

REVIEW

Advancing Genetic Theory and Application by Metabolic Quantitative Trait Loci Analysis

Daniel J. Kliebenstein¹

Genetics Graduate Group and Department of Plant Sciences, University of California, Davis, California 95616

This review describes recent advances in the analysis of metabolism using quantitative genetics. It focuses on how recent metabolic quantitative trait loci (QTL) studies enhance our understanding of the genetic architecture underlying naturally variable phenotypes and the impact of this fundamental research on agriculture, specifically crop breeding. In particular, the role of whole-genome duplications in generating quantitative genetic variation within a species is highlighted and the potential uses of this phenomenon presented. Additionally, the review describes how new observations from metabolic QTL mapping analyses are helping to shape and expand the concepts of genetic epistasis.

INTRODUCTION

Biochemistry and genetics have long been entwined in a mutualistic relationship beginning with Mendel's reliance upon metabolic phenotypes (anthocyanins and starch) to develop basic genetic theory. Biochemistry has advanced the understanding of DNA structure and metabolism, and genetic analyses have been equally crucial for the construction of biochemical pathways (Tatum et al., 1950; Watson and Crick, 1953; Yanofsky, 1960). The relationship between biochemistry and genetics has yielded great benefits to the broader scientific community. For instance, the analysis of genetic mutants in the *Escherichia coli* lactose and Trp biosynthetic pathways have shaped our current understanding of gene regulation (Yanofsky, 1960; Jacob and Monod, 1961). Even today, the integration of genetic and modern metabolic network theory are forging connections between genome-scale data and organismal phenotypes (Segre et al., 2005; Gjuvsland et al., 2007; Fu et al., 2009).

The combination of biochemistry and genetics has been of great utility in the study of plants. Biochemical phenotypes generated by defects in processes such as flavonol and Trp biosynthesis have been fundamental to developing basic genetic approaches within model plants (Koornneef et al., 1982; Last and Fink, 1988; Shirley et al., 1995). Genetic studies have also identified biochemical processes that proved recalcitrant to solely biochemical approaches (Humphreys and Chapple, 2002; Xie et al., 2003). Ecological and evolutionary genetic studies of plants have relied strongly on biochemical phenotypes, predominantly those manifested as variation in flower color, to further our understanding of how both biochemistry and genetics control fitness in wild plant populations (Brown and Clegg, 1984; Stanton et al., 1991; Holton et al., 1993).

The combination of biochemistry and genetics can also provide powerful insights into the origins and maintenance of natural

variation. Modern biochemical analyses allow quantification of metabolic diversity that can be associated with specific genetic markers, mRNA transcripts, and enzyme activities, allowing linkage between variation from genetic to biochemical levels that is more complicated for less-defined or more pleiotropic phenotypes, such as growth or disease susceptibility. The ability to analyze variation at different functional levels provides an opportunity to query the mechanisms by which this variation evolves and is maintained in populations and species.

Recent advances in genome sequencing and high-throughput phenotyping technologies allow plant biologists to revisit the study of quantitative genetics and how genetic variation is connected to phenotypic variation. The combination of biochemistry with quantitative genetics has a long history in plants, as one of the first identified QTL controls a biochemical phenotype, specifically, seed color in *Phaseolus* (Sax, 1923). More recently, these combined disciplines have expanded our fundamental understanding of how metabolites can provide biotic stress resistance in both *Zea mays* and *Brassica* (Mithen and Magrath, 1992; Magrath et al., 1993; Mithen and Toroser, 1995; Byrne et al., 1996, 1998; McMullen et al., 1998). The advent of broad-spectrum metabolite profiling technologies is enabling a fuller investigation into the influence of quantitative genetic variation on the plant metabolome and correspondingly how the fitness consequences of these metabolic changes alter the genetic architecture of the species. Several previous reviews cover the use of metabolic QTL to integrate across different levels of genomic information (sequence, transcript, and protein) to better understand biomass production, improve crop breeding, and obtain ecological inference about the corresponding selective pressure acting on these QTL (Fernie and Schauer, 2009; Keurentjes, 2009; Kliebenstein, 2009). These reviews also provide significant evidence for the myriad specific molecular events that can lead to altered phenotypes, including epigenetic, coding sequence, and regulatory polymorphisms. This review focuses on how recent metabolic QTL studies enhance our fundamental understanding of the genetic architecture underlying

¹ Address correspondence to kliebenstein@ucdavis.edu.
www.plantcell.org/cgi/doi/10.1105/tpc.109.067611

naturally variable phenotypes and the impact of this basic research on agriculture, specifically crop breeding.

METABOLIC QTL AND THE UNDERPINNINGS OF NATURAL GENETIC VARIATION

At a basic level, we presume that we understand the very underpinnings of qualitative and consequentially Mendelian genetics, concepts such as allelic dominance, heritability, and epistasis (Table 1). However, all of these concepts exist on quantitative scales, and there is little knowledge about the distributions of these genetic forces. Understanding these distributions and any relationship between them is essential to developing predictive approaches for the analysis of natural genetic variation, irrespective of the field of study. Due to its relative economy, metabolomics or metabolic profiling is becoming a popular phenotyping platform for quantitative genetic analysis of biochemistry at a genome-wide level.

In a multiyear study of tomato (*Solanum lycopersicum*) metabolism, 76% of metabolic QTL displayed a dominance relationship between the alleles, whereas 18% of the metabolic QTL had a semidominant relationship (Schauer et al., 2006, 2008). Interestingly, the distribution of dominance effects varied between metabolic classes, with sugar alcohol-linked QTL showing the lowest proportion of semidominance and organic acids showing the highest proportion of semidominance. This suggests that different phenotypic classes can have dissimilar dominance relationships among naturally variable alleles, possibly due to their different underlying molecular networks. Within gene expression QTL, *cis*-acting expression polymorphisms show greater additive effects, whereas *trans*-acting polymorphisms are more likely to show a dominant effect. One explanation for this discrepancy is that a *cis*-acting polymorphism by definition cannot affect the other allele in a heterozygous individual and thereby prevent it from being dominant. It remains to

be seen if this relationship differs between networks or related phenotypes as with metabolic traits (Stupar and Springer, 2006; Lemos et al., 2008).

Heritability may also differ among classes of metabolic traits. The majority of metabolite studies estimate broad-sense heritability (H^2), which combines all potential genetic and epigenetic contributions to the phenotype. Using this approach, an analysis of tomato metabolites showed that H^2 significantly varied among metabolites, revealing distinct clusters of high and low heritability metabolites (Schauer et al., 2008). This observation was echoed by an analysis of *Arabidopsis thaliana* secondary metabolites showing that two different classes of secondary metabolite, indolic and aliphatic glucosinolates, detected on the same phenotyping platform had distinctly different H^2 estimates in multiple mapping populations (Kliebenstein et al., 2001b; Wentzell et al., 2007). Interestingly, the transcripts encoding the biosynthetic enzymes for these metabolites had higher estimated heritability than the metabolites themselves (Wentzell et al., 2007; West et al., 2007). Thus, it is clear that heritability (e.g., genetic variation) can differ across molecular levels: transcript, metabolite, and physiology, for what is typically considered as the same phenotype, processes involved in glucosinolate accumulation. This difference in heritability could be due to each higher-order phenotype (transcript, metabolite, and physiology) integrating more and different inputs than the lower-order phenotypes. For instance, larger or more complex regulatory networks may affect a phenotype at the metabolic level than at the transcript level. Alternatively, each additional phenotypic level may accumulate additional stochastic noise, thus decreasing the estimate of heritability (Elowitz et al., 2002; Raser and O'Shea, 2004). Broader tests of measurable phenotypes in multiple species and mapping population are required to separate these possibilities.

Application

A major goal of combining genomic analysis with quantitative genetics applications, such as plant breeding, is to allow a predictive connection from genetic to phenotypic variation. This would allow crop developers to use genomic information efficiently to engineer a phenotypic change within the crop. However, this requires a deep understanding of basic genetic components like heritability and allelic dominance. Understanding how these basic variables differ among phenotypes will allow genomic information to be incorporated more efficiently into breeding schemes. A complication of this application is that broad-sense heritability is less useful for breeding applications than narrow-sense heritability, which is the genetic variation that additively contributes to the phenotype (Table 1). Broad- and narrow-sense heritability estimates can differ dramatically depending on the level of allelic dominance, multilocus epistasis, and parental influence on the phenotype (Falconer and Mackay, 1996; Lynch and Walsh, 1998). Studies have not yet thoroughly queried the relationship between these two measures of heritability on a genomic scale, which will be required to enable the full use of genomics to predict phenotypic consequences and allow for predictive engineering of crops.

Table 1. How Genetics Terms Are Defined within This Article

Term	Definition
Dominance	The phenotypic relationship between alleles of a single gene. Complete dominance indicates that phenotype associates with one allele dominates in a heterozygote, whereas semidominance indicates that the heterozygote phenotype is intermediate between the two alleles.
Epistasis	A nonadditive interaction of genetic variation at two or more loci upon a phenotype.
Heritability	This can be partitioned into the two following concepts that are not equivalent due to the action of dominance and epistasis.
Narrow-sense heritability	h^2 is the degree to which offspring resemble the parents.
Broad-sense heritability	H^2 is the effect of any aspect of the genotypic variation upon phenotypic variation.
QTL	A region of the genome where genotypic variation links to phenotypic variation.

METABOLIC QTL AND WHOLE-GENOME DUPLICATIONS

The study of metabolic QTL may shed light on evolutionary questions regarding the fate of duplicated genes. If a duplication event produces two copies of a gene that are afterwards maintained, these duplicated genes may diverge from the ancestral gene or from each other. This divergence can occur via neofunctionalization, where one of the duplicates obtains a novel function, or subfunctionalization, such that the duplicate copies obtain differential expression patterns in terms of tissue specificity or stress response (Ohno, 1970; Lynch and Conery, 2000). Interestingly, most duplicates from whole-genome duplications are lost over long time frames as represented by speciation events (Freeling, 2009), yet within species, these duplicates appear to play significant roles in QTL formation and maintenance (Kliebenstein, 2008). While further work is necessary to test if this is a biological discrepancy or an observational bias, the analysis of metabolic QTL has both identified examples of subfunctionalization and neofunctionalization and identified new roles for these evolutionary mechanisms within a species.

Neofunctionalization and Metabolic Diversity

The cloning of metabolic QTL, largely for secondary metabolites, has revealed several examples of neofunctionalization where duplicated genes have evolved independent biochemical functions (Kliebenstein et al., 2001a; Lambrix et al., 2001; de Meijer et al., 2003; Benderoth et al., 2006; Schnee et al., 2006; Burow et al., 2009). For example, a tandem duplication generated two genes in glucosinolate metabolism, *AOP2* and *AOP3*, that evolved different biochemical reactions using the same substrate (Kliebenstein et al., 2001a). *AOP2* maintained the ancestral function of converting a methylsulfinyl moiety to an alkenyl, while *AOP3* derived the ability to convert the methylsulfinyl to a hydroxyalkyl. Showing that QTL controlling metabolic diversity in secondary metabolites result directly from neofunctionalization suggests that gene duplication may facilitate plant evolution of new responses to a complex and changing environment, partly through increasing metabolic diversity (Ober, 2005). This agrees with recent work showing that duplicated genes are over-represented in responses to biotic and abiotic stress (Hanada et al., 2008; Kliebenstein, 2008).

Subfunctionalization and Expression Diversity

Duplicate genes may also subfunctionalize, developing expression patterns that differ spatially or temporally from ancestral expression patterns. A hallmark of subfunctionalization in metabolic QTL analysis is detection of tissue-specific QTL for an enzymatic activity, which is likely to have resulted from differential expression patterns between members of an encoding gene family (Sergeeva et al., 2004). These analyses can be extended to identify the genes controlling the biochemical reaction in specific tissues (Sergeeva et al., 2006). However, more detailed metabolic QTL experiments are beginning to show that gene duplicates diverging by tissue expression pattern may represent only the simplest scenario, while subfunctionalization can generate complex expression patterns that integrate development, on-

togeny, and environment such that duplicated genes may show highly conditional effects (Wentzell et al., 2008; Wentzell and Kliebenstein, 2008; Burow et al., 2009). In one example, the main metabolic QTL controlling secondary metabolite-mediated insect resistance within *Arabidopsis*, *ESP* and *ESM1*, result from two successive whole-genome duplications generating four highly similar genomic regions (Wentzell et al., 2008). Further endoreduplication and gene loss accompanied by promoter divergence has generated QTL conditional on tissue, ontogeny, and environmental influences (Lambrix et al., 2001; Zhang et al., 2006; Wentzell and Kliebenstein, 2008; Burow et al., 2009). Gene duplication, followed by sequence changes that altered expression patterns of duplicated genes, has apparently given the plant more precise control over its metabolic defenses. As such, subfunctionalization may be more prevalent in biotic stress responses. Given that the above observations pertain to precise regulatory variation among diverse genotypes, this highly integrated regulation would be difficult to study using single gene mutants from a single reference genotype but is readily accessible in natural populations. The frequency of these higher-order interactions remains to be determined.

Genetic Subfunctionalization

In addition to obtaining new functions or novel partitioning of function, metabolic QTL studies suggest that duplicate genes can obtain independent loss-of- or decrease-in-function polymorphisms, essentially a genetic subfunctionalization where genes have different levels of function in different genotypes. The simplest form of this genetic subfunctionalization is exemplified by the recent cloning of two enzymatic QTL that combine to form a lethal epistatic interaction (Bikard et al., 2009). In this example, a gene encoding a key enzyme for His biosynthesis, *HPA*, was duplicated within *Arabidopsis* and then divergent accessions independently lost one or the other copy of the gene. F₂ progeny generated from crosses of these divergent accessions showed segregation patterns where one-sixteenth of the progeny lacked both copies of the *HPA* gene and were thus nonviable (Bikard et al., 2009). A secondary metabolism family of flavin-monooxygenases shows hallmarks of genetic subfunctionalization (Li et al., 2008). However, it is still unknown how

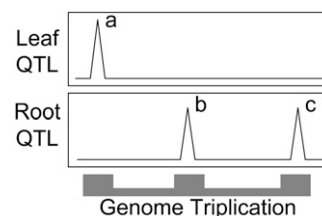


Figure 1. Homoeologous QTL.

Two hypothetical QTL results are shown for the same phenotype in the leaf and root. Peak locations represent the approximate genetic positions of the loci, with small letters labeling the three loci. The bar at the bottom shows the position of the three corresponding genomic regions arising via a triplication of a single original genomic fragment.

frequently genetic subfunctionalization may occur within larger gene families, especially given that gene duplicates show elevated levels of transcript abundance polymorphism (Kliebenstein, 2008).

Application

Understanding the complex contributions of gene duplication to plant metabolism may benefit efforts to clone individual metabolic QTL and provide precise markers for marker-assisted selection. Essentially, identification and cloning of a QTL for a given trait in a single tissue, condition, or genotype implicates homoeologous members of the same gene family that colocalize with the other QTL as candidate genes for control of the trait in other tissues or conditions (Figure 1). For an example, take a species where a gene has been triplicated via a whole-genome duplication and an ensuing segmental duplication of only one of the resulting copies (Figure 1). If a researcher identified the molecular basis of QTL^a that controls a phenotype in the shoot and root, the homoeologous duplicates of this gene might be considered strong candidates for control of root-specific QTL. As such, cloning one QTL would potentially accelerate identification of the further subfunctionalized duplicate genes with minimal investment of new resources. More broadly, whole-genome assessments of duplication patterns (Vision et al., 2000) could be analyzed in concert with genomic positions of QTL for a given trait to identify relationships between QTL incidence and genome duplication events.

METABOLIC QTL AND MOLECULAR MECHANISMS UNDERLYING GENETIC EPISTASIS

Epistasis is loosely defined as an interaction between alleles at two or more genes, and there is long-standing debate about the presence and relevance of epistasis in genetic studies (Fisher, 1930; Wright, 1931; Carlborg and Haley, 2004; Hill et al., 2008). In functional molecular studies, epistasis generally implies a regulatory or physical interaction between the genes and their products, specified here as functional epistasis (Phillips, 2008). By contrast, epistasis in genetic studies indicates that the link between allelic variation and phenotypic variation at one locus is dependent upon the allelic state at another locus or loci, specified here as genetic epistasis. In a quantitative genetics study, epistasis appears as a statistical interaction between loci that does not allow for immediate divination of the molecular causes.

One area where epistatic interactions between naturally variable loci is of key importance is the study of genome incompatibility in *Arabidopsis*, *Mimulus* spp, and other plants (Fishman and Willis, 2001; Bomblies et al., 2007; Sweigart et al., 2007; McDaniel et al., 2008). These are often loosely classified as Dobzhansky-Muller incompatibility loci whereby the two genes have diverged in function such that when reintroduced they cause a hybrid incompatibility that may be associated with speciation events (Dobzhansky, 1937; Muller, 1942). Furthermore, there is significant interest in the analysis of hybrid necroses, incompatibility combinations that alter plant defense mechanisms (Bomblies et al., 2007; Alcazar et al., 2009). Ge-

netic incompatibilities are not phenotypically limited but can result from any multilocus combination of alleles that generates negative effects on diverse phenotypes, such as chloroplast development, disease resistance, and pollen viability (Fishman and Willis, 2001; Bomblies et al., 2007; Sweigart et al., 2007; McDaniel et al., 2008). Similar arguments have been made to explain the positive impacts of genome mixing upon heterosis or hybrid vigor (Birchler et al., 2003). However, current data are insufficient to connect observations of genetic epistasis with molecular mechanisms consistent with solely a functional epistasis explanation.

It is in the arena of explaining the basis of genetic epistasis that the analysis of metabolic QTL displays its power to address basic genetic questions while remaining useful in applied settings such as crop breeding. Metabolic QTL analyses are beginning to identify three broad classes of molecular mechanism underlying genetic epistases in natural populations: functional epistasis, gene family epistasis, and diffuse epistasis. The following sections define these three postulated forms of epistasis and provide examples identified via metabolic QTL studies.

Functional Epistasis and Metabolite-Transcription Factor Communication

The most familiar form of epistasis is functional epistasis, where genes are connected within a biochemical or regulatory pathway. Cloned metabolic QTL reveal numerous instances where genetic epistasis is controlled by traditional functional epistasis. For example, in the maize (*Zea mays*) flavone biosynthetic pathway, multiple regulatory loci interact to determine the level of this defensive compound (Byrne et al., 1996, 1998; McMullen et al., 1998). In agreement, regulatory interactions that simultaneously and directly control transcriptional, enzymatic, and metabolite levels for a specific metabolite or pathway have been identified within *Arabidopsis* (Fu et al., 2009). These studies demonstrate how standard regulatory or enzymatic interactions can generate genetic epistasis through functional epistasis. However, there are also numerous instances where a given QTL does not control all three levels, suggesting that there are complex and potentially different networks controlling these three levels (Keurentjes et al., 2006, 2007, 2008).

One such potential complexity is that epistatic interactions are not limited to functional networks involving simple linear pathways. The *Arabidopsis* glucosinolate biosynthetic pathway presents an advanced example of the link between genetic and functional epistasis. Six QTL controlling glucosinolate accumulation show genetic epistasis (Kliebenstein et al., 2005; Kliebenstein, 2009) (Figure 2A). These six loci act together to control both the type and relative level of glucosinolate accumulation, and the underlying genes encode four enzymes and two transcription factors (Kliebenstein et al., 2005; Sønderby et al., 2007; Hansen et al., 2008; Kliebenstein, 2009) (Figure 2B). This work showed that the two transcription factor loci, *MYB28* and *MYB29*, directly bind promoters to regulate transcript abundance of the four enzyme-encoding genes, *GS-OX*, *AOP2*, *GS-OH*, and *MAM1*, as well as other genes encoding glucosinolate biosynthetic enzymes (Gigolashvili et al., 2007, 2008; Hirai et al., 2007) (Figure 2B). Whereas most models of functional epistasis are sufficiently

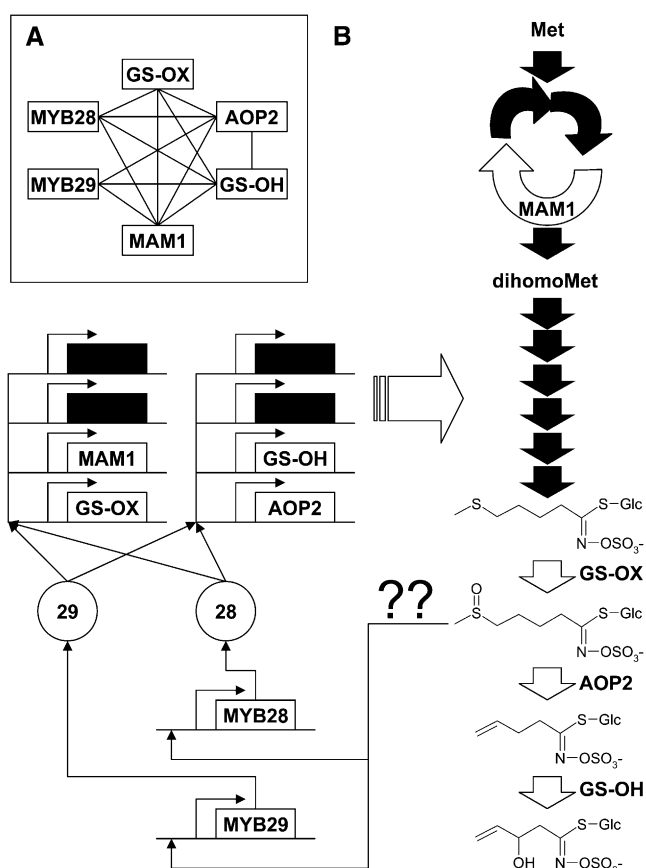


Figure 2. Functional Epistasis Underlying Complex Genetic Epistasis.

(A) The known genetic epistasis linking six cloned QTL controlling glucosinolate accumulation within *Arabidopsis* is shown. The names of the QTL are shown with lines indicating genetic epistatic interactions identified within any of the existing natural *Arabidopsis* populations.

(B) The current model of functional epistasis underlying the above genetic epistasis is shown. Small arrows show specific enzymes with names showing those enzymes that are the mechanistic basis of specific QTL in (A). Metabolic structures show the specific metabolites that accumulate from these enzymes. The line arrows show the proposed link between a glucosinolate metabolite and the control of transcription for the MYB28/MYB29 transcription factors. The double question mark shows that this is just one possible connection among many metabolite/gene interactions. The circles represent the MYB28/MYB29 proteins with line arrows showing the promoters with which these transcription factors interact. These genes encode the enzymes shown by the arrows for the metabolic pathway.

explained with transcription factors controlling enzymes, the glucosinolate QTL analyses showed that enzyme-encoding QTL altered transcript abundance for both the two MYB transcription factors and the other biosynthetic genes, suggesting an effect of these enzymes upon transcript levels (Kliebenstein et al., 2006; Wentzell et al., 2007). Transgenic introduction of one of these enzymes, *AOP2*, into a null background altered transcript levels for the transcription factors as well as the biosynthetic pathway, suggesting that one of the metabolic products of *AOP2* can modulate transcript levels of the *MYB28* and *MYB29* transcrip-

tion factors (Wentzell et al., 2007). Thus, this complex genetic epistasis involves signaling loops whereby transcription factors control expression of the biosynthetic pathway that in turn affect transcript accumulation of the transcription factors (Sønderby et al., 2007; Wentzell et al., 2007) (Figure 2). As such, the analysis of metabolic QTL is providing a richer image of functional epistasis than traditional linear relationship assumptions have allowed.

Genetic Epistasis within Gene Families

In contrast with functional epistasis, genetic epistasis does not inherently require a molecular functional interaction between genes. The formation of a gene family with similar enzymatic function, possibly via a series of duplication events, provides a possible mechanism for the generation of genetic epistasis in the absence of any functional interaction among loci. The simplest model for this form of genetic epistasis is a two-gene system with unlinked loss-of-function alleles, as in the above example of His biosynthesis (see Genetic Subfunctionalization) (Bikard et al., 2009). However, this form of genetic epistasis is not limited to gene families with only two members. It has been predicted that larger gene families whose members are all involved a single biochemical reaction could generate genetic epistasis affecting the related metabolites (Keightley, 1989). This arises because the amount of enzyme does not translate linearly into how much precursor or product is available (Figure 3). Instead, enzyme/metabolite relationships typically follow a curvilinear association governed by Michaelis-Menten kinetics. In the example model, an enzymatic activity is encoded by 10 genes, each containing a natural knockout polymorphism that accounts for five units of activity (Figure 3). This model generates a linear gradient of enzymatic activity across the different QTL combinations, which, following Michaelis-Menten kinetics, translates into an epistatic surface for metabolite accumulation (Figure 3). This epistatic relationship is not obvious, as most of the single and multiple gene combinations have apparently additive effects upon the metabolite (Figure 3, see range from 40 to 100 enzymatic units). As the number of knockouts increases, however, the organism experiences the enzyme kinetics represented on the left-hand side of the graph, and genetic epistasis can be detected among the loci (Figure 3, see range from 0 to 40 enzymatic units). Thus, theoretically it is possible for the existence of gene families to cause genetic epistasis. This further predicts that epistasis may not affect all trait levels (metabolite, transcript, and physiology) equally (Figure 3, compare enzyme to metabolite).

Recent metabolic QTL studies have validated the existence of this more complex model of gene family-based genetic epistasis in natural populations. A key step of glucosinolate metabolism in *Arabidopsis* is the oxidation of a methylthiol to a methylsulfinyl glucosinolate (Kliebenstein et al., 2001c; Hansen et al., 2007). This reaction is controlled by five different genes encoding similar enzymatic activities that also exhibit natural variation in transcript abundance (Kliebenstein et al., 2001c; Hansen et al., 2007; Kliebenstein, 2008; Li et al., 2008). Interestingly, no epistatic influence on transcript accumulation was detected for these loci, although production of the methylsulfinyl product exhibited epistatic interactions (Wentzell et al., 2007; Li et al.,

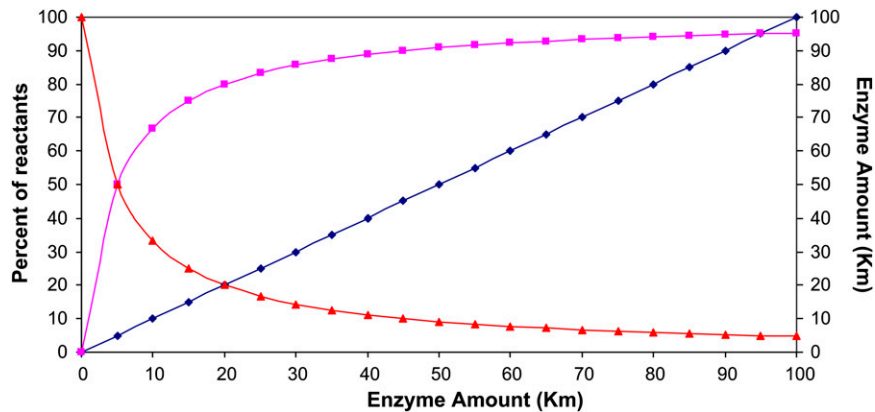


Figure 3. Gene Family Epistasis.

Michaelis-Menten kinetics controlling the relationship between gene-family polymorphisms and genetic epistasis is shown. The line with diamonds represents the impact of stacking 10 QTL each with two alleles adding a 5 unit (in K_m terms) effect upon the accumulation of the enzyme, thus providing a range from 0 units of activity to 100. Michaelis-Menten equations were then used to approximate the amount of product (squares) and substrate (triangles) that would likely accumulate within the plant under this simple model.

2008) as expected from the model (Figure 3). Further evidence that the existence of gene families can allow for genetic epistasis came from an analysis of expression QTL affecting all identified and predicted biochemical enzymes within *Arabidopsis* (Kliebenstein, 2008). This study showed that duplicated genes individually have an elevated level of transcript polymorphism due to sequence polymorphisms at or near their physical position. Thus, gene family members are more polymorphic than the average gene and the polymorphisms between individual genes can combine to generate a wide range of phenotypic diversity within the gene family's encoded function. Thus, gene families arising from duplication may be maintained in a genome not only through changes in function but also by the development of complex genetic interactions among gene family members.

Network Architecture and Diffuse Epistasis

A final form of epistasis is diffuse epistasis, and its existence is supported by the image of metabolic QTL provided by metabolic profiling (Whitlock et al., 1995). Diffuse epistasis arises from the control of many biological processes by networks containing complex interactions and multiple crosstalking signals. These highly connected biological networks create more robust links between input and output, protecting biological systems from fluctuations (Barkai and Leibler, 1997; Freeman, 2000; Yi et al., 2000; Dorogovtsev and Mendes, 2002; Spirin and Mirny, 2003; Boccaletti et al., 2006). Although these fluctuations are often considered to be random biological noise, they could also represent genetic variation that is dampened by the network (Figure 4). Networks are largely resistant to the effects of single or a few alterations, but they become increasingly likely to fail with the introduction of multiple unlinked defects of small to moderate effect (Albert et al., 2000; Albert and Barabasi, 2002; Agoston et al., 2005; Csermely et al., 2005). Where network structure is

known, these failures are to some extent predictable, as shown in Figure 4 for three unlinked genetic mutations that completely disrupt the input-output connection. The corresponding three genetic loci, although not identifying any direct functional epistatic link, would identify a diffuse interaction when viewed in combination. QTL mapping populations contain genotypes that represent a random assortment of hundreds to thousands of

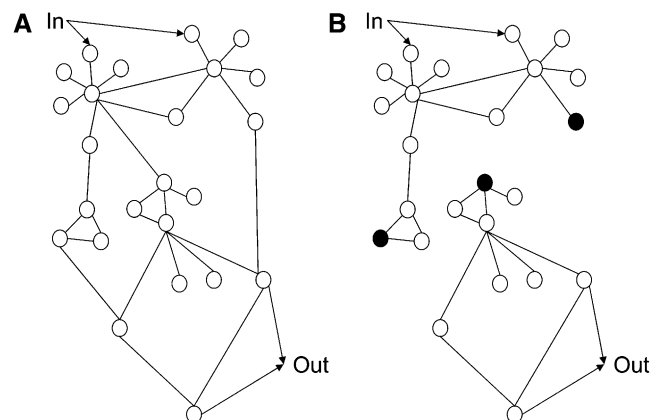


Figure 4. Power Network and Diffuse Epistasis.

A simplified network diagram showing the multiple independent paths linking any given input to any related output is shown. The circles can be considered either a gene or the protein encoded by the gene. Lines indicate a mechanistic connection such that information or metabolic flux can pass from one node to another.

(A) The fully functional network with all associated redundancy.

(B) The effects of introducing three polymorphisms (black circles) to the system and the unexpected disruption of the link between input and output.

independent genetic defects present within the parents and thus serve as an excellent tool to test for multilocus diffuse epistasis (Clark et al., 2007).

One impact of diffuse epistasis on plant metabolism is that the metabolic network is vulnerable to combinations of independent mutations. Loss of metabolic network function would then affect other phenotypes, such as growth. This link between metabolite QTL and growth has been observed in several *Arabidopsis* populations (Keurentjes et al., 2006; Meyer et al., 2007; Liseć et al., 2008; Fu et al., 2009). However, despite predictions that diffuse epistasis occurs in these systems, no significant epistasis was observed in some of these studies (Meyer et al., 2007; Liseć et al., 2008). A potential explanation of this lack of detectable epistatic effects is that diffuse epistasis involves combinations of polymorphisms that would cause extreme phenotypic values, which frequently are treated as statistical outliers and thus eliminated prior to analysis (Liseć et al., 2008). In another experiment querying metabolic QTL, diffuse epistasis appeared to control the tricarboxylic acid cycle (Rowe et al., 2008). Combined polymorphism in at least three loci caused a fourfold to eightfold increase in tricarboxylic acid metabolites as well as stunted plant growth. Only a single specific combination of alleles at the three loci showed a dramatic effect, consistent with expectations for diffuse epistasis (Rowe et al., 2008). Identifying the underlying genetic bases of this observed multilocus effect on metabolism will further test this concept.

Network collapse is most fully explored within the context of power distribution grids, specifically focused on the relation of grid architecture to the frequency of defects that will lead to network collapse (Dobson et al., 2005, 2007; Nedic et al., 2006; Ren and Dobson, 2008). While power distribution grids obviously differ from plant metabolic networks, similarities include the need to balance energy generation with energy use and storage. It may thus be possible to apply power grid distribution theory to our understanding of how complex genetic interactions can generate the plant phenotypic equivalent of network collapse. Power grid distribution theory suggests that the frequency of catastrophic system failures increases with the number of combined defects, with the practical implication that large mapping populations are required to query the potential for these effects (Dobson et al., 2005, 2007; Nedic et al., 2006; Ren and Dobson, 2008). Smaller populations do not allow sufficient sampling of genotypic combinations at more than three loci, causing ascertainment bias in small populations. For instance, in the example in Figure 4, the combination of alleles leading to a network collapse would only be identified in one out of 64 F₂ progeny, assuming that this network collapse was not lethal. As such, much larger populations are necessary to study the prevalence and biological impacts of diffuse epistasis.

CONCLUSIONS AND FUTURE PERSPECTIVES

Recent application of broad-spectrum metabolite profiling to the study of natural variation and quantitative genetics has provided unique insights into the basic underpinnings of plant genetics. These insights may inform numerous fields of research, ranging from gene evolution and network biology to the application of quantitative genetics in crop breeding. As this new information is

Application

It has been difficult to incorporate the effects of epistasis into breeding programs due to the increased population sizes required for statistical detection of epistases and the necessity of simultaneously breeding for multiple loci to enhance phenotypes controlled by multilocus interactions. However, recent advances in genomics and genotyping provide the ability to rapidly genotype massive populations, providing the potential to efficiently incorporate epistasis into breeding designs. Functional epistases will be relatively easy to apply to crop improvement once the genes involved in a pathway and their functional interactions are known.

Gene family and diffuse epistasis may be of less immediate practical value and may indeed present an unexpected barrier to some breeding regimes. In the example of gene-family epistasis, to increase the amount of the hypothetical substrate within the crop, a relatively straightforward approach would incorporate loss-of-function mutations at all 10 loci to block the reaction (Figure 4). While this might increase the content of the desired metabolite, it may also cause deleterious consequences if the downstream products have other roles in the plant. As such, gene family and diffuse epistasis raise the potential that the blind stacking of numerous QTL alleles associated with increased accumulation of a given metabolite, without regard for the molecular bases of the QTL and its genetic interactions, could lead to unexpected and possibly unpredictable deleterious consequences. This may occur even if the positive alleles come from different pathways (Figure 4).

Both gene family epistasis and diffuse epistasis represent a breakdown in the relationship between phenotypic consequence and the level of explained phenotypic variance within the population (Keightley, 1989). In this analysis, it was noted that there is not an inherent relationship between the amount of population variance explained by a QTL and the QTL's effect upon the phenotype. This relationship between phenotypic consequence and explained phenotypic variance rapidly breaks down with increases in the involved number of loci or the number of alleles per locus (Keightley, 1989). As such, although it is possible to alter phenotypes dramatically by stacking genes that participate in diffuse and gene-family epistases, large populations are required to first identify these complex higher-order genetic interactions. Additionally, the presence of more than two alleles at any locus in a population will necessitate even more dramatic increases in population size. In the example shown in Figure 4, it would likely take a population of at least 256 F₂ progeny to identify just four extreme phenotype individuals; exponential increases in population size could be necessary to provide the statistical support necessary to confidently pursue these results. Most QTL discovery populations are not massive enough for rigorous analysis of the tails of any phenotypic distribution, where genotypes revealing both gene family and diffuse epistasis would reside. Additionally, there is the chance that these genotypes will show a lethal interaction and may only be identifiable by multilocus segregation distortion (Bikard et al., 2009).

at present largely limited to two plant species, the expansion of these techniques into other species will test the generality of these observations and reveal the malleability of genetic architecture across phylogenetic distance. Given the intimate link between plant metabolism and physiology, it is likely that the combination of biochemistry and genetics will be as powerful in the future of quantitative genomics as it has proven to be for genetics in the past century.

ACKNOWLEDGMENTS

I thank Eva Chan, Heather Rowe, Adam Wentzell, Jason Corwin, and Rachel Kerwin for reading and commenting on drafts of this manuscript as well as for dealing with my ramblings. Funding for this work was provided by National Science Foundation Grant DBI 0820580 to D.J.K.

Received April 3, 2009; revised May 12, 2009; accepted June 4, 2009; published June 12, 2009.

REFERENCES

- Agoston, V., Csermely, P., and Pongor, S. (2005). Multiple weak hits confuse complex systems: A transcriptional regulatory network as an example. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **71**: 051909.
- Albert, R., and Barabasi, A.L. (2002). Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**: 47–97.
- Albert, R., Jeong, H., and Barabasi, A.L. (2000). Error and attack tolerance of complex networks. *Nature* **406**: 378–382.
- Alcazar, R., Garcia, A.V., Parker, J.E., and Reymond, M. (2009). Incremental steps toward incompatibility revealed by *Arabidopsis* epistatic interactions modulating salicylic acid pathway activation. *Proc. Natl. Acad. Sci. USA* **106**: 334–339.
- Barkai, N., and Leibler, S. (1997). Robustness in simple biochemical networks. *Nature* **387**: 913–917.
- Benderoth, M., Textor, S., Windsor, A.J., Mitchell-Olds, T., Gershenson, J., and Kroymann, J. (2006). Positive selection driving diversification in plant secondary metabolism. *Proc. Natl. Acad. Sci. USA* **103**: 9118–9123.
- Bikard, D., Patel, D., Le Mette, C., Giorgi, V., Camilleri, C., Bennett, M.J., and Loudet, O. (2009). Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. *Science* **323**: 623–626.
- Birchler, J.A., Auger, D.L., and Riddle, N.C. (2003). In search of the molecular basis of heterosis. *Plant Cell* **15**: 2236–2239.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.U. (2006). Complex networks: Structure and dynamics. *Phys. Rep.* **424**: 175–308.
- Bombliès, K., Lempe, J., Eppele, P., Warthmann, N., Lanz, C., Dangl, J.L., and Weigel, D. (2007). Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. *PLoS Biol.* **5**: 1962–1972.
- Brown, B.A., and Clegg, M.T. (1984). Influence of flower color polymorphism on genetic transmission in a natural population of the common morning glory, *Ipomoea purpurea*. *Evolution Int. J. Org. Evolution* **38**: 796–803.
- Burrow, M., Losansky, A., Muller, R., Plock, A., Kliebenstein, D.J., and Wittstock, U. (2009). The Genetic basis of constitutive and herbivore-induced ESP-independent nitrile formation in *Arabidopsis*. *Plant Physiol.* **149**: 561–574.
- Byrne, P.F., McMullen, M.D., Snook, M.E., Musket, T.A., Theuri, J.M., Widstrom, N.W., Wiseman, B.R., and Coe, E.H. (1996). Quantitative trait loci and metabolic pathways: Genetic control of the concentration of maysin, a corn earworm resistance factor, in maize silks. *Proc. Natl. Acad. Sci. USA* **93**: 8820–8825.
- Byrne, P.F., McMullen, M.D., Wiseman, B.R., Snook, M.E., Musket, T.A., Theuri, J.M., Widstrom, N.W., and Coe, E.H. (1998). Maize silk maysin concentration and corn earworm antibiosis: QTLs and genetic mechanisms. *Crop Sci.* **38**: 461–471.
- Carlborg, O., and Haley, C.S. (2004). Epistasis: Too often neglected in complex trait studies? *Nat. Rev. Genet.* **5**: 618–625.
- Clark, R.M., et al. (2007). Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* **317**: 338–342.
- Csermely, P., Agoston, V., and Pongor, S. (2005). The efficiency of multi-target drugs: The network approach might help drug design. *Trends Pharmacol. Sci.* **26**: 178–182.
- de Meijer, E.P.M., Bagatta, M., Carboni, A., Crucitti, P., Moliterni, V.M.C., Ranalli, P., and Mandolino, G. (2003). The inheritance of chemical phenotype in *Cannabis sativa* L. *Genetics* **163**: 335–346.
- Dobson, I., Carreras, B.A., Lynch, V.E., and Newman, D.E. (2007). Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization. *Chaos* **17**: 026103.
- Dobson, I., Carreras, B.A., Lynch, V.E., Nkei, B., and Newman, D.E. (2005). Estimating failure propagation in models of cascading blackouts. *Probability in the Engineering and Informational Sciences* **19**: 475–488.
- Dobzhansky, T. (1937). *Genetics and the Origin of Species*. (New York: Columbia University Press).
- Dorogovtsev, S.N., and Mendes, J.F.F. (2002). Evolution of networks. *Adv. Phys.* **51**: 1079–1187.
- Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. *Science* **297**: 1183–1186.
- Falconer, D.S., and Mackay, T.F.C. (1996). *Introduction to Quantitative Genetics*. (Essex, UK: Longman, Harlow).
- Fernie, A.R., and Schauer, N. (2009). Metabolomics-assisted breeding: a viable option for crop improvement? *Trends Genet.* **25**: 39–48.
- Fisher, R.A. (1930). *The Genetical Theory of Natural Selection*. (Oxford, UK: Clarendon).
- Fishman, L., and Willis, J.H. (2001). Evidence for Dobzhansky-Muller incompatibilities contributing to the sterility of hybrids between *Mimulus guttatus* and *M. nasutus*. *Evolution Int. J. Org. Evolution* **55**: 1932–1942.
- Freeling, M. (2009). Bias in plant gene content following different sorts of duplication: Tandem, whole-genome, segmental or by transposition. *Annu. Rev. Plant Biol.* **60**: 433–453.
- Freeman, M. (2000). Feedback control of intercellular signalling in development. *Nature* **408**: 313–319.
- Fu, J., et al. (2009). System-wide molecular evidence for phenotypic buffering in *Arabidopsis*. *Nat. Genet.* **41**: 166–167.
- Gigolashvili, T., Engqvist, M., Yatusевич, R., Muller, C., and Flugge, U.I. (2008). HAG2/MYB76 and HAG3/MYB29 exert a specific and coordinated control on the regulation of aliphatic glucosinolate biosynthesis in *Arabidopsis thaliana*. *New Phytol.* **177**: 627–642.
- Gigolashvili, T., Yatusевич, R., Berger, B., Müller, C., and Flüggé, U.I. (2007). The R2R3-MYB transcription factor HAG1/MYB28 is a regulator of methionine-derived glucosinolate biosynthesis in *Arabidopsis thaliana*. *Plant J.* **51**: 247–261.
- Gjuvsland, A.B., Hayes, B.J., Omholt, S.W., and Carlborg, O. (2007). Statistical epistasis is a generic feature of gene regulatory networks. *Genetics* **175**: 411–420.
- Hanada, K., Zou, C., Lehti-Shiu, M.D., Shinozaki, K., and Shiu, S.H. (2008). Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiol.* **148**: 993–1003.
- Hansen, B.G., Kerwin, R.E., Ober, J.A., Lambrix, V.M., Mitchell-Olds,

- T., Gershenzon, J., Halkier, B.A., and Kliebenstein, D.J.** (2008). A novel 2-oxoacid-dependent dioxygenase involved in the formation of the goiterogenic 2-hydroxybut-3-enyl glucosinolate and generalist insect resistance in *Arabidopsis*. *Plant Physiol.* **148**: 2096–2108.
- Hansen, B.G., Kliebenstein, D.J., and Halkier, B.A.** (2007). Identification of a flavin-monooxygenase as the S-oxygenating enzyme in aliphatic glucosinolate biosynthesis in *Arabidopsis*. *Plant J.* **50**: 902–910.
- Hill, W.G., Goddard, M.E., and Visscher, P.M.** (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* **4**: e1000008.
- Hirai, M., et al.** (2007). Omics-based identification of *Arabidopsis* Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc. Natl. Acad. Sci. USA* **104**: 6478–6483.
- Holton, T.A., Brugliera, F., Lester, D.R., Tanaka, Y., Hyland, C.D., Menting, J.G.T., Lu, C.Y., Farcy, E., Stevenson, T.W., and Cornish, E.C.** (1993). Cloning and expression of cytochrome-P450 genes controlling flower color. *Nature* **366**: 276–279.
- Humphreys, J.M., and Chapple, C.** (2002). Rewriting the lignin roadmap. *Curr. Opin. Plant Biol.* **5**: 224–229.
- Jacob, F., and Monod, J.** (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**: 318–356.
- Keightley, P.D.** (1989). Models of quantitative variation of flux in metabolic pathways. *Genetics* **121**: 869–876.
- Keurentjes, J.J.B.** (2009). Genetical metabolomics: closing in on phenotypes. *Curr. Opin. Plant Biol.* (In Press).
- Keurentjes, J.J.B., Fu, J.Y., de Vos, C.H.R., Lommen, A., Hall, R.D., Bino, R.J., van der Plas, L.H.W., Jansen, R.C., Vreugdenhil, D., and Koornneef, M.** (2006). The genetics of plant metabolism. *Nat. Genet.* **38**: 842–849.
- Keurentjes, J.J.B., Fu, J.Y., Terpstra, I.R., Garcia, J.M., van den Ackerveken, G., Snoek, L.B., Peeters, A.J.M., Vreugdenhil, D., Koornneef, M., and Jansen, R.C.** (2007). Regulatory network construction in *Arabidopsis* by using genome-wide gene expression quantitative trait loci. *Proc. Natl. Acad. Sci. USA* **104**: 1708–1713.
- Keurentjes, J.J.B., Sulpice, R., Gibon, Y., Steinhauser, M.-C., Fu, J., Koornneef, M., Stitt, M., and Vreugdenhil, D.** (2008). Integrative analysis of genetic variation in enzyme activities of primary carbohydrate metabolism reveal distinct modes of regulation in *Arabidopsis thaliana*. *Genome Biol.* **9**: R129.
- Kliebenstein, D., Lambrix, V., Reichelt, M., Gershenzon, J., and Mitchell-Olds, T.** (2001a). Gene duplication and the diversification of secondary metabolism: side chain modification of glucosinolates in *Arabidopsis thaliana*. *Plant Cell* **13**: 681–693.
- Kliebenstein, D., West, M., van Leeuwen, H., Loudet, O., Doerge, R., and St. Clair, D.** (2006). Identification of QTLs controlling gene expression networks defined a priori. *BMC Bioinformatics* **7**: 308.
- Kliebenstein, D.J.** (2008). A role for gene duplication and natural variation of gene expression in the evolution of metabolism. *PLoS ONE* **3**: e1838.
- Kliebenstein, D.J.** (2009). A quantitative genetics and ecological model system: Understanding the aliphatic glucosinolate biosynthetic network via QTLs. *Phytochem. Rev.* **8**: 243–254.
- Kliebenstein, D.J., Gershenzon, J., and Mitchell-Olds, T.** (2001b). Comparative quantitative trait loci mapping of aliphatic, indolic and benzylic glucosinolate production in *Arabidopsis thaliana* leaves and seeds. *Genetics* **159**: 359–370.
- Kliebenstein, D.J., Kroymann, J., Brown, P., Figuth, A., Pedersen, D., Gershenzon, J., and Mitchell-Olds, T.** (2001c). Genetic control of natural variation in *Arabidopsis thaliana* glucosinolate accumulation. *Plant Physiol.* **126**: 811–825.
- Kliebenstein, D.J., Kroymann, J., and Mitchell-Olds, T.** (2005). The glucosinolate-myrosinase system in an ecological and evolutionary context. *Curr. Opin. Plant Biol.* **8**: 264–271.
- Koornneef, M., Jorna, M.L., Derswan, D., and Karssen, C.M.** (1982). The isolation of abscisic acid (ABA) deficient mutants by selection of induced revertants in non-germinating gibberellin sensitive lines of *Arabidopsis thaliana* (L) Heynh. *Theor. Appl. Genet.* **61**: 385–393.
- Lambrix, V., Reichelt, M., Mitchell-Olds, T., Kliebenstein, D., and Gershenzon, J.** (2001). The *Arabidopsis* epithiospecifier protein promotes the hydrolysis of glucosinolates to nitriles and influences *Trichoplusia ni* herbivory. *Plant Cell* **13**: 2793–2807.
- Last, R.L., and Fink, G.R.** (1988). Tryptophan-requiring mutants of the plant *Arabidopsis thaliana*. *Science* **240**: 305–310.
- Lemos, B., Araripe, L.O., Fontanillas, P., and Hartl, D.L.** (2008). Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression. *Proc. Natl. Acad. Sci. USA* **105**: 14471–14476.
- Li, J., Hansen, B.G., Ober, J.A., Kliebenstein, D.J., and Halkier, B.A.** (2008). Subclade of flavin-monooxygenases involved in aliphatic glucosinolate biosynthesis. *Plant Physiol.* **148**: 1721–1733.
- Lisec, J., Meyer, R.C., Steinfath, M., Redestig, H., Becher, M., Witucka-Wall, H., Fiehn, O., Torjek, O., Selbig, J., Altmann, T., and Willmitzer, L.** (2008). Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations. *Plant J.* **53**: 960–972.
- Lynch, M., and Conery, J.S.** (2000). The evolutionary fate and consequences of duplicate genes. *Science* **290**: 1151–1155.
- Lynch, M., and Walsh, B.** (1998). *Genetics and Analysis of Quantitative Traits*. (Sunderland, MA: Sinauer Associates).
- Magrath, R., Herron, C., Giamoustaris, A., and Mithen, R.** (1993). The inheritance of aliphatic glucosinolates in *Brassica napus*. *Plant Breed.* **111**: 55–72.
- McDaniel, S.F., Willis, J.H., and Shaw, A.J.** (2008). The genetic basis of developmental abnormalities in interpopulation hybrids of the moss *Ceratodon purpureus*. *Genetics* **179**: 1425–1435.
- McMullen, M.D., Byrne, P.F., Snook, M.E., Wiseman, B.R., Lee, E.A., Widstrom, N.W., and Coe, E.H.** (1998). Quantitative trait loci and metabolic pathways. *Proc. Natl. Acad. Sci. USA* **95**: 1996–2000.
- Meyer, R.C., Steinfath, M., Lisec, J., Becher, M., Witucka-Wall, H., Törjek, O., Fiehn, O., Eckardt, A., Willmitzer, L., Selbig, J., and Altmann, T.** (2007). The metabolic signature related to high plant growth rate in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **104**: 4759–4764.
- Mithen, R., and Toroser, D.** (1995). Biochemical genetics of aliphatic glucosinolates in Brassica and *Arabidopsis*. In *Amino Acids and their Derivatives in Higher Plants*, Society for Experimental Biology Seminar Series No. 56, R.M. Wallsgrave, ed, (Cambridge, UK: Cambridge University Press.), pp. 261–275.
- Mithen, R.F., and Magrath, R.** (1992). Glucosinolates and resistance to *Leptosphaeria maculans* in wild and cultivated Brassica species. *Plant Breed.* **108**: 60–68.
- Muller, H.** (1942). Isolating mechanisms, evolution, and temperature. *Biol. Symp.* **6**: 71–124.
- Nedic, D.P., Dobson, I., Kirschen, D.S., Carreras, B.A., and Lynch, V.E.** (2006). Criticality in a cascading failure blackout model. *Int. J. Electr. Power Energy Syst.* **28**: 627–633.
- Ober, D.** (2005). Seeing double: Gene duplication and diversification in plant secondary metabolism. *Trends Plant Sci.* **10**: 444–449.
- Ohno, S.** (1970). *Evolution by Gene Duplication*. (New York: Springer-Verlag).
- Phillips, P.C.** (2008). Epistasis - The essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* **9**: 855–867.
- Raser, J.M., and O'Shea, E.K.** (2004). Control of stochasticity in eukaryotic gene expression. *Science* **304**: 1811–1814.
- Ren, H., and Dobson, I.** (2008). Using transmission line outage data to

- estimate cascading failure propagation in an electric power system. *IEEE Transactions on Circuits and Systems II-Express Briefs* **55**: 927–931.
- Rowe, H.C., Hansen, B.G., Halkier, B.A., and Kliebenstein, D.J.** (2008). Biochemical networks and epistasis shape the *Arabidopsis thaliana* metabolome. *Plant Cell* **20**: 1199–1216.
- Sax, K.** (1923). The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* **8**: 552–560.
- Schauer, N., Semel, Y., Balbo, I., Steinfath, M., Repsilber, D., Selbig, J., Pleban, T., Zamir, D., and Fernie, A.R.** (2008). Mode of inheritance of primary metabolic traits in tomato. *Plant Cell* **20**: 509–523.
- Schauer, N., et al.** (2006). Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat. Biotechnol.* **24**: 447–454.
- Schnee, C., Kollner, T.G., Held, M., Turlings, T.C.J., Gershenson, J., and Degenhardt, J.** (2006). The products of a single maize sesquiterpene synthase form a volatile defense signal that attracts natural enemies of maize herbivores. *Proc. Natl. Acad. Sci. USA* **103**: 1129–1134.
- Segre, D., DeLuna, A., Church, G.M., and Kishony, R.** (2005). Modular epistasis in yeast metabolism. *Nat. Genet.* **37**: 77–83.
- Sergeeva, L.I., Keurentjes, J.J.B., Bentsink, L., Vonk, J., van der Plas, L.H.W., Koornneef, M., and Vreugdenhil, D.** (2006). Vacuolar invertase regulates elongation of *Arabidopsis thaliana* roots as revealed by QTL and mutant analysis. *Proc. Natl. Acad. Sci. USA* **103**: 2994–2999.
- Sergeeva, L.I., Vonk, J., Keurentjes, J.J.B., van der Plas, L.H.W., Koornneef, M., and Vreugdenhil, D.** (2004). Histochemical analysis reveals organ-specific quantitative trait loci for enzyme activities in *Arabidopsis*. *Plant Physiol.* **134**: 237–245.
- Shirley, B.W., Kubasek, W.L., Storz, G., Bruggemann, E., Koornneef, M., Ausubel, F.M., and Goodman, H.M.** (1995). Analysis of *Arabidopsis* mutants deficient in flavonoid biosynthesis. *Plant J.* **8**: 659–671.
- Sonderby, I.E., Hansen, B.G., Bjarnholt, N., Ticconi, C., Halkier, B.A., and Kliebenstein, D.J.** (2007). A systems biology approach identifies a R2R3 MYB gene subfamily with distinct and overlapping functions in regulation of aliphatic glucosinolates. *PLoS ONE* **2**: e1322.
- Spirin, V., and Mirny, L.A.** (2003). Protein complexes and functional modules in molecular networks. *Proc. Natl. Acad. Sci. USA* **100**: 12123–12128.
- Stanton, M., Young, H.J., Ellstrand, N.C., and Clegg, J.M.** (1991). Consequences of floral variation for male and female reproduction in experimental populations of wild radish, *Raphanus sativus* L. *Evolution Int. J. Org. Evolution* **45**: 268–280.
- Stupar, R.M., and Springer, N.M.** (2006). Cis-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F-1 hybrid. *Genetics* **173**: 2199–2210.
- Sweigart, A.L., Mason, A.R., and Willis, J.H.** (2007). Natural variation for a hybrid incompatibility between two species of *Mimulus*. *Evolution Int. J. Org. Evolution* **61**: 141–151.
- Tatum, E.L., Barratt, R.W., Fries, N., and Bonner, D.** (1950). Biochemical mutant strains of *Neurospora* produced by physical and chemical treatment. *Am. J. Bot.* **37**: 38–46.
- Vision, T.J., Brown, D.G., and Tanksley, S.D.** (2000). The origins of genomic duplications in *Arabidopsis*. *Science* **290**: 2114–2117.
- Watson, J.D., and Crick, F.H.** (1953). The structure of DNA. *Cold Spring Harb. Symp. Quant. Biol.* **18**: 123–131.
- Wentzell, A.M., Boeye, I., Zhang, Z.Y., and Kliebenstein, D.J.** (2008). Genetic networks controlling structural outcome of glucosinolate activation across development. *PLoS Genet.* **4** (online), doi:10.1371/journal.pgen.1000234.
- Wentzell, A.M., and Kliebenstein, D.J.** (2008). Genotype, age, tissue, and environment regulate the structural outcome of glucosinolate activation. *Plant Physiol.* **147**: 415–428.
- Wentzell, A.M., Rowe, H.C., Hansen, B.G., Ticconi, C., Halkier, B.A., and Kliebenstein, D.J.** (2007). Linking metabolic QTL with network and cis-eQTL controlling biosynthetic pathways. *PLoS Genetics* **3**: e162.
- West, M.A.L., Kim, K., Kliebenstein, D.J., van Leeuwen, H., Michelmore, R.W., Doerge, R.W., and St.Clair, D.A.** (2007). Global eQTL mapping reveals the complex genetic architecture of transcript level variation in *Arabidopsis*. *Genetics* **175**: 1441–1450.
- Whitlock, M.C., Phillips, P.C., Moore, F.B.-G., and Tonsor, S.J.** (1995). Multiple fitness peaks and epistasis. *Annu. Rev. Ecol. Syst.* **26**: 601–629.
- Wright, S.** (1931). Evolution in Mendelian populations. *Genetics* **16**: 97–159.
- Xie, D.Y., Sharma, S.B., Paiva, N.L., Ferreira, D., and Dixon, R.A.** (2003). Role of anthocyanidin reductase, encoded by BANYULS in plant flavonoid biosynthesis. *Science* **299**: 396–399.
- Yanofsky, C.** (1960). The tryptophan synthetase system. *Microbiol. Mol. Biol. Rev.* **24**: 221–245.
- Yi, T.M., Huang, Y., Simon, M.I., and Doyle, J.** (2000). Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc. Natl. Acad. Sci. USA* **97**: 4649–4653.
- Zhang, Z.-Y., Ober, J.A., and Kliebenstein, D.J.** (2006). The gene controlling the quantitative trait locus *EPITHIOSPECIFIER MODIFIER1* alters glucosinolate hydrolysis and insect resistance in *Arabidopsis*. *Plant Cell* **18**: 1524–1536.