

## INVITED REVIEW

# Quantitative trait locus analyses and the study of evolutionary process

DAVID L. ERICKSON,\* CHARLES B. FENSTER,† HANS K. STENØIEN‡ and DONALD PRICES§

\*Laboratory of Analytical Biology, Smithsonian Institution, Suitland, MD 20746, USA; †Department of Biology, University of Maryland, College Park, USA; ‡Department of Plant Ecology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden; §Department of Biology, University of Hawaii at Hilo, Hilo, USA

## Abstract

The past decade has seen a proliferation of studies that employ quantitative trait locus (QTL) approaches to diagnose the genetic basis of trait evolution. Advances in molecular techniques and analytical methods have suggested that an exact genetic description of the number and distribution of genes affecting a trait can be obtained. Although this possibility has met with some success in model systems such as *Drosophila* and *Arabidopsis*, the pursuit of an exact description of QTL effects, i.e. individual gene effect, in most cases has proven problematic. We discuss why QTL methods will have difficulty in identifying individual genes contributing to trait variation, and distinguish between the identification of QTL (or marker intervals) and the identification of individual genes or nucleotide differences within genes (QTN). This review focuses on what ecologists and evolutionary biologists working with natural populations can realistically expect to learn from QTL studies. We highlight representative issues in ecology and evolutionary biology and discuss the range of questions that can be addressed satisfactorily using QTL approaches. We specifically address developing approaches to QTL analysis in outbred populations, and discuss practical considerations of experimental (cross) design and application of different marker types. Throughout this review we attempt to provide a balanced description of the benefits of QTL methodology to studies in ecology and evolution as well as the inherent assumptions and limitations that may constrain its application.

*Keywords:* adaptation, epistasis, evolution, linkage, speciation, QTL

*Received 11 January 2004; revision received 18 May 2004; accepted 18 May 2004*

## Introduction

Quantitative traits are those traits under polygenic control and often demonstrate continuous variation within or among populations (Falconer & Mackay 1995). The evolution of important life history, behavioural and morphological characters representing adaptive evolution is thought to reflect evolution at many loci (Fisher 1958; Lynch & Walsh 1998). Thus, evolutionary biologists have sought to examine the underlying genetic basis of those traits including the number of genes affecting complex traits, the relative effects of those genes, and their mode of gene expression that, *in*

*toto*, comprise genetic architecture (Cheverud & Routman 1993; Mackay 2001; Barton & Keightly 2002). The use of genetic markers to infer genetic architecture is termed quantitative trait locus (QTL) analysis, reflecting the interest in describing the genetic loci that contribute to a quantitative trait (Liu 1998; Lynch & Walsh 1998). In this review, we outline the goals and limits of QTL analysis, with special emphasis on how QTL experiments may be applied to questions in ecology and evolution.

The use of genetic markers in the analysis of quantitative traits is not new. Payne (1918) demonstrated that several of the loci that responded to selection for high scutellar bristle number in *Drosophila melanogaster* were closely linked to known markers on the first and third chromosomes. Sax (1923) was able to demonstrate linkage between a genetic

Correspondence: D.L. Erickson. Fax: 1 301-238-3059; E-mail: derickso@onyx.si.edu

**Table 1** Genetic markers employed in QTL analyses

Marker type	Variability	Cost	Speed to screen	Expression
AFLP	High	Low	Medium–Fast	Dominant (and infrequently codominant)
Microsatellites	Highest	Medium	Medium	Codominant
SNP	High–Medium	Medium	Medium	Codominant
EST	High	High	Fast	Codominant

AFLP and microsatellites (also termed SSR) are the most common markers used in the development of new linkage maps and QTL studies. AFLP are preferable for rapid map construction and genotyping of many individuals. Microsatellites are preferable to AFLP owing to codominance, but require a lot of front-end labour to generate. Indeed, laboratories just beginning a QTL mapping programme may be better off using AFLP because of the speed and cost of implementation. With the reduced cost and effort of sequencing, two marker classes may eclipse both AFLP and SSR. Single nucleotide polymorphisms (SNP) are point mutations that distinguish individuals or populations, and are often found by sequencing cDNA libraries. The markers are codominant, PCR based and are directly attributable to genes, and thus have some desirable properties (Tao & Boulding 2003). Expressed sequence tags (EST) are sequenced genes from cDNA libraries that exhibit some type of diagnostic polymorphism. EST-based markers may include SNP type polymorphisms, or may include small insertion-deletions or even SSR repeats that are diagnostic. SNP and EST type markers may be similar in development to SSR, but are gene based markers and can take advantage of targeted QTL analysis. EST mapping represents an approach to both map QTL and simultaneously map QTL effects down to individual genes (Lexer *et al.* 2004; Zhang *et al.* 2004).

marker (a seed colour polymorphism due to a single gene) and a quantitative trait, seed weight, in *Phaseolus vulgaris*. Despite these early forays into detailed descriptions of polygenic inheritance, until the last decade the practical application of QTL analyses was limited by the lack of polymorphic genetic markers (Lander & Botstein 1989; Doerge *et al.* 1997; however, see Shrimpton & Robertson 1988). Advancements in molecular marker technology (Table 1) and the parallel development of analytical software for the combined analysis of genetic and phenotypic data (Table 2) have resulted in the application of QTL analyses in medicine, agriculture and, increasingly, in ecology and evolution (Cheverud & Routman 1993; Mackay 2001). Indeed, there has been an explosive number of QTL analyses published that seek to identify the genetic basis of evolutionarily and ecologically relevant traits. Concordantly, there have been many informative reviews and perspectives discussing the utility of QTL approaches. These reviews have discussed the statistical underpinnings of QTL analyses (e.g. Jansen 1996; Doerge *et al.* 1997; Zeng *et al.* 1999; Flint & Mott 2001; Doerge 2002), the ability of QTL analyses to fulfil the promise of mapping phenotypic variation down to the gene or even nucleotide (e.g. Mitchell-Olds 1995; Nadeau & Frankel 2000; Mackay 2001; Gibson & Mackay 2002; Paran & Zamir 2003; Pruitt *et al.* 2003; Remington & Purugganan 2003), and the application of QTL methodology to questions in ecology and evolution (e.g. Cheverud & Routman 1993; Mitchell-Olds 1995; Mackay 2001; Mauricio 2001; Orr 2001; Barton & Keightly 2002; Boake *et al.* 2002).

We focus on how QTL analyses can improve our understanding of ecological and evolutionary processes and also point out where and why QTL analyses may be less fruitful than promised. In particular, we emphasize what types of questions QTL methods are likely to be useful for address-

ing in natural populations and discuss emerging questions and methods regarding the analysis of QTL in nonmodel systems. Questions in ecology and evolution often reflect a focus towards understanding patterns of biological diversity; accordingly research in the fields of ecology and evolution is overwhelmingly directed towards the use of nonmodel organisms. Because of this, we discuss techniques for analysing nontraditional experimental designs that can be applied to more natural population structures. We begin by discussing some of the methods and practical limitations associated with QTL experiments, which largely dictate the types of questions one can address with QTL analyses, and review current progress to ameliorate these limitations. We do not intend this to be a comprehensive review of the QTL literature, and we necessarily have chosen a subset of the available studies to represent the subjects we discuss.

### Factors affecting QTL analyses

Many factors will affect the power of a QTL experiment to identify the loci that underlie phenotypic traits. These factors include experimental design, marker type, number and sample size. We pay particular attention to those issues affecting researchers who work outside model systems, how QTL analyses may address questions of specific concern to ecologists and evolutionary biologists, and suggest methods and future areas of development that may aid QTL analyses in studies of ecology and evolution.

#### *Marker number, type and population sample size*

The effects of marker number, type and sample size have been addressed in a number of fine reviews and books (Doerge *et al.* 1997; Liu 1998; Lynch & Walsh 1998;

**Table 2** Overview of the development of some methods used in QTL detection

Method (reference)	Program	Significance
Interval mapping (Lander & Botstein 1989)	MAPMAKER/QTL	Localizes QTL into marker intervals
Composite interval mapping (Zeng 1994; Basten <i>et al.</i> 2001)	CIM in QTL CARTOGRAPHER	Employs adjacent markers and/or other QTL as cofactors
Multiple interval mapping (Zeng <i>et al.</i> 1999)	MIM in QTL CARTOGRAPHER	Searches for multiple QTL simultaneously, using a single test for a chromosome.
Bayesian interval mapping (Satagopan <i>et al.</i> 1996)	BMAPQTL, also BIM in QTL CARTOGRAPHER	Can estimate QTL effect and position separately, use of a prior may improve power.
Outbred QTL (Seaton <i>et al.</i> 2002)	QTL EXPRESS	Least squares regression in some outbred crossing designs including sibs and pedigrees.

Staying abreast of the most recent advances in QTL methodology can be a full-time job, and we do not propose to outline all the methods currently available. Rather, this table outlines some of the major developments in QTL analysis, and notes what each step was supposed to improve upon. The earliest methods for linking genotype with phenotype were based upon regression, where a positive correlation of genotype with phenotype was evidence of QTL linkage to a marker (Thoday 1961). An essential problem with single marker regression techniques is the confounding effect of QTL magnitude and position (Doerge *et al.* 1997). A major breakthrough was the development of the interval mapping approaches, which localize QTL to intervals between a pair of genetic markers (Lander & Botstein 1989). They employed a maximum likelihood estimator (which produces a LOD score) to determine the likelihood that a QTL is present within a given interval. This allowed the magnitude of effect to be distinguished from the distance of the QTL from the markers. Subsequent techniques such as composite interval mapping (CIM; Basten *et al.* 2001) and multiple interval mapping (MIM), have extended the interval mapping approach by incorporating searches at multiple marker intervals (Zeng 1994; Zeng *et al.* 1999). Further developments promise to increase the power and precision of QTL mapping analyses by employing Bayesian methodologies to estimate the number of QTL and their effect separately, thereby removing the confounding effect they have upon each other and allowing more accurate estimation of environmental and epistatic interactions among loci (Satagopan *et al.* 1996; Berry 1998; Sillanpaa & Arjas 2000; Sen & Churchill 2001; Borevitz *et al.* 2003; Yi *et al.* 2003). Lastly, a few methods have been developed including least squares regression and variance component methods to infer QTL in outbred line cross designs (Seaton 2002). These methods offer the possibility that QTL can be mapped in species where typical inbred line crosses cannot be readily conducted. Locations for software include: <http://statgen.ncsu.edu/qtlcart/index.php> for QTL CARTOGRAPHER; <http://www.stat.wisc.edu/~yandell/qtl/software> for Bayesian BMAPQTL; <http://qtl.cap.ed.ac.uk/> for QTL EXPRESS; <http://linkage.rockefeller.edu/soft/list.html> for MAPMAKER/sibs.

Patterson 1998; Doerge 2002). We briefly summarize some of the most salient points for the sake of context in the balance of this article. Essentially, the central issues in detecting QTL depend on the type of markers employed, their distribution (including coupling and repulsion phase), cross design and the magnitude of the QTL. In general, QTL studies employing traditional experimental designs and large sample sizes will readily identify QTL that are of large effect (i.e. where QTL effect exceeds 15%), however, identification of all or most loci contributing to a trait will be challenging at best. For the purpose of this discussion, we define QTL as chromosomal regions that are flanked by two markers delineating its position within the genome. Furthermore, we define QTL effect as the proportion of the genetic variance — as observed in a segregating population — that is explained by the QTL (alternatively, QTL effect can be defined in terms of the proportion of the difference between the parents). A rule of thumb in QTL experiments may be that experiments employing fewer than 300 individuals will have difficulty in estimating the true distribution of QTL effects. Under ideal conditions, a perfectly additive QTL exhibiting no dominance with an effect of 5% can be detected using 206 individuals in a  $F_2$  intercross,

using codominant markers, at a spacing of 5 cm. However, because of  $G \times E$  interactions, low heritability and incomplete accuracy in estimating both genotype and phenotype, it is suggested that 300 is a reasonable sample size to employ (Doerge *et al.* 1997). The type of experimental design (cross design) as well as the type of markers employed will affect this number and these issues have been addressed in some detail in Lynch & Walsh (1998) and Liu (1998). For example, under the conditions just mentioned, an experiment that employed a backcross rather than an  $F_2$  design would require double the sample size to infer QTL with equal precision (Lynch & Walsh 1998).

In terms of correctly estimating the magnitude of a QTL, a statistical problem associated with employing small sample sizes is exaggeration of the QTL effect, which has been termed the Beavis effect (Beavis 1994, 1998). When sample sizes fall far below 300, estimates of QTL effects will be exaggerated, and the power to identify small-effect QTL declines dramatically. The bias to inflate QTL effect is reduced as more and more small-effect QTL are identified (Xu 2003a), and is ultimately tied to increasing the overall power of the experiment. Thus an experiment with low power may not only fail to identify true QTL, it may also

falsely suggest QTL or greatly exaggerate the effect of those QTL that are correctly identified as having an effect.

This statistical artefact is less important to plant and animal breeders and human health researchers (who are most interested in QTL of large effect), but is more important to ecologists and evolutionary biologists who may seek to investigate genetic architecture in terms of addressing predictions based upon evolutionary theory. Consequently, it may be more fruitful for experimental designs to maximize sample sizes employed at the expense of generating highly saturated linkage maps as a first approximation to infer QTL.

The types of genetic marker will also have some effect on QTL resolution. We can classify markers based on dominant vs. codominant markers (see Table 1). As an example, in order to generate the same inference of linkage between markers that are 10 cm apart, an experiment employing an  $F_2$  intercross design with dominant markers in the repulsion phase must include nearly 20 times as many individuals as would an  $F_2$  intercross using codominant markers (Liu 1998; Table 6.24). Dominant markers, such as AFLP, will produce two genotypic classes in an  $F_2$  population cross rather than three classes due to dominance, such that it is not possible to distinguish between the heterozygote and dominant homozygote dominant classes. Owing to masking of the genotypic state, there are fewer observable recombination events within the marker interval, resulting in lower information content when dominant markers are used (Liu 1998). AFLP markers do allow one to construct linkage maps with wide genome coverage without engaging in extensive sequencing or marker development programmes. Finally, AFLP are also faster than individual codominant marker types because a single polymerase chain reaction (PCR) can derive multiple loci simultaneously. Codominant markers, such as microsatellites, single nucleotide polymorphisms (SNP) and increasingly expressed sequence tags (EST) (Table 1), offer much greater power to infer recombination between adjacent markers and have much improved information content (Liu 1998). However, their greater expense in development and application are balanced by the greater power to resolve QTL effect and position.

Lastly, the distribution of markers on a chromosome will affect the power to resolve both QTL position and effect. Most QTL mapping programs make use of marker intervals, and in doing so help to define the location of a QTL within a pair of markers. As more markers are added to an experiment, a more precise estimate of both QTL position and effect can be generated. However a balance between marker number, or more correctly the size of the intervals between adjacent markers along a chromosome, with sample size need be established. The prior estimates of 300 individuals should be appropriate for 10–15 cm marker intervals in most experimental designs. If one has many more markers and hence smaller marker intervals, the

number of recombination events between any pair of markers declines and the problem of un-replicated genotypes can arise. It has been suggested that QTL analyses can best be conducted in drafts. An initial draft would maximize sample size at the expense of marker density, and would identify broad intervals (~20 cm) containing putative QTL. More markers could then be included in the areas of interest, to refine QTL position effect issues in subsequent analyses. This approach can save time and money by avoiding the genotyping of areas where no QTL are suggested. The number of markers and sample sizes employed will depend upon the research questions, but many QTL experiments may provide the initial impetus to further explore quantitative genetic architecture.

### *Experimental design*

The experimental design employed, i.e. the type of cross employed, will have a significant impact on the ability to detect QTL. There are a number of crossing designs employed in QTL analysis and we briefly review some of these and comment on their applicability. We specifically contrast inbred line cross designs with what may be termed 'outbred' QTL designs, the latter of which may be broadly applicable in ecological and evolutionary contexts where inbred line cross designs are not feasible.

Of the different designs, the inbred line cross is generally the most powerful method because it increases linkage disequilibrium between the genetic markers used and the QTL (Doerge *et al.* 1997; Liu 1998; MacKay 2001). This design employs two individuals that are highly differentiated for both the trait(s) of interest and the molecular markers used. One or more crosses between these individuals generate a hybrid  $F_1$  generation that may then be crossed to form a recombinant intercross generation ( $F_2$ ) or backcrossed to either parent, or both, to make a backcross ( $BC_1$ ). However, inbred line crossing designs do have a number of practical and experimental constraints. For many researchers, the creation of a recombinant  $F_2$  population derived from an inbred line cross may be impractical. Generation times and the ability to handle the organism in question may limit the ability to implement these designs. In addition, inbred line crosses necessarily limit the number of alleles present at any single QTL location. Thus, the populations from which individuals are derived may contain multiple alleles at each QTL location, but because only one, presumably inbred and hence homozygous, individual is chosen from each population, a maximum of two alleles at each QTL location is included in the experiment. If one wishes to detect multiple QTL that may reside within one or more populations, then an inbred line cross design may be the wrong method. Experiments based on outcrossed designs or sib-pair methods may be more appropriate for many questions and organisms, and we discuss these later.

There are a variety of derivations of the inbred line cross design, including recombinant inbred lines (RIL) or near isogenic lines (NIL). These are both fixed recombinant lines, in which after 6–7 generations of selfing (RIL) or backcrossing (NIL) each 'line' or individual is fixed for a different set of recombinant markers from each parent. Such fixed recombinant lines have some desirable properties, such as the ability to use progeny testing in multiple environments to test for genotype–environment interactions, as well as improved detection of *epistasis*, and, in some cases, more precise estimates of QTL location and effect (Doerge *et al.* 1997; Doerge 2002). Likewise, designs using both backcrossing and intercrossing (or selfing in plants) can create mixtures of recombinant genomes that may facilitate mapping of some quantitative traits (Doebly *et al.* 1995b; Rieseberg *et al.* 1996; Liu 1998). The power of these methods is to increase recombination and control for the genetic background into which putative QTL are placed. Researchers who have the time and ability to employ such designs should seriously consider them. For the rest of us, some developing alternatives provide hope to pursue QTL detection in less malleable study systems.

We consider two general classes of QTL design that reside outside the inbred line cross models – pedigree and sib-pair methods, respectively. We generally describe these as 'outbred' designs because the parents used in the cross are not inbred and may be heterozygous at both marker loci and QTL. The advantage of an outbred QTL design includes the ability to capture more than two alleles per QTL location, the high levels of recombination in the sample population, and its application to systems in which highly manipulated inbred line crosses are not possible. In addition, questions concerning whether QTL derived from inbred line crosses represent variation between or within lines should be considered. It is possible that most evolutionarily important variation occurs within lines, and although some work has addressed this directly (Nagamine *et al.* 2003), outbred designs may be able to more readily discern such variation.

A QTL mapping programme using a pedigree in structured outbred populations follows the methods of complex disease mapping in humans (Almasy & Blangero 1998; Almasy *et al.* 1999) and agricultural populations (Haley *et al.* 1994; Knott *et al.* 1998; Nagamine & Haley 2001), although it is considerably more difficult, because obtaining pedigrees from natural populations presents greater obstacles (Groover *et al.* 1994; Slate *et al.* 1999; George *et al.* 2000). The pedigrees must include many individuals, and thus may span multiple generations, otherwise sample sizes may be too small to detect any linkage among markers and QTL. Typically, a three-generation pedigree is the starting point, and is referred to as a 'grandparent' design (Williams 1998). The power of these methods is strongly affected by missing data, particularly at the grandparent or parent

level. Methods that employ pedigrees in QTL detection estimate coefficients of identity by descent for marker loci calculated from the genotypes of the parents. Putative QTL are then inferred by identifying individuals with alleles identical by descent (IBD) that also share the same phenotype. However, ambiguity in estimating IBD and the confounding effect of missing genetic data reduce the power of these studies (Slate *et al.* 1999). For these reasons, very few studies on genetic architecture of fitness traits in wild, un-manipulated populations have been performed, although they have been employed with success in agricultural species such as wild boar and pigs (Knott *et al.* 1998). However, a method to map QTL in complex pedigrees has been described based on variance components analysis (George *et al.* 2000). Slate *et al.* (2002) used interval mapping and George *et al.*'s (2000) variance component analysis to map QTL for birth weight in wild, un-manipulated populations of red deer using a six-generation pedigree of > 350 animals. Evidence for segregating QTL was found on three linkage groups, one of which was significant at the genome-wide suggestive threshold. The authors argue that the QTL might be genuine, as two of the QTL were detected using alternate approaches making different assumptions in the underlying model, and also because birth weight QTL have been mapped at homologous sites in cattle (Davis *et al.* 1998; Stone *et al.* 1999; Grosz & MacNeil 2001). However, the QTL effects were likely upwardly biased, reflecting the limited sample sizes of specific families. Thus, application of these approaches employing organisms that have large family sizes may be most fruitful. Another way to improve the power of these methods is selective genotyping, in which individuals that are most highly differentiated are selected for inclusion in the study (Lynch & Walsh 1998). Lastly, methods that employ variance component or maximum likelihood models to detect QTL will require further analysis beyond identification of QTL to establish confidence intervals regarding the position of the QTL and the use of bootstrap or Markov chain-Monte Carlo (MCMC) simulations to estimate detection thresholds (Churchill & Doerge 1994; George *et al.* 2000). As with all the methods to search for QTL, a Bayesian methodology to search for QTL in pedigrees has been developed (Bink *et al.* 2002; Perez-Enciso 2003) which offers the advantage of accepting a wide array of experimental designs and marker information.

Sib-pair methods for QTL deduction have not, to our knowledge, been employed in an evolutionary or ecological context. However, the statistical underpinnings of these methods have been well investigated in the search for human QTL. The sib-pair method was first suggested by Haseman & Elston (1972), and employs the difference in trait value between pairs of relatives (typically sibs) in conjunction with estimation of IBD at sets of markers along a mapped chromosome. This approach uses the squared difference in phenotype between pairs of sibs in a regression

onto the set of alleles that are IBD for that sib pair (Drigalenko 1998), and has been used extensively in QTL discovery in humans (Elston *et al.* 2000). The power of this method is that one can take advantage of the naturally occurring family structure, where there are many small families that show variation for the trait of interest. This method may be particularly useful in estimating QTL segregating within populations, or possibly within zones of hybrid contact. If a set of relatives differs for some trait, QTL affecting trait differentiation can be detected through regression of IBD against trait differentiation. The advantage of this type of method is that many plant and animal systems show the pattern of a large number of small nuclear families that can be identified. Although the method has some very real limitations with regard to power to detect QTL (Amos & Elston 1989; Lynch & Walsh 1998), its utility increases with the size and number of sibships employed, and it may serve as a robust alternative to the pedigree method when the reconstruction of a pedigree is too difficult or the size of the offspring class (essentially  $F_2$  population size) is too small. The general method of sib-pair analysis has been extended to incorporate elements of interval mapping (Fulker & Cardon 1994) and multipoint (multimarker) designs (Fulker *et al.* 1995; Cardon 2000) which offer promise in investigating QTL detection in natural populations. Knott & Haley (1998) further extended sib-pair methodology using a multipoint mapping method, which can account for differences in recombination between sexes. Nonmodel species, including many birds and mammals in which it is possible to sample many discrete families containing two or more sibs, or long-lived plants that produce very large half-sib arrays, may be good candidates for the sib-pair method of QTL deduction. In past reviews of sib-pair methods (Weller *et al.* 1990; Lynch & Walsh 1998), the limits in power of QTL detection have been highlighted. The increased efficiency of genotyping individuals offers promise to allow sufficient sample sizes to be employed in examination of QTL with these methods in an evolutionary context. These methods will never have the power of inbred line cross methods, because the degree of linkage disequilibrium between marker and QTL is relatively weak. Accordingly, a description of all loci contributing to quantitative variation using these designs is unrealistic. However, it is very possible to address questions of the role of major genes, the role of genetic–environment interactions and in some cases the mode of gene action at QTL affecting a trait. Furthermore, synteny — conservation of gene order — among species or genera may lead to the opportunity to complement initial QTL experiments with candidate gene approaches as QTL within an interval may be matched to genes of known function in homologous chromosomal locations identified in related model systems. Thus outbred systems of QTL detection do not offer the full power of inbred line designs to reveal a complete description of

genetic architecture, but do offer the opportunity to ask some basic questions about gene number and effect, and may allow for further exploration through use of candidate genes or other evolving technologies.

#### *Extension of QTL to specific questions of genetic architecture*

Much of this review has considered some basic concepts of design that will affect resolution of QTL. However, there are a few considerations that must be mentioned directly to give a full account of the power of QTL to determine genetic architecture. These include distinguishing linkage from pleiotropy, measuring genotype–environment interaction ( $G \times E$ ), differentiating between dominant and overdominant gene action, and estimation of epistasis. In traditional quantitative genetics experiments, researchers often seek to distinguish the effect of the genetic variance from the effect of the environment on phenotypic variance, as well as the modes of gene action such as additive, dominance, and epistatic and pleiotropic effects.

*Mode of gene expression.* In addition to gene number and relative effect, the degree to which interactions play a role in phenotypic expression, at the level of alleles within loci (recessive, dominant, overdominant gene expression), with other loci (epistasis) or with the environment (genetic by environment interactions), has important implications for our understanding of a variety of evolutionary phenomena. Furthermore, many relevant evolutionary questions focus on the relative role of pleiotropy vs. linkage in the co-expression of multiple traits. Marker-assisted approaches would appear to have the obvious advantage over previous biometrical methods by permitting description of a range of effects associated with individual marker or flanking regions vs. a sum or average effect across the genotype. The development of increasingly sophisticated analytical approaches is rapidly allowing much greater insight into quantifying modes of gene expression. Below we discuss, in turn, the ability of QTL methods to quantify the various modes of gene expression.

*Dominance vs. overdominance basis of heterosis.* Our understanding of both the maintenance and evolution of mating systems will be greatly enhanced by a thorough understanding of the genetic basis of heterosis and its converse, inbreeding depression (Charlesworth & Charlesworth 1987, 1999; Uyenoyama and Waller 1991). If the expression of heterosis is due to dominant gene action, then recessive deleterious alleles should be relatively efficiently purged from a population over the course of inbreeding, facilitating the evolution of inbreeding mating systems.

Carefully conducted biometric studies reveal that recessive deleterious alleles contribute to inbreeding depression

(e.g. Dudash & Carr 1998; Willis 1999). However, a recent review of the QTL literature (Carr & Dudash 2003) indicates that overdominance-based heterosis is frequently initially observed, although later, more thorough analyses sometimes reveal dominance-based heterosis. The essential issue is whether QTL analysis can distinguish overdominance from pseudo-overdominance. Pseudo-overdominance is the situation in which two viability loci are closely linked in repulsion ( $++--/--++$ ) and a cross between lines manifests apparent overdominance (i.e. the heterozygotes appear to have the highest fitness), when in fact the wild-type (+) alleles are simply dominant to the deleterious alleles at the closely linked loci. This can be easily seen by associating the two genotypes with flanking markers (e.g.  $M1M1^{++}--M2M2 \times M1'M1'--++M2'M2'$ ), resulting in individuals manifesting the heterozygote flanking region,  $M1M1'M2M2'$ , having highest fitness. Thus QTL analysis of inbreeding depression will be sensitive to map coverage and the number of loci within flanking regions. Given that QTL-based analyses of inbreeding depression have been mostly conducted using artificially selected varieties, e.g. maize (Stuber *et al.* 1992) and rice (Li *et al.* 2001; Luo *et al.* 2001) one would expect a high degree of repulsion linkage of viability loci associated with the Hill–Robertson effect (Hill & Robertson 1966). Indeed, pseudo-overdominance has been implicated in the maize results, following analyses that take into account multiple QTL per chromosome (Cockerham & Zeng 1996), and incorporate fine-scale mapping (Graham *et al.* 1997). The detection of a large contribution of epistasis to heterosis in the rice studies suggests that there are many loci within the flanking regions and thus pseudo-overdominance, especially considering the highly selfing mating system of rice. Studies with different cultivars in pine (Kuang *et al.* 1999; Remington & O'Malley 2000a,b), provide evidence for a mostly dominance basis of heterosis. QTL studies of the basis of inbreeding depression are clearly needed in wild populations, but will require extra effort to develop adequate coverage (Fu & Ritland 1994; Karkkainen *et al.* 1999). More sophisticated analytical approaches that allow for the greater detection of multiple QTL per linkage group (see Table 2), will greatly contribute to our understanding of the relative role of dominance and overdominance in the expression of heterosis.

*Epistasis.* The relevance of epistasis to questions in ecology and evolution is discussed in more detail in the applications section, and here we limit the discussion to methods for quantifying its contribution to phenotype through QTL analysis. Theory can be used to predict the existence of epistatic QTL that have no significant marginal effects (Culverhouse *et al.* 2002), but empirical demonstration of this fact has been rare. There have been a number of recent methods proposed to infer the contribution of interaction

between markers which may help demonstrate epistasis in QTL studies: a one-dimensional scan that looks for the interaction of an allele with the genetic background (Jannink & Jansen 2001), simultaneous two-way searches at multiple, selected pairs of loci (Kao *et al.* 1999) and, most recently, a genome-wide method for the simultaneous mapping of all pairs of loci (Carlborg *et al.* 2003). These scans quantify the extent to which a QTL effect is dependent on the presence or absence of other QTL, i.e. the genetic background of the recombinant generation. However, because of the very many possible pair-wise interactions that must be considered [ $n * (n - 1) / 2$  possible pairs of markers where  $n$  = number of loci], very large sample sizes are necessary to detect interaction effects at even moderately significant levels of significance. Carlborg *et al.* (2003) employed a population of over 800 individuals, using a simultaneous genome-wide scan to detect high levels of interaction among markers that otherwise exhibited no marginal effects. Other studies that have identified a significant contribution of epistasis have either used large sample sizes (Li *et al.* 1997; Shook & Johnson 1999) or focus on specific candidate loci, or induced mutations to reduce the number of comparisons that must be made (Fijneman *et al.* 1996; Fedorowicz *et al.* 1998; Wade 2001). Because of the increased number of type 1 errors in estimating epistasis at many loci, higher thresholds of acceptance must be employed, and standard errors of 1 LOD score are inappropriate (Lander & Kruglyak 1995). However, the simultaneous methods for inferring epistasis reduce the problem of multiple tests and randomization tests can be used to estimate significance levels for interacting QTL (Carlborg & Andersson 2002).

$G \times E.$  Consideration of genetic interactions with the environment, or  $G \times E$  interactions, is important in QTL studies not only to understand how the genes interact with the environment, but also to correctly document the relative effect of QTL. Several studies have documented the importance of  $G \times E$  in shaping trait variability. Experiments that have identified QTL for resistance to a fungal pathogen (*Ustilago myadis*) in maize revealed that only a subset (~25) of the QTL is constant among all environments (Lubberstedt *et al.* 1998a). Similar experiments found that as many as 50% of the QTL were constant among experimental populations, but only for about half the populations compared (Lubberstedt *et al.* 1998b). In a study of QTL for date of bolting (the transition from vegetative to reproductive growth) in several natural field and laboratory environments, Weinig *et al.* (2002a) found substantial environmental-dependent expression of allelic variation in many QTL within *Arabidopsis thaliana*. This study used an RIL design, which allows for progeny testing of genetically identical individuals in multiple environments. They observed that most of the loci controlling variation in timing of bolting

differed not only among populations, but also between spring and autumn generations in the same geographical locations. The authors hypothesize that if the genetic potential for response to natural selection on reproductive life histories differs among seasonal cohorts, then phenotypes expressed in autumn and spring may potentially evolve independently in response to divergent selection across seasons. Ungerer *et al.* (2003) similarly used an RIL design with *A. thaliana* to investigate  $G \times E$  at QTL affecting inflorescence development. They found plasticity and  $G \times E$  for the majority of 13 inflorescence traits, and this was associated with variable effects of specific QTL. Pooled across traits, 27% and 52% of QTL exhibited QTL–environment interactions in two recombinant inbred mapping populations. Interestingly, the observed interactions were attributable to changes in the magnitude of QTL effects rather than changes in rank order (sign) of effects. This is in contrast to associated reaction norms exhibiting frequent changes in rank order. This shows that changes in rank orders of reaction norms need not require congruent patterns of QTL effects.  $G \times E$  at QTL has also been observed in *Drosophila melanogaster* and several crop species (see overview in Weinig *et al.* 2002), where the effects of QTL vary with the environment and the genetic background. It is important to take the possibility of  $G \times E$  into consideration when designing QTL experiments aimed at identifying factors associated with natural variability in given traits. Artificial and unrealistic captive environments or growth conditions may yield phenotypic variance and associated QTL effects not necessarily present in natural environments of organisms.

*Pleiotropy.* Pleiotropy is invoked in a number of models of evolution, particularly with regard to mechanisms of speciation. For example, sympatric speciation in insects may be facilitated by pleiotropic effects contributing to both feeding site and mate choice (Hawthorne & Via 2001). However, suffice it to say that making a definitive determination of pleiotropy is challenging. Unless polymorphisms within actual genes are employed, interval mapping methods employing neutral genetic markers that outline 5 cm+ intervals (that may contain hundreds of genes) can only suggest the possibility of pleiotropy. Candidate gene approaches may be a powerful method to directly demonstrate pleiotropy, but even then, deletion mapping or complementation approaches would need to be employed to definitively demonstrate its effect. It is far easier to falsify the hypothesis of pleiotropy than to make a definitive determination of its action. Indeed, the search for pleiotropy in some ways encapsulates the search for QTL in general. While the power to resolve QTL into discrete intervals, with known effect on variance in phenotype can be achieved, there is a profound difference between identification of one or more QTL intervals

and a complete description of the genetic architecture affecting phenotype. Although QTL represent a dramatic improvement over biometrical methods, and the technology is constantly advancing, prudence in interpretation of what the results of a QTL analysis mean is still the most important tool in estimating the contribution of QTL to phenotype.

*Significance thresholds and the problem of linkage.* We wish to briefly comment on two other issues in QTL mapping. First, how we decide on the appropriate threshold for accepting or rejecting a QTL as significant will have a profound effect on the estimation of genetic architecture. Historically, QTL were deemed significant if they exceeded a LOD score of 3.0, which is based upon assumptions of the distribution of QTL number and effect (Lynch & Walsh 1998). However, randomization or permutation methods are more robust for determining the threshold LOD score for acceptance of significant marginal QTL effects (Churchill & Doerge 1994; Doerge & Churchill 1996). These methods use data in a simulation to estimate the number and LOD scores of false positives. As mentioned previously, this threshold is even more important when epistatic QTL are considered. These methods do not rely on the assumptions of the number and distribution of QTL effect, and should provide LOD score thresholds that are more appropriate for each dataset used. Permutation tests have been incorporated into a number of QTL detection software packages (notably QTL CARTOGRAPHER) and should replace arbitrary estimates of QTL significance.

A second issue is the confounding relationship between QTL position and effect. Because hundreds of genes may reside within a 5 cm interval along a chromosome, mapping a QTL into such an interval leaves open the possibility that multiple linked QTL reside within that interval. Even in animal model systems of human disease, the issue of linked small-effect QTL limits the identification and cloning of important candidate loci (Mott & Flint 2002). At one level, it may not matter if one or more QTL reside in that interval, and we may wish to treat the linked QTL as a single integrated expression unit. Alternatively, if one wants to try to map to the gene or the nucleotide level, a candidate gene approach may help resolve the linkage/effect question. With the advance of genome sequencing, and the high degree of synteny among related groups of organisms, it may be possible to identify genes in model species at the approximate chromosomal location of your QTL (see below). With a bit of sequencing, it is possible to develop markers based on one or multiple candidate genes and then use them in a standard QTL analysis. Acceptance or rejection of these candidate loci, as well as comparison of their effect relative to the QTL effect of the entire interval may provide insight into contribution of linkage to QTL effect within a large marker interval.



## Applications of QTL analysis

Given the practical constraints that limit our ability to completely define the underlying genetic architecture of trait evolution, some types of questions are better suited to analysis using QTL approaches than others. Studies that focus on the relative distribution of effects and mode of gene action (e.g. pleiotropy vs. gene linkage, additive vs. nonadditive gene expression) may benefit more from QTL analysis than studies focusing on the absolute basis of gene expression (e.g. quantifying all genes contributing to trait divergence). Below, we outline a number of questions where QTL approaches may have the most value, and we use a few representative reports that demonstrate both the utility and limits of QTL analyses in ecology and evolution.

### *Adaptive differentiation*

Darwin's fundamental vision of evolution as a gradual process with natural selection acting on continuous variation was reformulated by the founders of neodarwinism as evolution reflecting the fixation of genes of uniformly small effect (Provine 1971). Most recently, Fisher's (1958) geometric vision of adaptation, which precludes an important role for mutations of large effect, has been challenged (Orr 1995). Models taking into account the distance of a population from the trait optimum and the probability of fixation of new mutations conclude that the evolution of a trait reflects the fixation of alleles which have an exponential distribution of effect, resulting in mutations of both major and minor effect contributing to evolution. Furthermore, the distribution of gene effects influencing a trait, more than the number of genes for that trait, has been shown to be important in determining short-term responses to selection on the trait (Lande 1975; Barton & Turelli 1987). The issue, then, is the extent to which QTL methods can truly address questions concerning the evolution of genetic architecture as it relates to the response to selection and adaptation. QTL mapping using crosses between differentiated taxa have recently been used to try to quantify the complete distribution and number of QTL effects that underlie phenotypic evolution (Moritz & Kadereit 2001; Gadau *et al.* 2002). Note that this approach is used as a proxy for the absolute identification of all loci and is used to characterize the contribution of segregating allelic variation to response to selection *at the within-population level*, a central issue of evolutionary genetics (Barton & Turelli 1987).

As expected, the detection of genes of major effect (or the rejection of major gene effects) is especially well suited to QTL approaches. For example, many QTL analyses appear to confirm prior studies based upon traditional biometrical approaches (Wright–Castle technique) with respect to general estimates of gene number underlying a trait. Notable

examples include Doebley and collaborators' (Doebley 1992) findings that five major loci are responsible for the domestication of maize from its wild relative teosinte, corresponding to estimates made by Beadle (1980) based upon recovery of Mendelian ratios. Other examples include the demonstration that genes of large effect are responsible for trait differentiation across two species of *Mimulus* with different pollination syndromes (Bradshaw *et al.* 1995, 1998), mirroring the results of Heisey *et al.* (1971). However, QTL studies simply identify flanking regions that harbour a locus or many linked loci contributing to trait variation. Whether a single mutational event, associated with a QTN, or many nucleotide changes at many closely linked loci are responsible for trait divergence has only been quantified for very few model organisms (see Remington *et al.* 2001). Recent findings (Fishman *et al.* 2002) that many QTL of mostly small effect are responsible for the trait differentiation underlying mating system evolution in *Mimulus*, confirm prior investigations (Macnair & Cumbes 1989; Fenster & Ritland 1994; Fenster *et al.* 1995). Likewise, Shaw (1996) also observed a highly polygenic basis to sexual selection characters in the cricket genus *Laupala*, and confirmed the result using a QTL approach (Shaw & Parsons 2002).

Some recent examples of the adaptive significance of QTL come from experiments that employed model species in field-based experiments. Verhoeven *et al.* (2004) observed that a number of relatively small-effect QTL in barley contributed to differentiation among populations without any individual loci demonstrating a counteracting fitness effect in different environments (i.e. no evidence of trade-offs in adaptation for individual QTL in different environments). However, another study using *Arabidopsis* RIL planted into the same environment in separate spring and autumn plantings did demonstrate a trade-off for individual QTL in different environments, as well as evidence that epistasis contributes to that trade off (Weinig *et al.* 2003a). Likewise, Lexer *et al.* (2003) found a number of QTL that demonstrated significant trade-offs, which seems to explain adaptive differentiation mediating habitat preference between a common sunflower species *Helianthus petiolaris* and a putative homoploid hybrid derivative *H. paradoxus*. Although these three studies may lack the power to resolve all QTL affecting the traits they examined, QTL analyses were able to specifically test hypotheses regarding the role of individual loci in contributing to adaptive differentiation by assessing trade-offs of individual QTL in different environments or genetic backgrounds.

Thus QTL approaches confirm instances at the extreme, where very few or very many loci underlie trait differentiation. Likewise, the effect of individual QTL on adaptation may be addressed by estimating the phenotypic effect,  $G \times E$  and mode of gene action. Because of the limits in power to detect all loci associated with linkage or small effect, QTL approaches cannot detect the full range of loci

contributing to differentiation. However, recent studies incorporating Bayesian approaches hold great promise for an accurate description of the full distribution of the gene effect (Xu 2003b). Certainly, one clear advantage of QTL methodologies is that they foster further, detailed molecular examination of the genetic basis of morphological evolution, as seen in studies of maize (Lukens & Doebley 1999). We further emphasize that where issues of linkage are of prime evolutionary importance (e.g. pleiotropy vs. linkage, Hawthorne & Via 2001; sex-limited expression of traits, Boake *et al.* 2002), QTL approaches have clear advantages over earlier methodologies. Furthermore, marker-based approaches can elucidate the genetic basis of within population variation as we discuss below.

### *QTLs and the signature of selection*

Orr (1998) developed a sign test that compares the number of plus alleles present in the high condition of a trait with a model of neutrality assuming either equal or differential allelic effects. Consequently, QTL data can provide evidence for the presence of directional selection, when one can demonstrate a polarity to allelic substitution (e.g. gain or loss of a trait relative to an ancestor). This approach has been used in such divergent organisms as sunflowers and Lake Malawi cichlids to help quantify the dominant selective agents responsible for the diversification of the respective organisms. Specifically, the overarching selective factors in sunflower domestication appear to be selection on achene size (Burke *et al.* 2002), whereas diversification of the Lake Malawi cichlids is strongly associated with coordinated selection on jaws and teeth resulting in the functional divergence of feeding behaviour (Albertson *et al.* 2003). A recent review of 84 QTL studies focused on domesticated vs. wild progenitor species, and intra- and interspecific differentiation in wild species confirmed that directional selection plays a prominent role in phenotypic divergence (Rieseberg *et al.* 2002).

Examination of the effects of QTL may also have interesting implications for our understanding of the evolution of complexity. The adaptive phenotype must represent a coordinated selected response of multiple traits, even if only applied to syndrome concepts. Although it has been extremely difficult to demonstrate coordinated selection with phenotypic selection analysis (Kingsolver *et al.* 2001), coordinated selection was readily demonstrated by QTL analysis in the above Lake Malawi cichlids example. Another interesting case is the use of QTL to test concepts of pleiotropy as they relate to organismal organization, i.e. modularity (Wagner & Altenberg 1996). A QTL analysis of crosses between inbred strains of mice demonstrated that pleiotropy is associated with functionality, providing support for the notion that mouse mandible evolution reflects modular organization (Mezey *et al.* 2000; Workman *et al.*

2002). Pleiotropy was also seen to contribute to quantitative variation in cattle (Schrooten & Bovenhuis 2002). Simultaneously demonstrating parallel evolutionary responses of two or more traits at the QTL level thus appears to be a powerful new tool for understanding phenotypic evolution. The motivation for some of our approaches to quantify the process underlying diversification has come full circle. QTL approaches can be used to not only understand the genetic architecture underlying the evolution of ecologically important traits, but also to determine the specific targets of selection and how they have contributed to adaptation. Lastly, the directionality of selection on individual QTL may be dependent on the environment, leading to heterogeneous effects on individual QTL throughout the lifespan of the organism (Weinig *et al.* 2003b).

### *Speciation genetics and epistasis*

Within the framework of the Biological Species Concept (Mayr 1942), the inherent property of a species is that it is genetically incompatible with other species. Thus one of the fundamental questions of speciation genetics focuses on the origin and description of these incompatibilities (Dobzhansky 1937; Muller 1940; Gravilets & Hastings 1996). This question can be viewed as the origin and consequences of negative epistasis (Whitlock *et al.* 1995; Fenster *et al.* 1997). Theoretical (Orr 1995) and laboratory empirical (Rice & Hostert 1993) studies indicate that genetic incompatibilities can arise rapidly. Indeed, recent results suggest that negative epistasis contributes frequently and substantially to natural population divergence (e.g. Galloway & Fenster 1999; Fenster & Galloway 2000; Wolf *et al.* 2000). Theory has also suggested that: (i) the opportunity for incompatibilities to evolve increases exponentially with increasing population genetic divergence, (ii) the loci involved in the interactions will differ between species and thus are not symmetrical, and (iii) higher order interactions (> 3 loci) are more likely to be involved in hybrid incompatibilities (Orr 1995). By documenting the number of interacting chromosome regions, QTL analyses allow us to quantify the number of loci contributing to incompatibility and, because it is a linkage-based analysis, identify asymmetries and higher order interactions. QTL experiments that employ an outbred design may have greater power to resolve these types of interactions because they can include many alleles at a QTL locus. This would increase the power to document interaction effects, and would also suggest that additive, main-effect QTL are robust because the inclusion of more alleles at all loci increases the likelihood of capturing all alleles that affect the trait.

An additional example of where QTL analyses contribute to our understanding of evolutionary processes at the species level is a series of studies that examined the number and distribution of factors (QTL) contributing to reproductive

isolation between two sunflower species (Kim & Rieseberg 1999; Lexer *et al.* 2003). The alternative genetic architectures represented by many distributed factors vs. relatively few, would lead to either a decreased likelihood of successful introgression or an enhanced likelihood of introgression, through linkage to relatively many or few sterility factors, respectively. By mapping the QTL that contribute to the sterility barrier isolating *H. annuus* and *H. debilis* ssp. *cucumerifolius* and those that contribute to neutral morphological traits differentiating the two species, it was found that only 2 (of 58) QTL contributed to the sterility barrier between the two parental species, suggesting that much of the genome between these species is permeable to introgression. This is in contrast to a similar study of *H. annuus* and *H. petiolaris* (Rieseberg *et al.* 1999) in which 21 sterility factors were detected by QTL associated with pollen sterility. This latter result was mirrored in a study of cotton, in which a large number of barriers to introgression was observed between interspecific populations of polyploid cotton, with genome-wide epistasis seeming to contribute to the impermeability of the two genomes (Jiang *et al.* 2000). Thus, the ability to resolve questions regarding the permeability of a genome and the linkage among different types of QTL (ex. QTL for sterility and QTL for morphological variation) is very appropriate for marker-assisted QTL analyses. QTL analysis represents a substantial improvement over non-QTL biometrical methods in such studies owing to their ability to discern linkage between sterility factors and other neutral characters that differ among species.

### Sexual selection

Sexually selected traits are some of the most conspicuous traits in nature and the theories proposed to explain their evolution have generated controversy. One aspect of the debate centres on the trade-offs that may occur between the sexually selected trait (e.g. male attractive traits) and other important fitness-related traits. The 'good genes' sexual selection theories propose that there is a positive genetic correlation among attractive male traits and other components of fitness such as offspring viability (Hamilton & Zuk 1989; Petrie 1994). In contrast, 'runaway' sexual selection theories posit that there may be a negative genetic correlation among the male traits and offspring viability (Lande & Arnold 1985). These theories generally proposed that these genetic correlations are due to the pleiotropic effect of individual genes that affect the two types of traits. We can classify these two models as referring to linkage disequilibrium (LD) between unlinked loci and pleiotropy, respectively. Classical quantitative genetic analyses and life-history studies have demonstrated support for each of these hypotheses (e.g. Welch 2003). However, the challenge remains to disentangle the environmental and genetic causes of phenotypic correlations that may

be generated in natural populations among exaggerated male traits and other component of fitness (e.g. Kokko 2001; Boake *et al.* 2002). A rigorous QTL approach in an experimental system could provide some insight into the causes of the genetic correlations, by testing between hypotheses of linkage and pleiotropy, and the sign of the correlation between QTL with significant marginal effects, but subject to the same limits discussed above

Several examples of QTL analyses of sexually selected traits suggest that it is possible to diagnose their genetic architecture. The interpulse interval (IPI) in *Drosophila melanogaster* is an important species-specific courtship signal. Gleason *et al.* (2002) demonstrated that there are three QTL that explain 54% of the genetic variation for this trait. However, the resolution of this study does not allow for a clear determination of the number of genes involved, as there may be more than one QTL associated with each marker. Interestingly, the locations of the identified QTL do not correspond with previously predicted candidate genes for IPI. This lack of correspondence between QTL and candidate genes also occurs for mating recognition among species that is related to mating preferences in *D. simulans* and *D. sechellia* (Civetta & Cantor 2003). Another QTL analysis of the colour difference between two cichlid species, *Labeotropheus fuelleborni* and *Metriachima zebra* indicates that QTL for a female colour pattern may be closely linked to a candidate gene (Streelman *et al.* 2003). QTL approaches to examine sexually selected traits have begun to address the genetic architecture of courtship traits with evidence for both few genes of large effect and many genes of small effect. The potential to examine pleiotropy of some of the genes of large effect for sexually traits with other fitness-related traits could begin to address the 'good genes' and 'runaway' sexual selection theories. When competing theories invoke alternate genetic architectures — here pleiotropy vs. LD between unlinked loci — QTL have an opportunity to resolve the questions.

### Comparative QTL mapping

Fortunately, the results of QTL analyses may be applied across taxa or environments to examine the constancy of specific loci in their effects on the phenotype. Thus results for well-characterized organisms may be extended for use in their wild relatives. Comparisons across different genera of crop cereals for QTL associated with domestication traits found similar QTL associated with these traits (Paterson *et al.* 1995). Relatively few QTL were observed within species when analysed independently, from 1 to 10 QTL/trait/species, with observed QTL explaining > 50% of the phenotypic variance for a given trait. There was concordance among these QTL among cereal taxa when tested using shared genetic markers. Flowering time variability in several Brassicaceae species has been shown, through comparative

QTL and association mapping, to be controlled by members of the *CONSTANS* gene family (Axelsson *et al.* 2001; Österberg *et al.* 2002). It is therefore hypothesized that major QTL detected in the different species could be the result of duplicated copies of the same ancestral gene, possibly the ancestor of *CONSTANS*.

In general, much of the concordance among populations, environments and taxa seems to be due to major genes that are conserved in action across speciation events and thus manifest little environmental or epistatic interaction. Unfortunately, more complex traits have demonstrated less constancy. Yan *et al.* (1998) examined QTL at different stages of development in cultivated rice (*Oryza sativa*) affecting plant height, and found that only two of nine QTL affected plant height at all developmental stages, with the other seven distributed among the nine stages tested. Whether these cases derived from human-imposed selection apply to instances of diversification via natural selection should be a research priority. Recent work demonstrates great synteny between *Arabidopsis thaliana* and its closely related outcrossing congener, *A. lyrata* (H. Kuittinen *et al.* manuscript submitted for publication). Thus QTL mapping of trait differentiation between natural populations of the outcrosser can take advantage of the well-studied *A. thaliana* genome to readily identify candidate loci. One remarkable example in which natural selection for genetic architecture has been demonstrated via comparative QTL is in comparisons of natural and synthetic hybrids in sunflower. Rieseberg *et al.* (2003) used comparative mapping in synthetic hybrids of two *Helianthus* species and compared QTL distributions with that in a putative ancient hybrid species derived from the same two parental species. They observed similar genetic architecture in the synthetic hybrids as in the ancient hybrids, suggesting that selection on introgressed segments resulted in specific beneficial combinations of genes being maintained, with other combinations selected against during early viability selection.

Comparative mapping may demonstrate the prevalence of common developmental pathways associated with specific adaptations. It may also indirectly help clarify a number of the evolutionary phenomenon cited above, and may be especially revealing of character evolution during adaptive radiations. It may be employed to demonstrate that the same loci are responding to selection in radiating taxa or that different genetic mechanisms underlay the evolution of trait diversity that is so prevalent during adaptive radiations.

## Future directions

### *Identification and cloning of individual loci vs. the complete description of QTL effects*

The mapping of QTL effects down to the nucleotide will be difficult, if not impossible, in nonmodel systems. In

fact, there are relatively few examples of the cloning of evolutionarily or ecologically important QTL (Remington *et al.* 2001). Indeed, even in animal model systems of human disease, the cloning of genes can be problematic because as researchers break large intervals into smaller segments, they have often found that large-effect QTL break down into several-linked QTL of smaller effect (Mott *et al.* 2000). In cases where QTL relevant to evolution have been cloned, it is often in plants, where NIL crossing designs have been employed. The cloning and description of the genes that determine maize domestication are a classic example of a system in which initial QTL analyses were followed up with extensive molecular work to clone and characterize the five major QTL contributing to the transition of the wild grass teosinte into maize (Doebley 1992; Doebley *et al.* 1995a,b; Lukens & Doebley 1999). While work on domestication in maize and a handful of other species represents what is possible, the ambitious goal of working from QTL to gene is challenging. In most instances, to map QTL down to the gene or nucleotide, one must employ approaches such as deletion mapping or fine-scale recombinational mapping with NILs (Frary *et al.* 2000; Yano *et al.* 2000; Remington *et al.* 2001). The cloning of QTL may become easier in the future owing to the use of candidate gene methods. The synteny among organisms and sequence conservation in many coding genes suggests information from model systems may help exploration of QTL effects in nonmodel systems. In addition, genetic marker systems based on EST (Table 1), which utilize sequence differences within genes, and can be targeted to include genes of interest, may be used to deduce QTL location and effect while also generating candidate genes for further analysis.

### *QTL and genomics*

At the finest level of resolution, we may seek to quantify how nucleotide changes contribute to quantitative variation, i.e. to detect quantitative trait nucleotides (QTNs). Detailed knowledge of the molecular basis of quantitative trait evolution will necessitate not only QTL mapping of candidate genes, but also the use of functional genomics and bioinformatics, including comparative mapping, cloning and various transgenic approaches in order to verify the effect of candidate genes on individual fitness (Buckler & Thornsberry 2002; Österberg *et al.* 2002). DNA microarray technology is a powerful method that can assess gene function and variation at many loci simultaneously, making a quantitative assessment of changes in gene expression, which can then be correlated to phenotypic differences. One can think of expression data for a single gene as a quantitative assessment of gene activity. Change in the expression of a particular gene may then be described as an expression-level polymorphism (ELP; Doerge 2002). ELP can be associated with regions of the genome, much like

QTL. Essentially, the same framework and machinery that exist for QTL mapping can now be superimposed on the gene-expression data, and can be used as a means to identify regions of a genome associated with the expression of groups of genes.

Microarray and QTL may actually inform each other, with QTL suggesting regions where known genes may be examined via microarray, or where microarray data may suggest candidate genes for use in QTL analysis (Liu *et al.* 2001; Gibson 2002; Wayne & McIntyre 2002). A limitation of microarray data is that changes in gene expression may not be attributable to allelic variation. Microarrays also only detect changes in expression levels, thus alternate alleles that mediate different phenotypes, but which have equal mRNA expression cannot be identified by microarrays alone. Recent novel approaches of controlling for the inherent noise associated with gene expression data (Stuart *et al.* 2003) will facilitate the discovery of novel genes contributing to a specific biological function. The combination of QTL and ELP approaches seems much more appropriate in model organisms, where the genome is sequenced. Thus, one may be able to perform a QTL analysis to identify every gene within a QTL interval based on sequence data, and then use microarrays or even transgenics to follow-up and identify specific loci that mediate the evolutionary changes that distinguish the parentals. The reverse is also true, where microarray data may suggest candidate genes that may then be investigated directly through QTL mapping. Thus QTL mapping and QTL approaches can complement each other and may speed the movement from QTL to gene.

An extension of using ELPs in QTL studies is to screen for DNA polymorphisms directly using oligonucleotide arrays. This method effectively sequences DNA at random, but in such a way that comparisons among individuals or groups can be made for the sake of mining for candidate genes. Borevitz *et al.* (2003) developed a method for screening a large number of genome-wide single-feature polymorphisms (SFPs) using direct hybridization of labelled genomic DNA. This technique offers several advantages to QTL analyses, as pointed out by Borevitz *et al.* (2003). In QTL analyses performed to date, recombination breakpoints have often been inferred between markers using interval-mapping approaches. However, array hybridization can precisely define recombination breakpoints, allowing QTL to be defined by intervals. Such a dense marker set is clearly an advantage for large RIL populations (Dupuis & Siegmund 1999). An additional advantage is that a single RI line can be completely genotyped with one hybridization, multiple loci do not need to be independently assayed. As the price of the oligochips employed is reduced, the benefits of increased resolution and higher throughput should make SFP genotyping very attractive for QTL mapping.

## Conclusions

QTL mapping methodologies represent an improvement over biometrical techniques in their ability to resolve the position and effect of individual QTL. These two parameters, position and effect, are central components of many models of how evolution proceeds in nature. The relative position (or linkage) of QTL is an important parameter that contributes to our understanding of many evolutionary processes (e.g. Barton & Hewitt 1985; Howard *et al.* 2002). Likewise, the distribution of effects of QTL can strongly influence the approach of a population to a multilocus fitness optimum (Orr 1995), the rate at which a population responds to selection (Lande 1975; Barton & Turelli 1987), and can elucidate the contribution of additive effects relative to dominance, epistatic and environmental influences upon phenotype. QTL analyses can rapidly identify candidate loci that may be examined further — as has been done in maize evolution (Doebley 1992). The limitations of QTL analyses to identify the position and effect of underlying QTL typically reside in the limitation in our ability to measure and genotype test populations of finite size. This is particularly true for loci that show interactions, where number of genotypic categories to examine is large, which can either exceed the sample population size or diminish the power to observe interaction effects as they occur. However, QTL analyses can empower one to develop more specific tests, based on different crosses or candidate gene approaches. For example, recent work in *Mimulus* demonstrated that a single gene — or closely linked loci — identified by QTL analysis, contributes to pollinator discrimination of related species based on flower colour (Bradshaw & Schemske 2003), perhaps allowing the quantification of selection acting on a single locus via a known selective agent. The technology of QTL mapping has already allowed evolutionary biologists to address central questions regarding the molecular basis of evolution in natural species, while further developments in the type and availability of genetic markers and statistical analyses will only broaden the scope of questions that can be addressed in the future. In the short-term, model systems may represent the best opportunity to utilize QTL approaches in the complete resolution of genetic architecture underlying evolutionary relevant traits.

## Acknowledgements

The manuscript was greatly improved by comments from P. Danley, L. Rieseberg, M. Rutter, K. Shaw and B. Walsh and the contributions of three anonymous reviewers. We appreciate the support of funding from the US NSF (9815780 to C. Fenster, 9972366 to D. Price), US NIH (5S06GM008073-24 to D. Price) and the Research Council of Norway (134800/410 to C. Fenster and H. Stenøien) and Swedish Research Council (grant no. 621-2002-5896 to H. Stenøien). This manuscript also benefited from interactions

with L. Zimmer at the Smithsonian Institution's Laboratory of Analytical Biology. All of the authors benefited from participation in the N. C. State Summer Institute for Genetics.

## References

- Albertson RC, Strelman JT, Kocher TD (2003) Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proceedings of the National Academy of Sciences of the USA*, **100**, 5252–5257.
- Almasy L, Blangero J (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. *American Journal of Human Genetics*, **62**, 1198–1211.
- Almasy L, Hixson JE, Rainwater DL *et al.* (1999) Human pedigree-based quantitative trait locus mapping: localization of two genes influencing HDL cholesterol metabolism. *American Journal of Human Genetics*, **64**, 1686–1693.
- Amos CI, Elston RC (1989) Robust methods for the detection of genetic linkage for quantitative data from pedigree. *Genetic Epidemiology*, **6**, 349–360.
- Axelsson T, Shavorskaya O, Lagercrantz U (2001) Multiple flowering time QTLs within several *Brassica* species could be the result of duplicated copies of one ancestral gene. *Genome*, **44**, 856–864.
- Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113–148.
- Barton NH, Keightley PD (2002) Understanding quantitative genetic variation. *Nature Reviews, Genetics*, **3**, 11–21.
- Barton NH, Turelli M (1987) Adaptive landscapes, genetic-distance and the evolution of quantitative characters. *Genetical Research*, **49**, 157–173.
- Basten CJ, Weir BS, Zeng ZB (2001) *QTL Cartographer*, Version 1.15. Department of Statistics, North Carolina State University, Raleigh.
- Beadle GW (1980) Ancestry of corn. *Scientific American*, **242**, 112–119.
- Beavis WD (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. In: *Proceedings of the Forty-Ninth Annual Corn and Sorghum Industry Research Conference* (ed. Wilkinson DB), pp. 250–266. American Seed Trade Association, Chicago, IL.
- Beavis WD (1998) QTL analyses: power, precision and accuracy. In: *Molecular Dissection of Complex Traits* (ed. Paterson AH), pp. 145–162. CRC Press, New York.
- Berry CC (1998) Computationally efficient Bayesian QTL mapping in experimental crosses. *ASA Proceedings of the Biometrics Section*, pp. 164–169. American Statistical Association, Alexandria, VA, USA.
- Bink MCAM, Uimari P, Sillanpaa MJ, Janss LLG, Jansen RC (2002) Multiple QTL mapping in related plant populations via a pedigree-analysis approach. *Theoretical and Applied Genetics*, **104**, 751–762.
- Boake CRB, Arnold SJ, Breden F *et al.* (2002) Genetic tools for studying adaptation and the evolution of behavior. *American Naturalist*, **160** (Suppl. S), S143–S159.
- Borevitz JO, Liang D, Plouffe D *et al.* (2003) Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Research*, **13**, 513–523.
- Bradshaw HD, Otto KE, Frewen BE *et al.* (1998) Quantitative trait loci affecting differences in floral morphology between two species of monkeyflower (*Mimulus*). *Genetics*, **149**, 367–382.
- Bradshaw HD, Schemske DW (2003) Allele substitution at a flower color locus produces a pollinator shift in monkeyflowers. *Nature*, **426**, 176–178.
- Bradshaw HD, Wilbert JR, Otto KG *et al.* (1995) Genetic mapping of floral traits associated with reproductive isolation in monkeyflowers (*Mimulus*). *Nature*, **376**, 762–765.
- Buckler ES, Thornsberry JM (2002) Plant molecular diversity and applications to genomics. *Current Opinion in Plant Biology*, **5**, 107–111.
- Burke JM, Tang MS, Knapp SJ *et al.* (2002) Genetic analysis of sunflower domestication. *Genetics*, **161**, 1257–1267.
- Cardon LR (2000) A sib-pair regression model of linkage disequilibrium for quantitative traits. *Human Heredity*, **50**, 350–358.
- Carlborg O, Andersson L (2002) Use of randomization testing to detect multiple epistatic QTLs. *Genetical Research*, **79**, 175–184.
- Carlborg O, Kerje S, Schutz K *et al.* (2003) A global search reveals epistatic interaction between QTL for early growth in the chicken. *Genome Research*, **13**, 413–421.
- Carr DE, Dudash MR (2003) Recent approaches into the genetic basis of inbreeding depression in plants. *Philosophical Transactions of the Royal Society, Series B*, **358**, 1071–1084.
- Charlesworth B, Charlesworth D (1999) The genetic basis of inbreeding depression. *Genetical Research*, **74**, 329–340.
- Charlesworth D, Charlesworth B (1987) Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics*, **18**, 237–268.
- Cheverud JM, Routman E (1993) Quantitative trait loci – individual gene effects on quantitative characters. *Journal of Evolutionary Biology*, **6**, 463–480.
- Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. *Genetics*, **138**, 963–971.
- Civetta A, Cantor EJJ (2003) The genetics of mating recognition between *Drosophila simulans* and *D-sechellia*. *Genetical Research*, **82**, 117–126.
- Cockerham CC, Zeng ZB (1996) Design III with marker loci. *Genetics*, **143**, 1437–1456.
- Culverhouse R, Suarez BK, Lin J *et al.* (2002) A perspective on epistasis: limits of models displaying no main effects. *American Journal of Human Genetics*, **70**, 461–471.
- Davis GP, Hetzel DJS, Corbet NJ *et al.* (1998) The mapping of quantitative trait loci for birth weight in a tropical beef herd. In: *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, Armidale, Australia*, **25**, 441–444.
- Dobzhansky T (1937) *Genetics of the Evolutionary Process*. Columbia University Press, New York.
- Doebley J (1992) Mapping the genes that made maize. *Trends in Genetics*, **8**, 302–307.
- Doebley J, Stec A, Gustus C (1995b) Teosinte branched1 and the origin of maize – evidence for epistasis and the evolution of dominance. *Genetics*, **141**, 333–346.
- Doebley J, Stec A, Kent B (1995a) Suppressor of sessile spikelets1 (SOS1) – a dominant mutant affecting inflorescence development in maize. *American Journal of Botany*, **82**, 571–577.
- Doerge RW (2002) Mapping and analysis of quantitative trait loci in experimental populations. *Nature Reviews Genetics*, **3**, 43–52.
- Doerge RW, Churchill GA (1996) Permutation tests for multiple loci affecting a quantitative character. *Genetics*, **142**, 285–294.
- Doerge RW, Zeng Z-B, Weir BS (1997) Statistical issues in the search for genes affecting quantitative traits in experimental populations. *Statistical Science*, **12**, 195–219.
- Drigalenko E (1998) How sib pairs reveal linkage. *American Journal of Human Genetics*, **63**, 1242–1245.
- Dudash MR, Carr DE (1998) The genetics underlying inbreeding depression in *Mimulus* with contrasting mating systems. *Nature*, **393**, 682–684.

- Dupuis J, Siegmund D (1999) Statistical methods for mapping quantitative trait loci from a dense set of markers. *Genetics*, **151**, 373–386.
- Elston RC, Buxbaum S, Jacobs KB, Olsen JM (2000) Haseman and Elston revisited. *Genetic Epidemiology*, **19**, 1–17.
- Falconer DS, Mackay TFC (1995) *Introduction to Quantitative Genetics*, 4th edn. Addison-Wesley Longman, Harlow, UK.
- Fedorowicz GM, Fry JD, Anholt RRH *et al.* (1998) Epistatic interactions between smell-impaired loci in *Drosophila melanogaster*. *Genetics*, **148**, 1885–1891.
- Fenster CB, Diggel PK, Ritland K, Barrett SCH (1995) The genetics of floral development differentiating 2 species of *Mimulus* (Scrophulariaceae). *Heredity*, **74**, 258–266.
- Fenster CB, Galloway L, Chao L (1997) Epistasis and its consequences for the evolution of natural populations. *Trends in Ecology and Evolution*, **12**, 282–286.
- Fenster CB, Galloway LF (2000) Population differentiation in an annual legume: genetic architecture. *Evolution*, **54**, 1157–1172.
- Fenster CB, Ritland K (1994) Quantitative genetics of mating system divergence in the yellow monkeyflower species complex. *Heredity*, **73**, 422–435.
- Fijneman RJA, deVries SS, Jansen RC *et al.* (1996) Complex interactions of new quantitative trait loci, *Sluc1*, *Sluc2*, *Sluc3*, and *Sluc4*, that influence the susceptibility to lung cancer in the mouse. *Nature Genetics*, **14**, 465–467.
- Fisher RA (1958) *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford.
- Fishman L, Kelly AJ, Willis JH (2002) Minor QTLs underlie floral traits associated with mating system divergence in *Mimulus*. *Evolution*, **56**, 2138–2155.
- Flint J, Mott R (2001) Finding the molecular basis of quantitative traits: successes and pitfalls. *Nature Reviews Genetics*, **2**, 437–445.
- Frary A, Nesbitt TC, Frary A *et al.* (2000) *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. *Science*, **289**, 85–88.
- Fu YB, Ritland K (1994) Marker-based inferences about fecundity genes contributing to inbreeding depression in *Mimulus guttatus*. *Genome*, **37**, 1005–1010.
- Fulker DW, Cardon LR (1994) A sib-pair approach to interval mapping of quantitative trait loci. *American Journal of Human Genetics*, **54**, 1092–1103.
- Fulker DW, Cherny SS, Cardon LR (1995) Multipoint interval mapping of quantitative trait loci, using sib pairs. *American Journal of Human Genetics*, **56**, 1224–1233.
- Gadau J, Page RE, Werren JH (2002) The genetic basis of the interspecific differences in wing size in *Nasonia* (Hymenoptera; Pteromalidae): major quantitative trait loci and epistasis. *Genetics*, **161**, 673–684.
- Galloway LF, Fenster CB (1999) The effect of nuclear and cytoplasmic genes on fitness and local adaptation in an annual legume. *Evolution*, **53**, 1734–1743.
- George AW, Visscher PM, Haley CS (2000) Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach. *Genetics*, **156**, 2081–2092.
- Gibson G (2002) Microarrays in ecology and evolution: a preview. *Molecular Ecology*, **11**, 17–24.
- Gibson G, Mackay TFC (2002) Enabling population and quantitative genomics. *Genetical Research*, **80**, 1–6.
- Gleason JM, Nuzhdin SV, Ritchie MG (2002) Quantitative trait loci affecting a courtship signal in *Drosophila melanogaster*. *Heredity*, **89**, 1–6.
- Graham GI, Wolff DW, Stuber CW (1997) Characterization of a yield quantitative trait locus on chromosome five of maize by fine mapping. *Crop Science*, **37**, 1601–1610.
- Gravrilets S, Hastings A (1996) Founder effect speciation: a theoretical reassessment. *American Naturalist*, **147**, 466–491.
- Groover A, Devey T, Fiddler M *et al.* (1994) Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. *Genetics*, **138**, 1293–1306.
- Grosz MD, MacNeil MD (2001) Putative quantitative trait locus affecting birth weight on bovine chromosome 2. *Journal of Animal Science*, **79**, 68–72.
- Haley CS, Knott SA, Elsen JM (1994) Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics*, **136**, 1195–1207.
- Hamilton WD, Zuk M (1989) Parasites and sexual selection – reply. *Nature*, **341**, 289–290.
- Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. *Behavior Genetics*, **2**, 3–19.
- Hawthorne DJ, Via S (2001) Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature*, **412**, 904–907.
- Heisey WM, Nobs MA, Bjorkman O (1971) Experimental studies on the nature of species. V. Biosystematics, genetics, and physiological ecology of the erythranthe section of *Mimulus*. *Carnegie Institute of Washington Publication*, **628**.
- Hill WG, Robertson A (1966) The effect of linkage on limits to artificial selection. *Genetical Research*, **8**, 269–294.
- Howard DJ, Marshall JL, Hampton DD *et al.* (2002) The genetics of reproductive isolation: a retrospective and prospective look with comments on ground crickets. *American Naturalist*, **159** (Suppl. S), S8–S21.
- Jannink JL, Jansen RC (2001) Mapping epistatic quantitative trait loci with one-dimensional genome searches. *Genetics*, **157**, 445–454.
- Jansen RC (1996) A general Monte Carlo method for mapping multiple quantitative trait loci. *Genetics*, **142**, 305–311.
- Jiang CX, Chee PW, Draye X *et al.* (2000) Multilocus interactions restrict gene introgression in interspecific populations of polyploid *Gossypium* (cotton). *Evolution*, **54**, 798–814.
- Karkkainen K, Savolainen O, Koski V (1999) Why do plants abort so many developing seeds: bad offspring or bad maternal genotypes? *Evolutionary Ecology*, **13**, 305–317.
- Kao CH, Zeng ZB, Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. *Genetics*, **152**, 1203–1216.
- Kim SC, Rieseberg LH (1999) Genetic architecture of species differences in annual sunflowers: implications for adaptive trait introgression. *Genetics*, **153**, 965–977.
- Kingsolver JG, Gomulkiewicz R, Carter PA (2001) Variation, selection and evolution of function-valued traits. *Genetica*, **112**, 87–104.
- Knott SA, Haley CS (1998) Simple multiple-marker sib-pair analysis for mapping quantitative trait loci. *Heredity*, **81**, 48–54.
- Knott SA, Marklund L, Haley CS *et al.* (1998) Multiple marker mapping of quantitative trait loci in an outbred cross between wild boar and large white pigs. *Genetics*, **149**, 1069–1080.
- Kokko H (2001) Fisherian and ‘good genes’ benefits of mate choice: how (not) to distinguish between them. *Ecology Letters*, **4**, 322–326.
- Kuang H, Richardson T, Carson S, Wilcox P, Bongarten B (1999) Genetic analysis of inbreeding depression in plus tree 850.55 of *Pinus radiata* D. Don. I. Genetic map with distorted markers. *Theoretical and Applied Genetics*, **98**, 697–703.

- Lande R (1975) The maintenance of genetic variation by mutation in a polygenic character with linked loci. *Genetic Research*, **26**, 221–235.
- Lande R, Arnold SJ (1985) Evolution of mating preference and sexual dimorphism. *Journal of Theoretical Biology*, **133**, 651–664.
- Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, **121**, 185–199.
- Lander ES, Kruglyak I (1995) Genetic dissection of complex traits – guidelines for interpreting and reporting linkage results. *Nature Genetics*, **11**, 241–247.
- Lexer C, Lai Z, Rieseberg LH (2004) Candidate gene polymorphisms associated with salt tolerance in wild sunflower hybrids: implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. *New Phytologist*, **161**, 225–233.
- Lexer C, Welch ME, Durphy JL *et al.* (2003) Natural selection for salt tolerance quantitative trait loci (QTLs) in wild sunflower hybrids: Implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. *Molecular Ecology*, **12**, 1225–1235.
- Li ZK, Luo LJ, Mei HW *et al.* (2001) Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. *Genetics*, **158**, 1737–1753.
- Li ZK, Pinson SRM, Park WD *et al.* (1997) Epistasis for three grain yield components in rice (*Oryza sativa* L.). *Genetics*, **145**, 453–456.
- Liu BH (1998) *Genomics, Linkage Mapping and QTL Analysis*. CRC Press, Boca Raton, FL.
- Liu HC, Cheng HH, Tirunagaru V *et al.* (2001) A strategy to identify positional candidate genes conferring Marek's disease resistance by integrating DNA microarrays and genetic mapping. *Animal Genetics*, **32**, 351–358.
- Lubberstedt T, Klein D, Melchinger AE (1998a) Comparative QTL mapping of resistance to *Ustilago maydis* across four populations of European flint-maize. *Theoretical and Applied Genetics*, **97**, 1321–1330.
- Lubberstedt T, Melchinger AE, Fahr S *et al.* (1998b) QTL mapping in testcrosses of flint lines of maize. III. Comparison across populations for forage traits. *Crop Science*, **38**, 1278–1289.
- Lukens LN, Doebley J (1999) Epistatic and environmental interactions for quantitative trait loci involved in maize evolution. *Genetical Research*, **74**, 291–302.
- Luo LJ, Li ZK, Mei HW *et al.* (2001) Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. II. Grain yield components. *Genetics*, **158**, 1755–1771.
- Lynch M, Walsh B (1998) *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.
- Mackay TFC (2001) The genetic architecture of quantitative traits. *Annual Review Genetics*, **35**, 303–339.
- Macnair MR, Cumbes QJ (1989) The genetic architecture of interspecific variation in *Mimulus*. *Genetics*, **122**, 211–222.
- Mauricio M (2001) Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. *Nature Reviews Genetics*, **2**, 370–381.
- Mayr E (1942) *Systematics and the Origin of Species*. Columbia University Press, New York.
- Mezey JG, Cheverud JM, Wagner GP (2000) Is the genotype-phenotype map modular?: a statistical approach using mouse quantitative trait loci data. *Genetics*, **156**, 305–311.
- Mitchell-Olds T (1995) The molecular basis of quantitative genetic variation in natural populations. *Trends in Ecology and Evolution*, **10**, 324–328.
- Moritz DML, Kadereit JW (2001) The genetics of evolutionary change in *Senecio vulgaris* L.: a QTL mapping approach. *Plant Biology*, **3**, 544–552.
- Mott R, Flint J (2002) Simultaneous detection and fine mapping of quantitative trait loci in mice using heterogeneous stocks. *Genetics*, **160**, 1609–1618.
- Mott R, Talbot CJ, Turri MG *et al.* (2000) A method for fine mapping quantitative trait loci in outbred animal stocks. *Proceedings of the National Academy of the USA*, **97**, 12649–12654.
- Muller HJ (1940) Bearing of the *Drosophila* work on systematics. In: *The New Systematics* (ed. Huxley J), pp. 185–268. Clarendon Press, Oxford.
- Nadeau JH, Frankel WN (2000) The roads from phenotypic variation to gene discovery: mutagenesis versus QTLs. *Nature Genetics*, **25**, 381–384.
- Nagamine Y, Haley CS (2001) Using the mixed model for interval mapping of quantitative trait loci in outbred line crosses. *Genetical Research*, **77**, 199–207.
- Nagamine Y, Haley CS, Sewalem A, Visscher PM (2003) Quantitative trait loci variation for growth and obesity between and within lines of pigs (*Sus scrofa*). *Genetics*, **164**, 629–635.
- Orr HA (1995) The population-genetics of speciation – The evolution of hybrid incompatibilities. *Genetics*, **139**, 1805–1813.
- Orr HA (1998) Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics*, **149**, 2099–2104.
- Orr HA (2001) The genetics of species differences. *Trends in Ecology and Evolution*, **16**, 343–350.
- Österberg MK, Shavorskaya O, Lascoux M *et al.* (2002) Naturally occurring indel variation in the *Brassica nigra* COL1 gene is associated with variation in flowering time. *Genetics*, **161**, 299–306.
- Paran I, Zamir D (2003) Quantitative traits in plants: beyond the QTL. *Trends in Genetics*, **19**, 303–306.
- Paterson AH, Lin YR, Li ZK *et al.* (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science*, **269**, 1714–1718.
- Patterson AH (1998) *Molecular Dissection of Complex Traits*. CRC Press, New York.
- Payne F (1918) The effect of artificial selection on bristle number in *Drosophila ampelophila* and its interpretation. *Proceedings of the National Academy of Sciences of the USA*, **4**, 55–58.
- Perez-Enciso M (2003) Fine mapping of complex trait genes combining pedigree and linkage disequilibrium information: a Bayesian unified framework. *Genetics*, **163**, 1497–1510.
- Petrie M (1994) Improved growth and survival of offspring of peacocks with more elaborate trains. *Nature*, **371**, 598–599.
- Price DK, Burley NT (1994) Constraints on the evolution of attractive traits: selection in male and female zebra finches. *The American Naturalist*, **144**, 908–934.
- Provine WB (1971) *The Origins of Theoretical Population Genetics*. University of Chicago Press, Chicago.
- Pruitt RE, Bowman JL, Grossniklaus U (2003) Plant genetics: a decade of integration. *Nature Genetics*, **33** (Suppl. S), 294–304.
- Remington DL, O'Malley DM (2000a) Whole-genome characterization of embryonic stage inbreeding depression in a selfed loblolly pine family. *Genetics*, **155**, 337–348.
- Remington DL, O'Malley DM (2000b) Evaluation of major genetic loci contributing to inbreeding depression for survival and early growth in a selfed family of *Pinus taeda*. *Evolution*, **54**, 1580–1589.
- Remington DL, Purugganan MD (2003) Candidate genes, quantitative trait loci, and functional trait evolution in plants. *International Journal of Plant Science*, **164**, S7–S20.



- Remington DL, Ungerer MC, Purugganan MD (2001) Map-based cloning of quantitative trait loci: progress and prospects. *Genetical Research*, **78**, 213–218.
- Rice WR, Hostert EE (1993) Laboratory experiments on speciation – what have we learned in 40 years. *Evolution*, **47**, 1637–1653.
- Rieseberg LH, Raymond O, Rosenthal DM (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*, **301**, 1211–1216.
- Rieseberg LH, Sinervo B, Linder CM, Ungerer M, Arias DM (1996) Role of gene interactions in hybrid speciation: evidence from ancient and experimental hybrids. *Science*, **272**, 741–745.
- Rieseberg LH, Whitton J, Gardner K (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*, **152**, 713–727.
- Rieseberg LH, Widmer A, Arntz AM *et al.* (2002) Directional selection is the primary cause of phenotypic diversification. *Proceedings of the National Academy of Sciences of the USA*, **99**, 12242–12245.
- Satagopan JM, Yandell BS, Newton MA, Osborn TC (1996) Markov chain Monte Carlo approach to detect polygene loci for complex traits. *Genetics*, **144**, 805–816.
- Sax K (1923) The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics*, **8**, 552–560.
- Schrooten C, Bovenhuis H (2002) Detection of pleiotropic effects of quantitative trait loci in outbred populations using regression analysis. *Journal of Dairy Science*, **85**, 3503–3513.
- Seaton G, Haley CS, Knott SA, Kearsey M, Visscher PM (2002) QTL EXPRESS: mapping quantitative trait loci in simple and complex pedigrees. *Bioinformatics*, **18**, 339–340.
- Sen S, Churchill GA (2001) A statistical framework for quantitative trait mapping. *Genetics*, **159**, 371–387.
- Shaw KL (1996) Polygenic inheritance of a behavioral phenotype: interspecific genetics of song in the Hawaiian cricket genus *Laupala*. *Evolution*, **50**, 256–266.
- Shaw KL, Parsons YM (2002) Divergence of mate recognition behavior and its consequences for genetic architectures of speciation. *American Naturalist*, **159**, S61–S75.
- Shook DR, Johnson TE (1999) Quantitative trait loci affecting survival and fertility-related traits in *Caenorhabditis elegans* show genotype–environment interactions, pleiotropy and epistasis. *Genetics*, **153**, 1233–1243.
- Shrimpton AE, Robertson A (1988) The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster*. 2. Distribution of 3rd chromosome bristle effects within chromosome sections. *Genetics*, **118**, 445–459.
- Sillanpaa MJ, Arjas E (2000) Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. *Genetics*, **148**, 1373–1388.
- Slate J, Pemberton JM, Visscher PM (1999) Power to detect QTL in a free-living polygynous population. *Heredity*, **83**, 327–336.
- Slate J, Visscher PM, MacGregor S *et al.* (2002) A genome scan for quantitative trait loci in a wild population of red deer (*Cervus elaphus*). *Genetics*, **162**, 1863–1873.
- Stone RT, Keele JW, Shackelford SD *et al.* (1999) A primary screen of the bovine genome for quantitative trait loci affecting carcass and growth traits. *Journal of Animal Science*, **77**, 1379–1384.
- Stuart RO, Bush KT, Nigam SK (2003) Changes in gene expression patterns in the ureteric bud and metanephric mesenchyme in models of kidney development. *Kidney international*, **64**, 1997–2008.
- 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**, 823–839.
- Streelman JT, Albertson RC, Kocher TC (2003) Genome mapping of the orange blotch colour pattern in cichlid fishes. *Molecular Ecology*, **12**, 2465–2471.
- Tao WJ, Boulding EG (2003) Associations between single nucleotide polymorphisms in candidate genes and growth rate in Arctic charr (*Salvelinus alpinus* L.). *Heredity*, **91**, 60–69.
- Thoday JM (1961) Location of polygenes. *Nature*, **191**, 368–369.
- Ungerer MC, Halldorsdottir SS, Purugganan MD, Mackay TF (2003) Genotype–environment interactions at quantitative trait loci affecting inflorescence development in *Arabidopsis thaliana*. *Genetics*, **165**, 353–365.
- Uyenoyama MK, Waller DM (1991) Coevolution of self-fertilization and inbreeding depression 2. Symmetrical overdominance in viability. *Theoretical Population Biology*, **40**, 47–77.
- Verhoeven KJF, Vanhala TK, Biere A *et al.* (2004) The genetic basis of adaptive population differentiation: a quantitative trait locus analysis of fitness traits in two wild barley populations from contrasting habitats. *Evolution*, **58**, 270–283.
- Vos P, Hogers R, Bleeker M, Reijmans M *et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.
- Wade MJ (2001) Epistasis, complex traits, and mapping genes. *Genetica*, **112**, 59–69.
- Wagner GP, Altenberg L (1996) Complex adaptations and the evolution of evolvability. *Evolution*, **50**, 967–976.
- Wayne ML, McIntyre LM (2002) Candidate gene discover: combining mapping and microarrays in the genetic analysis of an adaptive trait in *Drosophila*. *Proceedings of the National Academy of Sciences of the USA*, **99**, 14903–14906.
- Weinig C, Dorn LA, Kane NC *et al.* (2003b) Heterogeneous selection at specific loci in natural environments in *Arabidopsis thaliana*. *Genetics*, **165**, 321–329.
- Weinig C, Stinchcombe JR, Schmitt J (2003a) QTL architecture of resistance and tolerance traits in *Arabidopsis thaliana* in natural environments. *Molecular Ecology*, **12**, 1153–1163.
- Weinig C, Ungerer MC, Dorn LA *et al.* (2002) Novel loci control variation in reproductive timing in *Arabidopsis thaliana* in natural environments. *Genetics*, **162**, 1875–1884.
- Welch AM (2003) Genetic benefits of a female mating preference in gray tree frogs are context-dependent. *Evolution*, **57**, 883–893.
- Weller I, Kashi Y, Soller M (1990) Power of daughter and grand-daughter designs for determining the linkage between marker loci and quantitative traits in dairy cattle. *Journal of Dairy Science*, **73**, 2525–2537.
- Whitlock MC, Phillips PC, Moore FBG, Tonsor SJ (1995) Multiple fitness peaks and epistasis. *Annual Review of Ecology and Systematics*, **26**, 601–629.
- Williams CG (1998) QTL mapping in outbred pedigrees. In: *Molecular Dissection of Complex Traits* (ed. Paterson AH), pp. 81–94. CRC Press, New York.
- Willis JH (1999) The role of genes of large effect on inbreeding depression in *Mimulus guttatus*. *Evolution*, **53**, 1678–1691.
- Wolf JB, Brodie ED, Wade MJ (2000) *Epistasis and the Evolutionary Process*. Oxford University Press, New York.
- Workman MS, Leamy LJ, Routman EJ *et al.* (2002) Analysis of quantitative trait locus effects on the size and shape of mandibular molars in mice. *Genetics*, **160**, 1573–1586.
- Xu S (2003a) Theoretical basis of the Beavis effect. *Genetics*, **165**, 2259–2268.

- Xu S (2003b) Estimating polygenic effects using markers of the entire genome. *Genetics*, **163**, 789–801.
- Yan JQ, Zhu J, He CX *et al.* (1998) Molecular dissection of developmental behavior of plant height in rice (*Oryza sativa* L.). *Genetics*, **150**, 1257–1265.
- Yano M, Katayose Y, Ashikari M *et al.* (2000) *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabi-dopsis* flowering time gene *CONSTANS*. *Plant Cell*, **12**, 2473–2483.
- Yi N, Xu S, Allison DB (2003) Bayesian model choice and search strategies for mapping interacting quantitative trait loci. *Genetics*, **165**, 867–883.
- Zeng ZB (1994) Precision mapping of quantitative trait loci. *Genetics*, **136**, 1457–1468.
- Zeng ZB, Kao CH, Basten CJ (1999) Estimating the genetic architecture of quantitative traits. *Genetical Research*, **74**, 279–289.
- Zhang WK, Wang YJ, Luo GZ *et al.* (2004) QTL mapping of ten agronomic traits on the soybean (*Glycine max* L. Merr.) genetic map and their association with EST markers. *Theoretical and Applied Genetics*, **108**, 1131–1139.

## Appendix

**AFLP:** amplified fragment length polymorphisms. Vos *et al.* (1995) described what is essentially a polymerase chain reaction (PCR)-based multiallelic random fragment length polymorphism (RFLP). Individual fragments of DNA are generated through DNA restriction and then PCR amplified to produce very many PCR fragments of different size. The difference in the number and size of the resulting fragments can be interpreted as different loci.

**Genetic architecture:** A description of the number and mode of action (degree of dominance, epistasis or pleiotropy) of loci affecting a trait.

**IBD:** alleles that are identical in state (either size or sequence) and which are derived from a single, common ancestor.

**Pleiotropy:** The phenomenon where a single locus affects multiple traits.

**Epistasis:** The phenomena where the expression of a gene or an allele is affected by alleles present at other loci.

**LOD score:** A LOD score is the likelihood (or log of the odds) that a QTL resides at a specific marker interval. The

---

David Erickson is a postdoctoral fellow at the Smithsonian Institution. He is interested in questions of genetic architecture affecting fitness as well as issues in ecological genetics including gene flow and phylogeography. Charlie Fenster is in the Department of Biology at the University of Maryland. His research focuses on quantifying the modes of microevolutionary process, including gene flow, natural selection and mutation, using plants as model systems as well as determining the genetic architecture underlying fitness differentiation and floral evolution. Hans Stenoien is an assistant professor at Uppsala University. His research on evolution at the molecular level spans from primitive land plants, mosses, to the angiosperm *Arabidopsis thaliana*, addressing questions focused at quantifying gene flow and targets of natural selection including the flowering time pathway in *A. thaliana*. Donald K. Price is an evolutionary biologist at the University of Hawaii with interests in the genetics and ecology of behaviors and speciation in birds and insects. Currently he is investigating quantitative genetics of Hawaiian picture-winged flies and the genetics of Nene, the last remaining Hawaiian goose.

---

LOD score is actually a constant times the likelihood ratio statistic, which itself is the ratio of the likelihood that a QTL resides within a given marker interval divided by the likelihood that the observed genotype-phenotype relationship occurs solely by chance.

**Microarray:** A method to simultaneously measure the direct expression of many genes via levels of messenger RNA (mRNA). This is essentially a multilocus Northern blot, where hybridization of two sets of mRNA (usually a control RNA set and an experimental set), reveal differences in mRNA level, and hence gene expression at many loci simultaneously.

**Microsatellite:** PCR-based markers that reflect the number of repetitive DNA motifs between a pair of primers (e.g. a *GA10* allele vs. a *GA12* allele that is 4 bp larger). Tend to be highly polymorphic and exhibit codominant expression.

**RAPD:** randomly amplified polymorphic DNA. This is a PCR-based marker in which small (~10 bp) primers that are not specific to a chromosomal region are used to amplify randomly chosen segments of genomic DNA. Polymorphism is detected through the presence/absence of a PCR fragment at a given size.