## Invited Review

# Identifying the genes underlying quantitative traits: a rationale for the QTN programme

Young Wha Lee[1,†], Billie A. Gould[1,†] and John R. Stinchcombe[1,2]*

[1] Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, ON, Canada M5S 3B2
[2] Centre for the Analysis of Genome Evolution and Function, University of Toronto, Toronto, ON, Canada M5S 3B2

**Abstract.** The goal of identifying the genes or even nucleotides underlying quantitative and adaptive traits has been characterized as the 'QTN programme' and has recently come under severe criticism. Part of the reason for this criticism is that much of the QTN programme has asserted that finding the genes and nucleotides for adaptive and quantitative traits is a fundamental goal, without explaining *why* it is such a hallowed goal. Here we outline motivations for the QTN programme that offer general insight, regardless of whether QTNs are of large or small effect, and that aid our understanding of the mechanistic dynamics of adaptive evolution. We focus on five areas: (i) vertical integration of insight across different levels of biological organization, (ii) genetic parallelism and the role of pleiotropy in shaping evolutionary dynamics, (iii) understanding the forces maintaining genetic variation in populations, (iv) distinguishing between adaptation from standing variation and new mutation, and (v) the role of genomic architecture in facilitating adaptation. We argue that rather than abandoning the QTN programme, we should refocus our efforts on topics where molecular data will be the most effective for testing hypotheses about phenotypic evolution.

**Keywords:** Adaptation; ecological genomics; ecologically important traits; genetic variation; phenotypic evolution; population genomics; QTL; QTN; quantitative genetics; vertical integration.

## Introduction

A dominant goal of contemporary evolutionary genetics has been to describe, at the molecular level, the loci responsible for adaptations and complex phenotypes. While this goal first gained steam in non-model systems with the widespread adoption of quantitative trait locus (QTL) mapping in the late 1990s, the recent explosion of sequencing technologies is likely to only heighten interest in this goal. New sequencing technologies, analytical approaches and statistical methods now offer the prospect of detecting the nucleotide variants responsible for variation in quantitative traits—an agenda Rockman (2012)

termed the QTN programme. It is easy to understand the excitement: adaptive phenotypes previously only understood statistically with variances, covariances and linear algebra might soon be understood as a collection of single-nucleotide polymorphisms (SNPs) whose inheritance, allele frequencies and evolutionary dynamics could be understood with the basic principles of population and transmission genetics.

Many, but not all, papers in contemporary ecological and evolutionary genetics take it for granted that describing the molecular basis of adaptations is a worthy goal— or simply assert that it is a fundamental goal—without providing an explanation for *why it is a worthwhile goal*.

[†] These authors contributed equally, and order was determined by coin flip.

* Corresponding author's e-mail address: john.stinchcombe@utoronto.ca

An important consideration for any QTN-based research is: What are the benefits of knowing the *specific* genes, genetic regions or nucleotides responsible for adaptive phenotypes above and beyond the generic knowledge that there must be *some unknown* genes, genetic regions or nucleotides that contribute? Here we aim to provide answers to this question, and as such a rationale for the QTN programme.

## Criticisms of the QTN Programme

Rockman (2012) strongly criticizes several aspects of the QTN programme, largely on the basis of logical flaws in how the results are interpreted. Most of the methods for detecting QTNs—QTL mapping, candidate gene association studies, genome-wide association studies (GWAS) and others—are strongly biased towards detecting nucleotide polymorphisms with a large effect on phenotype and overestimating their phenotypic consequences (i.e. the Beavis effect, see Slate 2013). As such, he argues that generality from QTN studies is likely to be elusive: the methods we use predispose us to find large-effect variants, giving a biased picture of their importance. Rockman argues persuasively that neither theory nor data support the notion that large-effect variants are the primary loci of adaptation, and that existing data suggest that many large-effect loci are likely to be qualitatively different at the molecular level than small-effect loci. Accordingly, Rockman argues that if the loci that we are able to find are a biased sample of the universe of QTNs, any conclusions we make about the evolutionary and genetic phenomena based on them are also likely to be biased.

Travisano and Shaw (2013) criticize the QTN programme on a more fundamental level. They argue that searching for QTNs holds little value in that such research focuses on patterns (e.g. genomic signatures of selection or statistical associations between sequence variants and phenotype) rather than the evolutionary process itself. Travisano and Shaw (2013) argue that intensive molecular studies '… have not altered fundamental understanding of the relationship between genotype and phenotype …' nor yielded the ability to predict organismal evolution in response to selection. They argue that QTN searches are thus less worthy of investment than studies that are mechanistic in nature.

## What Now?

We suggest that the challenge posed by Rockman is one of properly interpreting the generality and inference space of QTN results. A dissecting scope and scanning electron microscope both provide enhanced magnification for studying biological features: one is useful for studying external morphology and the other for cellular and sub-cellular structures. The inability to describe cellular morphology with a dissecting scope does not imply that we should discard it, but rather only use it for tasks where it is appropriate. Are the existing QTNs—which Rockman argues are primarily (unrepresentative) large-effect loci—so well characterized that we can stop studying them? Below, we describe a rationale for why we should characterize the genetics, ecology and evolutionary dynamics of QTNs and what we can learn from them. While the smallest-effect, infinitesimal QTN might be beyond our reach right now (or perhaps, always), we seek to clarify what one can learn from any QTN and why they are worth pursuing.

We also endorse Travisano and Shaw's (2013) call for mechanistic studies of natural selection and adaptation that lead to changes in phenotypes. We see great value in this approach, and see no reason why it should not be paired with molecular genetic studies. The challenge, as we see it, is to design genetic studies to clarify aspects of the evolutionary process that are unavailable at the purely phenotypic level, and can enhance evolutionary interpretations.

## A Rationale for the QTN Programme

Numerous reasons to seek QTNs exist, and here we explore in detail those that are most compelling to us. Throughout, we focus on work in natural (rather than agricultural or experimental) systems, seeking to highlight plant examples whenever possible. A summary of the strengths and weaknesses of the various approaches can be found in Table 1, and a schematic on the relative precision of alternative approaches in Fig. 1. We see at least five compelling reasons to seek QTNs: (i) understanding adaptive phenotypes across many levels of biological organization, from the nucleotide to the ecological and community context, a feature we refer to as *vertical integration* (M. A. Bell, pers. comm.); (ii) understanding genetic parallelism and the role of pleiotropy in constraining adaptation; (iii) understanding the maintenance of genetic variation; (iv) understanding the role of standing genetic variation in adaptation; and (v) understanding the role of genomic architecture in adaptation.

### Vertical integration

Biologists of many stripes often repeat the famous Dobzhansky (1973) quote, 'nothing in biology makes sense except in the light of evolution', to refer to the unifying and explanatory power of evolution, across many levels of biological organization. And, it is indeed true

**Table 1.** Strengths and weaknesses of frequently used experimental methods in the QTN programme. LD, linkage disequilibrium.

| Approach (potential resolution) | Method | Advantages | Disadvantages | Examples |
|---|---|---|---|---|
| **Bi-parental crosses** | | | | |
| Fine mapping and positional cloning (QTN) | A QTL is introgressed into a homogeneous genetic background. Resulting lines segregate only within the QTL region (near isogenic lines). Recombinants are generated and tested for trait associations | • Very-small-effect variants can be resolved by progeny testing • Unhampered by the confounding factors and power issues arising in the population level association studies | • Time and labour intensive • Limited to measuring only two parental alleles per locus • Behaviour of QTLs as they are resolved into smaller regions is unpredictable—the association often disappears with increasing resolution or different genetic background • Linkage map required | Li *et al.* (2004) |
| Bulk segregant mapping (QTN) | A large recombinant bi-parental mapping population is created. Truncation selection is performed and the selected pools are sequenced and queried for shifts in allele frequency compared with the control | • Comparatively inexpensive as bulks can be sequenced in pools • Fast—QTNs are within reach with one generation of breeding in an F2 or backcross mapping population | • Large sample sizes mean that phenotyping is labour intensive • Limited to sampling two parental alleles • Resolution and power low for small-effect QTLs • QTN resolution requires a reference genome | Ehrenreich *et al.* (2010) |
| Nested association mapping (a few genes–a few cM) | Multiple parents are chosen and subject to a balanced crossing design that also seeks to maximize informative meioses. A high-resolution mapping population is created where all genomic segments have been shuffled relative to each other | • Allows population sampling while reducing the confounding effects of population structure • Rare variants accessible as their frequency is artificially increased • Can become a community resource | • Time and labour intensive to generate and maintain • Low general genomic resolution | McMullen *et al.* (2009) |
| Candidate gene association study (QTN) | A candidate gene is cloned starting with PCR primers based on a candidate gene sequence in another species. The gene is sequenced in natural population(s) using traditional Sanger or next-generation sequencing | • Fast—no need to generate mapping populations • Accessible for obligately outbreeding individuals • Information in literature on gene of interest | • Need prior knowledge of candidate genes • False positives due to unknown background factors or unaccounted for LD • Power low for small-effect and/or rare QTNs | Harjes *et al.* (2008) |
| Genome-wide association study (QTN) | Large population samples are either genotyped with a set of high-density markers or whole genomes are sequenced. Statistical models seek to associate genetic variants with trait variation while accounting for potentially confounding factors | • Large representative population samples • Accessible for obligately outbreeding individuals | • Expensive to sequence/genotype • False negatives in the process of accounting for multiple testing • False positives due to population or kinship structure • Power low for small-effect and/or rare QTNs • Requires a reference genome | Gudbjartsson *et al.* (2008) Lettre *et al.* (2008) Weedon *et al.* (2008) Atwell *et al.* (2010) |
| Transcriptomics (gene) | Expression levels of many/most genes in the transcriptome are measured using RNA sequencing, microarrays or other approaches. Expression variation for each transcript is associated with phenotype | • Less expensive and labour intensive than other approaches • Data specific to expressed portions of the genome • Tractable in non-model species | • Produces many significant targets • No estimate of effect size • Control for background genetic effects can be difficult | Ness *et al.* (2011) Liu *et al.* (2013) |
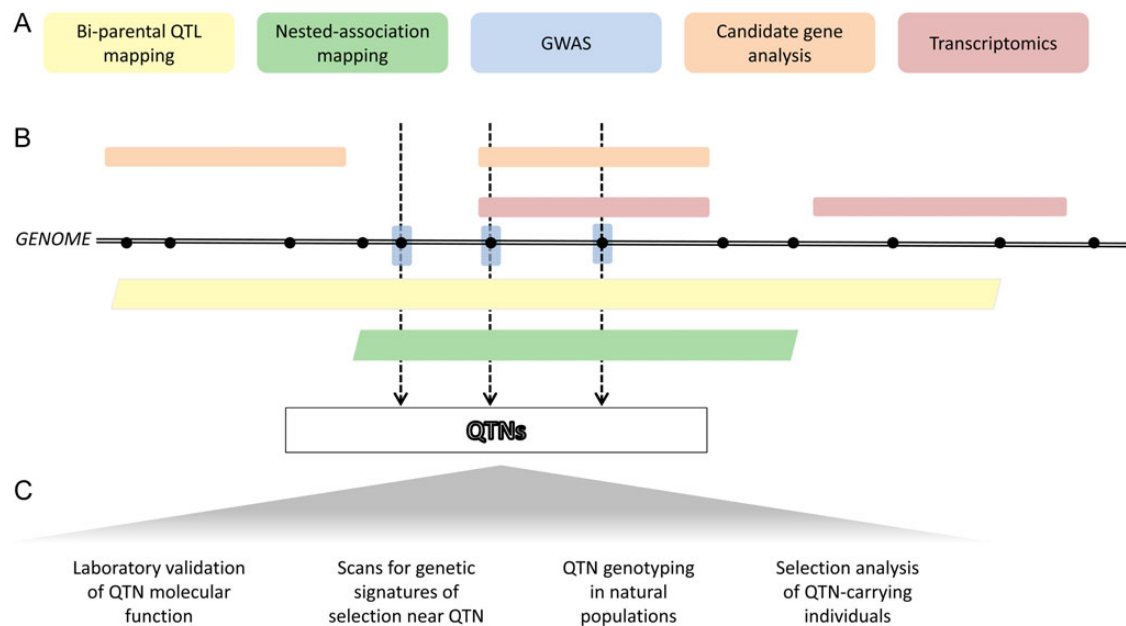
**Figure 1.** Schematic of common approaches to the QTN programme, with relative resolution of the methods. Major approaches used to identify QTN underlying ecologically important traits are listed in colour-coded boxes (A) (also described in Table 1). Each approach implicates genomic targets (i.e. single genetic polymorphisms, whole genes or genomic regions) that potentially underlie variation in the trait of interest. These targets are shown in relation to their position in the genome in (B), in the same colour as the corresponding method in (A). In (B), dots along the genome represent genetic variants (most often SNPs), which are used as markers in QTL, GWAS and candidate gene studies. Rectangles represent coding regions (genes), and parallelograms represent larger genomic regions. Variants in the genome that are implicated by multiple studies that use different methods are our best candidates for true QTN. These are highlighted with dotted arrows in (B). Following discovery of potential QTN, further analyses can then be undertaken (C) to verify the influence of each QTN on organismal phenotype and to explore their population genetic and ecological dynamics in natural settings.

that many features of biology, from the cellular and the molecular to broad-scale ecological patterns, can be understood from an evolutionary perspective. Yet, how many examples do we have where we understand the evolution of adaptation from the level of genetic variation at causal nucleotides, to heritability in natural populations, to the strength of contemporary selection, to how patterns of genetic variation and selection depend on geography and ecological context? We have a plethora of examples where a particular feature in this hierarchy of biological organization is understood, but not the rest of it. For example, in many systems we now have an exquisite understanding of genetic mechanisms and developmental processes underlying phenotypes (reviewed by Stern and Orgogozo 2008), but with far less knowledge of their ecological function and context. Likewise, many of the well-studied systems of evolutionary ecology have provided great detail on the ecological mechanisms of selection, as well as patterns of genetic variation and natural selection, in total ignorance of the molecular genetic basis of the traits or the developmental pathways producing them. Nearly 60 years after the discovery of DNA and almost 100 years after the evolutionary synthesis, it is remarkable that we have so few examples we can

point to of complete vertical integration in our understanding of adaptive phenotypes.

In large part, what the QTN programme strives for is to generate these much-needed, empirically complete studies of ecological adaptation. Often this involves the identification of QTN for traits that have been studied at the phenotypic and ecological levels for decades, and has the potential to reveal unpredicted complexities of the evolutionary process. For example, recently the genetic basis of flower colour evolution was uncovered in two closely related wildflowers, *Phlox drummondii* and *Phlox cuspidatum*. In sympatric populations, an unusual dark red morph of *P. drummondii* occurs (both species are normally light pink), and work over the decades supported the hypothesis that the phenotype evolved as a result of classical reinforcement (selection for reduced interspecies hybridization) (Erbe and Turner 1962; Levin 1972, 1985).

In 2011 Hopkins and Rausher used a candidate gene approach to uncover QTN linked to flower colour in these species. They uncovered one mutation that is tightly associated with pigmentation (pink vs. red) and another for colour intensity (light vs. dark). They then returned to 39 natural populations, sequenced the

QTN-containing loci and found nucleotide diversity patterns indicative of recent, strong natural selection at the red colour locus. No signature was found at the intensity locus and thus it seemed red flower colour had been verified as the target of selection (Hopkins *et al.* 2011). However, motivated by the identification of two separate QTN, they then generated plants carrying different combinations of the QTN for colour and intensity, introduced them back into natural settings and observed pollinator visitation to gauge selection pressures on the alleles. Unexpectedly, they found that pollinators imposed selection for intensity (darkness) but not colour itself (Hopkins and Rausher 2012). Vertical integration through QTN research in this system has revealed a more complex story than would otherwise have been appreciated as a result of either ecological or molecular studies in isolation. The reason red flower colour has evolved in conjunction with recent selection for dark colour intensity remains a mystery, perhaps to be solved by continued genetic work.

Similarly, vertical integration through a study of QTNs has expanded our understanding of adaptation and reproductive isolation in the common monkey flower (*Mimulus guttatus*) (Wright *et al.* 2013). In this species, populations growing on contaminated mine soils have evolved a high level of copper tolerance, and the trait co-occurs with tissue necrosis in hybrid offspring between on- and off-mine populations. The discovery that a single genomic region (QTL) was linked to variation in both traits made it seem probable that reproductive isolation between populations had been driven by genetic pleiotropy at one selected locus (Macnair and Christie 1983). However, continued genetic dissection of the QTN underlying the QTL showed that, despite known similarities between the genes that confer both heavy metal tolerance and hybrid necrosis in *Arabidopsis*, not one but two separate, physically linked loci control variation in the two traits. Genetic hitchhiking during a selective sweep at the copper tolerance locus captured alleles producing hybrid incompatibility between on- and off-mine populations. Only a detailed QTN approach was able to reveal the mechanistic connection between the evolution of adaptation and reproductive isolation.

The list of cases where ecologically important QTNs or genes have been identified and studied to some extent in natural settings continues to grow [e.g. *Boechera stricta* (Prasad *et al.* 2012), peppered moths (van't Hof *et al.* 2011; Cook and Saccheri 2013), various species of mice (Nachman *et al.* 2003; Hoekstra *et al.* 2006; Linnen *et al.* 2013), *Heliconius* butterflies (Baxter *et al.* 2010), stickleback fish (Colosimo *et al.* 2005) and many others (McKay and Stinchcombe 2008; Stern and Orgogozo 2008; Manceau *et al.* 2010; Martin and Orgogozo 2013)]. However, we argue on two fronts that discovery of QTN

under selection is still a much-needed area of research. First, as Rockman noted, the phenotypes that have been linked to QTN thus far are heavily biased toward traits that are qualitative in nature (i.e. pink vs. red flower colour) and have a very simple genetic basis. Relatively few examples exist where selection on QTN for truly quantitative traits has been demonstrated. Effectively detecting QTN underlying complex traits is a challenging goal in ecological genetics, but one that is not out of reach. Genome-wide association studies in cultivated plants and *Arabidopsis* have already made great strides in this area through the use of large sample sizes and marker sets that are saturating across the genome (Atwell *et al.* 2010; Brachi *et al.* 2010). Second, rarely have any QTN been shown empirically to change in frequency through more than a few generations in natural populations, on time scales pertinent to the evolution of traits through ecological selection (but see Barrett and Schluter 2008), thus pairing the QTN approach with the direct observation of the evolutionary process advocated by Travisano and Shaw. Such a pairing requires analysis of preserved specimens, genetic sampling of QTNs in populations across years or other creative approaches. We predict that some of the most empirically complete examples of vertical integration yet to come will track traits and their underlying QTN through time in natural populations, providing unique examples of microevolution in action.

## Genetic parallelism and pleiotropy

QTN research has revealed captivating stories in the study of the underlying mechanisms of convergent phenotypic evolution across related organismal groups. Identifying genes and QTNs responsible for similar but evolutionarily independent phenotypic shifts in different taxa is the only way to directly test for the extension of convergence down to the nucleotide level. The degree of genetic parallelism has direct implications for the role of pleiotropy and evolutionary constraint in adaptation (Wake 1991; Hoekstra 2006; Des Marais and Rausher 2008, 2010; Samis *et al.* 2012). This is to say, if the same genes or mutations produce a common phenotype (genetic parallelism), it suggests that either (i) there are limited genetic mechanisms for producing the same trait due to developmental or structural constraints (Wake 1991), or (ii) repeated changes to specific genes or regions are favoured because changes to those regions have fewer deleterious pleiotropic side-effects relative to others (Stern and Orgogozo 2008). In contrast, if the same phenotype can be produced by many genetic mechanisms (genetic non-parallelism), it suggests genetic heterogeneity underlying adaptive phenotypes (Kelly 2006; Travisano and Shaw 2013). Theoretical work confirms

the intuition that parallel evolution at the sequence level will be common when there are relatively few genetic mechanisms to produce the adaptive phenotype, and much more rare when there are many genetic routes to adaptation (Orr 2005).

There are still very few studies of the molecular basis of parallel phenotypic changes across taxa in wild plants, with one striking exception: the study of evolutionary transitions in flower colour. It has long been noted that within genera, two types of flower colour changes, from pigmented flowers to white flowers and from blue to red flowers, have occurred independently in many angiosperm clades. The molecular genetic basis of these changes has now been determined in at least 10 separate taxa (reviewed by Rausher 2008; Hopkins and Rausher 2011; Smith and Rausher 2011). Researchers have found that the evolution of parallel shifts in flower colour in the wild is often controlled by changes in the exact same loci in distant clades. Examples include comparisons of flower colour shifts between and within snapdragons, morning glories, monkey flowers, columbines and petunias, among others. While the nature of specific QTNs in these loci varies between species, evolution of flower colour is indeed parallel at the level of genes in many cases. Remarkable patterns of convergence in the evolution of morphological traits at the genetic level have also been recently discovered in cultivated plants (Ryan and Delhaize 2010; Lin *et al.* 2012), insects (Reed *et al.* 2011), fish (Colosimo *et al.* 2005; Hohenlohe *et al.* 2010; Colombo *et al.* 2013) and other vertebrates (reviewed by Hoekstra 2006). The observation of genetic parallelism in many cases, despite the known complexity of gene and protein networks that culminate in the realization of phenotypic traits, provides support for the role of pleiotropy in constraining evolution, or at least restricting it to certain loci.

## Maintenance of standing genetic variation

Even before mechanistically examining the alleles underlying adaptation, we might consider another question of fundamental importance in evolutionary biology: Why is there such an abundance of heritable variation in nature? Can the QTN programme help explain how it is maintained? The maintenance of standing genetic variation is one remaining major evolutionary question that lacks a significant body of empirical evidence to distinguish between alternative hypotheses (Lewontin 1974), despite a rich body of theory that provides genetic models of the underlying evolutionary dynamics. Such models can be classified into two categories that differ in the role of selection and its effect on variation. Under mutation–selection balance, the genetic variation observed in populations reflects an equilibrium between mutation that

introduces deleterious variation and purifying selection that depletes it (reviewed by Johnson and Barton 2005). The variants maintained in the population in this case are expected to be composed mostly of rare, recessive alleles. In contrast, balancing selection hypotheses propose that negative and positive selection acting antagonistically on the same site for different components of the trait, or in different genetic or environmental contexts, result in the maintenance of both alleles. Balancing selection predicts polymorphisms at intermediate, sometimes stable frequencies. There are numerous scenarios that lead to this result: frequency-dependent selection (Hori 1993), overdominance induced by antagonistic pleiotropy (opposing effects on different fitness components; Rose 1982; Hedrick 1999) and variable selection on genotypes in heterogeneous spatial or temporal environments (Levene 1953). For reviews of the relevant theory, see Hedrick *et al.* (1976), Hedrick (1986) and Mitchell-Olds *et al.* (2007).

Before QTN studies, empirical work on the maintenance of standing variation fell naturally under the domain of quantitative genetics. While a few traits with a Mendelian genetic basis had been identified to be under balancing selection (e.g. cichlid handedness maintained by frequency-dependent selection; Hori 1993), most traits are quantitative. Estimates of heritability and artificial selection experiments have shown that abundant genetic variation exists for almost any given trait (Houle 1992; Merila and Sheldon 1999; Hansen *et al.* 2011), but there has been limited success in explaining the maintenance of that variation. Manipulative experiments have attempted to confirm the contributions of rare and/or recessive deleterious alleles (Charlesworth and Hughes 1999; Kelly and Willis 2001; Kelly 2003; Charlesworth *et al.* 2007) with equivocal results. For example, Kelly and Willis (2001) examined whether rare, recessive mutations accounted for genetic variation in flower size in the common monkey flower (Kelly and Willis 2001). In brief, inbred and outbred populations were subject to artificial selection, and changes in mean flower size and directional dominance (the direction of how the trait changes with inbreeding) were estimated. Rare recessive alleles will result in a larger change in the directional dominance of the trait relative to the change in the mean (Kelly 1999). They found that rare recessive alleles were not enough to explain the existing genetic variation in flower size, suggesting that intermediate-frequency alleles contribute to flower size variation, implicating balancing selection on flower size (and the alleles responsible for it). These results suggest a more complicated interplay of selective forces than previous experiments that showed the population carried a large amount of partially recessive genetic load (Willis 1999*a*, *b*, *c*), suggesting mutation–selection

balance. While tantalizing, these methods that rely on the estimation of net inbreeding load or variance components do not disentangle the contributions of mutation, negative selection, positive selection and balancing selection acting on many different loci that affect traits, and will not result in a comprehensive understanding of the balance of forces in the maintenance of variation. In contrast, identifying and directly studying selection on small- and large-effect QTN for flower size and other traits in the field has this potential.

By identifying the causal sites underlying standing genetic variation, the QTN programme provides the opportunity to examine how selection acts on standing variation (and by inference, on variation in the associated traits), thereby connecting theory, quantitative genetics and population genetics. For instance, consider the prediction that QTNs maintained by mutation–selection balance will be rare, partially recessive and deleterious. An empirically testable hypothesis is that individuals with extreme phenotypes carry a greater number of mutations affecting that trait, a hypothesis that is now supported by QTN studies of human disease (Ji *et al.* 2008; Diogo *et al.* 2013). In these studies, candidate genes identified through GWAS were surveyed through deep population resequencing to identify rare putatively functional mutations. When 'case' and 'control' groups of individuals were compared, the case population was enriched in rare mutations. While the related notion that individuals with extreme phenotypes carry more deleterious mutations figures prominently in quantitative genetic models of apparent stabilizing selection (McGuigan *et al.* 2011), as of yet we know of no explicit tests of these predictions for quantitative traits thought to be under selection in nature. Comparisons of the number of functional mutations (genome wide or in the subset of genes in the molecular pathway determining the trait) carried by individuals with extreme phenotypes and those in the centre of the phenotypic distribution could be accomplished without bias due to QTN effect size. If a relationship was found, the implication is that at least some genetic variance is due to neutral or deleterious variants with pleiotropic effects on the focal trait.

The QTN programme also has the potential to determine the importance of the maintenance of genetic variation due to balancing selection, as knowledge of the causal QTNs enables the testing of specific theoretical models. Direct evidence includes a demonstration of heterozygote advantage at the selected locus or the absence of a genotype that has maximum fitness in all environments and/or genetic backgrounds. In the special case of frequency-dependent selection, it requires demonstration that the minor allele is always advantageous. Empirically detecting this is a challenging endeavour, and

recent years have seen progress on detecting fitness trade-offs at the scale of QTL (reviewed in Lowry 2012), including within-population QTLs (Mojica *et al.* 2012), though fitness effects attributed to QTLs may be confounded by linkage. The best examples are the culmination of years of fieldwork (Johnston *et al.* 2013) or benefit from a strong history of molecular genetics that implicate candidate genes for further experiments in the field (Tian *et al.* 2003; Todesco *et al.* 2010). These studies are classic examples of 'vertical integration'.

While such cases of 'vertical integration' are of great value, the QTN programme also offers an alternative route to testing the general importance of balancing selection for maintenance of variation in quantitative traits. Population genomic scans for the molecular imprint of balancing selection (see Charlesworth 2006) combined with a systematic search for QTNs affecting a variety of traits thought to be components of fitness variation in nature are an additional method of evaluating the role of balancing selection. Diagnostic molecular population genetic signatures of balancing selection include increased diversity in sites linked to QTN under long-term balancing selection, the existence of shared segregating haplotypes across related species or a common haplotype harbouring little variation for an allele that recently increased in frequency (see Charlesworth 2006 for a review). Dense genotype data have facilitated genome scans in human populations (Sabeti *et al.* 2002; Andres *et al.* 2009; Leffler *et al.* 2013), but without the overlap of candidate genes it is unknown what traits are affected by the regions under selection (Stinchcombe and Hoekstra 2008). While differentiating the signal of selection from neutral variation in these data sets is a challenge, the ability to combine a catalogue of QTNs where we have estimates of trait heritability with molecular population genetic tests for balancing selection would provide exciting advances in our understanding of the maintenance of standing genetic variation in the wild.

### Role of standing genetic variation in adaptation

Evolutionary responses to selection will differ greatly depending on whether the substrate of adaptation comes from either new mutations or standing variation. In the first case, adaptive divergence would be dependent on the rate of occurrence of new beneficial mutations. However, if the standing variation provides the substrate of adaptation, the evolutionary dynamics are determined by the factors that convert the standing variation into beneficial alleles, or change the magnitude of the selection coefficient. Consequently, a key question— regardless of the initial allele frequencies—is what changes to the ecological environment or selective regime make alleles sufficiently beneficial that they

increase in frequency and fix. If the alleles that fix had previously been maintained by balancing selection, evolutionary responses will be influenced, at least in part, by how standing variation is maintained.

An elegant example of how to combine QTN and molecular population genetic scans for selection was provided by a recent study using European-wide and local *Arabidopsis* accessions (Fournier-Level *et al.* 2011), and is an example of the way forward to a more general assessment of the contribution of standing variation to adaptation. Fixation events in the course of adaptation are expected to leave a characteristic molecular footprint on linked sites, where adaptation from new mutation is accompanied by the signature of a 'hard sweep', i.e. a steep decline in nucleotide diversity around the fixed site characterized by an excess of rare alleles close to the selected site. Fixation events from the standing variation are accompanied by the signature of a 'soft sweep', where a lesser decline in diversity around the fixed site is also associated with more intermediate-frequency alleles due to selection acting on multiple haplotypes containing the causal QTN (Przeworski *et al.* 2005). Fournier-Level *et al.* grew a large collection of accessions in four geographically distinct regions that spanned the climatic range limits in Europe, and QTNs for fertility and viability were identified in each region. Interestingly, the authors found that the alleles increasing fitness generally did not overlap between regions. Fournier-Level *et al.* also found that some of the beneficial alleles were naturally more abundant locally than the geographic distribution of a set of control SNPs. These results are consistent with local adaptation; however, there was limited evidence for the signal of recent 'hard' selective sweeps from new mutations around those sites in the local samples. These results imply that local adaptation was due to the differences between regions in the standing variation, and that evolutionary dynamics within and among different populations were not limited by the occurrence of new mutations.

The power to detect QTNs is determined by how much phenotypic variance is explained by the QTN, which is in turn determined by the effect size and allele frequency of that QTN. Thus the identification of QTNs is unavoidably limited by genetic architecture. The empirical results to date suggest that the genetic architecture of plant traits often includes major-effect alleles, with the caveat that plant GWAS to date have focused on crop species that underwent population bottlenecks and strong artificial selection, landraces with histories of introgression, and selfers such as *Arabidopsis*. Plant GWAS to date typically have smaller sample size (<1000) and yet successfully identify handfuls of loci that explain significant trait variance and overlap previously identified candidate genes (Aranzana *et al.* 2005; Chan *et al.* 2011; Famoso *et al.*

2011; Filiault and Maloof 2011; Huang *et al.* 2011; Morris *et al.* 2013; Rosas *et al.* 2013; Stanton-Geddes *et al.* 2013). For example, a GWAS of rice landraces ($n = 517$) identified QTNs for 14 traits that explained on average 36 % of trait variation (Huang *et al.* 2010). Similarly, a GWAS for 107 traits in *Arabidopsis* ($n = 96$ or 192 individuals depending on the trait) identified numerous common major-effect alleles underlying trait variation (Atwell *et al.* 2010; also see Bergelson and Roux 2010 for an overview). While most of these identified associations remain to be substantiated by further experiments (for example, in crossing designs or functional analysis), these data to date suggest that it is possible to identify QTNs in plants that explain a significant proportion of the phenotypic variance in natural populations. These results, however, do not obviate the contribution of small-effect loci underlying heritability in plants; for example, a 70-generation selection experiment for higher and lower oil content in maize showed that the sustained response was due to many small-effect loci (Laurie *et al.* 2004).

While the utility of GWAS for identifying genetic sources of trait variation has been demonstrated, we suggest the QTN programme expand to include work that focuses on sampling within (rather than between) populations, to identify the QTNs that are of the most evolutionary and ecological significance. The QTNs that we are able to identify are those that explain a significant portion of phenotypic variance for the trait. The contribution of a given biallelic locus to the additive genetic variance in a trait is given by $2pq\alpha^2$, where $p$ and $q$ are the frequencies of alternative alleles, and $\alpha$ is the average effect (Lynch and Walsh 1998). Consequently, common alleles will be the primary drivers of the initial selection response and will be of the most evolutionary interest, at least until rare alleles that also affect fitness increase in frequency. In addition, large-effect alleles—which might be in the minority of all alleles affecting a trait, or that are ultimately involved in the selection response—can make significant contributions to the genetic variance and evolutionary response, simply because of their effect size. To date most GWAS in plants have been composed of global samples that seek to maximize genetic and phenotypic variance, and identify causal associations while controlling for broad-scale population structure. An implicit assumption in these experimental designs is that causal alleles are widespread across many populations. However for quantitative traits, the same phenotypes in different populations may have a different genetic basis. Such global samples will cause even common alleles in local populations to be rare in the global sample if they are of limited geographic distribution, as would be expected if local adaptation is common. Thus except for globally common alleles, the frequency of

any given allele is lowered, and the power to detect locally significant QTNs may be diluted. The QTNs identified in such global samples will also not necessarily be the ones underlying local selection responses. Even with power limitations, if the QTN programme includes studies aimed at understanding the selection response, and not simply finding genes, even those few QTNs we identify will lead to insight into the evolutionary process. As within-population sampling becomes more common, one area of future development will likely be methods for controlling for close kinship between individuals (e.g. Manichaikul *et al.* 2010).

## Understanding the role of genomic architecture in adaptation

The QTN programme can also clarify the role of aspects of genomic architecture—chromosomal inversions, translocations, 'supergenes' in areas of suppressed or restricted recombination—in adaptation (Kirkpatrick and Barton 2006; Scoville *et al.* 2009; Yeaman 2013). In particular, the potential role of supergenes in adaptive phenotypes has been suspected since the dawn of plant genetics, and a robust set of theoretical predictions and population genetics has developed in these areas. Moreover, turning these suspicions into testable hypotheses, and evaluating our theoretical predictions, is something that can only be achieved with a molecular approach.

Early work on the genetics of heterostyly is illustrative (e.g. Bateson and Gregory 1905; Gregory *et al.* 1923; De Winton and Haldane 1933, 1935; Mather 1950). Heterostylous plants are polymorphic for the reciprocal arrangement of anthers and stigmas along with pollen size and other floral characteristics, with successful fertilization usually only possible between anthers and stigmas at the same level (long styles and pollen from long-level anthers; short styles and pollen from short-level anthers; see Barrett 1992 and Barrett and Shore 2008 for more details). Current evidence suggests that heterostyly functions to increase the proficiency of cross-pollination, promoting disassortative mating between the floral morphs (Kohn and Barrett 1992; Lloyd and Webb 1992). Genetic crossing designs strongly suggest one (distyly) or two (tristyly) diallelic loci governing the style–stamen polymorphisms. The complex morphological and physiological components of the heterostyly syndrome, along with observations that recombinants are rare, have suggested supergene control involving a co-adapted linkage group (Lewis and Jones 1992).

There are numerous challenging evolutionary questions regarding the evolution of heterostylous pollination systems. How did these intricate, seemingly co-adapted systems evolve? Are initially unlinked loci captured by chromosomal inversions or translocations, or must the loci already be loosely linked? How do selection and recombination interact in creating and breaking up these supergene complexes? How much pleiotropy is required for diverse features of floral morphology to be inherited together? These questions have received intense investigation from the theoretical population genetics perspective (Charlesworth and Charlesworth 1976; Charlesworth 1979; B. Charlesworth and D. Charlesworth 1979; D. Charlesworth and B. Charlesworth 1979). The construction of linkage maps and estimation of recombination rates and signatures of selection in genomic regions affecting floral phenotypes would help test the 'supergene' hypothesis. Consequently, the QTN programme—or at a minimum, molecular markers, linkage maps and molecular population genetic inference—can clarify the genetic basis and evolutionary forces responsible for polymorphisms that directly influence plant mating and fitness. For example, the generation of linkage maps can reveal whether marker order near style length QTL regions is reversed in different populations, suggesting chromosomal inversions. If inversions do not explain transmission patterns, molecular population genetic data on patterns of polymorphism, linkage disequilibrium (LD) and recombination can reveal whether loci influencing heterostyly are in regions of suppressed recombination. Similarly, population genetic analyses can test whether these regions harbour elevated divergence and LD indicating the long-term maintenance of a supergene cluster. We see these research questions as ones in which classical approaches and theory have led to clear, testable predictions that can be answered by the QTN programme, and where the conclusions are robust to issues of effect size and bias (since the predictions do not hinge on detection of small-effect QTNs). To date this approach has been successfully applied to other systems where supergenes have been invoked (mimetic butterflies: Joron *et al.* 2006, 2011; Counterman *et al.* 2010; shell colour and banding in snails: Richards *et al.* 2013).

Another area where the QTN (or QTL) programme is necessary for evaluating the role of genomic architecture in adaptation is in studies on the role of chromosomal inversions in local adaptation. Theoretical work by Kirkpatrick and Barton (2006) showed that chromosomal inversions could 'capture' locally advantageous haplotypes and spread quite rapidly. In this manner, several genes or loci that lead to local adaptation would be preserved together, largely because of the suppressed recombination that often accompanies inversions. These loci are not required to interact epistatically, genetic drift is not required, and there is no trough in the adaptive landscape to traverse (Kirkpatrick and Barton 2006). Kirkpatrick and Barton's theory was a significant advance in our expectations for the evolutionary dynamics of inversions, which previously

assumed that they would be generally deleterious due to meiotic imbalance or disrupted genes at the breakpoints (Kirkpatrick 2010). For an overview of the role of chromosomal inversions in diverse topics in evolutionary biology, see Kirkpatrick (2010) and Kirkpatrick and Kern (2012). The discovery of a positively selected chromosomal inversion in humans (Stefansson *et al.* 2005) was in fact an early basic science 'spin-off' discovery facilitated by human GWAS studies (Visscher *et al.* 2012).

One of the best examples of a locally adaptive inversion is in the yellow monkey flower, *M. guttatus*, which inhabits a broad geographic range in western North America. Inland ecotypes tend to be annual, while coastal ecotypes are perennial, and the two types also differ significantly in flowering time. The divergent flowering time between inland and coastal ecotypes generates strong natural selection when the opposite type is experimentally transplanted: late-flowering, perennial coastal plants fail to reproduce before droughts in inland habitats, while early-flowering, annual inland plants fail to capitalize on the extended growing season provided by the coastal climate. Initial observations (Hall *et al.* 2010) suggested an area of suppressed recombination in a recombinant inbred line mapping population made from an inland × coastal cross. Subsequent work by Lowry and Willis showed that marker order was reversed between inland × inland and coastal × coastal crosses, and that recombination (as reflected by genetic map distances) was suppressed in inter-ecotype crosses, but not within ecotypes. Finally, ambitious ecological genetic field experiments showed that the inversion influenced flowering time, morphological characteristics and fitness in the field (Lowry and Willis 2010). The design used by Lowry and Willis (2010) allowed them to estimate the relative contribution of the inversion to reproductive isolation between the two ecotypes. In this case, a detailed QTN- or QTL-based approach again revealed fundamental insights into the evolutionary process—local adaptation and reproductive isolation due to chromosomal inversions—that were unattainable without the aid of molecular tools.

## Conclusion

The continuing advance of next-generation sequencing technology and its subsequent drop in price has opened the flood gates for research into the genes underlying trait variation in nature. Concomitant with this rush, some of the deeper reasons why we are interested in this information sometimes get lost, and often the caveats and limitations of QTN work are brushed aside in the wake of excitement over the seemingly endless possibilities. We need to reconvene, refocus and then redouble our efforts to gain insight into the evolutionary process, using genetic data where it can be the most effective and allow greatest testing of evolutionary hypotheses. Refocusing and planning efforts in the QTN programme must take into careful consideration not only these broader contexts, but the experimental resolution of available methods, their inference space and generality (Rockman 2012), and whether the outcome will generate a starting point for the essential mechanistic work advocated by Travisano and Shaw (2012). As evolutionists we see not the end to the utility of QTN work but an exciting future in which we are able to directly address how ecological and genetic factors interact over evolutionary time to generate and maintain phenotypic diversity.

## Contributions by the Authors

All authors conceived the ideas, wrote the paper, edited manuscript drafts and responded to reviewer comments.

## Conflicts of Interest Statement

None declared.

## Literature Cited

Andres AM, Hubisz MJ, Indap A, Torgerson DG, Degenhardt JD, Boyko AR, Gutenkunst RN, White TJ, Green EJ, Bustamante CD, Clark AG, Nielsen R. 2009. Targets of balancing selection in the human genome. *Molecular Biology and Evolution* **26**:2755–2764.

Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, Jakob K, Lister C, Molitor J, Shindo C, Tang C, Toomajian C, Traw B, Zheng H, Bergelson J, Dean C, Marjoram P, Nordborg M. 2005. Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLoS Genetics* **1**:e60.

Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, Jiang R, Muliyati NW, Zhang X, Amer MA, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, Ecker JR, Faure N, Kniskern JM, Jones JD, Michael T, Nemri A, Roux F, Salt DE, Tang C, Todesco M, Traw MB, Weigel D, Marjoram P, Borevitz JO, Bergelson J, Nordborg M. 2010. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**:627–631.

Barrett RDH, Schluter D. 2008. Adaptation from standing genetic variation. *Trends in Ecology and Evolution* **23**:38–44.

Barrett SCH. 1992. Heterostylous genetic polymorphisms: model systems for evolutionary analysis. In: Barrett SCH, ed. *Evolution and function of heterostyly*. Berlin: Springer, 1–29.

Barrett SCH, Shore JS. 2008. New insights on heterostyly: comparative biology, ecology and genetics. In: Franklin-Tong V, ed. *Self-incompatibility in flowering plants: evolution, diversity and mechanisms*. Berlin: Springer, 3–32.

Bateson W, Gregory RP. 1905. On the inheritance of heterostylism in *Primula*. *Proceedings of the Royal Society London Series B* **76**: 581–586.

Baxter SW, Nadeau NJ, Maroja LS, Wilkinson P, Counterman BA, Dawson A, Beltran M, Perez-Espona S, Chamberlain N, Ferguson L, Clark R, Davidson C, Glithero R, Mallet J, McMillan WO, Kronforst M, Joron M, ffrench-Constant RH, Jiggins CD. 2010. Genomic hotspots for adaptation: the population genetics of Müllerian mimicry in the *Heliconius melpomene* clade. *PLoS Genetics* **6**:e1000794.

Bergelson J, Roux F. 2010. Towards identifying genes underlying ecologically relevant traits in *Arabidopsis thaliana*. *Nature Reviews Genetics* **11**:867–879.

Brachi B, Morris GP, Borevitz JO. 2010. Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biology* **12**:232.

Chan EK, Rowe HC, Corwin JA, Joseph B, Kliebenstein DJ. 2011. Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*. *PLoS Biology* **9**:e1001125.

Charlesworth B, Charlesworth D. 1979. The maintenance and breakdown of distyly. *The American Naturalist* **114**:499–513.

Charlesworth B, Hughes KA. 1999. The maintenance of genetic variation in life history traits. *Evolutionary Genetics: From Molecules to Morphology* **1**:369–392.

Charlesworth B, Miyo T, Borthwick H. 2007. Selection response of means and inbreeding depression for female fecundity in *Drosophila melanogaster* suggest contributions from intermediate frequency alleles to quantitative trait variation. *Genetical Research* **89**:85–91.

Charlesworth D. 1979. The evolution and breakdown of Tristyly. *Evolution* **33**:486–498.

Charlesworth D. 2006. Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics* **2**:e64.

Charlesworth D, Charlesworth B. 1976. Theoretical genetics of Batesian mimicry. II. Evolution of supergenes. *Journal of Theoretical Biology* **55**:305–324.

Charlesworth D, Charlesworth B. 1979. A model for the evolution of distyly. *The American Naturalist* **114**:467–498.

Colombo M, Diepeveen ET, Muschick M, Santos ME, Indermaur A, Boileau N, Barluenga M, Salzburger W. 2013. The ecological and genetic basis of convergent thick-lipped phenotypes in cichlid fishes. *Molecular Ecology* **22**:670–684.

Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* **307**:1928–1933.

Cook L, Saccheri I. 2013. The peppered moth and industrial melanism: evolution of a natural selection case study. *Heredity* **110**:207–212.

Counterman BA, Araujo-Perez F, Hines HM, Baxter SW, Morrison CM, Lindstrom DP, Papa R, Ferguson L, Joron M, Ffrench-Constant RH, Smith CP, Nielsen DM, Chen R, Jiggins CD, Reed RD, Halder G, Mallet J, McMillan WO. 2010. Genomic hotspots for adaptation: the population genetics of Müllerian mimicry in *Heliconius erato*. *PLoS Genetics* **6**:e1000796.

de Winton D, Haldane JBS. 1933. The genetics of *Primula sinensis*. II. Segregation and interaction of factors in the diploid. *Journal of Genetics* **27**:1–44.

de Winton D, Haldane JBS. 1935. The genetics of *Primula sinensis*. III. Linkage in the diploid. *Journal of Genetics* **31**:67–100.

Des Marais DL, Rausher MD. 2008. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* **454**: 762–765.

Des Marais DL, Rausher MD. 2010. Parallel evolution at multiple levels in the origin of hummingbird pollinated flowers in *Ipomoea*. *Evolution* **64**:2044–2054.

Diogo D, Kurreeman F, Stahl EA, Liao KP, Gupta N, Greenberg JD, Rivas MA, Hickey B, Flannick J, Thomson B, Guiducci C, Ripke S, Adzhubey I, Barton A, Kremer JM, Alfredsson L, Consortium of Rheumatology Researchers of North America, Rheumatoid Arthritis Consortium International, Sunyaev S, Martin J, Zhernakova A, Bowes J, Eyre S, Siminovitch KA, Gregersen PK, Worthington J, Klareskog L, Padyukov L, Raychaudhuri S, Plenge RM. 2013. Rare, low-frequency, and common variants in the protein-coding sequence of biological candidate genes from GWASs contribute to risk of rheumatoid arthritis. *American Journal of Human Genetics* **92**:15–27.

Dobzhansky T. 1973. Nothing in biology makes sense except in the light of evolution. *American Biology Teacher* **35**: 125–129.

Ehrenreich IM, Torabi N, Jia Y, Kent J, Martis S, Shapiro JA, Gresham D, Caudy AA, Kruglyak L. 2010. Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* **464**:1039–1042.

Erbe L, Turner BL. 1962. A biosystematic study of the *Phlox cuspidata*–*Phlox drummondii* complex. *American Midland Naturalist* **67**:257–281.

Famoso AN, Zhao K, Clark RT, Tung CW, Wright MH, Bustamante C, Kochian LV, McCouch SR. 2011. Genetic architecture of aluminum tolerance in rice (*Oryza sativa*) determined through genome-wide association analysis and QTL mapping. *PLoS Genetics* **7**: e1002221.

Filiault DL, Maloof JN. 2011. A genome-wide association study identifies variants underlying the *Arabidopsis thaliana* shade avoidance response. *PLoS Genetics* **8**:e1002589.

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM. 2011. A map of local adaptation in *Arabidopsis thaliana*. *Science* **334**:86–89.

Gregory RP, de Winton D, Bateson W. 1923. Genetics of *Primula sinensis*. *Journal of Genetics* **13**:219–253.

Gudbjartsson DF, Walters GB, Thorleifsson G, Stefansson H, Halldorsson BV, Zusmanovich P, Sulem P, Thorlacius S, Gylfason A, Steinberg S, Helgadottir A, Ingason A, Steinthorsdottir V, Olafsdottir EJ, Olafsdottir GH, Jonsson T, Borch-Johnsen K, Hansen T, Andersen G, Jorgensen T, Pedersen O, Aben KK, Witjes JA, Swinkels DW, den Heijer M, Franke B, Verbeek AL, Becker DM, Yanek LR, Becker LC, Tryggvadottir L, Rafnar T, Gulcher J, Kiemeney LA, Kong A, Thorsteinsdottir U, Stefansson K. 2008. Many sequence variants affecting diversity of adult human height. *Nature Genetics* **40**: 609–615.

Hall MC, Lowry DB, Willis JH. 2010. Is local adaptation in *Mimulus guttatus* caused by trade-offs at individual loci? *Molecular Ecology* **19**:2739–2753.

Hansen T, Pélabon C, Houle D. 2011. Heritability is not evolvability. *Evolutionary Biology* **38**:258–277.

Harjes CE, Rocheford TR, Bai L, Brutnell TP, Kandianis CB, Sowinski SG, Stapleton AE, Vallabhaneni R, Williams M, Wurtzel ET, Yan J, Buckler ES. 2008. Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science* **319**:330–333.

Hedrick PW. 1986. Genetic polymorphism in heterogeneous environments: a decade later. *Annual Review of Ecology and Systematics* **17**:535–566.

Hedrick PW. 1999. Antagonistic pleiotropy and genetic polymorphism: a perspective. *Heredity* **82**:126–133.

Hedrick PW, Ginevan ME, Ewing EP. 1976. Genetic polymorphism in heterogeneous environments. *Annual Review of Ecology and Systematics* **7**:1–32.

Hoekstra HE. 2006. Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity* **97**:222–234.

Hoekstra HE, Hirschmann RJ, Bundey Ra, Insel PA, Crossland JP. 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* **313**:101–104.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* **6**:e1000862.

Hopkins R, Rausher MD. 2011. Identification of two genes causing reinforcement in the Texas wildflower *Phlox drummondii*. *Nature* **469**:411–414.

Hopkins R, Rausher MD. 2012. Pollinator-mediated selection on flower color allele drives reinforcement. *Science* **335**:1090–1092.

Hopkins R, Levin DA, Rausher MD. 2011. Molecular signatures of selection on reproductive character displacement of flower color in *Phlox drummondii*. *Evolution* **66**:469–485.

Hori M. 1993. Frequency-dependent natural selection in the handedness of scale-eating cichlid fish. *Science* **260**:216–219.

Houle D. 1992. Comparing evolvability and variability in quantitative traits. *Genetics* **130**:195–204.

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang QF, Li J, Han B. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics* **42**:961–967.

Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, Li W, Guo Y, Deng L, Zhu C, Fan D, Lu Y, Weng Q, Liu K, Zhou T, Jing Y, Si L, Dong G, Huang T, Lu T, Feng Q, Qian Q, Li J, Han B. 2011. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature Genetics* **44**: 32–39.

Ji W, Foo JN, O'Roak BJ, Zhao H, Larson MG, Simon DB, Newton-Cheh C, State MW, Levy D, Lifton RP. 2008. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nature Genetics* **40**:592–599.

Johnson T, Barton NH. 2005. Theoretical models of selection and mutation on quantitative traits. *Philosophical Transactions of the Royal Society* **360**:1411–1425.

Johnston SE, Gratten J, Berenos C, Pilkington JG, Clutton-Brock TH, Pemberton JM, Slate J. 2013. Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature* **502**: 93–95.

Joron M, Papa R, Beltrán M, Chamberlain N, Mavárez J, Baxter S, Abanto M, Bermingham E, Humphray SJ, Rogers J, Beasley H, Barlow K, ffrench-Constant RH, Mallet J, McMillan WO, Jiggins CD. 2006. A conserved supergene locus controls colour pattern diversity in *Heliconius* butterflies. *PLoS Biology* **4**:e303.

Joron M, Frezal L, Jones RT, Chamberlain NL, Lee SF, Haag CR, Whibley A, Becuwe M, Baxter SW, Ferguson L, Wilkinson PA, Salazar C, Davidson C, Clark R, Quail MA, Beasley H, Glithero R, Lloyd C, Sims S, Jones MC, Rogers J, Jiggins CD, ffrench-Constant RH. 2011. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477**:203–206.

Kelly JK. 1999. An experimental method for evaluating the contribution of deleterious mutations to quantitative trait variation. *Genetics Research Cambridge* **73**:263–273.

Kelly JK. 2003. Deleterious mutations and the genetic variance of male fitness components in *Mimulus guttatus*. *Genetics* **164**: 1071–1085.

Kelly JK. 2006. Geographical variation in selection, from phenotypes to molecules. *The American Naturalist* **167**:481–495.

Kelly JK, Willis JH. 2001. Deleterious mutations and genetic variation for flower size in *Mimulus guttatus*. *Evolution* **55**:937–942.

Kirkpatrick M. 2010. How and why chromosome inversions evolve. *PLoS Biology* **8**:e1000501.

Kirkpatrick M, Barton N. 2006. Chromosome inversions, local adaptation and speciation. *Genetics* **173**:419–434.

Kirkpatrick M, Kern A. 2012. Where's the money? Inversions, genes, and the hunt for genomic targets of selection. *Genetics* **190**: 1153–1155.

Kohn JR, Barrett SCH. 1992. Experimental studies on the functional significance of heterostyly. *Evolution* **46**:43–55.

Laurie CC, Chasalow SD, LeDeaux JR, McCarroll R, Bush D, Hauge B, Lai C, Clark D, Rocheford TR, Dudley JW. 2004. The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. *Genetics* **168**:2141.

Leffler EM, Gao Z, Pfeifer S, Segurel L, Auton A, Venn O, Bowden R, Bontrop R, Wall JD, Sella G, Donnelly P, McVean G, Przeworski M. 2013. Multiple instances of ancient balancing selection shared between humans and chimpanzees. *Science* **339**:1578–1582.

Lettre G, Jackson AU, Gieger C, Schumacher FR, Berndt SI, Sanna S, Eyheramendy S, Voight BF, Butler JL, Guiducci C, Illig T, Hackett R, Heid IM, Jacobs KB, Lyssenko V, Uda M; Diabetes Genetics Initiative; FUSION; KORA; Prostate, Lung Colorectal and Ovarian Cancer Screening Trial; Nurses' Health Study; SardiNIA, Boehnke M, Chanock SJ, Groop LC, Hu FB, Isomaa B, Kraft P,

Peltonen L, Salomaa V, Schlessinger D, Hunter DJ, Hayes RB, Abecasis GR, Wichmann HE, Mohlke KL, Hirschhorn JN. 2008. Identification of ten loci associated with height highlights new biological pathways in human growth. *Nature Genetics* **40**: 584–591.

Levene H. 1953. Genetic equilibrium when more than one ecological niche is available. *The American Naturalist* **87**:331–333.

Levin DA. 1972. The adaptedness of corolla-color variants in experimental and natural populations of *Phlox drummondii*. *The American Naturalist* **106**:57–70.

Levin DA. 1985. Reproductive character displacement in *Phlox*. *Evolution* **39**:1275–1281.

Lewis D, Jones DA. 1992. The genetics of heterostyly. In: Barrett SCH, eds. *Evolution and function of heterostyly*. Berlin: Springer, 129–150.

Lewontin RC. 1974. *The genetic basis of evolutionary change*. New York: Columbia University Press.

Li J, Thomson M, McCouch SR. 2004. Fine mapping of a grain-weight quantitative trait locus in the pericentromeric region of rice chromosome 3. *Genetics* **168**:2187–2195.

Lin Z, Li X, Shannon LM, Yeh C-T, Wang ML, Bai G, Peng Z, Li J, Trick HN, Clemente TE, Doebley J, Schnable PS, Tuinstra MR, Tesso TT, White F, Yu J. 2012. Parallel domestication of the Shattering1 genes in cereals. *Nature Genetics* **44**:720–724.

Linnen CR, Poh Y-P, Peterson BK, Barrett RDH, Larson JG, Jensen JD, Hoekstra HE. 2013. Adaptive evolution of multiple traits through multiple mutations at a single gene. *Science* **339**:1312–1316.

Liu J, Sturrock R, Benton R. 2013. Transcriptome analysis of *Pinus monticola* primary needles by RNA-seq provides novel insight into host resistance to *Cronartium ribicola*. *BMC Genomics* **14**:884.

Lloyd DG, Webb CJ. 1992. The selection of heterostyly. In: Barrett SCH, ed. *Evolution and function of heterostyly*. Berlin: Springer, 179–207.

Lowry D. 2012. Commentary: local adaptation in the model plant. *New Phytologist* **194**:888–890.

Lowry D, Willis JH. 2010. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biology* **8**:e1000500.

Lynch M, Walsh B. 1998. *Genetics and the analysis of quantitative traits*. Sunderland, MA, USA: Sinauer Associates.

Macnair MR, Christie P. 1983. Reproductive isolation as a pleiotropic effect of copper tolerance in *Mimulus guttatus*? *Heredity* **50**: 295–302.

Manceau M, Domingues VS, Linnen CR, Rosenblum EB, Hoekstra HE. 2010. Convergence in pigmentation at multiple levels: mutations, genes and function. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* **365**: 2439–2450.

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. 2010. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**:2867–2873.

Martin A, Orgogozo V. 2013. The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution* **67**: 1235–1250.

Mather K. 1950. The genetical architecture of heterostyly in *Primula sinensis*. *Evolution* **4**:340–352.

McGuigan K, Rowe L, Blows MW. 2011. Pleiotropy, apparent stabilizing selection and uncovering fitness optima. *Trends in Ecology and Evolution* **26**:22–29.

McKay JK, Stinchcombe JR. 2008. Ecological genomics of model eukaryotes. *Evolution* **62**:2953–2957.

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C, Brown P, Browne C, Eller M, Guill K, Harjes C, Kroon D, Lepak N, Mitchell SE, Peterson B, Pressoir G, Romero S, Oropeza Rosas M, Salvo S, Yates H, Hanson M, Jones E, Smith S, Glaubitz JC, Goodman M, Ware D, Holland JB, Buckler ES. 2009. Genetic properties of the maize nested association mapping population. *Science* **325**:737–740.

Merila J, Sheldon BC. 1999. Genetic architecture of fitness and non-fitness traits: empirical patterns and development of ideas. *Heredity* **83**:103–109.

Mitchell-Olds T, Willis JH, Goldstein D. 2007. What evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Reviews Genetics* **8**:845–856.

Mojica JP, Lee YW, Willis JH, Kelly JK. 2012. Spatially and temporally varying selection on intrapopulation quantitative trait loci for a life history trade-off in *Mimulus guttatus*. *Molecular Ecology* **21**: 3718–3728.

Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera-Lizarazu O, Brown PJ, Acharya CB, Mitchell SE, Harriman J, Glaubitz JC, Buckler ES, Kresovich S. 2013. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sciences of the USA* **110**:453–458.

Nachman MW, Hoekstra HE, D'Agostino SL. 2003. The genetic basis of adaptive melanism in pocket mice. *Proceedings of the National Academy of Sciences of the USA* **100**:5268–5273.

Ness R, Siol M, Barrett S. 2011. De novo sequence assembly and characterization of the floral transcriptome in cross- and self-fertilizing plants. *BMC Genomics* **12**:298.

Orr HA. 2005. The probability of parallel evolution. *Evolution* **59**: 216–220.

Prasad KVSK, Song B-H, Olson-Manning C, Anderson JT, Lee C-R, Schranz ME, Windsor AJ, Clauss MJ, Manzaneda AJ, Naqvi I, Reichelt M, Gershenzon J, Rupasinghe SG, Schuler MA, Mitchell-Olds T. 2012. A gain-of-function polymorphism controlling complex traits and fitness in nature. *Science* **337**: 1081–1084.

Przeworski M, Coop G, Wall JW. 2005. The signature of positive selection on standing genetic variation. *Evolution* **59**:2312–2323.

Rausher MD. 2008. Evolutionary transitions in floral color. *International Journal of Plant Sciences* **169**:7–21.

Reed RD, Papa R, Martin A, Hines HM, Counterman BA, Pardo-Diaz C, Jiggins CD, Chamberlain NL, Kronforst MR, Chen R, Halder G, Nijhout HF, McMillan WO. 2011. Optix drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science* **333**: 1137–1141.

Richards PM, Liu MM, Lowe N, Davey JW, Blaxter ML, Davison A. 2013. RAD-Seq derived markers flank the shell colour and banding loci of the *Cepaea nemoralis* supergene. *Molecular Ecology* **22**: 3077–3089.

Rockman MV. 2012. The QTN Program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* **66**:1–17.

Rosas U, Cibrian-Jaramillo A, Ristova D, Banta JA, Gifford ML, Fan AH, Zhou RW, Kim GJ, Krouk G, Birnbaum KD, Purugganan MD, Coruzzi GM. 2013. Integration of responses within and across *Arabidopsis* natural accessions uncovers loci controlling root

systems architecture. *Proceedings of the National Academy of Sciences of the USA* **110**:15133–15138.

Rose MR. 1982. Antagonistic pleiotropy, dominance, and genetic variation. *Heredity* **48**:63–78.

Ryan PR, Delhaize E. 2010. The convergent evolution of aluminium resistance in plants exploits a convenient currency. *Functional Plant Biology* **4**:275.

Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES. 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**: 832–837.

Samis KE, Murren CJ, Bossdorf O, Donohue K, Fenster CB, Malmberg RL, Purugganan MD, Stinchcombe JR. 2012. Longitudinal trends in climate drive flowering time clines in North American *Arabidopsis thaliana*. *Ecology and Evolution* **2**:1162–1180.

Scoville A, Lee YW, Willis JH, Kelly JK. 2009. Contribution of chromosomal polymorphisms to the G-matrix of *Mimulus guttatus*. *New Phytologist* **183**:803–815.

Slate J. 2013. From Beavis to beak color: a simulation study to examine how much QTL mapping can reveal about the genetic architecture of quantitative traits. *Evolution* **67**:1251–1262.

Smith SD, Rausher MD. 2011. Gene loss and parallel evolution contribute to species difference in flower color. *Molecular Biology and Evolution* **28**:2799–2810.

Stanton-Geddes J, Paape T, Epstein B, Briskine R, Yoder J, Mudge J, Bharti AK, Farmer AD, Zhou P, Denny R, May GD, Erlandson S, Yakub M, Sugawara M, Sadowsky MJ, Young ND, Tiffin P. 2013. Candidate genes and genetic architecture of symbiotic and agronomic traits revealed by whole-genome, sequence-based association genetics in *Medicago truncatula*. *PLoS One* **8**: e65688.

Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J, Baker A, Jonasdottir A, Ingason A, Gudnadottir VG, Desnica N, Hicks A, Gylfason A, Gudbjartsson DF, Jonsdottir GM, Sainz J, Agnarsson K, Birgisdottir B, Ghosh S, Olafsdottir A, Cazier J-B, Kristjansson K, Frigge ML, Thorgeirsson TE, Gulcher JR, Kong A, Stefansson K. 2005. A common inversion under selection in Europeans. *Nature Genetics* **37**:129–137.

Stern DL, Orgogozo V. 2008. The loci of evolution: how predictable is genetic evolution? *Evolution* **62**:2155–2177.

Stinchcombe JR, Hoekstra HE. 2008. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* **100**:158–170.

Tian D, Traw MB, Chen JQ, Kreitman M, Bergelson J. 2003. Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* **423**:74–77.

Todesco M, Balasubramanian S, Hu TT, Traw MB, Horton M, Epple P, Kuhns C, Sureshkumar S, Schwartz C, Lanz C, Laitinen RA, Huang Y, Chory J, Lipka V, Borevitz JO, Dangl JL, Bergelson J, Nordborg M, Weigel D. 2010. Natural allelic variation underlying a major fitness trade-off in *Arabidopsis thaliana*. *Nature* **465**: 632–636.

Travisano M, Shaw RG. 2013. Lost in the map. *Evolution* **67**:305–314.

van't Hof AE, Edmonds N, Dalíková M, Marec F, Saccheri IJ. 2011. Industrial melanism in British peppered moths has a singular and recent mutational origin. *Science* **332**:958–960.

Visscher PM, Brown MA, McCarthy MI, Yang J. 2012. Five years of GWAS discovery. *American Journal of Human Genetics* **90**:7–24.

Wake DB. 1991. Homoplasy: the result of natural selection, or evidence of design limitations? *The American Naturalist* **3**: 543–567.

Ward J, Ponnala L, Weber C. 2012. Strategies for transcriptome analysis in nonmodel plants. *American Journal of Botany* **99**: 267–276.

Weedon MN, Lango H, Lindgren CM, Wallace C, Evans DM, Mangino M, Freathy RM, Perry JR, Stevens S, Hall AS, Samani NJ, Shields B, Prokopenko I, Farrall M, Dominiczak A; Diabetes Genetics Initiative; Wellcome Trust Case Control Consortium, Johnson T, Bergmann S, Beckmann JS, Vollenweider P, Waterworth DM, Mooser V, Palmer CN, Morris AD, Ouwehand WH; Cambridge GEM Consortium, Zhao JH, Li S, Loos RJ, Barroso I, Deloukas P, Sandhu MS, Wheeler E, Soranzo N, Inouye M, Wareham NJ, Caulfield M, Munroe PB, Hattersley AT, McCarthy MI, Frayling TM. 2008. Genome-wide association analysis identifies 20 loci that influence adult height. *Nature Genetics* **40**:575–583.

Willis JH. 1999a. Inbreeding load, average dominance, and the mutation rate for mildly deleterious alleles in *Mimulus guttatus*. *Genetics* **153**:1885–1898.

Willis JH. 1999b. The role of genes of large effect on inbreeding depression in *Mimulus guttatus*. *Evolution* **53**:1678–1691.

Willis JH. 1999c. The contribution of male sterility mutations to inbreeding depression in *Mimulus guttatus*. *Heredity* **83**:337–346.

Wright KM, Lloyd D, Lowry DB, Macnair MR, Willis JH. 2013. Indirect evolution of hybrid lethality due to linkage with selected locus in *Mimulus guttatus*. *PLoS Biology* **11**:e1001497.

Yeaman S. 2013. Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences of the USA* **110**:E1743–E1751.