Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Plant Biology

Modeling QTL for complex traits: detection and context for plant breeding

Mark Cooper[1], Fred A van Eeuwijk[2,3], Graeme L Hammer[4], Dean W Podlich[1] and Carlos Messina[1]

The genetic architecture of a trait is defined by the set of genes contributing to genetic variation within a reference population of genotypes together with information on their location in the genome and the effects of their alleles on traits, including intra-locus and inter-locus interactions, environmental dependencies, and pleiotropy. Accumulated evidence from trait mapping studies emphasizes that plant breeders work within a trait genetic complexity continuum. Some traits show a relatively simple genetic architecture while others, such as grain yield, have a complex architecture. An important advance is that we now have empirical genetic models of trait genetic architecture obtained from mapping studies (multi-QTL models including various genetic effects that may vary in relation to environmental factors) to ground theoretical investigations on the merits of alternative breeding strategies. Such theoretical studies indicate that as the genetic complexity of traits increases the opportunities for realizing benefits from molecular enhanced breeding strategies increase. To realize these potential benefits and enable the plant breeder to increase rate of genetic gain for complex traits it is anticipated that the empirical genetic models of trait genetic architecture used for predicting trait variation will need to incorporate the effects of genetic interactions and be interpreted within a genotype–environment–management framework for the target agricultural production system.

Addresses
[1] Pioneer Hi-Bred International, 7250 NW 62nd Ave, Johnston, IA 50131, United States
[2] Biometris, Wageningen UR, Wageningen, P.O. Box 100, 6700 AC, The Netherlands
[3] Centre for BioSystems Genomics, P.O. Box 98, 6700 AB, Wageningen, The Netherlands
[4] The University of Queensland, School of Land, Crop, and Food Sciences, Agricultural Production Systems Research Unit (APSRU), Brisbane, Qld. 4072, Australia

Corresponding author: Cooper, Mark (mark.cooper@pioneer.com), van Eeuwijk, Fred A (fred.vaneeuwijk@wur.nl), Hammer, Graeme L (g.hammer@uq.edu.au), Podlich, Dean W (dean.podlich@pioneer.com) and Messina, Carlos (charlie.messina@pioneer.com)

Current Opinion in Plant Biology 2009, **12**:231–240

This review comes from a themed issue on
Genome studies and molecular genetics
Edited by Masahiro Yano and Roberto Tuberosa

1369-5266/$ – see front matter

DOI 10.1016/j.pbi.2009.01.006

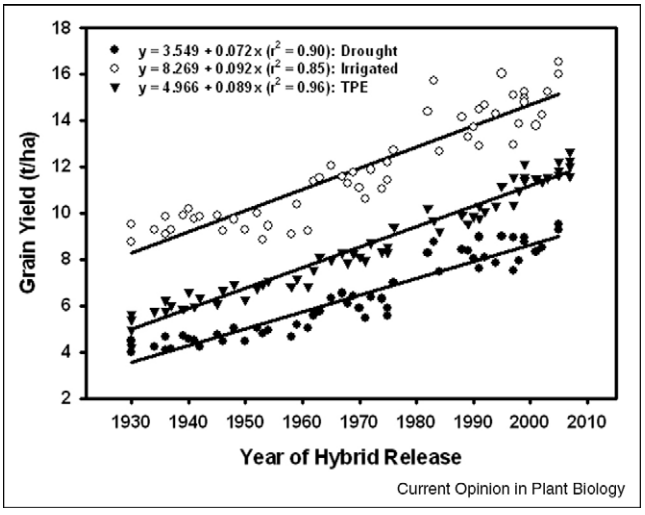
Introduction

There is long-standing interest in understanding the genetic architecture of yield and other complex traits of crop plants [1•,2,3]. We adopt a definition of trait genetic architecture used elsewhere [1•,2], which includes the number of genes, the number of alleles for the genes, and their effects on trait phenotypes. In studying the effects of alleles of genes on trait variation it is useful to distinguish between the simple additive case, where the effects of the alleles show no interactions and are consistent in combination with the genotypes of other genes and across environments and where there are no pleiotropic effects, and the more complex nonadditive cases where such interactions exist. Here, for the non-additive effects we will give specific emphasis to the importance of epistasis, genotype-by-environment interactions (GEIs), and pleiotropy. In developing a model of trait genetic architecture for application in plant breeding we are interested in determining whether a genetic model of trait variation based on the assumption of additive genetic effects is adequate and understanding the potential improvements that can be achieved for predicting trait variation by including the effects of genetic interactions within the model.

For most plant breeders grain yield is an important target trait for long-term genetic improvement (Figure 1). Yield is assumed to be influenced by multiple component traits, each with their own genetic architecture. For the majority of the 20th century yield of maize and other crops was improved using selection strategies motivated by quantitative genetic principles that were largely developed in the first half of the 20th century [4,5]. The foundations for the quantitative genetic models that underpin this classical quantitative genetics selection theory are based largely on cumulative effects of multiple independent Mendelian loci. There has always been debate about the relative importance of the various types of genetic interactions for genetic variation within populations and their contributions to selection response. Similar debates continue within the broader context of evolutionary genetics [6•,7].

Plant breeders have demonstrated that when nonadditive genetic effects contribute to the genetic variation that is important for trait genetic improvement, these effects and the associated genetic variation can be utilized to achieve genetic gain for the traits through appropriately

Figure 1



Long-term response to selection for grain yield for a sequence of successful maize hybrids released by Pioneer Hi-Bred International from the 1930s to 2007. Grain yield was measured in a sequence of experiments conducted from 1990 to 2007. The hybrids were grown side-by-side under different plant populations to represent the management conditions of the different decades. The sequence commences with open-pollinated varieties popular in the central US corn-belt in the 1920s and 1930s. Three environment-types are distinguished: irrigated environments, drought affected environments, and on-farm conditions representing the target population of environments (TPE) of the breeding program. For further discussion see [8].

designed breeding methods. For example, early in the 20th century heterosis was identified as an important source of nonadditive genetic variation for yield and agronomic traits of maize [4]. The genetic architecture of heterosis is still a topic of research today [8]. Utilization of this source of nonadditive genetic variation by maize breeders was enabled by the creation of heterotic groups of germplasm and design of hybrid breeding strategies [4,8]. Commercial breeding companies developed proprietary heterotic groups of germplasm to take advantage of this source of nonadditive genetic variation and generations of maize breeders have improved hybrid grain yield by conducting cycles of pedigree breeding within the heterotic groups that were created. Much of the genetic gain for hybrid yield that was realized from such long-term breeding strategies (e.g. Figure 1) was achieved without a detailed understanding of the genetic architecture of yield.

For a range of traits, including yield, recent empirical evidence obtained from mapping quantitative trait loci (QTL) indicates the influence of many QTL and the importance of both additive and nonadditive QTL effects. Theoretical considerations and empirical QTL mapping studies have provided some insight into the translation of trait genetic architecture into genetic variation and selection response [6•,7,9,10], but a lot remains to be learned, especially about the design of effective molecular breeding strategies for quantitative traits. In this context, we discuss the relationships between genetic architecture, genetic variation, and selection response from the perspective of two plant breeding questions. Firstly, how should the plant breeder detect and quantify the effects of QTL for traits to create predictive models that can be applied to enhance genetic gains, both in the short-term and long-term? Secondly, what components of trait genetic architecture primarily contribute to selection response in a plant-breeding program?

Trait dissection: QTL models and trait genetic architecture

Advances in mixed model statistical methodology for mapping QTL have enabled studies of the genetic architecture of traits. Within the mixed model framework the trait phenotype is modeled to be a sum of fixed and random terms. The joint modeling of fixed and random terms is the basis of the 'mixed' terminology. A pragmatic rule for the differentiation between fixed and random terms is that for random terms we can think it desirable to describe their behavior principally in terms of variances and correlations, whereas for fixed terms either we do not want to make such an assumption or we think such an assumption to be inadequate. Typically average plant performance in individual environments (environmental main effects) would be taken to represent fixed effects, whereas average genotypic performances across environments (genotypic main effects) would be taken random. Genotypic main effects will always include additive genetic effects, but they do not need to be restricted to just those types of effects. For example, for a collection of homozygous genotypes the genotypic main effects can include epistatic genetic effects as well. When genotypic main effects are random, genotypic differences in performance as conditioned by the environment (GEI) should also be random. Following these definitions, we can write the classical quantitative genetic model for the trait phenotype of a genotype i in an environment j as $\underline{P_{ij}} = \mu_j + \underline{G_i} + \underline{GE_{ij}}$, with μ_j representing a fixed intercept term for each environment, $\underline{G_i}$ the random genotypic main effect, and $\underline{GE_{ij}}$ the random GEI term. We will underline random terms. For convenience we will ignore microenvironmental noise. For the above model, the genetic variance, that is, genotype-related variance, in each environment will be $V_G + V_{GE}$, while the correlation between any pair of environments will be $V_G/(V_G + V_{GE})$. It is unlikely that for real life data the genetic variances across all environments will be homogeneous and neither will genetic correlations between environments always be constant. Therefore, a more general model is required. Thus, we write $\underline{P_{ij}} = \mu_j + \underline{GGE_{ij}}$, where we do not distinguish between genotypic main effects and GEI terms, but where we impose a structure on the variances and correlations governing the $\underline{GGE_{ij}}$ terms, so that the

genetic variances become environment specific and the genetic correlations unique for each pair of environments.

The construction of a QTL model entails the introduction of marker-based information in the above model so that the importance of the random $\underline{GGE}_{ij}$ term is reduced and the phenotype is predicted by marker-based predictors [11], called genetic predictors, in combination with estimated QTL allele effects. For example, for populations developed from inbred parent pairs, we can take the number of alleles coming from the first parent (Q versus q) as defining a genetic predictor for additive genetic QTL effects; this number can take the values 2 (QQ), 1 (Qq), and 0 (qq). Similarly genetic predictors can be defined for dominance and epistatic interaction effects within the model. More generally, instead of taking the number of alleles coming from a parent at marker positions as the basis for genetic predictors, we can work with the identity by descent probabilities of offspring individuals in relation to parents, predecessors in a pedigree, or founder individuals [12].

Restricting ourselves to genetic predictors for additive genetic QTL effects, that is, additive genetic predictors, let us denote the value of the qth additive genetic predictor for genotype i by x_{iq}. One way to understand QTL mapping is then to find a set of additive genetic predictors, Q, that adequately predicts the phenotype: $\underline{P}_{ij} = \mu_j + \Sigma_Q\{x_{iq}\alpha_q\} + \underline{GGE}_{ij}^*$, with $\Sigma_Q x_{iq}$ the set of selected additive genetic predictors, and α_q the allele substitution effect for the qth QTL for the case of main effects QTL assumed to have constant effects across environments. The random terms $\underline{GGE}_{ij}^*$ represent residuals whose variances and correlations across the environments should be smaller than those of the initial genetic terms $\underline{GGE}_{ij}$ before the introduction of QTL in the model for the phenotype. Various strategies have been proposed to identify an adequate QTL model within the mixed model framework [13••,14–17].

Assume that we are dealing with a single additive genetic QTL then $\underline{P}_{ij} = \mu_j + x_i\alpha + \underline{GGE}_{ij}^*$. However, when the QTL effect is not constant across environments we can define the QTL allele substitution effect to be environment dependent by fitting a separate QTL effect for each environment: $\underline{P}_{ij} = \mu_j + x_i\alpha_j + \underline{GGE}_{ij}^*$ (Figure 2) [13••]. The QTL-by-environment interaction (QEI) exemplified by the effects α_j constitutes the genetic basis of GEI. A further step in modeling the environmental dependence of the QEI effect is to regress these effects on environmental characterizations, such as stress indices, to create eco-physiological QTL models. When z_j stands for the value of a stress index in environment j, the QTL model becomes $\underline{P}_{ij} = \mu_j + x_i\alpha + x_iz_j + \underline{GGE}_{ij}^*$, with $x_i\alpha$ an intercept term expressing the QTL main effect and x_iz_j a regression of the QEI effects on the environmental characterization z_j [13••].

The modeling of GEI and QEI can be further extended, transcending the linear mixed model framework. When the phenotype across an environmental gradient can be described by a nonlinear function (e.g. a logistic function) the individual parameters of such functions can themselves be described in terms of QTL models [18,19]. Such functional QTL models are conceptually close to the idea of developing QTL models for the physiological parameters in crop growth simulation models, discussed in the next section. A further extension of the mixed model approach to QTL mapping is that of multi-trait modeling [20].

The mixed model QTL mapping framework described above for classical biparental populations is also applicable to other types of populations. We mention two common types: firstly, arbitrarily structured populations as for association mapping studies; secondly, multiple offspring populations connected through common parents (e.g. diallel designs, complex crosses, and pedigree populations). For association mapping, the principal problem is that spurious, that is nonlinkage based, marker-trait associations can occur for various population genetic reasons, such as drift, selection, and population substructure. An effective strategy to deal with the problem of such spurious associations is to impose structure on the variance–covariance matrix for the genetic effects in an appropriate mixed model. This structure should reflect genetic relatedness, and can be calculated either from known pedigree relationships [21], or, alternatively, from observed marker scorings [22]. In association mapping, the construction of genetic predictors is simple in the sense that it is restricted to marker positions. No form of interval mapping is possible in this context. For multiple connected populations [23], the calculation of genetic predictors requires heavier and more complicated calculations than for biparental populations, while the mixed model formulation needs to take into account the possible dependence of QTL effects on parents and populations, that is, QTL by genetic background interaction, a special form of epistatic interactions [24].

Thus, it is now possible to identify QTL for any trait that can be adequately phenotyped, including grain yield. The current view obtained from conducting such mapping studies is that the genetic architecture of yield and other traits is influenced by multiple QTL that can interact with each other and environmental conditions to influence single or multiple traits [2,13••,20,25•,26•]. Mapping traits in appropriate reference populations via the mixed model based genome scans described above provides a preliminary description of trait genetic architecture based on QTL rather than genes. The QTL and their alleles may themselves become targets for selection in a breeding program or may provide the starting point for further fine mapping and map-based cloning efforts designed to determine the genes and functional sequence

Figure 2

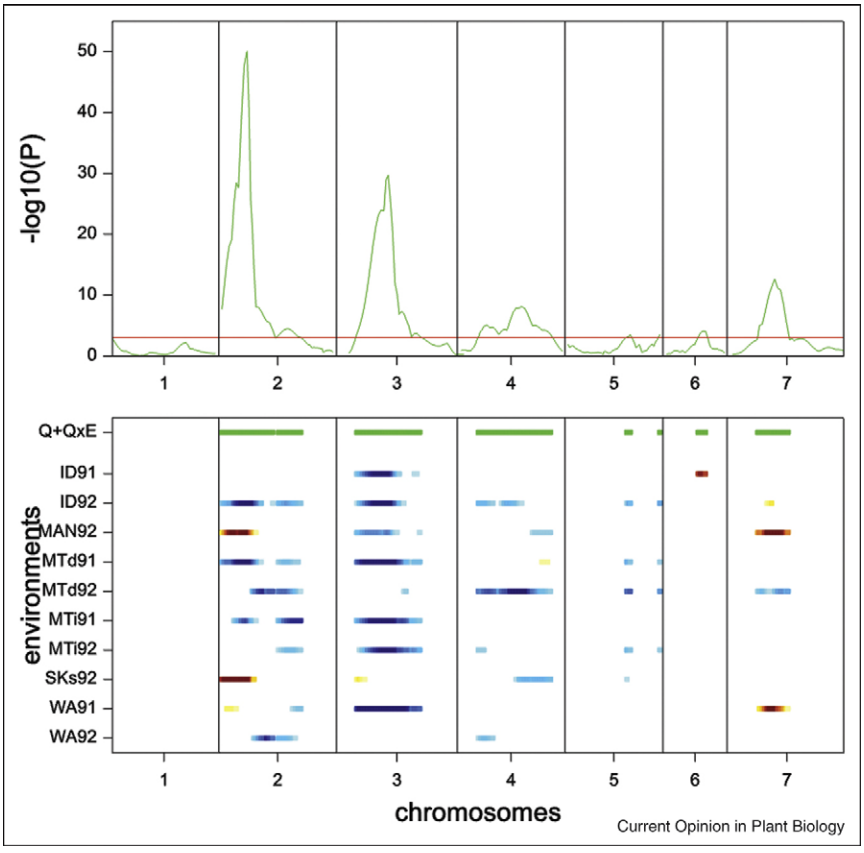


Illustration of a QTL mapping strategy based on searching for environment-dependent QTL allele effects, that is, QTL showing QTL-by-environment interaction (QEI). Trait is yield in barley (seven chromosomes), population is the biparental population Steptoe/Morex, evaluations for 10 trials in 1991 and 1992. Top panel. Genome-wide profile showing transformation of the *P*-value (−10 log(*P*-value)) resulting from the test for occurrence of a QTL effect in any of the 10 trials. The red horizontal line shows the threshold value for significance. QTL are found on all except the first chromosome. The strongest QTL appears on chromosome 2. Lower panel. Representation of QTL allele effects in individual environments. Red and yellow indicate strongly positive and moderately positive contributions from the allele coming from the Morex parent, while dark and light blue indicate the superiority of the allele coming from the Steptoe parent. The uppermost green line is a summary of the significance for the overall test on QTL presence as shown in the top panel. It can be seen that the QTL on chromosome 2 shows crossover interactions: in some environments the Morex parent delivers the superior allele, in other environments the Steptoe parent, and in yet other environments there is no indication for a QTL at all. The QTL on chromosome 3 shows also QEI, but in this case the positive effect always comes from the Steptoe parent, although the intensity changes between environments.

polymorphism ultimately underlying the QTL [1•,27]. Recent studies have begun to consider some implications of nonadditive genetic effects attributed to GEI, epistasis, and pleiotropy for mapping QTL and for their utilization in selection [9,28•,29••,30•,31].

While plant breeders can now select for favorable QTL alleles, it is important to recognize that the conventional selection that enabled much of the genetic gains realized to date (Figure 1) has operated on whole-plant trait performance and the success of new varieties and hybrids is judged on their whole-plant multi-trait performance across diverse environmental conditions and a range of farm management practices. Therefore, to complement the dissection of traits into sets of QTL representing components of trait genetic architecture a key question for the plant breeder is how to integrate the QTL information obtained from trait mapping studies to predict whole-plant trait phenotypes? In addition to the statistical prediction methods discussed above crop growth models can be used to integrate the effects of multiple QTL for multiple traits and predict the yield of genotypes given different environmental and management conditions [29••,30•,32•,33].

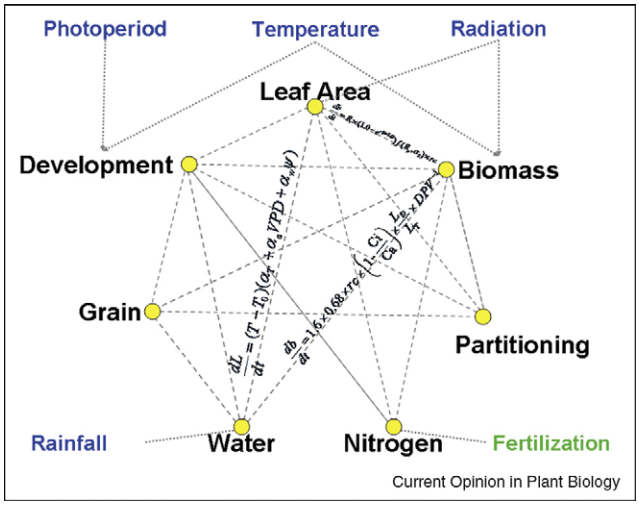
QTL integration: assessing impact of trait QTL on crop performance

Crop performance, and hence the trait phenotypes we measure and map, is an emergent consequence of the dynamic interplay of crop growth and development processes within an environmental context over the crop life cycle [33]. Crop growth models that are structured to capture these dynamic interactions of the physiological determinants of crop growth and development can be

employed to, first, analyze and aid understanding of the physiological and genetic basis of trait variation and thus inform phenotyping for QTL detection and second, predict consequences of genotype-by-environment-by-management interactions to assist crop improvement in general and molecular breeding in particular [29••,30•,33]. Many crop models estimate crop growth using environmental resource capture and conversion efficiency concepts for radiation and water, while allowing for influences of major nutrients such as nitrogen. The framework is explanatory and predictive as states of the crop system at any moment during the life cycle are predicted by integrating rates of the underpinning physiological processes (Figure 3). The vector of growth and development rate coefficients quantifies organ growth and development response to environmental conditions (e.g. temperature and water deficit) and to the state of the plant system itself. This abstract representation of the plant exposes points of connection between adaptive traits and genetic determinants, which become targets for mapping and QTL statistical modeling. Biological factors contributing to nonadditive genetic effects on crop states (e.g. leaf area) are implicit in the model structure and predicted as the consequence of the interplay among physiological processes and the environment [32•,33,34].

Studies aimed at detecting process QTL and integrating their effects for phenotypic prediction have employed a physiological modeling approach [35••,36,37]. By using an organ level model of environmental effectors (temperature, vapor pressure deficit, and plant water status) on leaf extension rate in maize, Reymond *et al.* [38] were able to identify QTL that overcame environment context dependencies associated with mapping QTL to leaf size. This enabled robust prediction of leaf size for novel combinations of QTL in a range of environments [35••] and identification of QTL with pleiotropic effects on silk extension and anthesis-silking interval [39]. Scaling QTL effects from leaf to canopy and from silk to kernel via nesting organ level effects in whole-plant models [40••] made it possible to map genotype to phenotype and investigate additive and nonadditive effects of trait QTL on drought adaptation at the crop level. This approach opens possibilities to study multi-trait adaptation landscapes and assess the impact of any detected QTL within the context of novel genotypes created by selection and the environments encountered in the target domain [29••,30•,32•,33].

Figure 3



Schematic representation of crop growth and development as captured in crop models. Nodes represent the system state at any time and edges represent interacting physiological processes. Some of the key environmental and management input factors are also shown.

Understanding traits and their targeted measurement in appropriate environments is a critical step in detecting relevant QTL and evaluating them for influence on crop yield. Within this context genotype-by-environment-by-management interactions must be understood to enable farmers to successfully grow the products of breeding programs. Studies in wheat [41] and maize [42] have highlighted the importance of root system architecture for water capture and drought adaptation. A modeling study in maize [42] suggested that enhanced water capture at depth associated with more compact, deeper root systems was more likely to explain a significant component of the historical yield advance in the environments of the US corn-belt (Figure 1) and its interaction with plant density than were changes in radiation capture associated with long-term trends in leaf erectness. In wheat, Manschadi *et al.* [41] identified associations of root system architecture with enhanced water extraction at depth and simulated observed phenotypic consequences on drought adaptation. They were able to associate variation in mature plant root system architecture with seedling seminal root angle [43], thus providing an avenue for germplasm screening and QTL detection. However, the ability for increased water capture (especially late in the crop cycle) to generate yield gain is dependent on the ability of the crop to set sufficient grain to allow continued filling. This trait order dependency for achieving high yield in drought prone environments reflects the crop physiology of an epistatic interaction on grain yield via contributions from component traits, each with their own genetic architecture.

Systems modeling and simulation to support plant breeding involves the ongoing assimilation of knowledge into a theoretical framework that can be used for quantitative predictions of crop yield. Modeling the physiology and genetics of complex traits requires a sufficiently detailed framework to ensure that important physiological linkages, trait interactions, and internal plant regulation are simulated implicitly, and bring about emergent patterns of crop growth and development that reflect genetic control and biological robustness [30•]. The next generation of crop

growth models is likely to horizontally integrate knowledge across species, and vertically integrate knowledge across levels of organization from gene-to-phenotype. Such models provide a framework to assimilate 'omics' information generated at different levels of gene-to-phenotype organization [44], continue to inform experimental investigation and development of statistical methodologies, and enable prediction of short-term and long-term consequences of plant breeding as trajectories on multi-trait adaptation landscapes for investigation, design, and optimization of breeding strategies [29••,30•,33].

Toward a framework for studying the quantitative genetics of gene networks

The study of gene networks and genotype-to-phenotype pathways via systems biology methodology provides another body of research that is complementary to the results obtained from QTL mapping studies [7,44]. Soon after the genome sequences of organisms became available the importance of genetic interactions in determining trait phenotypes was emphasized [45–47]. An emphasis in the study of genetic interactions in systems biology has been the description of the topology of gene networks, measuring statistical properties of the networks, and modeling genotype-to-phenotype network properties [48–50]. Motivated by the definition of networks of genes determining trait phenotypes, recent investigations have considered the implications of such gene networks for the expression of additive and nonadditive genetic variation for traits within a reference population of genotypes [51–53] and for selection response [9,28•,54•,55].

Although there is interest in quantifying the potential roles of gene networks in our current understanding of the genetic architecture of traits and their influence on the genetic variation within populations and selection response, we do not yet have a comprehensive theory for the quantitative genetics of gene networks. Some potential directions for developing such a theoretical framework have been considered. Kauffman [56••] proposed the *NK* model as a framework to study epistasis conditioned by gene networks; here N defines the number of genes and K defines the level of epistasis as a set of interactions between the N genes in the form of edges in a network graph. Different approaches have been applied to translate such network graph topologies into genotype-to-phenotype mappings for traits, including Boolean truth tables to simulate gene switching logic and sets of chemical equations. The crop growth models discussed above provide a relevant framework for connecting the networks of genes and QTL influencing traits to crop performance phenotypes. In addition to epistasis, both GEI and pleiotropy can be studied with the $E(NK)$ model [28•,54•,57•]; here E represents different environment-types that contribute to GEI. Many of the network properties investigated by Kauffman and his coworkers are not familiar to quantitative geneticists. However, there are working relationships that can be identified between the statistical properties of networks and some of the classical concepts of quantitative genetics [28•,54•,57•]. For example, the concept of an adaptation landscape, as defined by Kauffman [56••], can be considered as a multi-genic extension of the classical epistatic models of quantitative genetics. Further, the concept of additive genetic variance and covariance of relatives can be related to different autocorrelation measures used to describe the ruggedness of the trait adaptation landscapes associated with a gene network model [5,28•,56••]. Connections between the properties of these different frameworks for modeling genetic variation have been given consideration in the study of evolution [7,56••,58•]. Here we emphasize that it is also possible to apply the trait genetic models identified by QTL mapping, as discussed above, within the $E(NK)$ model framework (Box 1) to study plant breeding

Box 1 Modeling genetic complexity and selection response.

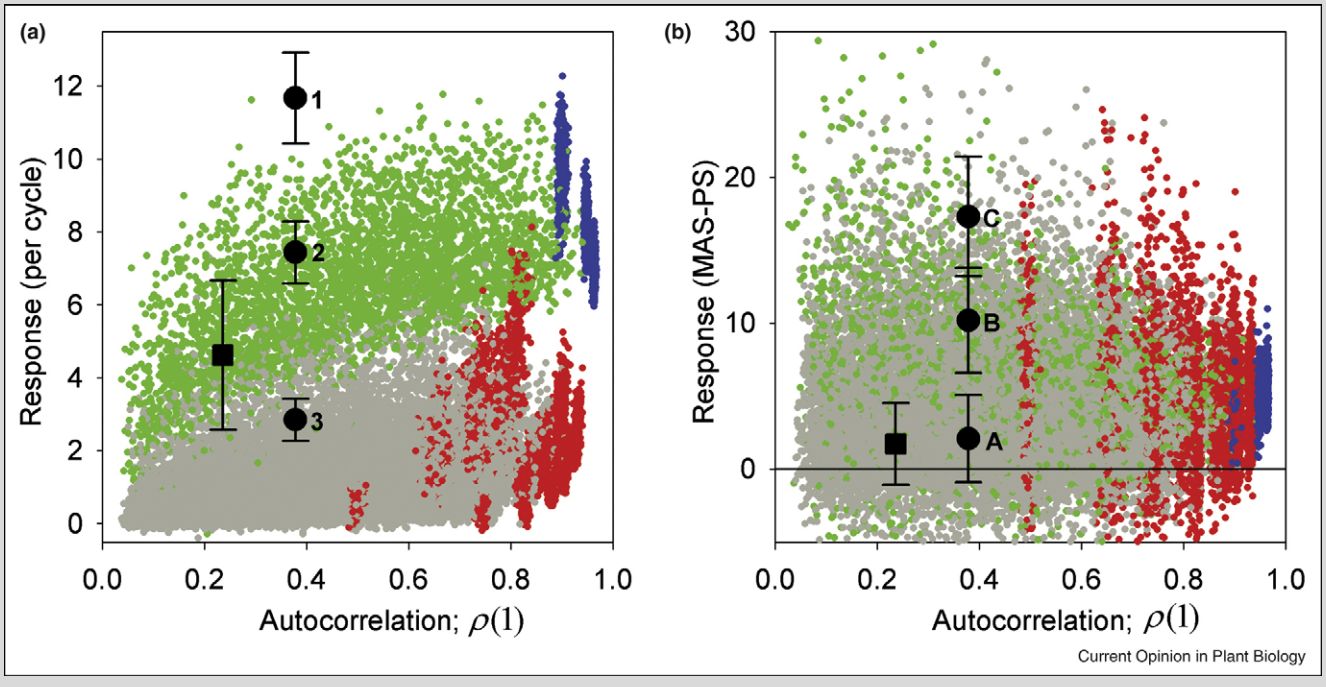
Models of trait genetic architecture obtained from QTL mapping can be used to parameterize the finite locus component of the $E(NK)$ model and within this framework simulation methods can be applied to examine the impact of trait genetic complexity on response to selection [9,28•,57•]. Summary results from some published examples [28•,29••,30•,33,42] are presented for purposes of illustration. A *complexity–response* graphical view comparing trait genetic complexity statistics, such as the autocorrelation, against response to selection statistics provides a summary of the results of simulation studies of breeding strategies parameterized using the available QTL models (Fig. Box 1). An ensemble of plausible genetic models was generated using the $E(NK)$ framework to represent different trait genetic architecture possibilities; genetic models were simulated by varying N to simulate different numbers of genes, varying E to simulate different levels of GEI, and varying K to simulate different levels of epistasis. Two empirical yield QTL models based on multiple trait contributions evaluated within a crop growth and development framework, were implemented within the $E(NK)$ framework and included in the complexity–response graphic (Fig. Box 1). The first is a sorghum-drought tolerance model [29••,30•] and the second a maize yield model for the US corn-belt [33,42]. For the purposes of demonstration both a phenotypic selection strategy and a marker-assisted selection strategy were conducted for five cycles for all genetic models and the responses to selection for both strategies were compared. For the maize example the scenario was extended to consider three different reference germplasm populations by varying the frequency of the favorable alleles for all QTL in the cycle 0 population: Case 1 favorable allele frequency = 0.1, Case 2 favorable allele frequency = 0.25, Case 3 favorable allele frequency = 0.5. For the ensemble of $E(NK)$ models for each of the N genes, and for the sorghum drought tolerance model for each of the QTL alleles, all alleles were defined to have a starting frequency of 0.5. For all reference populations the complexity of the trait genetic architecture was estimated as an autocorrelation of trait phenotypes for sequences of genotypes differing by 1 allele steps $\rho(1)$; the sequences were sampled from random walks in the genotype-to-phenotype state space defined by the genetic model. Response to selection was estimated as the difference in the population trait mean between cycle 5 and cycle 0. To construct the complexity–response plot the $\rho(1)$ autocorrelation was plotted against the response to selection for each genetic model (Fig. Box 1a). To compare the response obtained from phenotypic and marker-assisted selection the response difference between the two strategies was plotted against the $\rho(1)$ autocorrelation (Fig. Box 1b).

The background ensemble of *E*(*NK*) models generated a range of genetic model scenarios extending from relatively simple genetics, indicated by a high autocorrelation, to more complex genetics, indicated by a low autocorrelation. For the sorghum model $\rho(1) = 0.24$ and for the maize model $\rho(1) = 0.38$ (Fig. Box 1). As has been demonstrated previously for other examples [28•], the empirical QTL models for the two trait examples considered here reside within the ensemble of theoretical trait genetic model possibilities, as depicted on the complexity–response graph constructed applying the *E*(*NK*) model. Thus, the empirical genetic models can be considered as specific cases of the range of theoretical possibilities generated by varying the number of genes, levels of epistasis, GEI, and pleiotropy within the *E*(*NK*) framework. Given that the plant breeder works within such a trait genetic complexity continuum, the theoretical ensemble of genetic model possibilities, so constructed to include epistasis, GEI, and pleiotropy, can be used to obtain a preliminary view in the process of examining the relative merits of proposed molecular enhanced breeding strategies for a wide range of genetic models that represent a range in trait genetic architecture [9,28•]. The merits of the promising strategies can be examined in greater detail as the empirical trait genetic models are developed and refined over cycles of experiments [29••,30•,33,42].

For the sorghum model a positive response to phenotypic selection was predicted with a slight advantage associated with marker-assisted selection after five cycles. For the maize model a positive response to selection was predicted for all three reference populations, with a greater rate of response associated with lower frequencies of the favorable alleles in the cycle 0 reference populations. For Case 1 of the maize example (Fig. Box 1a; favorable allele frequency = 0.1) three QTL detection scenarios were considered (Fig. Box 1b): Case A = 25% of QTL detected, Case B = 50% of QTL detected, Case C = 75% of QTL detected. There was a slight advantage of the marker-assisted selection strategy over the phenotypic selection strategy when 25% of the QTL were detected (Fig. Box 1b, Case A). The advantage increased with an increase in the number of QTL detected (Fig. Box 1b, Case C > Case B > Case A).

Key observations obtained from the simulation experiment examples include, first, Phenotypic selection was effective over a wide range of simple and complex genetic models. In general the response was greater for the simpler genetic models, indicated by higher $\rho(1)$ autocorrelation. Second, marker-assisted selection demonstrated the potential to increase rates of genetic gain over phenotypic selection for a wide range of genetic models. Importantly, there was potential to realize a greater advantage from marker-assisted selection for the more complex genetic models including components of genetic variation attributed to epistasis, QEI, and pleiotropy, that is genetic models indicated by lower $\rho(1)$ autocorrelation. Third, simulation methodology, as demonstrated in the examples, can be used to model the impact of empirical genetic models obtained from QTL mapping studies within the context of short-term and long-term response to selection for breeding strategies.

Fig. Box 1



Complexity–response plot for an ensemble of *E*(*NK*) genetic models: (a) response to phenotypic selection (change in population mean per cycle after five cycles) against genetic model complexity measured by the $\rho(1)$ autocorrelation of the yield phenotype landscape. (b) The response difference (MAS − PS) between marker-assisted selection (MAS) and phenotypic selection (PS) at cycle five against genetic model complexity measured by the $\rho(1)$ autocorrelation of the yield phenotype landscape. Each point in the ensemble represents a different genetic model; blue = additive ($E = 1$, $K = 0$); red = epistatic effects only ($E = 1$, $K > 0$); green = gene-by-environment effects only ($E > 1$, $K = 0$); and gray = epistatic and gene-by-environment effects ($E > 1$, $K > 0$). Solid squares indicate the average projection (and standard deviation) onto the *complexity–response* plots of the $N = 15$ QTL sorghum drought tolerance model [29••,30•]. Solid circles indicate the average projection (and standard deviation) onto the *complexity–response* plots of the $N = 15$ QTL maize model for three reference populations [33,42]. In subfigure (a) three reference populations with different favorable allele frequencies were considered for the maize example: Case 1 = 0.1, Case 2 = 0.25, Case 3 = 0.5. In subfigure (b) for Case 1, where the favorable allele frequency was 0.1 in the reference population, three QTL detection scenarios were considered: Case A = 25% QTL detected, Case B = 50% QTL detected, Case C = 75% of QTL detected.

strategies for the genetic improvement of complex traits. Thus, the implications of the empirical trait genetic models identified by mapping trait QTL for both genetic variation and selection response can be studied within both the classical quantitative genetics framework [5] and the gene network framework proposed by Kauffman (Box 1, [9,28•,54•]). This approach has been utilized to examine the effectiveness of sorghum breeding strategies for drought prone environments [29••,30•]. The QTL for the traits contributing to drought tolerance in the sorghum example were defined within the context of a crop growth and development model and the implications for breeding for enhanced drought tolerance were evaluated by applying the $E(NK)$ framework and including the trait QTL as a subset of the N genes (Box 1, [29••,30•]). Possible trait combinations contributing to the observed long-term response to selection for yield of maize are also under investigation applying the same framework (Figure 1) [33,42].

Conclusion

What does the collective body of QTL mapping results mean for the field of quantitative genetics and plant breeding? For selection in plant breeding it is crucial to construct adequate prediction models for trait phenotypes. We have argued that the last decade of trait mapping experiments in different organisms indicates the presence of additive [10] and nonadditive genetic effects, including epistasis, QEI, and pleiotropy [13••,20,26•], that can influence trait genetic variation and selection response [28•]. This motivates consideration of genetic interactions and environmental dependencies when investigating the genetic architecture of traits [1•,6•,7] and the design of molecular enhanced plant breeding strategies [9,28•]. Our emphasis here is on using trait mapping methodologies to create genetic knowledge and apply that knowledge to enhance selection response over that which can be achieved by continuing the breeding strategies that have proven to be successful in the past (Figure 1). Additive QTL models provide the logical starting point. Different sources of genetic interactions and environmental dependencies can be iteratively included into the QTL models as the empirical results reveal their importance for prediction of trait variation in the reference populations of the breeding program.

There are two complementary avenues of research underway. Firstly, extensions of the classical quantitative genetics framework to include terms that capture properties of the genetic variation associated with epistasis, QEI, and pleiotropy. QTL mapping that enables testing for the presence of the interactions and trait physiology and modeling that explains some of them provide a logical experimental approach to support such investigations. Secondly, investigations to consider the feasibility of developing quantitative genetic models that are based on the concept of gene networks. Again mapping studies that allow definition of interactions at the QTL level provide a logical experimental approach to support such theoretical development. We encourage pursuit of both approaches and anticipate a merging of these efforts as we learn more about the genetic architecture of traits. Ultimately as geneticists and plant breeders we want a toolbox of genetic models that allows us to choose the appropriate modeling tools for the trait investigations at hand. The merits of the genetic models so constructed to incorporate epistasis, QEI, and pleiotropy can be systematically evaluated on a case-by-case basis for their ability to improve prediction of the trait phenotypes of new genetic combinations that can be created by plant breeding. Whenever prediction power can be improved plant breeders will have access to new molecular breeding tools that enhance the short-term and long-term outcomes achieved from the ongoing breeding programs by enabling practical utilization of both additive and nonadditive genetic effects of QTL as they are revealed for the genetic improvement of traits [9,28•].

Acknowledgements

Fred van Eeuwijk worked on this paper as part of the BB12 project on genotype-to-phenotype modeling of the Dutch CBSG2012 consortium (www.cbsg.nl). Figure 2 was kindly provided by Marcos Malosetti.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Mackay TFC: **The genetic architecture of quantitative traits**. • *Ann Rev Genet* 2001, **35**:303-339.
Comprehensive review of the issues involved in mapping trait QTL, results from model organisms, with emphasis on *Drosophila*, and some of the challenges in moving from the QTL level to the gene level and determining the sequence level polymorphisms that underlie QTL.

2. Holland JB: **Genetic architecture of complex traits in plants**. *Curr Opin Plant Biol* 2007, **10**:156-161.

3. Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD: **Resolution of quantitative traits into Mendelian factors using a complete linkage map of restriction fragment length polymorphisms**. *Nature* 1988, **335**:721-726.

4. Hallauer AR, Miranda Fo JB: *Quantitative Genetics in Maize Breeding*. edn 2. Iowa State University Press; 1988.

5. Lynch M, Walsh B: *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Inc; 1998.

6. Cheverud JM, Routman EJ: **Epistasis and its contribution to** • **genetic variance components**. *Genetics* 1995, **139**:1455-1461.
Important, thought provoking foundational paper that elegantly discusses the interplay between physiological and statistical epistasis and how physiological epistasis can contribute to the different components of genetic variation that can be detected applying classical quantitative genetics modeling concepts. An important result that is clearly demonstrated in this paper, but not often fully appreciated in many recent studies, is the context-dependent nature of any results obtained from the population-specific and environment-specific experiments conducted to study the genetic architecture of traits.

7. Phillips PC: **Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems**. *Nat Rev Genetics* 2008, **9**:855-867.

8. Duvick DN, Smith JSC, Cooper M: **Long-term selection in a commercial hybrid maize breeding program**. *Plant Breed Rev* 2004, **24(2)**:109-151.

9. Podlich DW, Winkler CR, Cooper M: **Mapping as you go: an effective approach for marker-assisted selection of complex traits**. *Crop Sci* 2004, **44**:1560-1571.

10. Hill WG, Goddard ME, Visscher PM: **Data and theory point to mainly additive genetic variance for complex traits**. *PLoS Genetics* 2008, **4**:1-10.

11. Jiang CJ, Zeng ZB: **Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines**. *Genetica* 1997, **101**:47-58.

12. Lange K: *Mathematical and Statistical Methods for Genetic Analysis*. Springer; 2002.

13. •• Boer MP, Wright D, Feng L, Podlich D, Luo L, Cooper M, van Eeuwijk FA: **A mixed model QTL analysis for multiple environment trial data using environmental covariables for QTLxE, with an example in maize**. *Genetics* 2007, **177**:1801-1803.
A comprehensive mixed-model analysis of a large bi-parental maize mapping study that incorporated terms in the mixed model to detect the presence of significant QEIs for grain yield and grain moisture. The results demonstrated that QEIs were present and could in part be explained by environmental covariables associated with experiments at different locations and in different years. Importantly, the form of the QTL-by-environment interactions that were detected demonstrated the presence of both changes in magnitude of QTL effects on trait phenotypes in different environments and crossover QTL effects, where different parental alleles of the QTL influenced the direction of the trait phenotypes in different environments.

14. Piepho HP: **A mixed model approach to mapping quantitative trait loci in barley on the basis of multiple environment data**. *Genetics* 2000, **156**:2043-2050.

15. Verbyla A, Eckermann PJ, Thompson R, Cullis B: **The analysis of quantitative trait loci in multi-environment trials using a multiplicative mixed model**. *Aust J Agr Res* 2003, **54**:1395-1408.

16. Piepho HP, Pillen K: **Mixed modelling for QTL × environment interaction analysis**. *Euphytica* 2004, **137**:147-153.

17. Gilmour AR: **Mixed model regression mapping for QTL detection in experimental crosses**. *Comput Stat Data Anal* 2007, **51**:3749-3764.

18. Ma CX, Casella G, Wu R: **Functional mapping of quantitative trait loci underlying the character process: a theoretical framework**. *Genetics* 2002, **161**:1751-1762.

19. Malosetti M, Visser RGF, Celis-Gamboa C, van Eeuwijk FA: **QTL methodology for response curves on the basis of non-linear mixed models, with an illustration to senescence in potato**. *Theor Appl Genet* 2006, **113**:288-300.

20. Malosetti M, Ribaut JM, Vargas M, Crossa J, van Eeuwijk FA: **A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.)**. *Euphytica* 2008, **161**:241-257.

21. Malosetti M, van der Linden CG, Vosman B, van Eeuwijk FA: **A mixed model approach to association mapping using pedigree information with an illustration to resistance for Phytophthora infestans in potato**. *Genetics* 2007, **175**:879-889.

22. Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB *et al.*: **A unified mixed-model method for association mapping that accounts for multiple levels of relatedness**. *Nat Genetics* 2005, **38**:203-208.

23. Cavanagh C, Morell M, Mackay I, Powell W: **From mutations to MAGIC; resources for gene discovery, validation and delivery in crop plants**. *Curr Opin Plant Biol* 2008, **11**:215-221.

24. Paulo MJ, Boer M, Huang X, Koornneef M, van Eeuwijk FA: **A mixed model QTL analysis for a complex cross population consisting of a half diallel of two-way hybrids in *Arabidopsis thaliana*: analysis of simulated data**. *Euphytica* 2008, **161**:107-114.

25. • Laurie CC, Chasalow SD, LeDeaux JR, McCarroll R, Bush D, Hauge B, Lai C, Clark D, Rocheford TR, Dudley JW: **The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel**. *Genetics* 2004, **168**:2141-2155.
Interesting preliminary analysis of the genetic architecture of a trait within the context of a maize population that has been subjected to long-term selection for divergent phenotypes. They identified multiple QTL for kernel oil concentration, with small additive QTL effects reported. Interesting comparison with [26•].

26. • Dudley JW: **Epistatic interactions in crosses of Illinois high oil × Illinois low oil and of Illinois high protein × Illinois low protein corn strains**. *Crop Sci* 2008, **48**:59-68.
Interesting preliminary analysis of the presence of QTL-by-QTL epistatic interactions as a component of the genetic architecture of traits that have been subjected to long-term selection in a maize population. The results reported in this paper complement and provide an interesting comparison with the results reported in [25•].

27. Salvi S, Tuberosa R: **To clone or not to clone plant QTLs: present and future challenges**. *Trends Plant Sci* 2005, **10**:297-304.

28. • Cooper M, Podlich DW, Smith OS: **Gene-to-phenotype models and complex trait genetics**. *Aust J Agric Res* 2005, **56**:895-918.
A preliminary description of theoretical models for studying the implications of gene-networks within a quantitative genetics framework. Potential applications of the proposed genetic models to the design of molecular enhance plant breeding strategies are discussed and examples presented.

29. •• Chapman S, Cooper M, Podlich D, Hammer G: **Evaluating plant breeding strategies by simulating gene action and dryland environment effects**. *Agron J* 2003, **95**:99-113.
The first in a series of publications reporting results from a comprehensive ongoing study of breeding strategies to improve sorghum drought tolerance for dryland environments in Australia. The paper describes an iterative genotype-to-phenotype model building process for linking the trait quantitative genetic models that can be identified by QTL mapping to the crop growth and development processes that are understood to be influenced by the putative drought tolerance traits. The combination of the quantitative genetic model and the crop growth model is used as a genotype-to-phenotype multi-trait model. Through simulation methodology the genotype-to-phenotype model of drought tolerance is used to provide integrated predictions of the contributions and value of traits and their QTL to the genetic improvement of sorghum grain yield in a target population of environments for a breeding program.

30. • Hammer GL, Chapman S, van Oosterom E, Podlich DW: **Trait physiology and crop modelling as a framework to link phenotypic complexity to underlying genetic systems**. *Aust J Agric Res* 2005, **56**:947-960.
Applying iterative model building strategies to advance a genotype-to-phenotype model of sorghum drought tolerance as new experimental data become available this paper provides an update of the earlier results reported in [29••].

31. Walsh B: **The struggle to exploit non-additive variation**. *Aust J Agric Res* 2005, **56**:873-881.

32. • Hammer G, Cooper M, Tardieu F, Welch S, Walsh B, van Eeuwijk F, Chapman S, Podlich D: **Models for navigating biological complexity in breeding improved crop plants**. *Trends Plant Sci* 2006, **11**:1360-1385.
Synopsis paper that summarizes the key concepts that emerged from a Symposium on the genetic improvement of complex traits that was held at the 4th International Crop Science Congress. This paper provides a useful entry point into the detailed papers that were contributed to the Symposium.

33. Messina C, Hammer G, Dong Z, Podlich D, Cooper M: **Modelling crop improvement in a G*E*M framework via gene-trait-phenotype relationships**. In *Crop Physiology: Applications for Genetic Improvement and Agronomy*. Edited by Sadras VO, Calderini DF. Academic Press, Elsevier; 2009, in press.

34. Tardieu F: **Virtual plants: modelling as a tool for the genomics of tolerance to water deficit**. *Trends Plant Sci* 2003, **8**:9-14.

35. •• Reymond M, Muller B, Leonardi A, Charcosset A, Tardieu F: **Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit**. *Plant Physiol* 2003, **131**:664-675.

Interesting presentation of results obtained from a bi-parental population QTL mapping study (based on recombinant inbred lines) that focused on mapping the genetic variation for coefficients of a model of plant response to environmental variables; the focus was on modeling maize leaf elongation rate in response to temperature, atmospheric vapor pressure deficit, and soil–plant water status. An outstanding feature of this study was the application of the results obtained from the QTL mapping study to an independent set of recombinant inbred lines from the same cross to use the QTL model to predict their leaf elongation rate responses to sets of environmental conditions that could be generated from the environmental variables included in the model. Thus, components of the leaf elongation responses of new genotypes to different environmental conditions were successfully predicted.

36. Messina CD, Jones JW, Boote KJ, Vallejos CE: **A gene-based model to simulate soybean development and yield responses to environment**. *Crop Sci* 2006, **46**:456-466.

37. Yin X, Struik PC, van Eeuwijk FA, Stam P, Tang J: **QTL analysis and QTL-based prediction of flowering phenology in recombinant inbred lines of barley**. *J Exp Bot* 2005, **56**:967-976.

38. Reymond M, Muller B, Tardieu F: **Dealing with the genotype × environment interaction via a modelling approach: a comparison of QTLs of maize leaf length or width with QTLs of model parameters**. *J Exp Bot* 2004, **55**:2461-2472.

39. Welcker C, Boussuge B, Bencivenni C, Ribaut JM, Tardieu F: **Are source and sink strengths genetically linked in maize plants subjected to water deficit? A QTL study of the responses of leaf growth and of Anthesis-Silking Interval to water deficit**. *J Exp Bot* 2007, **58**:339-349.

40. •• Chenu K, Chapman SC, Hammer GL, Mclean G, Ben Haj Salah, Tardieu F: **Short-term responses of leaf growth rate to water deficit scale up to whole-plant and crop levels: an integrated modelling approach in maize.**. *Plant, Cell Environ.* 2008, **31**:378-391.
Important experimental modeling paper that demonstrates integration of physiological modeling across scales of biological organization as a foundation to study and predict the effects of genetic variation at the organ level for complex traits at the whole-plant level. This paper demonstrates modeling concepts covered in earlier publications [33,34,35••].

41. Manschadi AM, Christopher J, deVoil P, Hammer GL: **The role of root architectural traits in adaptation of wheat to water-limited environments**. *Funct Plant Biol* 2006, **33**:823-837.

42. Hammer GL, Dong Z, McLean G, Doherty A, Messina C, Schussler J, Zinselmeier C, Paszkiewicz S, Cooper M: **Can changes in canopy and/or root system architecture explain historical maize yield trends in the US corn belt?** *Crop Sci.* 2009, **49**:299-312.

43. Manschadi AM, Hammer GL, Christopher JT, deVoil P: **Genotypic variation in seedling root architectural traits and implications for drought adaptation in wheat (*Triticum aestivum* L.)**. *Plant Soil* 2008, **303**:115-129.

44. Yuan JS, Galbraith DW, Dai SY, Griffin P, Stewart N Jr: **Plant systems biology comes of age**. *Trends Plant Sci* 2008, **13**:165-171.

45. Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh C, Minokawa T, Amore G, Hinman V, Arenas-Mena C *et al.*: **A genomic regulatory network for development**. *Science* 2002, **295**:1669-1678.

46. Komeda Y: **Genetic regulation of time to flower in *Arabidopsis thaliana***. *Annu Rev Plant Biol* 2004, **55**:521-535.

47. Lewis J: **From signals to patterns: space, time, and mathematics in developmental biology**. *Science* 2008, **322**:399-403.

48. Wagner A: **Estimating coarse gene network structure from large-scale gene perturbation data**. *Genome Res* 2002, **12**:309-315.

49. Barabási AL, Oltvai Z: **Network biology: understanding the cell's functional organization**. *Nat Rev Genetics* 2004, **5**:101-113.

50. Bornholdt S: **Less is more in modelling large genetic networks**. *Science* 2005, **310**:450-451.

51. Omholt SW, Plahte E, Øyehaug L, Xiang K: **Gene regulatory networks generating the phenomena of additivity, dominance and epistasis**. *Genetics* 2000, **155**:969-980.

52. Welch SM, Dong Z, Roe JL, Das S: **Flowering time control: gene network modelling and the link to quantitative genetics**. *Aust J Agric Res* 2005, **56**:919-936.

53. Gjuvsland AB, Hayes BJ, Omholt SW, Carlborg Ö: **Statistical epistasis is a generic feature of gene regulatory networks**. *Genetics* 2007, **175**:411-420.

54. • Cooper M, Podlich DW: **The *E*(*NK*) model: extending the *NK* model to incorporate gene-by-environment interactions and epistasis for diploid genomes**. *Complexity* 2002, **7**:31-47.
Technical paper that extends the *NK* model to the *E*(*NK*) model to enable joint modeling of genotype-by-environment interactions, epistasis, and pleiotropy.

55. Peccoud J, Vander Velden K, Podlich DW, Winkler C, Arthur L, Cooper M: **The selective values of alleles in a molecular network model are context-dependent**. *Genetics* 2004, **166**:1715-1725.

56. •• Kauffman SA: *The Origins of Order: Self-Organization and Selection in Evolution*Oxford University Press; 1993.
Tour de Force summary of many theoretical and experimental results that underpin the definition and utilization of the *NK* model by Stuart Kaufmann and his coworkers. This book represents an instructive starting point for those interested in applying the *NK* model to embark on research into the quantitative genetics of gene networks. A major emphasis of the book is on the evolution of genetic systems where epistasis is modeled as a property of gene networks.

57. • Podlich DW, Cooper M: **QU-GENE: a platform for quantitative analysis of genetic models**. *Bioinformatics* 1998, **14**:632-653.
The paper introduces a software tool for applying the *E*(*NK*) model to enable quantitative genetic analyses of different models of the genetic architecture of traits. This software was utilized in the examples reported in other publications reviewed here [28•,29••,30•,33].

58. • Gavrilets S: *Fitness Landscapes and the Origin of Species*. Princeton University Press; 2004.
Valuable reference that explains the application of fitness landscapes to the study of different questions in evolution. Recommended reading for anyone wanting to appreciate the strengths and weaknesses of applying fitness landscape concepts to the study of genetic systems.