MAPPING QUANTITATIVE TRAIT LOCI IN PLANTS: USES AND CAVEATS FOR EVOLUTIONARY BIOLOGY

Rodney Mauricio

Gregor Mendel was either clever or lucky enough to study traits of simple inheritance in his pea plants; however, many plant characters of interest to modern geneticists are decidedly complex. Understanding the genetic basis of such complex, or quantitative, traits requires a combination of modern molecular genetic techniques and powerful statistical methods. These approaches have begun to give us insight into understanding the evolution of complex traits both in crops and in wild plants.

Department of Genetics, Life Sciences Building, University of Georgia, Athens, Georgia 30602-7223, USA. e-mail: mauricio@uga.edu

The world has recently secured the first complete genetic blueprint of a plant. The mouse-ear cress, *Arabidopsis thaliana*, a small mustard plant, has recently joined the increasingly less exclusive club of organisms for which every gene has been sequenced[1–6]. In the next three years, the complete sequence of a plant of worldwide agricultural importance, rice (*Oryza sativa*), will be available[7]. Undoubtedly, the dawning of the genomics age will have a great impact on agriculture, human health and molecular genetics. This wealth of genetic information is beginning to have just as profound an impact on the study of evolutionary biology, particularly on understanding one of the most enduring problems in evolution and molecular biology — the genetic basis of complex traits.

The complexity of these phenotypic traits, particularly of those involved in adaptation, probably arises from segregation of alleles at many interacting loci (quantitative trait loci, or QTL), the effects of which are sensitive to the environment[8–10]. Recent and continuing advances in molecular genetics and statistical techniques make it possible to identify the chromosomal regions where these QTL are located. Ultimately, an understanding of adaptive evolution will require detailed knowledge of the genetic changes that accompany evolutionary change.

A needle in a haystack

Modern quantitative genetics was born from the fusion of Mendelism and biometry, the mathematical theory that surrounds the science of heredity (BOX 1). The conceptual basis for the genetic dissection of complex traits is relatively straightforward. At its most basic level, QTL mapping simply involves finding an association between a genetic marker and a phenotype that one can measure (FIG. 1). For example, if all the tall plants among 500 individual corn plants of varying height have a particular allele of a genetic marker, then there is a very high probability that a QTL for plant tallness is associated with this marker in this population of plants.

QTL mapping in both plants and animals involves just a few basic steps[11]. The primary requirement is for two parental strains that have differences between them in the alleles that affect variation in a trait. The parents need not be different in the mean phenotypic value of the trait as different allelic combinations can yield the same phenotypic mean. A polymorphic genetic map allows the two strains to be distinguished genetically. The more detailed the linkage map (that is, the greater the number of markers), the better the mapping resolution. The parental alleles are then shuffled by creating a large mapping population, in which the phenotype and the multilocus genotype of each individual are measured.

Box 1 | Quantitative genetics reborn

"… evolution is essentially a statistical problem," W. F. R. Weldon (1893)

The early history of evolutionary genetics focused on understanding complex traits, particularly those relating to humans, such as intelligence, temper and 'artistic faculty'[99]. Without the benefit of Mendel's ideas, Francis Galton and the mathematician, Karl Pearson, established that useful predictions of the evolutionary trajectory of complex traits could be made without recourse to an explicit understanding of inheritance[99]. This area of 'quantitative genetics' rested on the statistical properties of the MULTIVARIATE NORMAL DISTRIBUTION[8,18].

With the rediscovery of Mendel's theory of heredity in 1900, a conflict arose between this 'biometrical' school of quantitative genetics and the discrete genetics of the Mendelian school[99]. By 1910, it had been shown that continuous phenotypic variation could result from the action of the environment on the segregation of many Mendelian loci[99,100]. By 1918, Ronald Fisher convincingly reconciled the discrete inheritance of Mendelism with the biometrical approach[101]. Modern evolutionary quantitative genetics is largely based on the same statistical foundations that were laid by Pearson and Fisher[102–108].

For most of the twentieth century, quantitative genetics had a crucial role in both agriculture and evolutionary biology, but never seemed fully embraced by modern molecular genetics. There was (and perhaps still is) a widespread perception that quantitative genetics essentially ignored genetics, blanketing actual genes in what has been called a "statistical fog"[109].

Mapping quantitative trait loci (QTL) — the single or many genes that underlie quantitative phenotypes — might represent an important part of the final synthesis of molecular and quantitative genetics. By working from the phenotype to the genotype, QTL mapping uses statistical techniques to localize chromosomal regions that might contain genes contributing to phenotypic variation in a complex trait of interest. Working from the gene to the phenotype, molecular geneticists might be able to meet quantitative geneticists at some genetic Promontory Point.

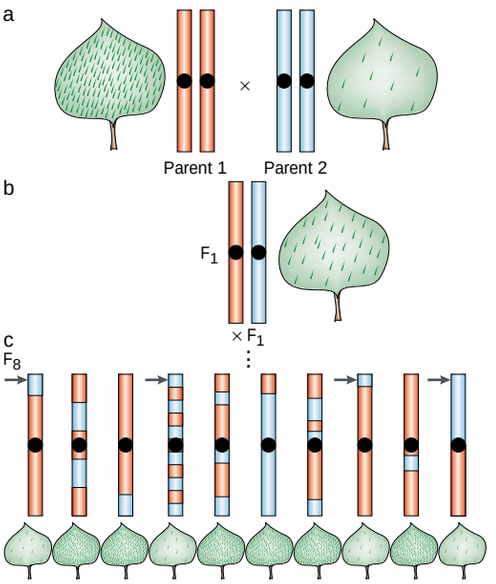


Figure 1 | **Principles of mapping quantitative trait loci.** The basic strategy behind mapping quantitative trait loci (QTL) is illustrated here for **a** | the density of hairs (trichomes) that occur on a plant leaf. Inbred parents that differ in the density of trichomes are crossed to form an F_1 population with intermediate trichome density. **b** | An F_1 individual is selfed to form a population of F_2 individuals. **c** | Each F_2 is selfed for six additional generations, ultimately forming several recombinant inbred lines (RILs). Each RIL is homozygous for a section of a parental chromosome. The RILs are scored for several genetic markers, as well as for the trichome density phenotype. In **c**, the arrow marks a section of chromosome that derives from the parent with low trichome density. The leaves of all individuals that have inherited that section of chromosome from the parent with low trichome density also have low trichome density, indicating that this chromosomal region probably contains a QTL for this trait.

In practice, several crossing schemes are used to generate this mapping population[12]. In all of these, the parents are mated to generate an F_1 population. In one approach, recombinant inbred lines can be created by selfing (self-fertilizing) each of the F_1 progeny for several (usually eight) generations (FIG. 1)[13]. In an 'F_2 design', the mapping population is generated by mating the F_1 progeny to each other (FIG. 2). In a 'backcross design', the mapping population is generated by crossing the F_1 progeny to either, or both, of the parents (FIG. 3). Several variations on these crossing schemes have been designed to maximize the shuffling of parental alleles[12,14]. Once all individuals in the mapping population are scored for phenotype and multilocus genotype, the actual QTL mapping can begin. The statistical tools at the foundations of QTL mapping have been used for many years (BOX 2). In fact, in 1923, Karl Sax mapped a QTL for seed size in the bean, *Phaseolus vulgaris*, by statistically associating it with a Mendelian locus for seed pigmentation[15].

Today, we generally have much more detailed genetic maps available. For example, *Arabidopsis thaliana* has 1,262 genetic markers, which consist of restriction fragment length polymorphisms (RFLPs) and single nucleotide polymorphisms (SNPs) that vary between the two parental lines of the most commonly used set of recombinant inbred lines (FIG. 1)[16,17]. Cereon Genomics has identified and made available a collection of 28,117 SNPs, which include 15,674 insertion/deletion polymorphisms that are polymorphic between the two parents of those same recombinant inbred lines.

Ultimately, QTL analysis yields a statistical description of the genes that underlie the phenotypes of interest. The 'statistical fog' is not completely lifted, but we can see the shadows of those genes.

MULTIVARIATE NORMAL DISTRIBUTION
The central limit theorem assures a normal (bell-shaped) distribution for a variable that is the summation of many independent, random inputs. This applies to single or multiple variables.

Figure 2 | **F_2 design: genetic mapping in monkeyflowers. a** | *Mimulus lewisii* and **c** | *Mimulus cardinalis* were crossed to produce **b** | a fertile F_1 progeny. The F_1 was self-pollinated to produce **d–l** | various F_2 progeny. The parent species differ in floral characteristics, including petal colour, COROLLA size and shape, presence or absence of NECTAR GUIDES, nectar volume, and concentration and position of ANTHERS and STIGMA. These floral characteristics are important in maintaining pollinator-induced reproductive isolation between these two species in the wild. (Adapted with permission from REF. 40 © (1999) National Academy of Sciences, USA.)

Genetics of evolution

QTL mapping is beginning to be embraced by evolutionary biologists as a means to answer various basic questions. Heritable phenotypic variation is the raw material of evolution, producing adaptations and organismal diversity. A comprehensive understanding of these evolutionary mechanisms will be enhanced by elucidating the molecular genetic basis of quantitative traits.

Several important evolutionary models are based on the key assumptions of quantitative genetics. These models assume that the complex phenotype of a trait is caused by the simultaneous segregation of a very large number of genes, each of which has a small, additive effect on the phenotype and interacts with other genes and with the environment[18]. QTL analysis allows us to test these assumptions. At one level, this is a 'straw man' — these assumptions cannot be entirely correct as there are only a finite number of genes available in a genome. (Conversely, one of the most important traits to an evolutionary biologist is 'fitness' — a very complex trait that is almost certainly controlled by many genes). On a practical level, once five or more loci contribute to a trait, it might be easier to model the evolution (or perhaps even function) of that trait with a mathematical abstraction rather than attempting to explicitly take into account each individual locus. Nevertheless, QTL mapping might indicate to evolutionary biologists the direction in which to proceed.

Genetics of adaptation. One of the most enduring controversies in evolutionary biology is the genetic basis of adaptation[19,20]. R. A. Fisher concluded that mutations in genes of very small effect were responsible for adaptive evolution[21]. H. Allen Orr and Jerry Coyne stated that "… the neo-Darwinian view has … triumphed, and the genetic basis of adaptation now receives little attention. Indeed, the question is considered so dead that few know the evidence responsible for its demise"[20]. Orr and Coyne re-examined the evidence for this Fisherian view and argued that both the theoretical and empirical basis for it were weak[20]. They encouraged evolutionary biologists to re-examine this research question by the 'genetic analysis of adaptive differences between natural populations or species'[20].

Plant evolutionary biologists have embraced this challenge and there are many examples of the use of QTL analysis to determine the genetic basis of traits that are thought to be of adaptive value. Several QTL have been identified for seed size, fruit size and seed number in *A. thaliana*[22]. Seed size is a very important adaptive trait in that large seeds often have higher fitness than small seeds[23]. Variation in seed size in *A. thaliana* was found to be determined by 11 QTL with relatively small additive effects. Seven of the seed size QTL localized to the same position as seed weight QTL, indicating possible pleiotropic actions of the underlying genes. So, QTL analysis can provide information about the mode of gene action.

QTL analysis has also been used to understand the genetic correlations among traits in natural populations. Floral traits, such as petal length and width, have long been used as an example of positively, genetically correlated traits[24]. In a study of floral morphology in *A. thaliana*, 18 QTL were found[25]. As in the study of seed size, 11 floral trait QTL were associated with more than one floral trait. So, the tight morphological integration of the flower is possibly due to pleiotropic gene action, or to tight linkage of loci[25].

The timing of flowering is another trait of importance to plants, as earlier reproduction often translates into higher fitness[26]. Three research groups have independently examined flowering time in *A. thaliana* using three unique crosses. A cross between the late-flowering strain, HM, and the early-flowering strain, WS, revealed that two unlinked QTL affect flowering

COROLLA
Collectively, the petals of a flower.

NECTAR GUIDES
Markings on the petals of flowers, often in contrasting colours or visible only in ultraviolet wavelengths, thought to act as directional beacons for pollinators, especially bees.

ANTHER
The pollen-bearing part of the male floral structure (stamen).

Figure 3 | **Backcross design: genetic mapping in the Louisiana irises. a** | *Iris fulva* (swamp iris) and **d** | *Iris brevicaulis* (blue iris) were crossed to produce an F_1. An F_1 backcrossed to *Iris brevicaulis* produced hybrids such as **b** |, and an F_1 backcrossed to *Iris fulva* produced hybrids such as **c** |. The parental species differ in floral characteristics, including petal and sepal colour, presence or absence of nectar guides, degree of petal reflexivity (petals of *I. fulva* curl down and inward, whereas those *of I. brevicaulis* angle upward), and anther position (the anthers of *I. fulva* extend out of the STYLAR TUBE, those of *I. brevicaulis* are inside the stylar tube). Backcross genotypes show a wide range of variation in these traits. These floral characteristics, as well as others involved in the reproductive isolation of these species, are now being analysed using quantitative trait loci mapping methods. (Courtesy of Amy Bouck, Department of Genetics, University of Georgia, USA.)

STIGMA
The region, usually the apex, of the gynoecium that receives pollen grains and on which the pollen germinates. The gynoecium is the seed-bearing organ of flowering plants, consisting of the stigma, style and ovary.

STYLAR TUBE
In the genus *Iris*, the styles (tubular columns of tissue arising from the top of the ovary) look like flower petals and are tightly appressed to the top of the actual petals, forming this tube between the petal and the style.

PHENOLOGY
The timing of periodic biological phenomena that are usually correlated with climactic conditions.

HYBRID ZONE
A region of reproduction between individuals of different species, usually occurring where the ranges of the species come together.

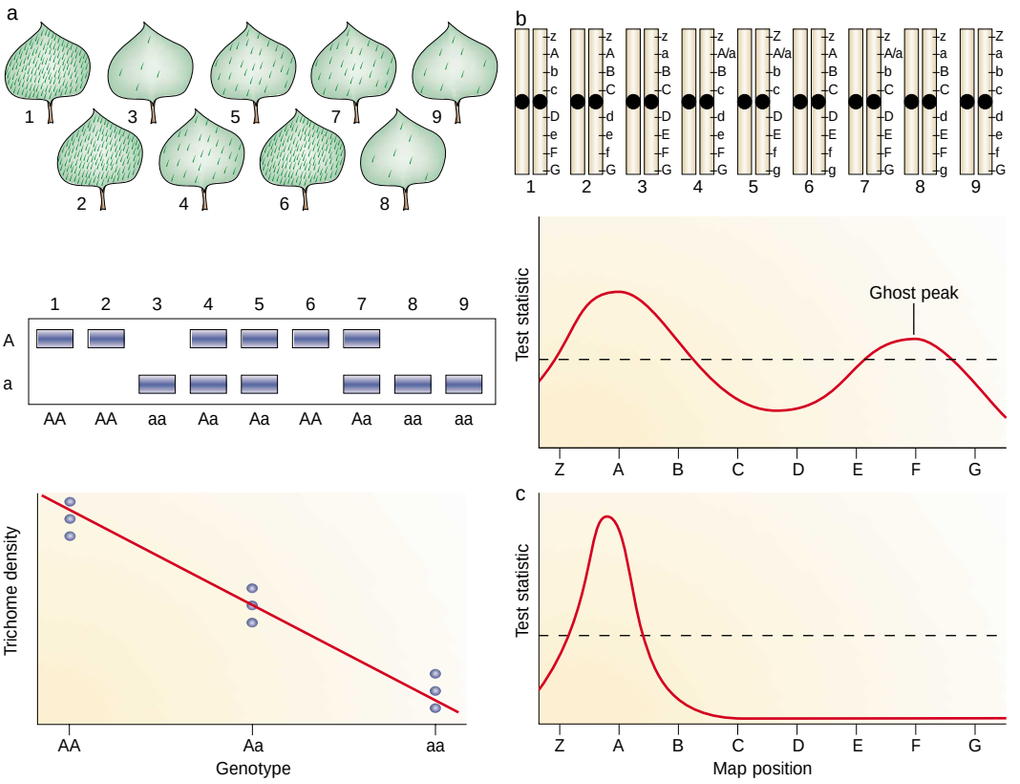
time on chromosome 5 (REF. 27). In a cross between the Columbia and Landsberg strains, two QTL, one on chromosome 1 and another on chromosome 2, were found[28]. Finally, a cross between the early-flowering strain, Li-5, and the late-flowering strain, Naantali, revealed at least seven QTL, with one QTL on chromosome 4 accounting for over 50% of the variation in flowering time[29]. The other 'minor' QTL were found on each of the five chromosomes of *A. thaliana*[29]. As this example illustrates, the QTL found for any particular character might be different depending on the parents used in the original cross.

Although *A. thaliana* has been a popular plant for studying the genetics of adaptation using QTL analysis, the power of the approach is best illustrated by considering studies of plants that are decidedly not annual weeds. QTL analysis has been successfully applied to studying the genetic basis of various important traits in forest trees, including pine, eucalyptus and poplar. QTL have been mapped for seedling height, leaf area and frost tolerance in the high altitude *Eucalyptus nitens*[30,31] and for frost hardiness and timing of bud set in natural populations of Scots pine (*Pinus sylvestris*)[32]. In studies of poplars, cottonwoods and aspens (*Populus* spp.), researchers have found QTL for bud set, bud flush, growth, form, PHENOLOGY and leaf shape[33,34].

Genetics of speciation. Extending QTL analysis to other plant systems is particularly important in addressing a second fundamental question in evolutionary biology: Are the genes that are variable in a population the same as those that cause divergence between populations and species?[35] As in the case of adaptive traits, there has been some controversy as to the role of the principal genes in speciation[36,37]. Recent QTL analyses use crosses between (necessarily) closely related species. By creating such hybrids, or by looking in natural HYBRID ZONES, the origin of species can be investigated at the genetic level.

Toby Bradshaw, Douglas Schemske and their colleagues pioneered this approach with their study of speciation in monkeyflowers (*Mimulus*)[38–40]. They crossed two species with contrasting floral traits. *Mimulus lewisii* is a bumble-bee-pollinated flower with pink petals, contrasting yellow nectar guides, a wide corolla opening, concentrated nectar, and inserted anthers and stigma (FIG. 2a). *Mimulus cardinalis* is pollinated by hummingbirds and has red petals, a narrow, tubular corolla, copious nectar and exserted anthers and stigma (FIG. 2c)[38]. The two species grow and flower together, but hybrids are not commonly observed in nature. However, they are completely interfertile when artificially mated (FIG. 2b). So, these two species are reproductively isolated owing to different preferences of the pollina-

Box 2 | Quantitative trait loci mapping methods



Several techniques exist for mapping quantitative trait loci (QTL) in the population; I have used the example of trichome density in leaves described in the figure to illustrate these methods. **a** | In the regression technique, the phenotype is correlated with each marker genotype[110] (the middle panel represents the differential migration of DNA on a gel). In this case, a single marker 'A' is scored. Individuals that are homozygous for the A allele have high trichome density, individuals that are homozygous for the 'a' allele have low trichome density, and heterozygotes have intermediate trichome density. A linear regression of trichome density on the number of A alleles shows a significant relationship between the marker and the phenotype, which indicates that a QTL for trichome density is probably linked to that marker. The simple regression method described is of limited use in localizing the chromosomal segment that contains a QTL. The method underestimates the effect of the QTL. The further the QTL is from the marker, the weaker the effect.

The interval mapping method uses a pair or two pairs of flanking markers at a time[83,111]. **b** | In this approach, the QTL is located within a chromosomal interval, defined by the flanking markers. The technique involves scoring a large number of markers, as illustrated in the top panel, and then assessing the probability that an interval between two markers is associated with a QTL that affects the trait of interest. The results of the analysis are plotted as a LIKELIHOOD-RATIO TEST STATISTIC against the chromosomal map position, measured in recombination units (Morgans). The dotted line represents a significance threshold above which a likelihood-ratio test provides a statistically significant fit to a model of the data. The best estimate of the location of the QTL is given by the chromosomal location that corresponds to the highest significant likelihood ratio. Although the interval mapping method was an important advance, it too is statistically biased. In particular, QTL outside the interval under consideration can affect the ability to find a QTL within it[112]. In addition, false identification of a QTL can arise if other QTL are linked to the interval of interest (the false 'ghost peak' on the right)[112]. **c** | A third method, known as composite interval mapping, combines the interval mapping technique with multiple regression analysis[81,85,112]. Composite interval mapping assesses the probability that an interval between two markers is associated with a QTL that affects the trait of interest, as well as controlling for the effects of other markers on the trait (thus providing more accurate results). The results of the analysis are plotted as in **b**. In this method, the test statistic is independent of QTL in other regions of the chromosomes (although it is still biased if the QTL is in the interval immediately adjacent to the interval of interest)[85].

Zhao-Bang Zeng and his colleagues have extended this method to a multiple interval mapping technique. Multiple interval mapping combines multiple QTL mapping analysis with the analysis of genetic architecture by using an algorithm to search for number, positions, effects and interactions of significant QTL[113,114].

Several computer programs that carry out these mapping methods are freely available[12,115]. Unfortunately, most of these programs lack an easy-to-use interface and are not of commercial quality, something desperately needed for most users.

LIKELIHOOD-RATIO TEST STATISTIC
A maximum-likelihood method of hypothesis testing. The likelihood-ratio test statistic is twice the natural logarithm of the ratio of the maximum likelihood that the data fit the alternative hypothesis to the maximum likelihood that the data fit the null hypothesis.

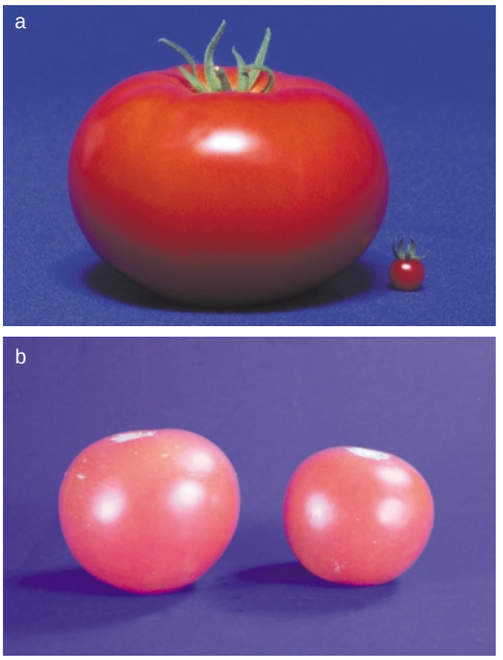


Figure 4 | **Genetic basis of phenotypic variation in fruit size in the tomato. a** | Fruits from the wild tomato species *Lycopersicon pimpinellifolium* (right) and the cultivated tomato 'giant red' (*Lycopersicon esculentum*). Reproduced with permission from *Science* **289**, 85–88 © (2000) The American association for the Advancement of Science. **b** | Phenotypic effect of a cosmid that contains the small-fruit allele of the *fw2.2* transgene in the Mogeor cultivar of *L. esculentum*. The fruit on the left is a tomato from the Mogeor line. When the same line carries a partially recessive large-fruit allele of *fw2.2* on a transgene, the weight of the fruit is reduced (right) on average by 15.5 g (REF. 54). (Courtesy of Steven Tanksley, Department of Plant Pathology, Cornell University, USA.)

tors. QTL for the floral traits that distinguish the monkeyflower species (and, presumably, pollinator preference) are therefore candidates for 'speciation' genes[38]. For each of eight floral traits studied, Bradshaw and colleagues found at least one QTL accounting for more than 25% of the phenotypic VARIANCE in floral morphology, and concluded that the evolution of reproductive isolation might involve genes of major effect[38]. It is still too early to conclude that 'speciation' genes will be commonly found in studies of reproductive isolation. Several similar studies are being conducted, including work on another pair of SYMPATRIC *Mimulus* species, *M. guttatus* and *M. nasutus*[41], on the natural hybrids of Louisiana irises (FIG. 3)[42,43] and on the floral nectar spurs of sympatric columbines[44].

A final example of the use of QTL analysis in studying MACROEVOLUTION is found in continuing work on sunflowers[45–47] and the Louisiana irises (FIG. 3)[42,43,48]. In both systems, natural hybrid populations exist in which the parents are adapted to different habitats. It has been suggested that hybrids can serve as a genetic bridge between isolated species, shuttling adaptive genes across the hybrid zone. Loren Rieseberg and his colleagues are studying the common weed sunflower, *Helianthus annuus*, which is postulated to have colonized Texas by acquiring advantageous alleles from the locally adapted, *Helianthus debilis*[45–47]. Michael Arnold and his colleagues are studying hybrids of a cross of *Iris brevicaulis* (FIG. 3d), which prefers dry, sunny habitats, and *Iris fulva* (FIG. 3a), which prefers wet, shady habitats[43,48]. By mapping QTL in the parents that control the ecological traits that are responsible for habitat preference and by examining hybrids, it might be possible to show explicitly that INTROGRESSIVE HYBRIDIZATION is an important evolutionary mechanism in plants.

Field of dreams

QTL mapping has enormous practical potential for agriculture. Although classical quantitative genetics and selective breeding have been enormously successful in agriculture, the process of breeding a crop plant with a particular desirable trait is a combination of luck, hard work, time and money. A detailed knowledge of the genes that affect traits such as yield, DEHISCENCE and resistance, could facilitate and speed the development of improved crops[49,50].

Gardeners accustomed to planting varieties of tomato know that they can have extraordinarily variable fruit sizes, ranging from a weight of a few grams to as much as half a kilogram (FIG. 4a). Examples of these varieties are Burpee's 'Big Boy', 'Big Girl' and 'Watermelon Beefsteak' at one extreme, and the many 'grape', 'currant' and 'cherry' varieties at the other extreme. Steven Tanksley and his colleagues used seven wild species of tomato and seven different crossing designs to locate QTL for fruit size and weight[51]. This work, starting in the 1980s has identified at least 28 different QTL for fruit weight. Some of these QTL have a significant effect on the phenotype, explaining over 20% of the phenotypic variance[51].

Over the past ten years, finer-scale mapping techniques were used to localize one of these important QTL, *fw2.2*, to a narrow chromosomal region[52,53]. Recently, the statistical fog was pierced by the cloning of the *fw2.2* QTL[54]. The *fw2.2* QTL corresponds to a single open reading frame in a transformed cosmid (called *ORFX*) that is expressed in the floral organs. When the wild-type allele of the gene was introduced into a cultivated tomato, the transformed plants produced small fruit of roughly the expected weight (FIG. 4b). This transformation experiment marks the movement of QTL mapping beyond a purely statistical association between a gene and the phenotype.

Beyond the obvious importance of QTL mapping for agriculture, some of the most striking examples of the use of QTL analysis to answer fundamental evolutionary questions come from the study of agricultural species. An evolutionary approach to QTL mapping has allowed new and powerful insights into the understanding of plant domestication. The work of John Doebley and colleagues on the evolution of maize (*Zea mays* ssp. *mays*) from its probable wild ancestor, teosinte (*Zea mays* ssp. *parviglumis*), is a landmark in the field of evolutionary QTL analysis[55–58]. For the past decade, QTL that are responsible for key steps in the

VARIANCE
A statistic that quantifies the dispersion of data about the mean. In quantitative genetics, the phenotypic variance (*V*p) is the observed variation of a trait in a population. *V*p can be partitioned into components, owing to genetic variance (*V*g), environmental variance (*V*e) and gene-by-environment correlations and interactions.

SYMPATRIC
Occurring in the same area without loss of identity from interbreeding.

MACROEVOLUTION
Evolution at or above the level of species.

INTROGRESSIVE HYBRIDIZATION
Incorporation of genes from one species into the gene pool of another species.

DEHISCENCE
The splitting open of a fruit.

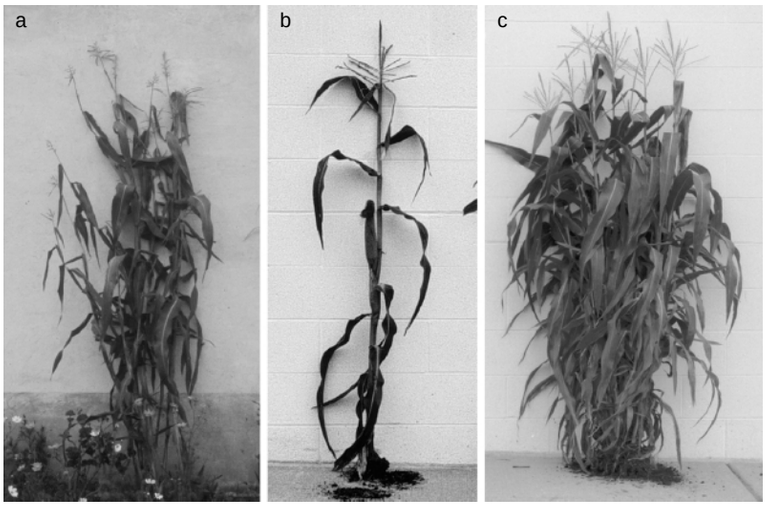


Figure 5 | **The evolution of apical dominance in maize (*Zea mays*). a** | The maize crops that are cultivated today are probably a domesticated form of the wild Mexican grass, teosinte. Note the bushy form of the teosinte, *Zea mays* ssp. *mexicana*, shown here. **b** | The single-stalk branching pattern of wild-type maize (Inbred A158). **c** | A maize plant that is mutant for the *teosinte branched 1* (*tb1*) gene. The *tb1* locus is likely to have had an important role in the evolution of maize plant architecture. (Reprinted with permission from REF. 61 © (1997) Macmillan Magazines Ltd.)

evolution of maize have been identified[59]. These include QTL for the distinct change in branching from the bushy, multi-stemmed teosinte to the single-stemmed maize varieties used in agriculture today (FIG. 5), and QTL for the change in fruit architecture between teosinte, with its encased kernels, to corn with its kernels exposed on the ear[58,60]. The actual gene that underlies the distinct morphological change in branching pattern, called *teosinte branched 1*, has been identified (FIG. 5)[61]. Furthermore, *teosinte branched 1* has been subjected to a molecular population genetic analysis to understand the evolutionary dynamics of the locus[62].

An evolutionary approach to QTL mapping can extend beyond single species, yielding more insight into the evolution of plant domestication. Diverse taxa in common taxonomic groups often share gene order over large chromosomal segments. 'Comparative QTL mapping' is possible because chromosomes of these different taxa can be aligned on the basis of common reference loci. The grass species sorghum, rice and maize, were each independently domesticated ~10,000 years ago. Each species has been selected to have large seeds, daylength-insensitive flowering and reduced fruit shattering. A small number of QTL were located for these traits[63]. Interestingly, the approximate location of the QTL for each trait mapped to roughly corresponding locations in each of the three species, despite 65 million years of reproductive isolation between these species[63]. This conservation of QTL location indicates that, 10,000 years ago, ancient farmers across three continents might have been independently selecting many of the same genes to obtain convergent phenotypes in each of the three species[63].

Finally, QTL analysis has been brought to bear on a biological problem that has stymied geneticists and evolutionary biologists for many years. Many flowering plant genomes are thought to be the product of one or more polyploidization events[64,65]. As such, many of the world's crop plants are polyploids, resulting either from complete duplication of their own genomes or from interbreeding with another species with a failure of complete disjunction at meiosis[66]. One of the most interesting applications of QTL mapping to evolutionary biology is the exploration of the joining of these genomes[67].

Andrew Paterson and his colleagues have been exploring such a polyploid event in species of domesticated cotton (*Gossypium*)[68,69]. Present-day varieties of cultivated cotton are tetraploid, but are derived from two distinct diploid parental species[68]. The most important trait to a cotton farmer is the length and quality of the seed fibre. Interestingly, the QTL that contribute to fibre quality in domesticated cotton derive from the diploid parent species that possesses no spinnable fibre on its seeds. This indicates a possible non-additive interaction between the two parental genomes with respect to seed fibre quality (BOX 3). This application of QTL analysis shows that the merger of genomes of divergent evolutionary histories can produce 'unique avenues' for selection[68,69].

A step off the bandwagon

The use of QTL mapping represents a rebirth for quantitative genetics. Quantitative geneticists have been successful at both developing robust evolutionary theory and providing practical benefits to agriculture, despite treating genes as a black box. QTL mapping represents a promising link between this statistical approach and an explicit understanding of the molecular basis of variation in complex traits. Although I have highlighted examples of the power and promise of QTL mapping for evolutionary biology, there are numerous and important caveats to keep in mind.

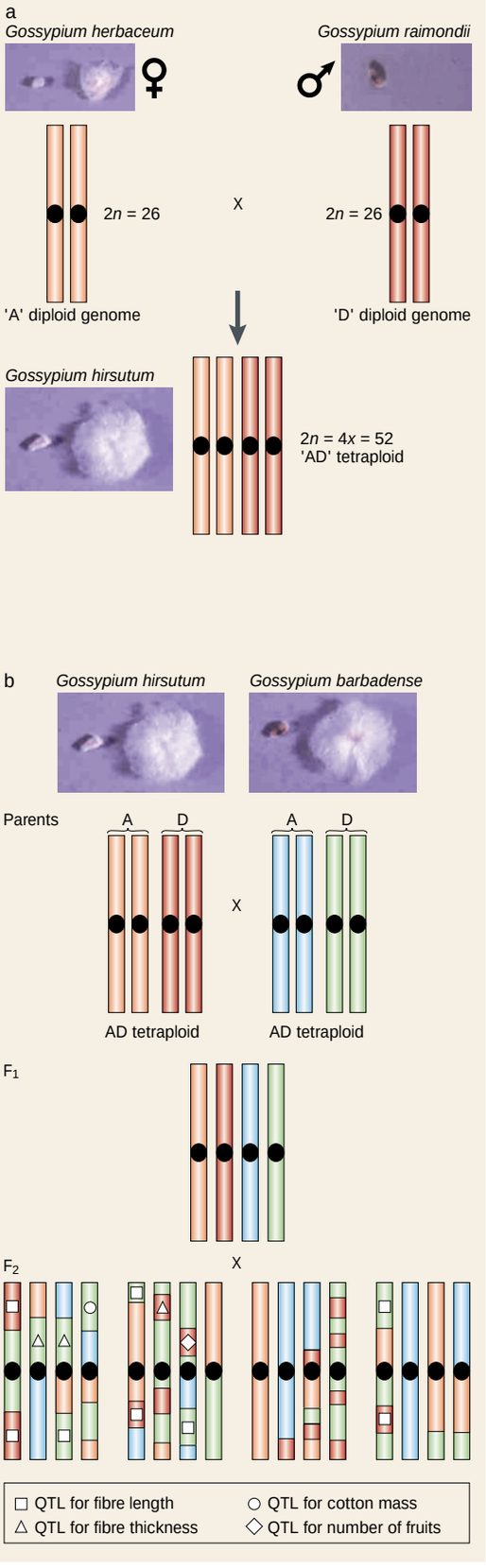
Maps and markers. Genetic maps are time consuming and expensive to construct. Despite this, the number of markers is rarely the limiting factor in QTL mapping experiments; the number of progeny examined is, however, crucial[70,71]. Each progeny represents an opportunity to identify a unique recombination event between markers. Any QTL that exists between two completely linked markers will segregate with both markers and will be indistinguishable from either. Because recombination rates vary across the genome[72], some QTL will be harder or easier to detect depending on their genetic location. For example, the centromeric regions, which are known to be suppressed for recombination[72] and to be poor for genetic variation[73], are a veritable black hole for QTL mapping: even two markers that are physically far apart will seem to be genetically close together near the centromere. Localization of a gene in these regions can be difficult.

Box 3 | Proposed formation of cultivated polyploid cotton (*Gossypium*)

Most of the world's cotton crop is made up of two tetraploid ($2n = 4x = 52$) species, *Gossypium hirsutum* ('Upland' cotton) and *Gossypium barbadense* ('Pima', 'Sea Island' or 'Egyptian' cotton)[68]. These species have been cultivated to produce long, spinnable fibres on their seeds: in each panel, the seed, with fibre removed, is shown on the left and the fibre from that one seed, if any, is shown on the right. In the figure, **a** | both tetraploid species are thought to have arisen by the hybridization of two diploid ancestors: a maternal Old World diploid species (denoted the 'A' genome) and a paternal New World diploid species (called the 'D' genome)[68]. So, the resulting tetraploid has an 'AD' genome.

The possible 'D' genome parents of the tetraploid are two, extant, neotropical species, *Gossypium raimondii* or a sister species of *Gossypium gossypioides* ($2n = 26$). The possible 'A' genome ancestors are two extant Old World species, *Gossypium arboreum* or *Gossypium herbaceum* ($2n = 26$). Interestingly, when one looks at the most distinctive and important trait to a cotton farmer, fibre, in both wild and domesticated cotton species, only the 'A' genome diploid and the 'AD' tetraploid taxa produce seeds that are covered in long, spinnable fibres (**a**)[68]. In Asia, a domesticated 'A' genome diploid is still bred and cultivated for its fibre[68]. However, although wild 'D' genome diploid species produce hairy seeds, none produce spinnable fibre and none has ever been successfully domesticated for fibre production[68]. Of course, humans have successfully domesticated and selected 'AD' tetraploids for high yield and quality of fibre. Therefore, the interaction of the 'A' and 'D' genomes in the tetraploid domesticated species produces higher quality and higher quantity fibre than is found in either of the diploid ancestors, even the currently domesticated 'A' genome diploids.

In **b**, an F_2 mapping population was created that was derived from a cross of the two species of cultivated 'AD' tetraploids for which they had developed detailed genetic maps[116]. These two species have very different seed fibres, and quantitative trait loci (QTL) that distinguished the parental types for various fibre characteristics were sought[68]. Remarkably, most of the QTL that influenced fibre quality and yield were located on the portion of the genome contributed by the 'D' genome (hypothetical data illustrated)[68]. Remember that the 'D' portion of the genome is derived from an ancestor that has no spinnable fibre. So, most of the genetic variation available for improvement of cultivated cotton fibres apparently comes from the parent without fibre. Perhaps thousands of years of selection on 'A' genome diploids exhausted genetic variation for fibre quality, but that selection was absent from the 'D' genome diploids because their seed fibres were not desirable to ancient farmers. (Images courtesy of Andrew Paterson, Applied Genetic Technology Center, Departments of Crop and Soil Sciences, Botany, and Genetics, University of Georgia, USA, and Thea Wilkins, Department of Agronomy and Range Science, University of California, Davis, USA.)



Moving from QTL to genes. A QTL is almost never an actual genetic locus. A QTL is a chromosomal segment, potentially encompassing many hundreds of individual loci, most of which have nothing to do with the phenotypic trait of interest. An actual locus that contributes to a phenotype is a veritable needle in a haystack of QTL. Conversely, a QTL might contain many genes that contribute to the phenotype of interest.

Although there have been examples of QTL mapping that yield an actual locus, these examples are rare[54,61]. Remember that it took more than ten years to winnow a single fruit size QTL to the actual gene. There are 28 known fruit size QTL in tomato[51]. Although it is doubtful that it will take 270 more years to find all the genes that influence fruit size in tomato, finding those genes will be time consuming and expensive. In the near future, only in those organisms for which genetic information is abundant will we be able to find the actual genes that underlie the phenotypes of interest.

Even in model organisms, the ability to move from QTL to gene will not be easy. In *A. thaliana*, the estimated genetic map is 586 centiMorgans (cM) and the physical size is ~125 megabases. On average, there are 213 kilobases of DNA and ~50 genes per cM in *A. thaliana*[1,72]. Even in the best QTL studies, many QTL are defined by markers that are more than 10 cM apart. Sorting through 500 genes requires time and money. Even if a genome project has identified each of the genes in that interval, proving that any particular gene is responsible for variation in a trait of interest is not a trivial exercise.

Association studies hold some promise in assessing the correlations between specific genetic variants (usually SNPs) and trait differences on a population level[74,75]. The most commonly used approach searches for differences in allele frequency between individuals with a particular phenotype and unrelated control individuals. However, many statistical caveats accompany such studies and they have, to date, been plagued by numerous spurious (false-positive) correlations[74,76].

Along similar lines, a 'candidate gene' approach might help in linking QTL with particular genes. In this approach, a gene known to be in a particular pathway, or have a predicted function, can be related to genes already known to have specific phenotypic effects, and will be considered to be a gene correlated to a QTL. As many genes will probably be included within the chromosomal boundaries of the QTL, it will still be difficult to provide convincing molecular evidence, such as complementation, that a candidate gene is the locus that contributes to the trait under study.

Those biologists working with less genetically endowed organisms might be able to lever the genetic information from model organisms by taking advantage of homology. In this way, a 'reverse quantitative genetics' approach could be fruitful in that one could ask how much phenotypic variation in a non-model organism is explained by the homologue of a gene with a similar phenotype in a model organism.

Experimental and statistical concerns. QTL mapping is a statistical approach and evolutionary biologists must be aware of the inherent limitations and biases of the statistical procedures themselves[11,77]. For example, QTL analyses assume that the distribution of trait values are normally distributed. Important experimental considerations that are involved in implementing these statistical tests include the heritability of the trait being mapped, the precision with which the trait can be measured and the size of the mapping population[11]. Shortcomings in any of these areas can undermine the accuracy and power of the QTL analysis.

The limits of QTL detection are determined by several factors, including recombination, the number of progeny in the mapping population and the number of markers[70,71]. QTL mapping always underestimates the number of genes that are involved in controlling a trait because only genes of sufficiently large phenotypic effect will be detectable as QTL[70,71,78,79]. Imagine staring out at the African savanna with very poor binoculars — you will see the elephants (QTL of large effect), but perhaps not the gazelles (QTL of small effect). With a progeny population of fewer than 500 individuals, regardless of marker density, there is little statistical power to identify QTL of small effect[71].

William Beavis summarized the results of several QTL mapping experiments on yield and height of maize, including replicate studies of the same crosses[70,71]. Although the same QTL were detected across studies, some of those detected were unique to each cross. Even the replicate studies did not detect the same QTL.

Another form of sampling bias, the use of only a few divergent parental lines in a minimum number of environmental conditions, can lead to the underestimation of the number of genes and their effects. For example, Paterson and colleagues did a mapping experiment for tomato in three environments. Of the 29 QTL detected, 4 were detected in all three environments, 10 in two environments and 15 in only one environment[80]. Clearly, the environment will be shown to have an important role in QTL mapping. At one level, the environment might complicate our efforts to map QTL, but understanding the interaction of QTL with the environment will be crucial to our understanding of gene function and evolution.

It is also possible that the QTL of large phenotypic effect that we see are an artefact of the strong directional selection often used to create the phenotypically divergent parental lines that are used for mapping[19]. Strong selection can fix alleles that normally segregate in the base population. In addition, artificial selection might create repeated bottlenecks through which only a sample of segregating alleles can pass. Only segregating alleles can be detected. So, fewer QTL will be detected and those that are eventually detected might explain an inflated portion of the phenotypic variance.

In addition to the estimation of the number of QTL, the magnitude of QTL effects might also be biased by small sample sizes[70,71]. In his study on QTL experiments in maize, Beavis showed that in all studies, one or a few

QTL of large effect were identified, along with several QTL of small effect. This distribution was more skewed in experiments that used small numbers of progeny. The fewer the progeny, the higher were the estimated effects of the largest QTL identified.

Finally, the issue of significance testing is still incompletely resolved. The statistical tests for assessing if a QTL actually exists are many and not independent. So, QTL mapping will yield a significant number of false QTL (ghost peaks). One commonly used solution to this problem is to use a conservative threshold value to reduce the probability of false positives[81,82]. For human data, it has been estimated that the threshold value should be a LOD SCORE of 3.3 (REF. 82). This value indicates that the probability that a QTL occurs in a particular interval is over 1,000 times more likely than the null hypothesis that no QTL exists in the interval. MONTE CARLO SIMULATIONS[83–85] and PERMUTATION TESTS[86,87] are two other approaches that are used to explicitly determine if a QTL is significant. BAYESIAN APPROACHES to QTL mapping are being introduced[32,88].

Future challenges

Given the caveats described above, knocking down the straw man of quantitative genetics (many genes of small, additive effect) might be more difficult than initial efforts have led us to believe. After all, how much phenotypic variation does a QTL have to explain before we call it a QTL of 'large' effect? Does the quantitative genetic model really predict that we will find no single QTL that explains any significant amount of the total phenotypic variation? Given the numerous caveats that apply to QTL mapping studies, the number of genes estimated by QTL mapping should be viewed as a hypothesis of genetic architecture. Furthermore, it is crucial that evolutionary biologists define their questions with the caveat that they might never actually find genes. In agriculture, having a QTL might be enough to serve in a marker-assisted selection programme. In how many evolutionary studies will knowing only a relatively large chromosomal region be informative? The challenge for evolutionary biologists will be to think carefully about how understanding the genetic basis of complex traits will inform their studies, especially if those conclusions rest on knowing the actual genes underlying the trait of interest. When is identifying a large chromosomal segment interesting enough to justify an expensive and time-consuming hunt for QTL?

Children of the corn

At some level, all geneticists, from molecular geneticists to population geneticists, are interested in finding the connection between the gene and the phenotype. Understanding this connection is most difficult for the cases of complex traits, such as most human diseases and many examples of adaptive evolution. QTL mapping holds some promise in helping us to make this connection. With the advent of genomics, geneticists see a path through the fog and there is growing awareness that an understanding of human disease will require an understanding of quantitative traits. Genomics will greatly assist in providing numerous markers and more complete maps. Genomic techniques might also make it possible to create larger progeny arrays as the cost of genotyping will probably decrease markedly.

Theoretically, this QTL mapping approach is as applicable to animals, including humans, as it is to plants. However, there are some significant advantages to studying complex traits in plants. Plants are easy to replicate and one can generate several parental lines and large progeny arrays. The variable that is often limiting in QTL studies is the cost of generating and maintaining large numbers of progeny. For humans, the number of progeny is necessarily small. Generating inbred lines in plants is generally possible. Because animals tend to be outcrossing, inbreeding can be a problem.

QTL mapping in populations or species of outbreeding organisms faces additional statistical and biological challenges. In these cases, the parents used in the mapping cross, either from controlled crosses or natural populations, are not necessarily fixed for alternate alleles, as is the case with inbred parental lines. The parents might be polymorphic for segregating alleles. Several experimental and statistical solutions to these challenges have been suggested for outbred species[89–92]. Statistical approaches using pedigrees are being developed[93,94] that should be applicable to humans. In highly heterozygous organisms, QTL mapping can be done in the F_1 generation itself, on the basis of SIMPLEX SEGREGATION of polymorphic markers[95,96]. In addition, new approaches have recently been presented for QTL mapping in polyploids and even in hybrid zones[97,98].

Finally, in a plant system, one can easily assess QTL in several realistic ecological conditions. Parents can be taken out of the field and offspring can be grown back in the field. If our ultimate goal is to understand how genes form complex phenotypes, we must come to realize that the environment has a crucial role. Replicated microarray studies that can simultaneously assess genome-wide gene expression and can be used on field experiments might eventually be a tool that geneticists and ecologists find invaluable. Understanding how the environment interacts with genes to yield phenotypes might be the most significant challenge to all geneticists.

LOD SCORE
(Base 10 'logarithm of the odds', or 'log-odds'.) A method of hypothesis testing. The logarithm of the ratio between likelihoods under the null and alternative hypotheses.

MONTE CARLO SIMULATION
The use of randomly generated or sampled data and computer simulations to obtain approximate solutions to complex mathematical and statistical problems.

PERMUTATION TEST
A method of hypothesis testing. In these tests, an empirical distribution of a test statistic is obtained by permuting the original sample many times. Each permuted sample is considered to be a sample of the population under the null hypothesis.

BAYESIAN APPROACH
An alternative statistical method that allows the use of prior information to evaluate the posterior probabilities of different hypotheses.

SIMPLEX SEGREGATION
Segregation in polyploids. Segregation with no crossovers of the simplex genotype Aaaa would result in a gametic ratio of 1/2 Aa and 1/2 aa.

Links

FURTHER INFORMATION *Arabidopsis thaliana* | rice | Mendelism | R. A. Fisher | Francis Galton | Karl Pearson | *Arabidopsis* Biological Resource Centre (ABRC) | Cereon Genomics *Arabidopsis* SNP collection | Nottingham *Arabidopsis* Stock Centre: Columbia × Landsberg RI lines | Rockefeller University's collection of genetic analysis software | The *Arabidopsis* Information Resource (TAIR) | The Institute for Genomic Research | Rodney Mauricio's lab

ENCYCLOPEDIA OF LIFE SCIENCES Adaptation genetics | Quantitative genetics

1. Kaul, S. *et al.* Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
2. Theologis, A. *et al.* Sequence and analysis of chromosome 1 of the plant *Arabidopsis thaliana*. *Nature* **408**, 816–820 (2000).
3. Lin, X. Y. *et al.* Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* **402**, 761–768 (1999).
4. Salanoubat, M. *et al.* Sequence and analysis of chromosome 3 of the plant *Arabidopsis thaliana*. *Nature* **408**, 820–822 (2000).
5. Mayer, K. *et al.* Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. *Nature* **402**, 769–777 (1999).
6. Tabata, S. *et al.* Sequence and analysis of chromosome 5 of the plant *Arabidopsis thaliana*. *Nature* **408**, 823–826 (2000).
7. Adam, D. Now for the hard ones. *Nature* **408**, 792–793 (2000).
8. Bulmer, M. G. *The Mathematical Theory of Quantitative Genetics* (Clarendon, Oxford, 1985).
9. Falconer, D. S. & Mackay, T. F. C. *Introduction to Quantitative Genetics* (Addison–Wesley–Longman, Harlow, 1996).
10. Lynch, M. & Walsh, B. *Genetics and the Analysis of Quantitative Traits* (Sinauer Associates, Sunderland, Massachusetts, 1997).
11. Tanksley, S. D. Mapping polygenes. *Annu. Rev. Genet.* **27**, 205–233 (1993).
12. Liu, B.-H. *Statistical Genomics: Linkage, Mapping and QTL Analysis* (Boca Raton, Florida, USA, 1998).
An excellent overview of QTL mapping techniques.
13. Burr, B. & Burr, F. A. Recombinant inbreds for molecular mapping in maize: theoretical and practical considerations. *Trends Genet.* **7**, 55–60 (1991).
14. Moreno-Gonzalez, J. Efficiency of generations for estimating marker-associated QTL effects by multiple regression. *Genetics* **135**, 223–231 (1993).
15. Sax, K. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* **8**, 552–560 (1923).
The first mapping of a QTL by association with a Mendelian marker.
16. Lister, C. & Dean, C. Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. *Plant J.* **4**, 745–750 (1993).
17. Cho, R. J. *et al.* Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nature Genet.* **23**, 203–207 (1999).
18. Barton, N. H. & Turelli, M. Evolutionary quantitative genetics: how little do we know? *Annu. Rev. Genet.* **23**, 337–370 (1989).
19. Lande, R. The response to selection on major and minor mutations affecting a metrical trait. *Heredity* **50**, 47–65 (1983).
20. Orr, H. A. & Coyne, J. A. The genetics of adaptation: a reassessment. *Am. Nat.* **140**, 725–742 (1992).
The landmark paper that exhumed the argument that genes of major effect could be important in adaptation.
21. Fisher, R. A. *The Genetical Theory of Natural Selection* (Dover, New York, 1958).
22. Alonso-Blanco, C., Blankestijn-de Vries, H., Hanhart, C. J. & Koornneef, M. Natural allelic variation at seed size loci in relation to other life history traits of *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **96**, 4710–4717 (1999).
23. Stanton, M. L. Seed variation in wild radish: effect of seed size on components of seedling and adult fitness. *Ecology* **65**, 1105–1112 (1984).
24. Fisher, R. A. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* **7**, 179–188 (1936).
25. Juenger, T., Purugganan, M. D. & Mackay, T. F. C. Quantitative trait loci for floral morphology in *Arabidopsis thaliana*. *Genetics* **156**, 1379–1392 (2000).
26. Kelly, M. G. & Levin, D. A. Directional selection on initial flowering date in *Phlox drummondii* (Polemoniaceae). *Am. J. Bot.* **87**, 382–391 (2000).
27. Kowalski, S. P., Lan, T. H., Feldmann, K. A. & Paterson, A. H. QTL mapping of naturally-occurring variation in flowering time of *Arabidopsis thaliana*. *Mol. Gen. Genet.* **245**, 548–555 (1994).
28. Mitchell-Olds, T. Genetic constraints on life history evolution — quantitative trait loci influencing growth and flowering in *Arabidopsis thaliana*. *Evolution* **50**, 140–145 (1996).
29. Kuittinen, H., Sillanpää, M. J. & Savolainen, O. Genetic basis of adaptation: flowering time in *Arabidopsis thaliana*. *Theor. Appl. Genet.* **95**, 573–583 (1997).
30. Byrne, M. *et al.* Identification and mode of action of quantitative trait loci affecting seedling height and leaf area in *Eucalyptus nitens*. *Theor. Appl. Genet.* **94**, 674–681 (1997).
31. Byrne, M., Murrell, J. C., Owen, J. V., Williams, E. R. & Moran, G. F. Mapping of quantitative trait loci influencing frost tolerance in *Eucalyptus nitens*. *Theor. Appl. Genet.* **95**, 975–979 (1997).
32. Hurme, P., Silanpää, M. J., Arjas, E., Repo, T. & Savolainen, O. Genetic basis of climatic adaptation in Scots pine by Bayesian quantitative trait locus analysis. *Genetics* **156**, 1309–1322 (2000).
33. Bradshaw, H. D. Jr & Stettler, R. Molecular genetics of growth and development in *Populus*. IV. Mapping QTLs with large effects on growth, form, and phenology traits in a forest tree. *Genetics* **139**, 963–973 (1995).
34. Frewen, B. E. *et al.* Quantitative trait loci and candidate gene mapping of bud set and bud flush in *Populus*. *Genetics* **154**, 837–845 (2000).
35. Lewontin, R. C. *The Genetic Basis of Evolutionary Change* (Columbia University Press, New York, 1974).
36. Gottlieb, L. D. Genetics and morphological evolution in plants. *Am. Nat.* **123**, 681–709 (1984).
An early, but underappreciated, paper that raised the issue that genes of major effect could be important in speciation.
37. Coyne, J. A. & Lande, R. The genetic basis of species differences in plants. *Am. Nat.* **126**, 141–145 (1985).
38. Bradshaw, H. D. Jr, Wilbert, S. M., Otto, K. G. & Schemske, D. W. Genetic mapping of floral traits associated with reproductive isolation in monkeyflowers (*Mimulus*). *Nature* **376**, 762–765 (1995).
One of the first attempts in natural populations to identify genes that are important in speciation by QTL mapping.
39. Bradshaw, H. D. Jr, Otto, K. G., Frewen, B. E., McKay, J. K. & Schemske, D. W. Quantitative trait loci affecting differences in floral morphology between two species of monkeyflower (*Mimulus*). *Genetics* **149**, 367–382 (1998).
40. Schemske, D. W. & Bradshaw, H. D. Jr. Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*). *Proc. Natl Acad. Sci. USA* **96**, 11910–11915 (1999).
41. Kelly, A. J. & Willis, J. H. Polymorphic microsatellite loci in *Mimulus guttatus* and related species. *Mol. Ecol.* **7**, 769–774 (1998).
42. Cruzan, M. B. & Arnold, M. L. Ecological and genetic associations in an *Iris* hybrid zone. *Evolution* **47**, 1432–1445 (1993).
43. Arnold, M. L. Anderson's paradigm: Louisiana irises and the study of evolutionary phenomena. *Mol. Ecol.* **9**, 1687–1698 (2000).
44. Hodges, S. A. & Arnold, M. L. Floral and ecological isolation between *Aquilegia formosa* and *Aquilegia pubescens*. *Proc. Natl Acad. Sci. USA* **91**, 2493–2496 (1994).
45. Rieseberg, L. H., Sinervo, B., Linder, C. R., Ungerer, M. C. & Arias, D. M. Role of gene interactions in hydrid speciation: evidence from ancient and experimental hybrids. *Science* **272**, 741–745 (1996).
46. Kim, S.-C. & Rieseberg, L. H. Genetic architecture of species differences in annual sunflowers: implications for adaptive trait introgression. *Genetics* **153**, 965–977 (1999).
47. Rieseberg, L. H., Whitton, J. & Gardner, K. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* **152**, 713–977 (1999).
48. Hodges, S. A., Burke, J. M. & Arnold, M. L. Natural formation of *Iris* hybrids: experimental evidence on the establishment of hybrid zones. *Evolution* **50**, 2504–2509 (1996).
49. Stuber, C. W. Mapping and manipulating quantitative traits in maize. *Trends Genet.* **11**, 477–481 (1995).
50. Young, N. D. A cautiously optimistic vision for marker-assisted breeding. *Mol. Breed.* **5**, 505–510 (1999).
51. Grandillo, S., Ku, H. M. & Tanksley, S. D. Identifying the loci responsible for natural variation in fruit size and shape in tomato. *Theor. Appl. Genet.* **99**, 978–987 (1999).
52. Alpert, K. B., Grandillo, S. & Tanksley, S. D. *fw-2.2* — a major QTL controlling fruit weight is common to both red-fruited and green-fruited tomato species. *Theor. Appl. Genet.* **91**, 994–1000 (1995).
53. Alpert, K. B. & Tanksley, S. D. High-resolution mapping and isolation of a yeast artificial chromosome contig containing *fw2.2*: a major fruit weight quantitative trait locus in tomato. *Proc. Natl Acad. Sci. USA* **93**, 15503–15507 (1996).
54. Frary, A. *et al. fw2.2*: A quantitative trait locus key to the evolution of tomato fruit size. *Science* **289**, 85–88 (2000).
A landmark in QTL analysis: the first molecular characterization of a locus that was originally identified entirely by QTL mapping. References 51–53 are the prologue to this milestone.
55. Doebley, J. & Stec, A. Genetic analysis of the morphological differences between maize and teosinte. *Genetics* **129**, 285–295 (1991).
56. Doebley, J. Mapping the genes that made maize. *Trends Genet.* **8**, 302–307 (1992).
57. Doebley, J. & Stec, A. Inheritance of the morphological differences between maize and teosinte: comparison of results for two F_2 populations. *Genetics* **134**, 559–570 (1993).
58. Doebley, J., Stec, A. & Gustus, C. *Teosinte branched 1* and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics* **141**, 333–346 (1995).
59. White, S. & Doebley, J. Of genes and genomes and the origin of maize. *Trends Genet.* **14**, 327–332 (1998).
60. Dorweiler, J., Stec, A., Kermicle, J. & Doebley, J. *Teosinte glume architecture 1*: a genetic locus controlling a key step in maize evolution. *Science* **262**, 233–235 (1993).
61. Doebley, J., Stec, A. & Hubbard, L. The evolution of apical dominance in maize. *Nature* **386**, 485–488 (1997).
62. Wang, R. L., Stec, A., Hey, J., Lukens, L. & Doebley, J. The limits of selection during maize domestication. *Nature* **398**, 236–239 (1999).
References 55–62 chronicle a decade of work on understanding the genetics of the domestication of modern maize.
63. Paterson, A. H. *et al.* Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science* **269**, 1714–1718 (1995).
An excellent example of the power of comparative QTL mapping.
64. Ramsey, J. & Schemske, D. W. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Syst.* **29**, 467–501 (1998).
65. Soltis, P. S. & Soltis, D. E. The role of genetic and genomic attributes in the success of polyploids. *Proc. Natl Acad. Sci. USA* **97**, 7051–7057 (2000).
66. Hilu, K. W. Polyploidy and the evolution of domesticated plants. *Am. J. Bot.* **80**, 1494–1499 (1993).
67. Paterson, A. H. *et al.* Comparative genomics of plant chromosomes. *Plant Cell* **12**, 1523–1539 (2000).
68. Jiang, C. X., Wright, R. J., El-Zik, K. M. & Paterson, A. H. Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). *Proc. Natl Acad. Sci. USA* **95**, 4419–4424 (1998).
69. Wright, R. J., Thaxton, P. M., El-Zik, K. M. & Paterson, A. H. D-subgenome bias of *Xcm* resistance genes in tetraploid *Gossypium* (cotton) suggests that polyploid formation has created novel avenues for evolution. *Genetics* **149**, 1987–1996 (1998).
70. Beavis, W. D. The power and deceit of QTL experiments: lessons from comparative QTL studies. *Proc. Corn and Sorghum Industry Res. Conf., Am. Seed Trade Assoc., Washington DC*, 255–266 (1994).
71. Beavis, W. D. in *Molecular Dissection of Complex Traits* (ed. Paterson, A. H.) 145–162 (CRC, Boca Raton, Florida, 1998).
References 70 and 71 provided the first and most influential caveats for the use of QTL analysis in both agriculture and evolutionary biology.
72. Copenhaver, G. P., Browne, W. E. & Preuss, D. Assaying genome-wide recombination and centromere functions with *Arabidopsis* tetrads. *Proc. Natl Acad. Sci. USA* **95**, 247–252 (1998).
73. Begun, D. J. & Aquadro, C. F. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356**, 519–520 (1992).
74. Risch, N. J. Searching for genetic determinants in the new millennium. *Nature* **405**, 847–856 (2000).
75. Weiss, K. M. & Terwilliger, J. D. How many diseases does it take to map a gene with SNPs? *Nature Genet.* **26**, 151–157 (2000).
76. Cardon, L. R. & Bell, J. I. Association study designs for complex diseases. *Nature Rev. Genet.* **2**, 91–99 (2001).
77. Doerge, R. W., Zeng, Z. B. & Weir, B. S. Statistical issues in the search for genes affecting quantitative traits in experimental populations. *Stat. Sci.* **12**, 195–219 (1997).
An excellent overview of the statistical issues involved in QTL analysis.
78. Melchinger, A. E., Utz, H. F. & Schon, C. C. Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. *Genetics* **149**, 383–403 (1998).
79. Otto, S. P. & Jones, C. D. Detecting the undetected: estimating the total number of loci underlying a quantitative trait. *Genetics* **156**, 2093–2107 (2000).
80. Paterson, A. H. *et al.* Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. *Genetics* **127**, 181–197 (1991).
81. Jansen, R. C. Interval mapping of multiple quantitative trait loci. *Genetics* **135**, 205–211 (1993).
82. Lander, E. & Kruglyak, L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genet.* **11**, 241–247 (1995).
83. Lander, E. S. & Botstein, D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199 (1989).
In this paper, interval mapping with molecular

markers to map QTL was first proposed for species in which many morphological markers were unavailable. A maximum-likelihood statistical approach for QTL mapping was also developed.
84. Knott, S. A. & Haley, C. S. Maximum likelihood mapping of quantitative trait loci using full-sib families. *Genetics* **132**, 1211–1222 (1992).
85. Zeng, Z.-B. Precision mapping of quantitative trait loci. *Genetics* **136**, 1457–1468 (1994).
86. Churchill, G. A. & Doerge, R. W. Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963–971 (1994).
87. Doerge, R. W. & Churchill, G. A. Permutation tests for multiple loci affecting a quantitative character. *Genetics* **142**, 285–294 (1996).
88. Sillanpää, M. J. & Arjas, E. Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. *Genetics* **148**, 1373–1388 (1998).
89. Haley, C. S., Knott, S. A. & Elsen, J.-M. Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* **136**, 1195–1207 (1994).
90. Hoeschele, I., Uimari, P., Grignola, F. E., Zhang, Q. & Gage, K. M. Advances in statistical methods to map quantitative trait loci in outbred populations. *Genetics* **147**, 1445–1457 (1997).
91. Sillanpää, M. J. & Arjas, E. Bayesian mapping of multiple quantitative trait loci from incomplete outbred offspring data. *Genetics* **151**, 1605–1619 (1999).
92. Pérez-Enciso, M. & Varona, L. Quantitative trait loci mapping in F_2 crosses between outbred lines. *Genetics* **155**, 391–405 (2000).
93. George, A. W., Visscher, P. M. & Haley, C. S. Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach. *Genetics* **156**, 2081–2092 (2000).
94. Hoeschele, I. in *Handbook of Statistical Genetics* (eds Balding, D., Bishop, M. & Cannings, C.) 599–644 (John Wiley & Sons Ltd, London, 2001).
95. Liu, S.-C., Lin, Y.-R., Irvine, J. E. & Paterson, A. H. in *Molecular Dissection of Complex Traits* (ed. Paterson, A. H.) 95–101 (CRC, Boca Raton, Florida, 1998).
96. Ming, R. *et al.* Detailed alignment of *Saccharum* and *Sorghum* chromosomes: comparative organization of closely related diploid and polyploid genomes. *Genetics* **150**, 1663–1682 (1998).
97. Doerge, R. W. & Craig, B. A. Model selection for quantitative trait locus analysis in polyploids. *Proc. Natl Acad. Sci. USA* **97**, 7951–7956 (2000).
98. Rieseberg, L. H. & Buerkle, C. A. Genetic mapping in hybrid zones. *Am. Nat.* (in the press).
99. Provine, W. B. *The Origins of Theoretical Population Genetics* (Chicago Univ. Press, Illinois, 1971).
A superb history of the early conflict between the 'Mendelians' and the 'biometricians', and its resolution.
100. East, E. M. A Mendelian interpretation of variation that is apparently continuous. *Am. Nat.* **44**, 65–82 (1910).
101. Fisher, R. A. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* **52**, 399–433 (1918).
102. Lande, R. The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genet. Res.* **26**, 221–235 (1976).
103. Lande, R. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution* **33**, 402–416 (1979).
104. Lande, R. & Arnold, S. J. The measurement of selection on correlated characters. *Evolution* **37**, 1210–1226 (1983).
105. Via, S. & Lande, R. Genotype–environment interaction and the evolution of phenotypic plasticity. *Evolution* **39**, 505–522 (1985).
106. Turelli, M. & Barton, N. H. Dynamics of polygenic characters under selection. *Theor. Popul. Biol.* **38**, 1–57 (1990).
107. Barton, N. H. & Turelli, M. Natural and sexual selection on many loci. *Genetics* **127**, 229–255 (1991).
108. Turelli, M. & Barton, N. H. Genetic and statistical analyses of strong selection on polygenic traits: what, me normal? *Genetics* **138**, 913–941 (1994).
109. Robertson, A. in *Heritage from Mendel* (ed. Brink, A.) 265–280 (Wisconsin Univ. Press, Madison, Wisconsin, 1967).
110. Haley, C. S. & Knott, S. A. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**, 315–324 (1992).
111. Thoday, J. M. Location of polygenes. *Nature* **191**, 368–370 (1961).
112. Zeng, Z.-B. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proc. Natl Acad. Sci. USA* **90**, 10972–10976 (1993).
113. Kao, C. H., Zeng, Z.-B. & Teasdale, R. D. Multiple interval mapping for quantitative trait loci. *Genetics* **152**, 1203–1216 (1999).
114. Zeng, Z.-B., Kao, C. H. & Basten, C. J. Estimating the genetic architecture of quantitative traits. *Genet. Res.* **74**, 279–289 (1999).
115. Basten, C. J., Weir, B. S. & Zeng, Z.-B. *QTL Cartographer: A Reference Manual and Tutorial for QTL Mapping* (Department of Statistics, North Carolina Univ. Press, Raleigh, North Carolina, 1995).
116. Reinisch, A. J. *et al.* A detailed RFLP map of cotton, *Gossypium hirsutum* x *Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome. *Genetics* **138**, 829–847 (1994).

Acknowledgements
I thank M. Arnold, R. Baucom, A. Bouck, C. Goodwillie, A. Johnson, A. Paterson, L. Rieseberg and J. Willis for unpublished material, helpful discussions and constructive comments on the manuscript.