5
		F2 (144)	3			34	5
		F2 (144)	3			45	5
		F2 (260)	7	35	4.0	67	18
		F2 (187)	11	27	4.0		20
	tiller number	F2 (260)	1			24	18
	leaf length	F2 (260)	7	17	5.0	57	18
	glume hardness	F2 (260)	5	42	6.0	72	18
	grain yield	F2 (1930)	13*	4	0.6		19
		F2 (1776)	18*	5	0.3		19
		F2 (187)	13	17	4.0		20
		F3BC1 (264)	8	15	6	59	53
		F3BC1 (264)	6	21	8	61	53
	number of ears	F2 (1930)	10*	5	0.4		19
		F2 (1776)	9*	4	0.2		19
Common bean	nodule number	F2/F3 (70)	4	17	11	50	39
Mung bean	seed weight	F2 (58)	4	33		50	22
Cowpea	seed weight	F2 (58)	2	37	32.9	53	22
Wheat	pre-harvest sprout	RI (78)	4	14	10.0	37	3
Mice	epilepsy	BC1 (87)	2			50	46
Rats	hypertension	F2 (115)	2	19	11	30	27

*Does not take into account linked molecular markers

account for > 20% of the phenotypic variation in a population (Table 1) and values as high as 42% have been reported for single QTL (18). Since the percent phenotypic variance includes environmental variance, the actual genetic variance attributable to major QTL is even larger (16, 53).

QTL with major effects have been identified for most characters studied, but most by far of QTL reported are those of smaller effects (Figure 5). This result is logical if one considers the possibility that most segregating genes in a population probably have some effect on most characters—albeit very small in most cases. The smallest effect a QTL can have and still be

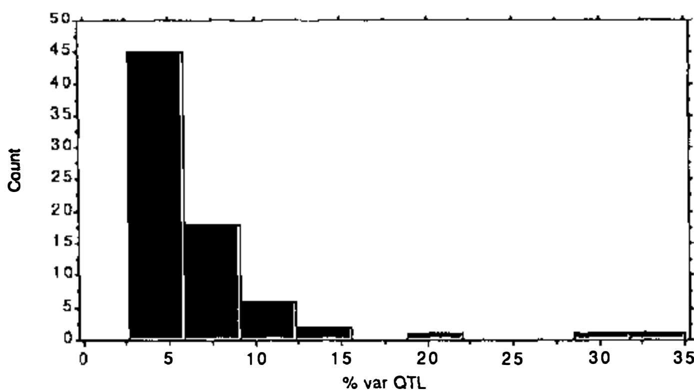


Figure 5 Histogram depicting the percentage of total phenotypic variance attributable to individual QTL derived from a published study in tomato (16). In this study 74 significant QTL were detected for 11 different quantitatively inherited characters. While "major" QTL were detected for most characters (i.e. those accounting for more than 10% of the phenotypic variance), most QTL were of much smaller effect, causing an asymmetrical distribution skewed towards QTL of smaller effects. QTL with individual effects less than 2% were not reported, presumably because the population size was not large enough to statistically detect such QTL.

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

detected by the marker method depends on a number of factors: (*a*) The map distance from nearest marker to the QTL: the closer a QTL is to a marker, the smaller the effect that QTL can have and still be detected statistically because the effects of QTL closer to the marker are less confounded by recombination events between the marker and the QTL. (*b*) Size of segregating population: the larger the population size, the more likely the effects of lesser QTL will reach statistical significance. (*c*) Heritability of trait: the larger the environmental effect on the character (i.e. lower the heritability), the less likely a QTL will be detected. (*d*) Probability criteria used for declaring a QTL effect significant: higher probability thresholds reduce the chances of spurious QTL being reported, but also reduce the chances of detecting QTL with smaller effects. In maize using an F2 population size of 1700 individuals and a probability threshold of 0.05, QTL contributing as little as 0.3% of the phenotypic variance were reported (20). In experiments with smaller population sizes and higher probability thresholds, QTL that explain less than 3% of the phenotypic variance are not normally detected (Table 1).

It is difficult to determine how efficient an experiment has been in identifying the QTL responsible for a trait. One measurement of success has been the cumulative phenotypic variance attributable to the combination of all significant QTL. Where complete molecular maps have been used

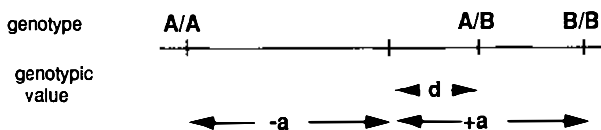
for quantitative studies, the values have ranged from as high as 95% (18) to less than 10% (16), with the average approximately 30–40% (Table 1).

The bias towards detecting QTL with larger effects means that it is unlikely that one will ever detect, map and characterize all of the polygenes affecting a character in any given segregating population. Theoretically, this may be a problem, but is only a minor practical limitation. Using the marker approach, one is likely to identify and characterize QTL making the largest contribution to the phenotype and it is likely to be these QTL that one would want to further characterize and ultimately to clone. For plant and animal breeders, the major QTL are of greatest interest to manipulate in breeding schemes via association with molecular markers.

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org
by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

Gene Action

In diploids, the two alleles at a genetic locus can interact in a number of ways to produce the phenotype of the individual, often referred to as gene action. In classical genetics, which is normally confined to macromutations, alleles are normally either dominant or recessive. However, in natural populations (and especially in loci underlying quantitative variation) alleles are seldom absolutely recessive or dominant. Instead, the gene action for specific alleles ranges from complete dominance to complete recessiveness to overdominance (heterozygotes exceed either parental homozygote), and everything in between. Quantitative geneticists have devised several parameters to describe such continuous gene action. The dominance/additivity (d/a) statistic is the most common and describes the degree to which the heterozygous genotype resembles the parental homozygotes (Figure 6). A d/a value of 1.0 is considered to be complete dominance and the heterozgote would be identical to the greater parental homozygote (Figure 6). A value of −1.0 is considered complete recessiveness and the heterozygote would



a = additive effect (of a single allele) = (B/B - A/A)/2
d = dominance deviation = A/B - [(A/A + B/B)/2)]
d/a = degree of dominance
(-1 < d/a < 1)

Figure 6 Statistics for individual QTL that can be estimated from marker-aided studies of quantitative traits. The dominance/additivity (d/a) statistic is most commonly used to describe gene action for QTL (see text for discussion).

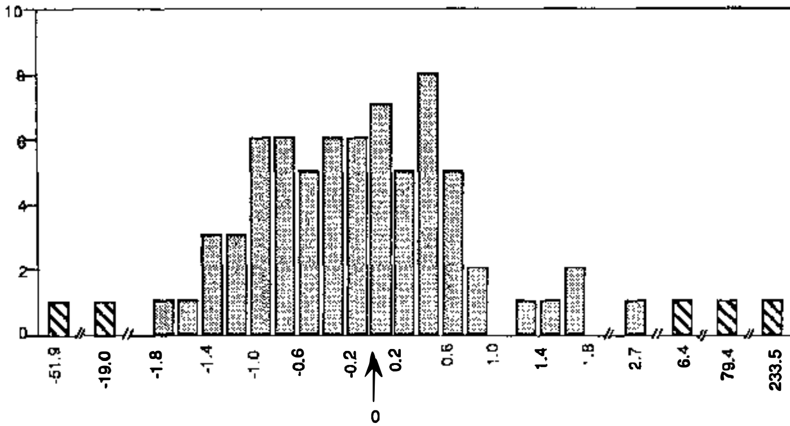


Figure 7 Histogram of individual QTL d/a values derived from a published study in tomato (16). Note the approximate normal distribution of gene action ranging from overdominance (striped bars on right) to underdominance (striped bars on left), with most QTL showing additivity (1≤d/a≤1).

be identical to the lesser homozygote. A value of 0.0 applies to cases where the heterozgote is exactly intermediate between the two homozygotes—a situation referred to as additive gene action. Values greater than 1.0 or less than −1.0 imply overdominance/underdominance.

The histogram in Figure 7 displays the estimated gene action for 74 individual QTL identified in a single mapping study in tomato (16). The continuous distribution of gene action observed in this study is typical of that reported in a variety of other QTL mapping studies (18, 20, 41, 53). This wide range of gene actions for naturally occurring polygenes is in sharp contrast to the dominance/recessiveness exhibited by the majority of alleles utilized in laboratory studies. This can probably be attributed to the fact that most alleles studied in laboratories are derived from either spontaneous or induced mutations, which result in loss of a gene product or production of a defective gene product. Such mutations would be expected to behave in a recessive manner. In nature, such loss-of-function mutations would likely be deleterious or lethal and would be weeded out by natural selection. The type of variation maintained in nature is more likely to be for allelic forms that modify, but do not eliminate, the gene product.

Epistasis

For polygenic traits, the potential number of epistatic interactions (two-locus, three-locus, etc) is enormous. In classical quantitative genetic studies, the genetic variance attributable to epistasis has often been referred to as

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org
by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

genotype-by-genotype interaction (21). In the past it was seldom possible to determine whether such interactions were occurring between specific polygenes or to measure the exact nature of specific interactions. The ability to map and characterize individual QTL via molecular mapping techniques raises the possibility of detecting and characterizing interactions among specific QTL. One simple statistical method for determining two-locus interactions of QTL is through the use of two-way analysis of variance (ANOVA). Once the significant QTL affecting a trait have been detected (using methods discussed previously), any two significant, unlinked QTL can be used as the independent variables and the quantitative character as the dependent variable in a two-way ANOVA. The "interaction" factor calculated in the two-way ANOVA will be an estimate of the interaction between the two QTL in determining the phenotype. A significant interaction factor suggests that the effect rendered by the two QTL together is not simply a sum of their independent effects.

Theoretically, all possible two-way, three-way, and up to *n*-way interaction among QTL can be measured through *n*-way ANOVA or related statistics. Practically, there are several problems (some very serious) with this approach:

1. Segregating population sizes are usually too small for accurate estimates of multi-locus interactions. To measure interaction between two loci, all possible genotypes for those two loci must appear in the population in sufficient frequencies to allow statistical comparisons. Taking the simple case of two independent QTL (A and B) segregating in an F2 population, there would be nine possible two-locus genotypes and the four rarest classes (double homozygotes, e.g. AABB, AAbb, etc) would each occur in an expected 1/16 frequency. If the original population size had been 500 individuals, the expected number of individuals in each of these rare classes would be 31. This number of individuals is a relatively small sample on which to estimate a phenotypic effect and with which to compare that effect with the effects observed in other similarly rare classes. To have a good measure of the phenotypic effect of each class (and thus to gain a more accurate measure of epistasis) would require a much larger population. This is only for the two-way interactions. Three-way interactions would result in 27 genotypic classes in an F2 for which phenotypic effects must be compared and the rarest class would occur at a frequency of 1/64. Segregating populations of thousands of individuals would be required to obtain enough individuals in each class for adequate statistical comparisons.
2. The number of potential multilocus interactions is very large, requiring many statistical tests, some of which will, by chance, reach statistical

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

significance. If N significant QTL were detected for some trait, there would be $N!/2(N-2)!$ possible two-way interactions. In the case of 10 QTL, the number would be 45. Using a significance level of 0.05, one would expect two spurious interactions to be reported. One solution is to raise the significance threshold to avoid reporting spurious interactions. However, as the threshold is increased, it is less and less likely that real interactions will be detected—a problem exacerbated by small population sizes, as discussed above. Higher level interactions are even more problematic statistically. With 10 QTL there would be 240 possible three-way interactions!

3. Interactive effects will be underestimated with simple statistical approaches like n-way ANOVA due to recombination between the QTL and the linked molecular marker on which the ANOVAs are based. The problem can be mitigated by using higher density molecular linkage maps for detecting and analyzing QTL such that most QTL are closely linked with at least one molecular marker.

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org
by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

Despite the limitations of estimating epistatic effects among specific loci, an interesting pattern is emerging from molecular marker studies of quantitative traits. In most studies conducted thus far the number of statistically significant QTL interactions has normally been close to the number expected to occur by chance (16, 20, 53, 55). While a few of the specific QTL interactions have had very high probability values (suggesting that they are real and not spurious), most have been near the probability threshold, which would be consistent with them being a statistical artifact. These results would suggest that strong epistatic interactions (i.e. the type observed between some macromutant alleles) are the exception and not the rule for naturally occurring polygenes. These conclusions are supported by the few studies in which individual QTL have been genetically isolated from other QTL in nearly isogenic lines and have been shown to continue producing their same individual effects (16).

Because of the limitations listed above, it seems unlikely that it will be possible to produce detailed descriptions of QTL by QTL interactions using primary segregating populations (e.g. F2, backcross, etc). More likely, it will be necessary for geneticists and breeders to construct nearly isogenic lines with single QTL and combinations of QTL that can be replicated in experimental designs to allow a more precise measurement and description of epistatic interactions.

Environment by QTL Interactions

The phenotype of an individual is conditioned not only by its genotype, but by the interaction of the genotype with the environment. Quantitative

geneticists have long recognized the importance of genotype by environment interactions and have documented numerous cases of such interactions (21, 24, 38). These results suggest that QTL important in one environment may not be as important in determining the phenotype in another environment. Two recent studies involving QTL mapping of the same populations in different environments have provided a test of this prediction.

In maize, seven agronomic traits, including grain yield, were measured in the same population in six different environments in the United States (53). QTL detected in one environment were frequently detected in the other environments, suggesting very little environment by QTL interaction. A related study in tomato examined segregating populations derived from the same original cross in three environments (two in the United States and one in the Middle East) (41). Forty-eight percent of the QTL were detected in at least two of the environments, with the remainder being detected in single environments. The QTL showing the largest effects in one environment were also more likely to be detected in another environment. The two US sites shared more common QTL with each other than they did with the Middle East sites, consistent with major differences in both the climate and horticultural practices.

While it is premature to draw any sweeping conclusions about environment by QTL interactions, the two studies described above suggest that a substantial proportion of QTL affecting a quantitative trait in one environment will be active in other environments and that this is especially true for QTL with major effects. This would be an especially welcome result for plant and animal breeders trying to use linked markers to transfer potentially valuable QTL into modern varieties/breeds in the hope that they will make positive contributions to the phenotype in many diverse environments.

High Resolution Mapping of QTL

In the genetics of diploids, complementation is the acid test for whether or not two genes are allelic. Complementation depends upon the availability of dominant/recessive alleles. Since very few QTL demonstrate such allelic interactions, complementation tests are of little value in determining whether a QTL is comprised of a single gene or several linked genes affecting the character of interest. Geneticists must therefore resort to other strategies for determining the genic composition of a QTL. If the QTL is the result of two or more linked genes, it should be possible to separate those genes by crossing-over. This entails high-resolution mapping of QTL. Unfortunately, high-resolution mapping of QTL in primary generations suffers from the same limitation as does measuring epistasis—segregating population sizes are too small to allow such fine mapping. Fine mapping of QTL involves comparing the means of individuals recombinant for molecular markers in the vicinity of

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

the QTL with individuals that are nonrecombinant. The classes of limiting size will be the recombinant classes. For example, in a backcross population of 500 individuals one would expect approximately 13 individuals in each of the two recombinant classes using two markers that are 5 cM apart. For markers 1 cM apart there would be approximately three individuals in each recombinant class. Current strategies for mapping QTL depend on comparing the means of recombinant and nonrecombinant classes (31, 55). Because of the small proportion of recombinant individuals, the statistical tests are not very powerful and given the practical sizes of segregating populations currently used for QTL studies, mapping resolution of QTL has been limited to approximately 10–20 cMs—inadequate for distinguishing between single gene versus multigene composition of individual QTL.

Paterson et al (42) proposed a solution to this problem in which recombinant individuals are identified in primary generations and selectively multiplied in subsequent generations so that the recombinant classes occur at near equal frequency with the nonrecombinant classes, increasing the power for the statistical comparisons among classes. In the ideal situation, a series of nearly isogenic lines (NILs), differing in recombination in the QTL regions, would be compared for the quantitative trait being mapped. Such comparisons can potentially allow placement of a QTL to a very small interval on the map (Figure 8). NILs have the advantage of being genetically uniform throughout the rest of their genomes, which makes it easier to study the effects of a single QTL. Unfortunately, it is often time consuming to develop NILs. As an alternative, earlier generations can also be used in a similar manner for high-resolution mapping of QTL—albeit with less efficiency (42).

If linked molecular markers are used as a means of selecting for recombinants in the vicinity of QTL for fine mapping, then the number of available molecular markers in the chromosomal region of interest can become a limiting factor. High density molecular linkage maps are being developed for a number of organisms and may provide a solution to this problem. In addition, methods of selectively identifying molecular markers in targeted regions of the genome have been reported and may provide the degree of marker saturation necessary for high-resolution mapping of QTL (23, 35, 37).

Recombination can be used as evidence for multigene composition of a QTL, but cannot be used directly to prove a single gene composition. If a recombinant separates a QTL into two separate QTL, then the multigene hypothesis is supported. Failure to do so could be due to the genes being so tightly linked that the appropriate recombinants were not recovered in the population examined. However, the more recombinants examined in an area, the more likely that multiple genes would be separated. Recovering crossovers at intervals of 0.1 cM (or even 0.01 cM), which might be

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

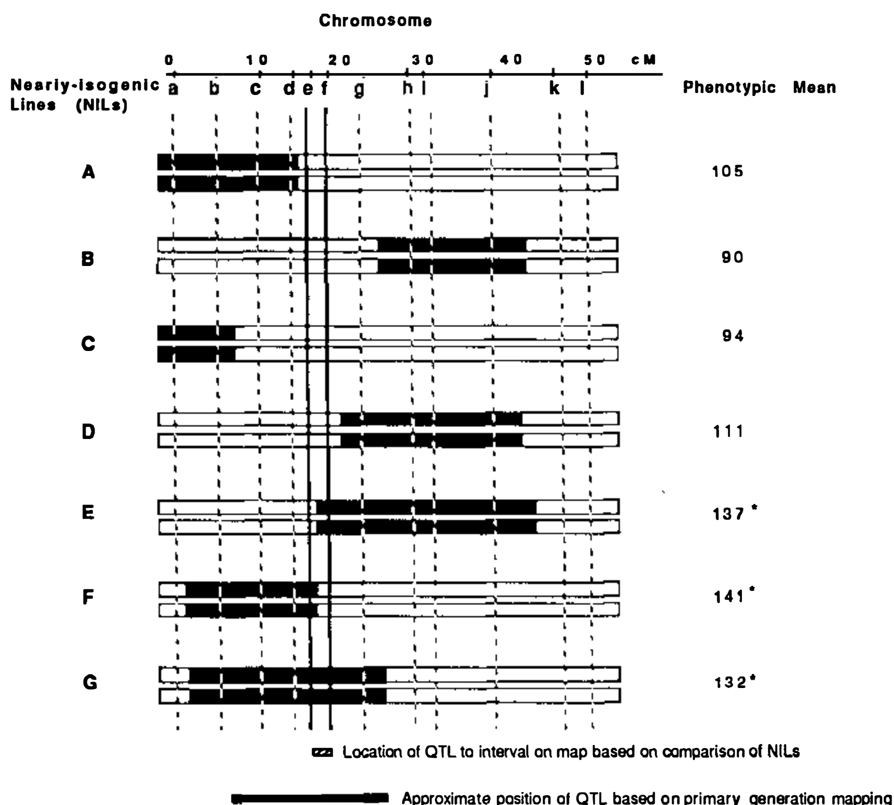


Figure 8 High-resolution mapping of a QTL with nearly isogenic lines (NILs). In primary generations (e.g. F2, BC1, RI), the map position of individual QTL is only roughly approximated (shown by hatched bar at bottom of figure). Once the approximate position of a QTL is known, NILs with crossovers in the vicinity of the QTL can be constructed using linked marker loci (a,b,c, etc). In this theoretical example, seven NILs have been created by backcrossing to the parent depicted by the white chromosomal segments. Dark chromosomal segments come from the other parent. By comparing the phenotypic values of the NILs, the QTL can be mapped more accurately. In this instance, E,F,G have phenotypic values significantly (indicated by asterisks) greater than other NILs indicating that the QTL resides in the interval bounded by marker e and f. If additional markers were available in the e,f interval, a second generation of NILs could be created with crossovers in the interval for even finer mapping of QTL.

necessary for studying QTL composition, can be a time-consuming and labor-intensive task. To have a 95% chance of recovering at least one crossover in a 0.1 cM interval would require approximately 3000 individuals. For some species, it will be difficult to generate such a large number of segregating progeny and even in the species where sufficient numbers can be generated, it is potentially a time-consuming and expensive task to screen

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

through these numbers with molecular probes. New sampling procedures for selectively isolating recombinants in specific regions of the genome (10), combined with advances in molecular marker technology, are likely to overcome this problem and in the near future, high-resolution mapping of QTL should be realized.

Relationship between Genes Controlling Polygenic and Monogenic Traits

A long-standing question in quantitative genetics is whether the loci controlling qualitative variation (e.g. those with macromutant alleles) are the same as those controlling quantitative variation (e.g QTL). It is apparent from previous work that the alleles for quantitative variation assume a wider range of allelic interactions and have a smaller individual effect than alleles for qualitative variation. However, as pointed out by Robertson (47), it is possible that alleles for qualitative mutants are simply loss-of-function alleles at the same loci underlying quantitative variation. Theoretically, QTL mapping studies can provide a test of this hypothesis. If a gene contributing to quantitative variation is allelic to a gene controlling qualitative variation, then these genes should map to the same locus along the chromosome. For some organisms (e.g. maize, Drosophila), many of the major qualitative loci controlling morphological variation have been mapped with a high degree of precision on genetic maps and these locations should be predictive of the locations of polygenes mapped for the same character.

This hypothesis has been tested in maize. Plant height is a quantitative trait with known qualitative mutants—many of which have been mapped in maize. Beavis and colleagues (5) attempted to test the relationship of qualitative mutants to quantitative variation by mapping QTL for plant height in four maize F2 populations and comparing the map position of those QTL with previously known positions of qualitative variations for the same character. The results showed a general concordance in map positions of QTL and major genes affecting height and is therefore consistent with the hypothesis. Unfortunately, the QTL were located on the map with a low degree of resolution, raising the possibility that the QTL are linked, but not identical to the qualitative loci—a problem acknowledged by the authors. Until QTL are mapped to higher degrees of precision and/or cloned, it will be difficult to prove that the particular QTL actually correspond to known loci defined by macromutant alleles.

Orthologous Polygenes

The advent of technologies for mapping genomes directly at the DNA level has opened the door for comparative genetic mapping among sexually incompatible species. Using a common set of RFLP probes, comparative

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org
by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

linkage maps have now been constructed for tomato-potato-pepper (6, 54, 56), maize-sorghum-rice-wheat (1, 2, 25, 66), and humans-cattle-mice (15, 69). Comparative linkage maps provide a basis for interpreting genetic information among divergent species. For species connected by comparative genetic maps, it should be possible to compare the map positions of QTL for the same or similar characters. Coincidence of map positions would support the hypothesis that loci underlying natural quantitative variation have been conserved during long periods of evolutionary divergence (i.e. they are orthologous genes). In this case, breeders might be able to predict the positions of important QTL (e.g. for growth rates in animals or yield in plants) in one species based on mapping studies from different species.

At present, too few QTL mapping studies have been published for species connected by comparative linkage maps to draw any general conclusions regarding the hypothesis of conserved QTL among divergent species. Perhaps the best evidence for orthologous QTL comes from mung bean and cowpeas (22). In this study the researchers showed that the single most significant QTL for determining seed weight in these two distinct species maps to the same chromosomal locus in both genomes and that the chance occurrence of such coincidental mapping is very unlikely.

Currently, QTL mapping studies are being conducted in rats and mice in the expectation that loci related to congenital diseases are conserved in mammals and that comparative linkage maps will provide a basis for extrapolating the results from studies with rats and mice to humans (14, 27).

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org
by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

Heterosis

Outcrossing (reproduction by mating of two different individuals versus self-fertilization of a single individual) is the commonest mode of reproduction in nature and is enforced by a number of mechanisms, including sex chromosomes in animals (and in some plants) and genetic self-incompatibility in plants. Outcrossing promotes heterozygosity and is often associated with heterosis or hybrid vigor, whereas self-fertilization promotes homozygosity and is associated with inbreeding depression in many organisms.

The genetic basis of heterosis has been debated for many years and is still not resolved. Most geneticists agree that heterosis is at least in part due to dominance (masking of deleterious recessive alleles in heterozygous individuals). More controversial is the role of overdominance in heterosis. Overdominance refers to the situation where individuals heterozygous at a particular locus are superior compared with individuals homozygous for either allele—implying a synergistic interaction between the gene products encoded by the two alleles. In the past, individual loci responsible for hybrid

vigor were hard to identify and characterize, making it difficult to resolve the issue of dominance versus overdominance. Recent QTL mapping studies have mapped some of the genes controlling heterosis and are shedding some light on this issue.

The most comprehensive study mapping loci for heterosis was reported by Stuber et al in maize (53). Progeny families derived from a highly heterotic F1 hybrid were evaluated for grain yield and a variety of other traits. The most notable result was that the majority of QTL detected for yield demonstrated overdominant gene action, which would support overdominance as an important factor in heterosis in this species. The limitation to this and virtually all studies involving the mapping of QTL in primary generations is that, since the QTL could not be mapped to exact points in the chromosomes, the QTL demonstrating overdominant gene action may actually be composites of two or more loci with dominant and recessive alleles in coupling which would mimic overdominance—a phenomenon sometimes referred to as pseudo-overdominance (13). High-resolution mapping of individual heterotic QTL should help resolve the issue of overdominance versus pseudo-overdominance since recombination between dominant and recessive loci would be expected with the latter. Ultimately it should be possible to clone heterotic QTL using map-based cloning techniques, and thus allow the molecular basis of this important phenomenon to be determined.

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org
by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

Transgressive Variation

Transgression is defined genetically as the appearance of individuals in segregating populations that fall beyond the parental phenotypes (usually with respect to quantitatively inherited characters) and are often observed in offspring of both intraspecific and interspecific matings. There are several potential causes of transgression including de novo mutation and unmasking of recessive deleterious alleles due to inbreeding (45). However, the cause most often proposed for transgression is accumulation in certain progeny of complementary alleles at multiple loci inherited from the two parents (45).

QTL mapping studies have provided direct evidence for the basis of transgression. Most QTL mapping studies reported thus far have involved analysis of the segregating progeny derived from crossing two individuals significantly different with respect to one or more quantitative traits. One would normally expect the parent with the higher value for a quantitative trait to possess a higher proportion of positive alleles with regard to that character. QTL mapping studies have largely borne out that expectation. At the same time, most QTL mapping studies also report the detection of QTL with allelic effects opposite to those predicted by the phenotype of the parent from which they originated (16, 18, 43). Such "complementary

QTL" would be expected in cases where transgressive segregation is observed in the progeny. What is somewhat surprising is the relatively high frequency with which complementary QTL have been found to occur---especially in interspecific crosses. A good example can be found in tomato.

An F2 population, derived from a cross between two different tomato species, was evaluated for 11 quantitative traits. Transgressive segregation was observed for eight of these traits. Mapping using RFLP markers uncovered 74 QTL that accounted for a large proportion of the genetic variance for the 11 characters, and 36% of those QTL had allelic effects opposite to those predicted by the parental phenotypes (16). In some instances, over half of the QTL detected for any given character were of a complementary nature. These complementary QTL were directly related to the appearance of transgressive individuals in the F2. The finding that different species contain such a high proportion of complementary QTL supports hypotheses that interspecific hybridization can rapidly lead to new races or species with characteristics or adaptations that exceed those of the parental species, due to recombination of a relatively few complementary QTL (16). These results also suggest a strategy for more efficient use of wild plant germplasm to improve domesticated crops (16).

PROGRESS AND PROSPECT

Our ability to map and characterize genes underlying quantitatively inherited traits has advanced considerably in the past 10 years due to the use of molecular marker techniques. We have learned that, while there are many genes influencing the phenotype of quantitative characters in segregating populations, these genes vary tremendously in both the magnitude and nature of their effects. Experimental results have also shown that single "major" QTL often account for 10–50% of the phenotypic variation in segregating populations. Further mapping studies are likely to reveal whether the same QTL reported in one population also account for the majority of variation in other, independent, populations. Ultimately, these results will be expanded to include the identification of orthologous QTL among different species. Conservation of QTL among species may provide opportunities for plant and animal breeders to use QTL mapping information from one species in the design and execution of breeding studies in another.

Finally, the ability to conduct high-resolution mapping experiments with QTL should help determine whether QTL are single genes or clusters of tightly linked genes and whether overdominance plays a significant role in conditioning heterosis. High-resolution mapping also sets the stage for map-based cloning of genes underlying quantitative variation—an event that will usher in the age of molecular quantitative genetics.

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org
by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

ACKNOWLEDGMENTS

I thank Drs. G Churchill, D Matthews, S McCouch, A Paterson, N Young, D Zamir for helpful comments, and S David for assistance in preparing the manuscript. Research in my laboratory was supported by grants the National Research Initiative Competitive Grants Program, USDA no. 91-37300-6418, the Binational Agricultural Research and Development Fund (no. IS-1822-90C) and The Rockefeller Foundation.

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org
by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

Literature Cited

1. Ahn S, Anderson JA, Sorrells ME, Tanksley SD. 1993. Homoeologous relationships of rice and wheat chromosomes. *Theor. Appl. Genet.* In press
2. Ahn S, Tanksley SD. 1993. Comparative linkage maps of the rice and maize genomes. *Proc. Natl. Acad. Sci. USA*. In press
3. Anderson JA, Sorrells ME, Tanksley SD. 1993. RFLP analysis of genomic regions associated with resistance to pre-harvest sprouting in wheat. *Genomics*. In press
4. Bateson W. 1907. The progress of genetics since the rediscovery of Mendel's paper. *Progr. Rei Bot.* 1:368
5. Beavis WD, Grant D, Albertsen M, Fincher R. 1991. Quantitative trait loci for plant height in four maize populations and their associations with qualitative genetic loci. *Theor. Appl. Genet.* 83:141–45
6. Bonierbale MW, Plaisted RL, Tanksley SD. 1988. RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics* 120:1095–103
7. Botstein D, White RL, Skolnick M, Davis RW. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32:314–31
8. Breese EL, Mather K. 1957. The organization of polygenic activity within a chromosome in Drosophila. I. Hair characters. *Heredity* 11:373–95
9. Burr B, Burr FA. 1991. Recombinant inbreds for molecular mapping in maize: theoretical and practical considerations. *Trends Genet.* 7:55–60
10. Churchill GA, Giovannoni JJ, Tanksley SD. 1993. Pooled-sampling makes high-resolution mapping practical with DNA markers. *Proc. Natl. Acad. Sci. USA* 90:16–20
11. Coe EH, Hoisington DS, Neuffer MG. 1990. Linkage map of corn (maize) (*Zea mays* L.) (2N=20). In *Genetic Maps*, ed. SJ O'Brien, pp. 6.39–6.67. New York: Cold Spring Harbor Lab. Press
12. Copeland NG, Jenkins NA. 1991. Development and applications of a molecular genetic linkage map of the mouse genome. *Trends Genet.* 7:113–18
13. Crow JF. 1952. Dominance and overdominance. In *Heterosis*, ed. JW Gowen, pp. 282–97. Iowa State Coll. Press
14. Darling SM, Abbott CM. 1992. Mouse models of human single gene disorders I: non-transgenic mice. *BioEssays* 14: 359–66
15. Davisson MT, Lalley PA, Peters J, Doolittle DP, Hillyard AL, Searle AG. 1991. Report of the comparative committee for human, mouse and other rodents (HGM11). *Cytogenet Cell Genet.* 58:1152–89
16. DeVicente MC, Tanksley SD. 1993. QTL analysis of transgressive segregation in an interspecific tomato cross. *Genetics*. 134:585–96
17. Dietrich W, Katz H, Lincoln SE, Shin HS, Friedman J, et al. 1992. A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics* 131: 423–47
18. Doebley J, Stec A. 1991. Genetic analysis of the morphological differences between maize and teosinte. *Genetics* 129:285–95
19. Edwards MD, Helentjaris T, Wright S, Stuber CW. 1992. Molecular-marker-facilitated investigations of quantitative trait loci in maize. *Theor. Appl. Genet.* 83:765–74
20. Edwards MD, Stuber CW, Wendel JF. 1987. Molecular-marker-facilitated investigations of quantitative trait loci in maize. I. Numbers, genomic distribution and types of gene action. *Genetics* 116:113–25

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org
by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.

21. Falconer R. 1960. *Introduction to Quantitative Genetics*. New York: Ronald Press
22. Fatokun CA, Menancio-Hautea DI, Danesh D, Young ND. 1992. Evidence for orthologous seed weight genes in cowpea and mung bean based on RFLP mapping. *Genetics* 132:841–46
23. Giovannoni JJ, Wing RA, Ganal MW, Tanksley SD. 1991. Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. *Nucleic Acids Res*. 23:6553–58
24. Haldane JBS. 1946. The interaction of nature and nurture. *Ann. Eugen.* 13: 197–205
25. Hulbert SH, Richer TE, Axtell JD, Bennetzcn JL. 1990. Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes. *Proc. Natl. Acad. Sci. USA* 87:4251–55
26. Hunter RL, Markert CL. 1957. Histochemical demonstration of enzymes separated by zone electrophoresis in starch gels. *Science* 125:1294–95
27. Jacob HJ, Lindpaintner K, Lincoln SE, Kusumi K, Bunker RK, et al. 1991. Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypertensive rat. *Cell* 67: 213–24
28. Landc R. 1981. The minimum number of genes contributing to quantitative variation between and within populations. *Genetics* 99:541–53
29. Lander ES, Botstein D. 1986. Mapping complex genetic traits in humans: new methods using a complete RFLP linkage map. *Cold Spring Harbor Symp. Quant. Biol.* 51:49–62
30. Lander ES, Botstein D. 1986. Strategies for studying heterogeneous genetic traits in humans by using a linkage map of restriction fragment length polymorphisms. *Proc. Natl. Acad. Sci. USA* 83:7353–57
31. Lander ES, Botstein D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185–99
32. Law CN. 1967. The location of factors controlling a number of quantitative characters in wheat. *Genetics* 56:445–61
33. Lewontin RC, Hubby JL. 1966. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* 54:595–609
34. Martin B, Nienhuis J, King G, Schaefer A. 1989. Restriction fragment length polymorphisms associated with water use efficiency in tomato. *Science* 243: 1725–28
35. Martin GB, Williams JGK, Tanksley SD. 1991. Rapid identification of markers linked to a Pseudomonas resistance gene in tomato by using random primers and near-isogenic lines. *Proc. Natl. Acad Sci. USA* 88:2336–40
36. Mather K, Jinks JL. 1971. *Biometrical Genetics*. Ithaca, NY: Cornell Univ. Press
37. Michelmore RW, Paran I, Kesselli RV. 1991. Identification of markers linked to disease resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions using segregating populations. *Proc. Natl. Acad. Sci. USA* 88:9828–32
38. Moll RH, Cockerham CC, Stuber CW, Williams WP. 1978. Selection responses, genetic-environmental interactions, and heterosis with recurrent parent selection for yield in maize. *Crop Sci.* 18:641–45
39. Nodari RO, Tsai SM, Guzman P, Gilbertson RL, Gepts P. 1993. Toward and integrated linkage map of common bean. III. mapping genetic factors controlling host-bacteria interactions. *Genetics* 134:341–50
40. Park YC. 1997. Theory for the numbcr of genes affecting quantitative characters 2. Biases from drift, inequality of gene effects, linkage disequilibrium and epistasis. *Theor. Appl. Genet.* 50:163–72
41. Paterson AH, Damon S, Hewitt JD, Zamir D, Rabinowitch HD, et al. 1991. Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. *Genetics* 127:181–97
42. Paterson AH, Deverna JW, Lanini B, Tanksley SD. 1990. Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes in an interspecies cross of tomato. *Genetics* 124:735–42
43. Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD. 1988. Resolution of quantitative traits into Mendelian factors using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335: 721–26
44. Reiter RS, Williams JG, Feldman KA, Rafalsk JA, Tingey SV, Scolnik PA. 1992. Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random

amplified polymorphic DNAs. *Proc. Natl. Acad. Sci. USA* 89:1477–81
45. Rick CM, Smith PG. 1953. Novcl variation in tomato species hybrids. *Am. Nat.* 88:359–73
46. Rise ML, Frankel WN, Coffin JM, Seyfried TN. 1991. Genes for epilepsy mapped in the mouse. *Science* 253: 669–73
47. Robertson DS. 1985. A possible technique for isolating genomic DNA for quantitative traits in plants. *J. Theor. Biol.* 117:1–10
48. Sax K. 1923. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* 8:552–60
49. Serikawa T, Kuramoto T, Hilbert P, Mori M, Yamada J, et al. 1992. Rat gene mapping using PCR-analyzed microsatellites. *Genetics* 131:701–21
50. Smith HH. 1937. The relation between genes affecting size and color in certain species of Nicotiana. *Genetics* 22:361
51. Smithies O. 1955. Zone electrophoresis in starch gels. *Biochem. J.* 61:629
52. Stuber CW, Edwards MD, Wendel JF. 1987. Molecular-marker facilitated investigations of quantitative trait loci in maize. II. Factors influencing yield and its component traits. *Crop Sci.* 27:639–48
53. Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander ES. 1992. Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. *Genetics* 132:823–39
54. Tanksley SD, Bernatzky R, Lapitan NL, Prince JP. 1988. Conservation of gene repertoire but not gene order in papper and tomato. *Proc. Natl. Acad. Sci. USA* 85:6419–23
55. Tanksley SD, Medina-Filho H, Rick CM. 1982. Use of naturally-occurring enzyme variation to detect and map gene controlling quantitative traits in an interspecific backcross of tomato. *Heredity* 49:11–25
56. Tanksley SD, Ganal MW, Prince JP, DVicente MC, Bonierbale MW, et al. 1992. High density molecular linkage maps of the tomato and potato genomes. *Genetics* 132:1141–60
57. Taylor BA. 1976. Genetic analysis of susceptibility to isoniazid-induced seizures in mice. *Genetics* 83:373–77
58. Thoday JM. 1961. Location of polygenes. *Nature* 191:368–70
59. Thompson JN Jr. 1975. Quantitative variation and gene number. *Nature* 258:665–68
60. Thompson JN Jr, Thoday JM, eds. 1979. *Quantitative Genetic Variation.* New York: Academic
61. Vallejos CE, Sakiyama NS, Chase CD. 1992. A molecular marker-based linkage map of *Phaseolus vulgaris* L. *Genetics* 131:733–40
62. Vallejos CE, Tanksley SD. 1983. Segregation of isozyme markers and cold tolerance in an interspecific backcross of tomato. *Theor. Appl. Genet.* 66:241–47
63. Weber JL, May PE. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* 44:388–96
64. Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, et al. 1992. A second-generation linkage map of the human genome. *Nature* 359:794–801
65. Weller JI, Soller M, Brody T. 1988. Linkage analysis of quantitative traits in an interspecific cross of tomato (*Lycopersicon esculentum* x *Lycopersicon pimpinellifolium)* by means of genetic markers. *Genetics* 118:329–39
66. Whitkus R, Doebley J, Lee M. 1992. Comparative genome mapping of sorghum and maize. *Genetics* 132:1119–30
67. Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. 1990. Oligonucleotide primers of arbitrary sequence amplify DNA polymorphisms which are useful as genetic markers. *Nucleic Acids Res.* 18:6531–35
68. Winkelman DC, Hodgetts RB. 1992. RFLPs for somatotropic genes identify quantitative trait loci for growth in mice. *Genetics* 131:929
69. Womack JE. 1990. In *Mapping the Genomes of Agriculturally Important Animals,* ed. JE Womack, pp. 19–24. New York: Cold Spring Harbor Lab. Press

Annu. Rev. Genet. 1993.27:205-233. Downloaded from arjournals.annualreviews.org
by UNIVERSITY OF WARWICK on 02/23/10. For personal use only.