A dynamic framework for quantifying the genetic architecture of phenotypic plasticity

Zhong Wang*, Xiaoming Pang*, Yafei Lv, Fang Xu, Tao Zhou, Xin Li, Sisi Feng, Jiahan Li, Zhikang Li and Rongling Wu

Submitted: 16th December 2011; Received (in revised form): 10th February 2012

Abstract

Despite its central role in the adaptation and microevolution of traits, the genetic architecture of phenotypic plasticity, i.e. multiple phenotypes produced by a single genotype in changing environments, remains elusive. We know little about the genes that underlie the plastic response of traits to the environment, their number, chromosomal locations and genetic interactions as well as environment impact on their effects. Here we review key statistical approaches for analyzing the genetic variation of phenotypic plasticity due to genotype—environment interactions and describe the implementation of a dynamic model to map specific quantitative trait loci (QTLs) that affect the gradient expression of a quantitative trait across a range of environments. This dynamic model is distinct by incorporating mathematical aspects of phenotypic plasticity into a QTL mapping framework, thereby better unraveling

Corresponding author. Rongling Wu, Center for Computational Biology, Beijing Forestry University, Beijing 100083, China, E-mail: rwu@bjfu.edu.cn and Center for Statistical Genetics, Pennsylvania State University, Hershey, PA 17033, USA, E-mail: rwu@phs.psu.edu; Zhikang Li, Institute of Crop Sciences/National Key Facility for Crop Gene Resources and Genetic Improvement, Chinese Academy of Agricultural Sciences, Beijing 100081, China, E-mail: lizhk@caas.net.cn

*These authors contributed to this work equally.

Zhong Wang obtained his PhD in Engineering Mechanics at Dalian University of Technology in 2000. He found and managed a software company in Japan from 2000 to 2008. He is Visiting Scholar in the Center for Computational Biology at Beijing Forestry University. He writes computer software for statistical genetic models.

Xiaoming Pang obtained his PhD in Pomology at Huazhong Agricultural University in 2002 and then worked in the Institute of Genetics and Developmental Biology of the Chinese Academy of Sciences and the National Institute of Fruit Tree Science of Japan. Since 2006, he has been Associate Professor of Tree Breeding at Beijing Forestry University. His research interest focuses on the utilization of molecular genetics and biotechnologies to study population genetic diversity and map quantitative trait loci in fruit trees.

Yafei Lv is a Master’s student in Forest Genetics and Tree Breeding in the Center for Computational Biology at Beijing Forestry University. His research focuses on the statistical genetics of quantitative traits in forest trees.

Fang Xu is a Master’s student in Forest Genetics and Tree Breeding in the Center for Computational Biology at Beijing Forestry University. His research focuses on the population genetics of forest trees.

Tao Zhou is a PhD candidate in Forest Genetics and Tree Breeding in the Center for Computational Biology at Beijing Forestry University. His research focuses on the study of the genetic mechanisms for adaptive response to environmental stress in forest trees.

Xin Li is a Master’s student in Quantitative Genetics and Bioinformatics in the Center for Computational Biology at Beijing Forestry University. Her research focuses on the development of statistical and computational algorithms for bioinformatics.

Sisi Feng is a Master’s student in Forest Genetics and Tree Breeding in the Center for Computational Biology at Beijing Forestry University. Her research focuses on the use of molecular biotechnologies to study the population genetics of forest trees.

Jiahan Li obtained his PhD in Statistics at the Pennsylvania State University in 2011. He is Assistant Professor of Applied and Computational Mathematics and Statistics at the University of Notre Dame. His research interest is in the development of new statistical methodologies for high-dimensional genetic problems.

Zhikang Li obtained his PhD in Genetics at the University of California, Davis, in 1989. He is Chief Scientist, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences. His expertise is on the genetic mapping of quantitative traits, marker-assisted breeding and functional genomics as well as the application of these technologies to rice breeding and evolutionary genetics.

Rongling Wu obtained his PhD in Quantitative Genetics at the University of Washington in 1995. He is Changjiang Scholars Professor of Genetics and Director of the Center for Computational Biology at Beijing Forestry University and Director of the Center for Statistical Genetics at Pennsylvania State University. His interest is to unravel the genetic roots for the outcomes of biological traits by dissecting the traits into their biochemical and developmental pathways.
the quantitative attribute of trait response to the environment. By testing the curve parameters that specify environment-dependent trajectories of the trait, the model allows a series of fundamental hypotheses to be tested in a quantitative way about the interplay between gene action/interaction and environmental sensitivity. The model can also make the dynamic prediction of genetic control over phenotypic plasticity within the context of changing environments. We demonstrate the usefulness of the model by reanalyzing a QTL data set for rice, gleaning new insights into the genetic basis for phenotypic plasticity in plant height growth.

Keywords: phenotypic plasticity; genotype–environment interaction; QTL; genetic architecture; statistical model; mathematical curve

INTRODUCTION
Because a single phenotype can rarely confer high fitness in all situations, organisms are equipped with a particular internal regulatory machinery for altering their phenotypes to cope with heterogeneities in the environment [1, 2]. The capacity of such phenotypic alterations induced by the environment, called phenotypic plasticity, is thought to be under genetic control; e.g. some genotypes may be more sensitive to environmental change than others [3, 4]. Understanding the genetic basis of phenotypic plasticity that guides varying responses for different genotypes to the environment has been the subject of a long-standing debate in biology [5–14]. To study the degree to which genotypes vary in response to changing environments, the same genotypes are grown under multiple environmental conditions [15, 16]. Traditional analysis of (co)variance approaches are then used to test genetic, environmental and their interaction effects, and estimate the variances due to these effects and their relative contributions to the total phenotypic variance [4, 14]. Such approaches, although simple and popularly used, have a significant limitation in studying the dynamic processes of genetic control over a range of environments, thus incapable of using genetic information to predict the phenotypic changes of a trait arising from per unit change of an environmental factor.

A quantitative framework has now been proposed to describe phenotypic responses to different environments by using various forms of mathematical curves [6, 7, 17–20]. Under this framework, variability of phenotypic values produced by the same genotype when exposed to different environments, i.e. reaction norms, can be explained by mathematical parameters that define response curves. Thus, studying the genetic variation and evolution of reaction norms becomes a statistical issue of estimating and testing these mathematical parameters. As compared to traditional approaches by which genotypic means are plotted and compared over ordered or unordered environmental states [15], this framework displays several important advantages. First, it provides a quantitative description of the change of reaction norms per unit change of an environmental factor, thus displaying a full capacity to capture each and every subtle variability in reaction norm trajectories. Second, it allows biological principles underlying phenotypic plasticity to be incorporated into the analysis of genotype–environment interactions (GEIs). For example, there is a universal law that the metabolic rate of an organism increases with increasing temperature but decreases right after temperature reaches a certain point [20]. The implementation of mathematical aspects of this law will foster the biological relevance and interpretations of results. Third, mathematical modeling enables a parsimonious number of parameters to describe environment-dependent phenotypes, increasing statistical power for identifying GEIs.

In this article, we describe and review a statistical framework model for unraveling the genetic basis of phenotypic plasticity by incorporating the impact of the graded change of an environmental factor. For continuous environments, such as temperature, photoperiod or nutrient availability, an appropriate mathematical equation of biological relevance is used to describe reaction norms as a function of the environmental factor; e.g. a quadratic function quantifies the thermal performance of an insect [20, 21]. In discrete environments, such as different sexes, races or host species for polyphagous insects, phenotypic plasticity can still be viewed as a graded response by using the environmental index as the independent variable [17, 22]. The model incorporates biologically meaningful mathematical equations into a statistical setting for genetic mapping based on molecular linkage maps. By testing differences in a set of mathematical parameters among genotypes at specific genes (or called quantitative trait loci, QTLs), the model can discern the significance of QTLs that trigger genetic effects on reaction norm trajectories. By studying the dynamic behavior of genetic control
exerted by individual QTLs across a continuum of an environmental factor or index, we will not only be able to study the genetic basis of variation in a complex trait, but also gain insight into the genetic architecture of plastic reactions.

MODEL DESCRIPTION
Experimental design
The genetic study of phenotypic plasticity should be based on an appropriate experimental design in the field. Consider a mapping population, derived from the cross between two contrasting inbred lines, which contains n progeny that can be replicated. Such a population can be composed of doubled haploids (DH), recombinant inbred lines, both commonly used in Arabidopsis and crops [23], or hybrid clones derived from cuttings or tissue culture available to many forest trees [10, 11]. Here, we assume the use of a DH population for which there are two homozygous genotypes each for an allele inherited from a different parent at each locus [23]. A high-density genetic linkage map has been constructed for this DH population.

The mapping population is planted in J environments, in each of which each progeny is replicated in K blocks and with L copies in a plot within each block in a randomized complete block design. Consider a quantitative trait of interest, such as plant height or grain yield, which is measured for each individual in each environment. Our interest is to analyze whether these progeny change their phenotypic values over the environment, and if so, whether there exist differences in environment-dependent change, i.e. phenotypic plasticity, among the progeny and whether specific genes contribute to these inter-progeny differences. We will describe several statistical approaches for addressing these questions and particularly review one that displays a pattern of variation in phenotypic plasticity over a range of discrete environments.

Traditional models for genetic variation in phenotypic plasticity
As a ubiquitous phenomenon, studies of variation in phenotypic plasticity among genotypes, expressed as GEI, have been a long-standing focus of developmental and evolutionary biology [2, 4, 6, 8, 10, 12, 14–16]. For a group of genotypes, there may be a particular pattern of variation in phenotypic plasticity.

In general, GEI can be classified into three types, which are described using four DH genotypes and two environments as an example (Figures 1–3) from the multi-site trial described above. These three types are (i) no interaction, (ii) non-crossover interaction and (iii) crossover interaction. Type 1 can be expressed as canalization, in which each genotype is stable across the environment (Figure 1A), or parallel alteration, in which all genotypes change their phenotypes uniformly over the environment (Figure 1B). In type 2, the amount and/or direction of response to the environment are different among genotypes, but the rank of genotypes is environment-invariant. It has two cases, in one of which non-crossover interactions are purely due to genotypic differences in the amount of phenotypic plasticity (Figure 2A), and in the other case both the amount and direction of phenotypic plasticity contribute to non-crossover interactions (Figure 2B). Type 3 shows changes of the rank of genotypes between different environments, which may have the same magnitude of genotypic differentiation between different environments (Figure 2B) or have different magnitudes of genotypic differentiation between environments (Figure 3A) or have different magnitudes of genotypic differentiation between environments (Figure 3B and C). The cases in Figure 3B and C are different in terms of genetic variation in the amount and/or direction of phenotypic plasticity. The former is merely due to the amount, whereas the latter due to both the amount and direction.

**Estimating overall genetic variation by analysis of variance**
Traditional quantitative genetic theory has well been developed to estimate and test the amount of GEI by incorporating an analysis of variance (ANOVA) for experimental data collected from the above multisite trial. The phenotypic value of the trait for individual i from progeny j within block k of environment j is expressed as

$$P_{ijkl} = \mu + G_i + E_j + B_{kij} + GEI_{ij} + GBI_{ikj} + e_{ijkl}$$

(1)

where $\mu$ is the overall mean, $G_i$ is the genetic effect due to the ith progeny, $E_j$ is the macroenvironment effect due to the jth site, $B_{kij}$ is the microenvironmental effect due to the kth block within site j, $GEI_{ij}$ and $GBI_{ikj}$ are the GEI effects at the levels of macroenvironment and microenvironment, respectively, and $e_{ijkl}$ is the residual.

The ANOVA model is used to estimate variance components due to each of these effects by using the structure of Table 1 [10, 11]. Under the assumption
that all these effects are random, the variances due to different effects are estimated as

\[ \hat{s}_G^2 = \frac{1}{LKJ} (MS_G + MS_{GBI} - MS_H - MS_{GEI}) \]  (2)

\[ \hat{s}_E^2 = \frac{1}{LKn} (MS_E + MS_{GBI} - MS_H - MS_{GEI}) \]  (3)

\[ \hat{s}_B^2 = \frac{1}{Ln} (MS_B - MS_{GBI}) \]  (4)

\[ \hat{s}_{GEI}^2 = \frac{1}{LK} (MS_{GEI} - MS_{GBI}) \]  (5)

\[ \hat{s}_{GBI}^2 = \frac{1}{L} (MS_{GBI} - MS_e) \]  (6)

\[ \hat{s}_e^2 = MS_e \]

The estimated genetic variance \( \hat{s}_G^2 \) is due to the overall genetic effect contributed by all underlying loci. The estimates \( \hat{s}_E^2 \) and \( \hat{s}_B^2 \) are the overall variances due to the effects of macroenvironment and microenvironment, respectively [11]. \( \hat{s}_{GEI}^2 \) and \( \hat{s}_{GBI}^2 \) are the estimated variances due to interactions between genotype and macroenvironment and between genotype and microenvironment, respectively. Thus, \( \hat{s}_{GEI}^2 \) can quantify the amount of genetic variance in phenotypic plasticity to macroenvironment, whereas \( \hat{s}_{GBI}^2 \) describes the genetic variance in response to microenvironment changes [11].

We can now use these estimated variance components to explain the types of GEI. In Figure 1A, there exists genetic variance, but there are no environmental variance and GEI variance. In Figure 1B, there are genetic variances and environmental variance, but GEI variance does not exist. In Figures 2 and 3, all these variance components may exist.

Genetic correlations of phenotypic values between different environments can also be used to quantify the genetic variance in phenotypic plasticity [7].
Figure 1A and B, where there is no GEI, such a genetic correlation is equal to one and four genotypic values are on a straight line if environment 1 is plot against environment 2. Across-environment genetic correlations in Figure 2A and B should be less than one, indicating the existence of GEI. In Figure 3, where there are crossovers, across-environment correlations can be negative (Figure 3A), or are less than one (Figure 3B and C).

Molecular dissection by a likelihood model

Variation in phenotypic plasticity can be attributed to individual QTLs if a genetic linkage map is available [8]. Although ANOVA can be used to dissect the phenotypic variance into different QTLs, it is less powerful for locating the chromosomal positions of QTLs than a maximum likelihood approach.

Let $y_i = (y_{ij1}, \ldots, y_{ijg})$ denote the phenotypic data of a trait for genotype $i$ in environment $j$. Assume that the genetic architecture of the trait involves actions and interactions of $m$ QTLs (forming $2^m$ distinct genotypes for the DH population) which interact with the environment in a complicated network to determine its final phenotype. At each QTL, there are two homozygotes $AA$ (coded as 1) and $aa$ (coded as 2). For any one progeny, it should arises from one (and only one) of the $2^m$ possible QTL genotypes. Thus, the distribution of phenotypic data is expressed as the $J$-component mixture probability density function, i.e.

$$y_i \sim f(y_i; \omega_i, \mu, \Sigma) = \sum_{h_1=1}^{2} \cdots \sum_{h_m=1}^{2} \omega_{h_1,\ldots,h_m} f(y_i; \mu_{h_1,\ldots,h_m}, \Sigma),$$

where $\omega_i = (\omega_{h_1,\ldots,h_m})_{h_1,\ldots,h_m=1}^{2}$ is a vector of mixture proportions associated with different QTL
genotypes \{h_1, ..., h_m\} for progeny

\[ \mathbf{g}_i = \{h_{i1}, ..., h_{im}\} \]

contains residual variances and covariances among

\[ \mathbf{D} \]

environments which are common for all QTL

innovations from marker genotypes, thus mixture pro-

portions \( \mathbf{u}_i \) are actually the conditional probabilities

Figure 3: Reaction norms of a phenotypic trait for four genotypes reared in different environments: Crossover

GEI. In this case genotypes that perform better in one environment do worse in the other. (A) Genotypes have
different directions of environment-dependent changes, with the genotype mean of phenotypic plasticity being
zero. The trait values are scattered above and below the diagonal, with between-environment correlation being greater than 1.

(B) Genotypes have the same direction of phenotypic plasticity, although its amount is different. The trait values are scattered above and below the diagonal, with between-environment correlation being greater than 1.

(C) Genotypes are different in both the direction and amount of phenotypic plasticity over two environments. The trait values are scattered above and below the diagonal, with between-environment correlation being greater than 1.
of QTL genotypes, conditional upon the marker genotype of a particular progeny \( i \). To derive such conditional probabilities for different types of mapping populations, see ref. [24].

The probability density function of genotype \( \{ h_1 \ldots h_m \} \), \( f_{m \ldots h_m}(y; \mu_{h_1 \ldots h_m}, \Sigma) \), is assumed to be multivariate normally distributed with \( J \)-dimensional mean vector

\[
\mu_{h_1 \ldots h_m} = (\mu_{h_1}, \ldots, \mu_{h_1 \ldots h_m})
\]

and \((J \times J)\) covariance matrix \( \Sigma \). The likelihood based on a mixture model containing \( 2^m \) QTL genotypes can be written as

\[
L(\Theta | y) = \prod_{i=1}^{n} \left( \sum_{b_{h_1 \ldots h_m} = 0}^{2} \omega_{b_{h_1 \ldots h_m}} f_{b_{h_1 \ldots h_m}}(y; \mu_{b_{h_1 \ldots h_m}}, \Sigma) \right)
\]

where \( \Theta \) is a vector of unknown parameters including the mixture proportions, QTL genotype specific mean vectors, and (co)variances.

In practice, the technique for monitoring the genetic effects of two QTLs at the same time has been available [16], but statistical modeling of multiple QTLs and high-order epistatic interactions is quite premature. Let two QTLs \( A \) and \( B \) explain the test of GEI. The two QTLs form a \( 2 \times 2 \) table for a DH population with genotypic values in environment \( j \) expressed as

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Mean square</th>
<th>Expected mean square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotype</td>
<td>( n-1 )</td>
<td>( MS_G )</td>
<td>( \sigma_a^2 + L\sigma_{GB}^2 + L\sigma_{GB}^2 + L\sigma_{GB}^2 )</td>
</tr>
<tr>
<td>Environment</td>
<td>( J-1 )</td>
<td>( MS_E )</td>
<td>( \sigma_e^2 + L\sigma_{GB}^2 + L\sigma_{GB}^2 + L\sigma_{GB}^2 )</td>
</tr>
<tr>
<td>Block/environment</td>
<td>( (K-1)J )</td>
<td>( MS_{B} )</td>
<td>( \sigma_f^2 + L\sigma_{GB}^2 + L\sigma_{GB}^2 )</td>
</tr>
<tr>
<td>GEI</td>
<td>( (n-1)(J-1) )</td>
<td>( MS_{GEI} )</td>
<td>( \sigma_f^2 + L\sigma_{GB}^2 + L\sigma_{GB}^2 )</td>
</tr>
<tr>
<td>GBI</td>
<td>( (n-1)(K-1) )</td>
<td>( MS_{GMB} )</td>
<td>( \sigma_f^2 + L\sigma_{GB}^2 )</td>
</tr>
<tr>
<td>Error</td>
<td>( (L-1)KJN )</td>
<td>( MS_e )</td>
<td>( \sigma_e^2 )</td>
</tr>
</tbody>
</table>

To detect whether there exists GEI at the individual QTL level, we formulate the null hypotheses as follows:

\[
H_0 : a_{ij} \equiv a_{1j}, \quad j = 1, \ldots, J
\]

\[
H_0 : a_{ij} \equiv a_{2j}, \quad j = 1, \ldots, J
\]

\[
H_0 : a_{ij} \equiv a_{12j}, \quad j = 1, \ldots, J
\]

for testing the significance of additive genetic \times environment interactions and (additive \times additive) \times environment interactions at QTL \( A \) and \( B \), respectively.

**Limitations**

A traditional ANOVA approach is powerful for estimating the magnitude of GEI using a variance component \( \sigma_{GB}^2 \) regardless of how large the number of progeny and the number of environments are. However, it fails to dissect the contributions of individual QTLs to GEI. Although ANOVA based on individual markers can make such dissection, we cannot estimate the change of genetic variance across the environment.

A likelihood approach based on Equation (8), if aimed to directly estimate individual elements in the genotypic vector (7) across different environments and individual elements in covariance matrix \( \Sigma \), has significant limitations:

First, it will be computationally prohibited when the number of environment is more than three.

**Table 1: Analysis of variance for GEIs in a multisite trial by assuming that all the effects are random**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Mean square</th>
<th>Expected mean square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotype</td>
<td>( n-1 )</td>
<td>( MS_G )</td>
<td>( \sigma_a^2 + L\sigma_{GB}^2 + L\sigma_{GB}^2 + L\sigma_{GB}^2 )</td>
</tr>
<tr>
<td>Environment</td>
<td>( J-1 )</td>
<td>( MS_E )</td>
<td>( \sigma_e^2 + L\sigma_{GB}^2 + L\sigma_{GB}^2 + L\sigma_{GB}^2 )</td>
</tr>
<tr>
<td>Block/environment</td>
<td>( (K-1)J )</td>
<td>( MS_{B} )</td>
<td>( \sigma_f^2 + L\sigma_{GB}^2 + L\sigma_{GB}^2 )</td>
</tr>
<tr>
<td>GEI</td>
<td>( (n-1)(J-1) )</td>
<td>( MS_{GEI} )</td>
<td>( \sigma_f^2 + L\sigma_{GB}^2 + L\sigma_{GB}^2 )</td>
</tr>
<tr>
<td>GBI</td>
<td>( (n-1)(K-1) )</td>
<td>( MS_{GMB} )</td>
<td>( \sigma_f^2 + L\sigma_{GB}^2 )</td>
</tr>
<tr>
<td>Error</td>
<td>( (L-1)KJN )</td>
<td>( MS_e )</td>
<td>( \sigma_e^2 )</td>
</tr>
</tbody>
</table>
Although we can always test GEI between each pair of environments, there will be too many interactions to be tested if the number of environment increases, from which it is difficult to comprehend the interpretation of results.

Second, such a treatment does not consider biological principles of phenotypic plasticity. For example, the thermal performance of an insect under a range of temperature involves particular development and regulatory mechanisms underlying body metabolism, which can be described by a quadratic function quantifies \[7, 18, 20, 21\]. Traditional mapping approaches do not incorporate such a biologically meaningful mathematical function, leaving mapping and biology to be disjointed. For discrete environments, the traditional approach treats them as being equally spaced, which is largely improper. For example, it is impossible that Philippines, China, India and Thailand are even spaced from each other in terms of environmental factors that influence plant height growth in rice.

Third, it is not parsimonious because of too many parameters being estimated. It is likely that phenotypic values across a series of environmental states display an autocorrelation structure. Parsimonious modeling of this correction helps to increase statistical power and robustness.

Fourth, because of these three reasons above, results from the traditional treatment are difficult to be synthesized and interpreted, thereby affecting the construction of a comprehensive picture of the genetic architecture of phenotypic plasticity for a complex trait.

A dynamic modeling framework for GEI

Considering the fact that the development of a trait over time has an underlying biological basis which can be quantified by a mathematical equation, a dynamic model, called functional mapping, has been developed for mapping the QTLs that govern developmental trajectories and pattern of the trait \[25\]. Korol and group have proposed a similar mechanistic approach for identifying QTLs that control the plastic response of a trait across a range of environments \[17, 19\]. We argue that functional mapping and Korol’s approach are biologically more relevant and meaningful and statistically more powerful for revealing the genetic architecture of phenotypic plasticity than traditional non-mechanistic approaches.

To explain our idea, we use an example for rice mapping. A population of 123 DH lines in rice was grown in 11 locations from each of which plant heights were measured. We calculated the environmental index (EI) for each location, defined as the difference of the mean of all DH lines in this location from the mean of all DH lines over all locations \[22\]. Figure 4 illustrates phenotypic trajectories of plant heights across 11 ordered environments determined by normalized EIs. By statistical tests, an exponential equation was found to adequately describe plant height trajectories as a function of EI. Thus, we use such an exponential equation to specify genotype-specific trajectories, expressed as

\[\mu_{h_1...h_m} = \alpha_{h_1...h_m} \beta_{h_1...h_m}^{x_j}\]

where \(\mu_{h_1...h_m}\) is the genotypic mean of plant height for QTL genotype \(\{h_1...h_m\}\) in environment \(j\), \(x_j\) is the EI of environment \(j\), \(\alpha_{h_1...h_m}\) is the constant (intercept) of environment-dependent trajectory for genotype \(\{h_1...h_m\}\), and \(\beta_{h_1...h_m}\) is the exponential coefficient of environment-dependent trajectory for genotype \(\{h_1...h_m\}\). By estimating and comparing only two parameters \((\alpha_{h_1...h_m}, \beta_{h_1...h_m})\) (rather than \(\mu_{h_1...h_m}\) means) for each genotype, we will be able to determine the dynamic pattern of genetic control over environment-dependent trajectories.

For some traits whose environment-dependent expression does not obey an explicit mathematical function, nonparametric approaches, such as B-spline, can be used. In many quantitative genetic analyses, a computation-efficient nonparametric
approach based on Legendre orthogonal polynomials (LOP) has been used for mean and variance modeling [26–28]. Since the LOP are orthogonal to each other and integrate to 0 in the interval [−1, 1], they have been applied to nonparametric regression [29], with parameter estimates possessing favorable asymptotic properties [30, 31].

The covariance matrix among phenotypic values measured at different environments may follow a structure. Many different approaches have been developed to model the covariance structure. The simplest approach is the first-order autoregressive (AR(1)) model by which variances and covariances are considered as being stationary over time. According to the AR(1) model, the variance in environment $j$ is expressed as $\sigma_j^2 \equiv \sigma^2$ and the covariance between environments $j_1$ and $j_2$ expressed as $\sigma(j_1, j_2) \equiv \sigma \rho_j^{x_{j_1}x_{j_2}}$ where $\rho$ is the correlation and $x_{j_1}$ and $x_{j_2}$ are the EI in environment $j_1$ and $j_2$, respectively. If the stationarity assumptions do not hold, we will need to use a nonstationary approach, such as a structured antidependence (SAD) model [32, 33] and autoregressive moving average (ARMA) [34], for the covariance structure. In some cases, nonparametric or semiparametric approaches are a better choice [35]. Thus, instead of estimating all elements in the covariance matrix, we estimate the parameters that model the covariance structure. Zimmerman and Nunez-Anton [32] discussed the procedures and criteria for model selection in covariance structure.

**Estimation and tests**

Many algorithms can be implemented to estimate the parameters, $\Theta$, contained in the likelihood (8). The $\Theta$ includes QTL positions, the parameters that model genotypic mean vectors over a range of environments, and the parameters that model the covariance structure. In general, the EM algorithm provides an overarching platform within which other computing algorithms, such as the simplex approach [36], are embedded to estimate $\Theta$.

After all the parameters are estimated, we can formulate several important hypothesis tests. First, we need to test whether there is genetic control over the complex trait studied. This can be done by

$$H_0 : (\alpha_{h_1...h_n}, \beta_{h_1...h_n}) \equiv (\tilde{\alpha}, \tilde{\beta}) \quad \text{vs} \quad H_1 : \text{Not all equalities in the } H_0 \text{ are held (} h_1, \ldots, h_m = 1, 2)$$

(16)

from which a log-likelihood ratio (LR) test statistics is calculated and $(\tilde{\alpha}, \tilde{\beta})$ are the estimates of exponential parameters under the $H_0$ (there is no QTL). The critical threshold for claiming the significance of genetic control can be determined from permutation tests [37]. After significant genetic control is detected, we will test many biologically meaningful hypotheses.

The second hypothesis is about the significance of the additive, dominant, and epistatic genetic effects. Consider a pair of QTLs $A$ and $B$. From Equations (9)–(11), the absence of additive effects and additive × additive epistatic effects between the two QTLs should meet the following constraints, respectively:

$$\alpha_{11}\beta_{11}^A + \alpha_{10}\beta_{10}^A = \alpha_{01}\beta_{01}^A + \alpha_{00}\beta_{00}^A$$

(17)

$$\alpha_{11}\beta_{11}^B + \alpha_{10}\beta_{10}^B = \alpha_{01}\beta_{01}^B + \alpha_{00}\beta_{00}^B$$

(18)

$$\alpha_{11}\beta_{11}^A + \alpha_{00}\beta_{00}^A = \alpha_{10}\beta_{10}^A + \alpha_{01}\beta_{01}^A$$

(19)

The likelihoods under the null hypotheses with each of these constraints (and then the corresponding LR values) are calculated to test whether each of these effects is equal to zero. The critical thresholds for the significance tests can be determined on the basis of simulation studies.

No matter whether these genetic effects are significant, the third test needs to be performed about the significance of GEIs expressed at the additive genetic effect of QTL $A$ and $B$ and their additive × additive epistatic interaction. These tests should meet the following constraints, respectively,

$$\alpha_{11}\beta_{11}^A + \alpha_{10}\beta_{10}^A - (\alpha_{01}\beta_{01}^A + \alpha_{00}\beta_{00}^A)$$

$$= \alpha_{11}\beta_{11}^B + \alpha_{10}\beta_{10}^B - (\alpha_{01}\beta_{01}^B + \alpha_{00}\beta_{00}^B)$$

(20)

$$\alpha_{11}\beta_{11}^A + \alpha_{00}\beta_{00}^A - (\alpha_{10}\beta_{10}^A + \alpha_{01}\beta_{01}^A)$$

$$= \alpha_{11}\beta_{11}^B + \alpha_{00}\beta_{00}^B - (\alpha_{10}\beta_{10}^B + \alpha_{01}\beta_{01}^B)$$

(21)

$$\alpha_{11}\beta_{11}^A + \alpha_{00}\beta_{00}^A - (\alpha_{10}\beta_{10}^A + \alpha_{01}\beta_{01}^A)$$

$$= \alpha_{11}\beta_{11}^B + \alpha_{00}\beta_{00}^B - (\alpha_{10}\beta_{10}^B + \alpha_{01}\beta_{01}^B)$$

(22)

for any $j_1, j_2 = 1, \ldots, J$.

If the null hypothesis under constraint (17) is rejected, this means that there is a significant additive × environment interaction over a continuum of environments at QTL $A$. The same is true for QTL $B$. If the null hypothesis under constraint (19) is rejected, this means that an (additive × additive epistatic) × environment interaction is significant over a range of environments. Likewise, one can test
whether a particular genetic effect interacts with a set of environments of interest. The critical thresholds for each of these GEI tests can be determined on the basis of simulation studies.

MODEL DEMONSTRATION

We suggest that a dynamic model that incorporates the mechanistic response of phenotypes across environments can better capture and interpret the genetic basis of phenotypic plasticity. Below, we show the application of this model by reanalyzing a published data set for QTL mapping for GEI in rice.

Field trial

Two rice varieties, Indica IR64 and Japonica Azucena, and their 123 DH lines were planted in seven different locations of four Asian countries (Philippines, China, India and Thailand), spanning from 13.5° to 31.5°N in latitude and from 76° to 121.5°E in longitude [16]. At one location in China, the same experiment was conducted in early and late growing seasons. Likewise, at the International Rice Research Institute of the Philippines, the experiment was performed in wet and dry seasons, in which an additional experiment was repeated with a subset of DH lines (82) in a different dry season but receiving two treatments, one under well-watered aerobic conditions and the other under water stress. Thus, a total of 11 environments were considered for genetic mapping of QTL–environment interactions. At each location, five representative plants of each line in a plot were randomly chosen to measure plant heights.

A complete linkage map for this DH population was constructed with a total of 176 markers [23]. This map, covering all 12 rice chromosomes, has a total genome size of 1711.2 cM and an average distance of 10.4 cM between adjacent markers.

Variation in reaction norms

By plotting plant heights over the normalized E1, reaction norms for the mapping population and their parents shorter IR64 and taller Azucena can be visualized as environment-dependent trajectories (Figure 4). The plant height for each line depends on the environmental growing conditions. Overall, all lines increase their plant height growth with increasing E1, although a particular line has decreasing growth in a certain range of E1. Parent Azucena is consistently taller over all environments than parent IR64, with the progeny displaying substantial variation in plant height trajectories. Crossovers in environment-dependent trajectories among different progeny implicate the existence of GEIs. While most progeny have plant heights intermediate between the two parents, some transgressive segregants (i.e., those that are taller than the taller parent or shorter than the shorter parent) were identified, suggesting the complexity of inheritance of height genes.

Environment-dependent expression of QTLs and QTL–environment interactions

The mechanistic mapping model incorporating Equation (15) was used to scan the rice genome for the detection of significant QTLs and their epistatic interactions for the phenotypic plasticity of plant height. Based on the test (16), the model identified six significant pairs of QTLs (A–F) distributed on chromosomes 1, 2, 3, 4, 5, 10 and 11 (Figure 5), which are

<table>
<thead>
<tr>
<th>Pair</th>
<th>QTL A Marker Interval</th>
<th>Chr</th>
<th>QTL B Marker Interval</th>
<th>Chr</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>U10–110–W1–35</td>
<td>1</td>
<td>RG71–RG157</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>RZ730–RG810</td>
<td>1</td>
<td>RG574–RZ394</td>
<td>3</td>
</tr>
<tr>
<td>C</td>
<td>RZ678–RZ574</td>
<td>3</td>
<td>RZ565–RZ675</td>
<td>4</td>
</tr>
<tr>
<td>D</td>
<td>RZ678–RZ574</td>
<td>3</td>
<td>CI195–RZ2174</td>
<td>10</td>
</tr>
<tr>
<td>E</td>
<td>RZ565–RZ675</td>
<td>4</td>
<td>RG556–RZ556</td>
<td>5</td>
</tr>
<tr>
<td>F</td>
<td>RG403–RG13</td>
<td>5</td>
<td>RG103–RG1109</td>
<td>11</td>
</tr>
</tbody>
</table>

At each pair, there are four homozygous genotypes $AABB$, $AAbb$, $aabb$, and $aabb$, assuming that the capital alleles inherit from the IR64 parent and the lowercase alleles from the Azucena parent. The estimated parameters of the power equation (15) were used to draw environment-dependent trajectories for plant height (Figure 6).

One of the most significant advantages of the dynamic model is that it can quantify the pattern and process of the genetic variation in phenotypic plasticity to changing environments. Using constraints (17)–(19) as the null hypothesis, we tested whether each pair of QTLs affects plant height across environments additively or in an additive epistatic manner. Each QTL was found to trigger an additive effect and also interact with other QTLs to determine plant height across environments. We used constraints (20)–(22) to test the significance of GEIs at the additive and epistatic level. Visualized by environment-dependent curves, the additive genetic
effects at each QTL (Figure 7A) and additive × additive epistatic effects at each pair of QTLs (Figure 7B) can be tested about whether they are significant on the phenotypic plasticity of plant height. Such effects are significant if their environment-dependent curves are significantly unparallel to the horizontal dash line (with a zero effect) in Figure 7.

Except for QTLs located in marker intervals [RZ565–RZ675] on chromosome 4, [RG403–RG13] on chromosome 5 and [C1195–R2174] on chromosome 10 (labeled as 6, 8 and 9, respectively) with an effect curve in parallel to the dash zero line, all QTLs display significant additive–environment interactions for plant height (Figure 7A). Also, based on the environmental slopes of the curves, we can quantify the degree of additive–environment interactions. For example, QTLs located in marker intervals [U10-110–W1-35] on chromosome 1, [RZ574–RZ394] on chromosome 3 and [RZ678–RZ574] on chromosome 3 (labeled as 1, 4 and 5, respectively) are more sensitive to environmental changes than QTLs located in marker intervals [RG556–RZ556] on chromosome 5 and [RG247–RG103] on chromosome 11 (labeled as 7 and 10, respectively) because the former have a larger slope than the latter. The QTL located in marker interval [RG171–RG157] on chromosome 2 (labeled as 3) has a similar slope to QTLs 1, 4 and 5, but it triggers a larger additive effect on plant height on the latter. Unlike these QTLs, the QTL located in marker interval [RZ730–RG810] on chromosome 1 (labeled as 2) exerts an increasing negative additive effect with environment because of directional change of the effects of parental alleles. At each QTL pair, there exists a significant additive × additive–environment interaction effect on plant height because no effect curves are parallel to dash zero line (Figure 7B). Pairs 5 and 9, 1 and 3, and 8 and 10 trigger different directions of interactions from pairs 2 and 4, 6 and 7, and 5 and 6.

DISCUSSION

The phenotype of an organism is not only controlled by its genes, but also by the environment where it grows. A growing body of evidence shows that the extent to which environment drives phenotypic changes, known as phenotypic plasticity [3, 4, 5–12, 14], is also under genetic control. In a previous study measuring variation in activity level for each of the roughly 6000 genes found in yeast across a range of stressful environments, Li et al. [13] found that some genes varied enormously in their expression levels from one environment to the next, whereas others were relatively constant. However, without proper analysis and modeling, we would not been able to determine how many of these plastic genes are involved in the phenotypic plasticity of a complex trait.

There has been a vast amount of literature on the genetic studies of phenotypic plasticity. A majority of these studies treat the environment as being discrete and use regression models to estimate the variance due to GEIs as a criterion to assess the genetic basis of phenotypic plasticity [15]. Founded on the established theory by several authors who view plastic
Figure 6: Environment-dependent trajectory curves of plant heights for four genotypes at each pair of QTLs which are derived from marker intervals: (A) [U10 - I10 - W1 - 35] on chromosome 1 and [RG171 - RG157] on chromosome 2, (B) [RZ730 - RG810] on chromosome 1 and [RZ574 - RZ394] on chromosome 3, (C) [RZ678 - RZ574] on chromosome 3 and [RG565 - RG675] on chromosome 4, (D) [RZ678 - RZ574] on chromosome 3 and [C1195 - RZ74] on chromosome 4, (E) [RZ565 - RZ675] on chromosome 4 and [RG556 - RZ556] on chromosome 5, and (F) [RG403 - RGI3] on chromosome 5 and [RG103 - RGI109] on chromosome 11. Capital alleles are inherited from the shorter IR64 parent, whereas lowercase inherited from the taller Azucena parent. Each lighter line represents a DH progeny line.
response as a curve [6, 7, 18–20], we described and reviewed a dynamic model that can identify the genetic architecture of how individual genes interact with the environment to determine complex traits using QTL mapping. Different from traditional discrete approaches in the literature, the model is able to detect QTLs that regulate the phenotypic change of a trait due to a graded change of an environmental factor. Given the fact that the evolution of quantitative traits is a continuous process [14], the new model is in a better position to be incorporated into the context of evolutionary biology by quantifying and predicting the genetic variation of phenotypic plasticity.

The dynamic model has several key biological meanings by embracing biological principles underlying phenotypic plasticity into the model using robust mathematical equations. First, it can identify genetic control for the origin of phenotypic plasticity; i.e. under what range of environmental conditions a QTL channels the phenotypic change of a trait through an environmental gradient.

Second, the model has power to detect QTLs responsible for the rate of plastic response to environmental changes within any range of environments. Third, the model enables the integration of the genetic mechanisms for phenotypic variation in a complex trait and for the dynamic process of how the trait changes its phenotype over changing environments. For example, if two phenotypes at a QTL are found to differ but parallel over a range of environments, the model suggests that this QTL controls phenotypic variation but does not exhibit GEIs. If two different genotypes do not parallel, however, the underlying QTL does not only affect variation in trait mean, but also regulate the plastic change of the trait. Because of different dynamic features, we infer that, compared with the former QTL, the latter QTL is more likely to direct the trait into evolution and speciation if the environment extends beyond a threshold. In sum, the model will help to gain a general and comprehensive view of the genetic control of phenotypic plasticity to specific environmental cues and enrich our understanding of fundamental issues in evolutionary biology.

**Key Points**

- The environment-dependent change of a genotype, called phenotypic plasticity, plays a central role in maintaining an organism's adaptability to changing environments.
- Genetic analysis of phenotypic plasticity can shed light on the variation and evolution of organismic adaptation and provide useful information to guide the selection of superior crop genotypes for particular environments.
- We describe and review a statistical framework for comprehending the genetic control mechanisms of phenotypic plasticity and predicting phenotypic changes of a quantitative trait over discrete environments.
FUNDING
NSF/IOS–0923975, Changjiang Scholars Award and ‘Thousand-person Plan’ Award.

References