Functional mapping of ontogeny in flowering plants

Xiyang Zhao*, Chunfa Tong*, Xiaoming Pang*, Zhong Wang, Yunqian Guo, Fang Du and Rongling Wu

Submitted: 26th May 2011; Received (in revised form): 27th July 2011

Abstract

All organisms face the problem of how to perform a sequence of developmental changes and transitions during ontogeny. We revise functional mapping, a statistical model originally derived to map genes that determine developmental dynamics, to take into account the entire process of ontogenetic growth from embryo to adult and from the vegetative to reproductive phase. The revised model provides a framework that reconciles the genetic architecture of development at different stages and elucidates a comprehensive picture of the genetic control mechanisms of growth that change gradually from a simple to a more complex level. We use an annual flowering plant, as an example, to demonstrate our model by which to map genes and their interactions involved in embryo and postembryonic growth. The model provides a useful tool to study the genetic control of ontogenetic growth in flowering plants and any other organisms through proper modifications based on their biological characteristics.

Keywords: functional mapping; growth; ontogeny; developmental stage; quantitative trait loci

Corresponding author. Rongling Wu. Center for Computational Biology, National Engineering Laboratory for Tree Breeding, Key Laboratory of Genetics and Breeding in Forest Trees and Ornamental Plants, Beijing Forestry University, Beijing 100083, China. Tel: +0086-10-62336283; Fax: +0086-10-62336283; E-mail: rwu@bjfu.edu.cn.

*X* These authors contributed equally to this work.

Xiyang Zhao obtained his PhD in Forest Genetics and Breeding at Beijing Forestry University in 2010 and is at present a post-doctoral researcher in the Center for Computational Biology at Beijing Forestry University. His research interest is to study the genetic and physiological mechanisms of plant growth and development and use this knowledge to develop an efficient breeding strategy for tree breeding.

Chunfa Tong obtained his PhD in Forest Genetics at Nanjing Forestry University in 2001 and is now Professor of Forest Genetics and Genomics at the same university. From 2009 to present, he was a visiting scholar in the Center for Statistical Genetics, The Pennsylvania State University. His research interest is in the methodological and algorithmic development of genetic mapping for complex traits.

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.

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 a post-doctoral researcher in the Center for Statistical Genetics at the Pennsylvania State University. He writes computer software for statistical genetic models.

Yunqian Guo obtained her PhD in Cell Biology at Tsinghua University in 2008 and is now Lecturer of Cell Biology in the Center for Computational Biology at Beijing Forestry University. She uses an integrative computational and experimental approach to understanding plant growth and development at the molecular and cellular level.

Fang Du obtained her PhD in evolutionary genetics at Lanzhou University and Cotutelle de THÉSE in The University of Bordeaux I in 2010. She is Lecturer of Tree Genetics in the Center for Computational Biology at Beijing Forestry University. Her present research projects include the use of computational and experimental approaches to study population, ecological and evolutionary genetics in forest trees.

Rongling Wu obtained his PhD in Quantitative Genetics at the University of Washington in 1995. He is Changjiang Scholars Professor of Genetics and the Director of the Center for Computational Biology at Beijing Forestry University. His interest is to unravel the genetic roots for the outcome of a biological trait by dissecting the trait into its biochemical and developmental pathways. He uses multidisciplinary tools, integrating genetics, molecular biology, statistics, mathematics, computer sciences and engineering, to solve genetic problems.
INTRODUCTION

Genes, expressed through their actions and interactions with a variety of developmental and environmental factors, are thought to be the driving force of organismic growth [1, 2]. Traditional approaches for studying the genetic control of growth are based on quantitative genetics by which the temporal pattern of genetic variation is analyzed and tested [3–5]. With recent developments of high-throughput genotyping techniques, it has been possible to dissect growth variation into the underlying genes throughout the genome and to estimate these genes individually or through combination [6, 7]. Considering its dynamic feature, a large body of present work integrates growth equations to map genes or called quantitative trait loci (QTLs) responsible for growth processes [8–13]. These models that have been packed into a series of toolkits called functional mapping [14] have proven to be powerful for addressing many conceptual and empirical problems, including those of when and where specific suites of genes control organ growth at different stages of development [15].

Still largely missing, however, is a complete quantitative framework that specifies how genes determine ontogenetic growth in the lifetime of an organism. Previous work for functional mapping focused on the identification of QTLs for a particular phase of development using a mathematical model for growth trajectories during this specific phase. Thus, identified QTLs from this approach cannot be used to elucidate the whole landscape of ontogenetic growth and development. Indeed, there exists a biological phenomenon that pervades all kingdoms of life, by which all organisms undergo a sequence of changes and transitions in development during their ontogeny [2]. For example, plant development includes a broad spectrum of processes, i.e. the formation of a complete embryo from a zygote, seed germination, the elaboration of a mature vegetative plant from the embryo, the formation of flowers, fruits and seeds, and many of the plant’s responses to its environment [16, 17]. Each of these processes is fundamental to determine the size, shape and production of a plant. For this reason, knowledge of the genetic basis of the variation in each process is important for understanding adaptive evolution in nature and deriving elite domestic crop varieties in agriculture.

In this article, we present a comprehensive framework for functional mapping to detect and identify ontogenetic QTLs that govern all developmental events in an annual flowering plant’s lifetime (Figure 1). To clearly demonstrate this framework, we focus on whole-body biomass growth spanning from embryo to adult, although the framework allows the modeling of other developmental events such as carbon allocation, morphological shape, structural allometry and reproductive behavior [18–20]. We first outline the principle of functional mapping, followed by a brief description of estimating genetic parameters within functional mapping. We provide a procedure to test how genes act and interact to determine the expression of whole-body biomass growth at various stages and its transition from one stage to next.

FUNCTIONAL MAPPING

Genetic design

Functional mapping is equipped with a capacity to detect the temporal pattern of genetic effects triggered by individual genes in an organism’s development. We use a backcross design for a flowering plant, e.g. maize (Figure 1), to describe the basic principle of functional mapping. Consider two inbred lines which are crossed to generate an F1 hybrid. By crossing the F1 with one of the original parents, a backcross population is produced. A specific gene with two alleles A and a is assumed to affect embryonic and post-embryonic growth trajectories. In the backcross, there are two segregating genotypes AA and aa, which will, through artificial self-pollination, produce three embryo genotypes, AA, Aa and aa, and one embryo genotype aA, respectively. A seed of higher plants contains a triploid endosperm, in addition to the diploid embryo, derived from double pollination. The endosperm of a self-pollinated seed of the backcross carries one of the four triploid genotypes AAA, AaA, Aaa and aA. Table 1 tabulates the genotypes of the backcross and the embryo and endosperm genotypes of seeds produced by the backcross through self-pollination.

In general, the lifecycle of a plant comprises approximately four distinct phases, seed formation, seed germination, vegetative growth and reproductive growth, in each of which biomass data are measured at multiple time points. After seeds are formed in the backcross plant, we will monitor the trajectory of seed growth. In the past, the measurement of seed size or biomass is based on oven-dried samples. However, this destructive approach makes it impossible to take several measurements on the same seed at different time points. A novel approach based on
Figure I: Life cycle of maize (Zea mays), demonstrating the process from flowering to seeding. The vegetative growth undertakes in the diploid sporophyte generation. When the plant is mature, meiosis occurs in the male (tassels) and female flowers (ears). The haploid microspores (male spores) develop into pollen grains, and the single surviving haploid megaspore (female spore) divides mitotically to form the embryo sac (megagametophyte). The egg forms in the embryo sac. Pollination leads to the formation of a pollen tube containing two sperm cells (the microgametophyte). Finally, double fertilization results in the formation of the diploid zygote, the first stage of the new sporophyte generation and the triploid endosperm cell. Adapted from Candela and Hake [16].

Table I: Compositions of genotypic values for seed development affected by a maternal, embryo and endosperm gene in a backcross design

<table>
<thead>
<tr>
<th>No.</th>
<th>Backcross Parent</th>
<th>Embryo</th>
<th>Endosperm</th>
<th>Genotypic composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ao (1/2 a)</td>
<td>AA (a)</td>
<td>AAA (1/2 a)</td>
<td>( \mu_1 = \mu + 3a + \frac{1}{2}d_1 + \frac{3}{2}d_2 + \frac{1}{2}d_3 )</td>
</tr>
<tr>
<td>2</td>
<td>Ao (1/2 a)</td>
<td>Aa (0)</td>
<td>Aaa</td>
<td>( \mu_2 = \mu + a + \frac{1}{2}d_2 )</td>
</tr>
<tr>
<td>3</td>
<td>Ao (1/2 a)</td>
<td>Aa (0)</td>
<td>Aaa (1/2 a)</td>
<td>( \mu_3 = \mu - \frac{1}{2}d_2 )</td>
</tr>
<tr>
<td>4</td>
<td>Ao (1/2 a)</td>
<td>aa (0)</td>
<td>aaa (1/2 a)</td>
<td>( \mu_4 = \mu - 2a - \frac{1}{2}d_1 - \frac{3}{2}d_2 + \frac{1}{2}d_3 )</td>
</tr>
<tr>
<td>5</td>
<td>aa (1/2 a)</td>
<td>aa (0)</td>
<td>aaa (1/2 a)</td>
<td>( \mu_5 = \mu - 3a + \frac{1}{2}d_1 + \frac{3}{2}d_2 + \frac{1}{2}d_3 )</td>
</tr>
</tbody>
</table>

Note. The mean values of joint maternal-embryo-endosperm genotypes are denoted as \( \mu_1, \mu_2, \mu_3, \mu_4 \) and \( \mu_5 \), respectively. Of their compositions, \( \mu \) is the overall mean, \( a \) is the additive effect of the gene, \( d_1 \) is the interaction between the maternal and embryo alleles within the gene, \( d_2 \) is the interaction between the maternal and endosperm alleles within the gene, and \( d_3 \) is the interaction between the embryo and endosperm alleles within gene.
image analysis has been proposed to infer seed size and biomass accurately as a non-destructive and fast alternative [21]. Suppose we have a series of seed size and biomass data for the self-pollinated progeny of the backcross at multiple time points from the formation of zygote to seed maturity (Figure 1).

The seed of the backcross is germinated and grown into a seedling, entering a post-embryonic stage. During the period of seed germination to the emergency of the first leaf (Figure 2), plant growth is fueled by nutrients reserved in the endosperm. The plant first grows the root stem, and then shoot that becomes leaves. Plant growth data are collected at multiple time points. To measure whole-plant biomass, non-destructive image analysis approaches can be used. After the emergency of the first true leaf, the seedling grows into a mature plant by using the energy produced from leaf photosynthesis (Figure 3). A series of vegetative growth data at multiple time points during this period are collected. The mature plant will produce flowers to enter the next generation of seed formation after pollination.

Below, we will model the pattern of how genes and genetic interactions determine the lifecycle of plants by considering different phases of development.

**Quantitative genetic models**

Various models are developed to study the genetic control of plant growth at different stages. The seed model specifies joint control of genes derived from the maternal, embryo and endosperm genomes [22–25]. Since seeds are seated in a maternal plant, their growth critically depends on the efficiency by which the plant transports nutrients and other essential materials into seeds. Also, a seed is composed of two mutually related but different tissues, embryo and endosperm, whose genomes contribute to seed growth in a different way.

For a specific seed, we can write genotypes at a gene A of interest based on its embryo and endosperm as well as the maternal plant from which the seed is derived (Table 1). Taken together, this will form a total of five groups of genotypes. The maternal backcross plants are classified into two genotypes Aa and aa which differ by an additive effect (denoted as a). The embryo genotypes are AA, Aa and aa, whose genotypic values are expressed as +a, 0a and –a, respectively. Similarly, genotypic values of four endosperm genotypes AAA, AAa, Aaa and aaa are denoted as +3a, +1a, –1a and –3a, respectively. Three genomes interact with each other to affect seed development, which is modeled by the interactions between the maternal and embryo (d1), the maternal and endosperm (d2), and the embryo and endosperm (d3). Table 1 gives the genetic compositions of each joint genotype from the maternal, embryo and endosperm.

Based on the joint genotypic values (Table 1), we solve each genetic composition as follows:

\[ a = \frac{1}{6}(\mu_1 - \mu_2) \]  
\[ d_1 = \frac{2}{9}(2\mu_1 - 9\mu_2 + 9\mu_3 - 3\mu_4 + \mu_5) \]  
\[ d_2 = \frac{1}{3}(\mu_1 - 6\mu_2 + 6\mu_3 + \mu_5) \]  
\[ d_3 = \frac{1}{3}(\mu_1 - \mu_2 - \mu_3 + \mu_4). \]

Seed germination may be affected only by embryo and endosperm genes because the seed has been removed from the maternal plant at this stage [26].

![Figure 2: Growth process for monocotylodeon and dicotyledon during seed germination. Seed structure for these two groups of plants is shown at the left panel. By absorbing moisture and nutrients from the endosperm inside the seed, a seed can germinate, sprout and grow into a seedling. At this phase, embryo and endosperm genes interact with each other to coordinate seedling growth. Adapted from Taiz and Zeiger [17].](http://bib.oxfordjournals.org/)

Downloaded from http://bib.oxfordjournals.org/ by guest on July 14, 2015
Figure 3: Different growth stages of a maize plant, including vegetative (V) and reproductive (R) stages. The V developmental milestones include emergence (VE), in which the coleoptile reaches the soil surface and elongates due to its exposure to sunlight; V1, in which the lowermost leaf has a visible leaf collar; V3, in which the plant has three leaf collars, whose growth purely relies on photosynthesis; V7, in which the plant has seven leaf collars and experiences rapid growth; V10, in which the plant equipped with 10 leaf collars has a rapidly-growing stalk and VT, in which the last branch of the tassel is visible. The R developmental milestones include RI, in which any silk is visible outside the husk; R2, in which kernels are white and resemble a blister in shape; R3, in which kernels are yellow on the outside with a milky white inner fluid; R4, in which starch is dough-like consistency; R5, in which kernels are dent in the outside the husk; R6, in which all kernels on the ear have reached maximum dry weight with physiological maturity. Adapted from [50].

However, imprinted within the seed, the maternal genes may also play an important role in plant growth during seed germination. Thus, it is interesting to test how maternal genes affect seed germination. Similarly, in vegetative and reproductive growth stages, imprinted influences of maternal and endosperm genes may be important and should be tested.

Dynamic models

Functional mapping can test dynamic changes of gene expression in a time course by incorporating the mathematical aspects of biological principles behind the process of growth. For n self-pollinated seeds of the backcross, we measure biomass growth for growing seeds (yi) at T1 time points, seedling biomass in the seed germination stage (y2) at T2 time points, vegetative biomass in the post-cotyledon stage (y3) at T3 time points and flower biomass (y4) at T4 time points. Let yi = [y11(t11), ..., y1i(t1T1)], y2i = [y21(t21), ..., y2i(t2T2)], y3i = [y31(t31), ..., y3i(t3T3)] and y4i = [y41(t41), ..., y4i(t4T4)] denote time-dependent biomass measures of seed i at the four stages, respectively. Let yT_i = [y1i; y2i; y3i; y4i]. The phenotypic values of yi, affected by a putative gene with five joint maternal–embryo–endosperm genotypes (Table 1), can be expressed as

\[ y_i = \sum_{j=1}^{5} \xi_{ij} \mu_j + X_j^T d + e_i, \]  \hspace{1cm} (5)

where \( \xi_{ij} \) is the dummy variable that is defined as 1 if progeny i carries genotype j and 0 otherwise; \( \mu_j \) is the mean vector of biomass at different time points of different stages; \( d = (d_1, \ldots, d_L) \) is the effect vector of L covariates, \( X_j \) is the design matrix for progeny j that reflect the covariate effects and \( e_i \) is the residual effect.

The likelihood of the data \( y_i = (y_{1i}; y_{2i}; y_{3i}; y_{4i}) \) is formulated as

\[ L(\Theta) = \prod_{j=1}^{5} \prod_{i=1}^{n_j} f(y_i | \mu_j, \Sigma), \]  \hspace{1cm} (6)

where \( \Theta \) is the unknown parameter vector to be estimated from the data (see below); \( n_j \) is the number of seeds for joint genotype j (j = 1, ..., 5) (Table 1); \( f(y_i | \mu_j, \Sigma) \) is a multivariate normal distribution for joint genotype j with \( (T_1 + T_2 + T_3 + T_4) \)-dimension mean vector containing three stages of development...
\[ \mu_j^T = (\mu_{1j}, \mu_{2j}, \mu_{3j}, \mu_{4j}) \]

\[ = [\mu_{1j}(t_1), \ldots, \mu_{1j}(t_{T_1}); \mu_{2j}(t_1), \ldots, \mu_{2j}(t_{T_2}); \mu_{3j}(t_3), \ldots, \mu_{3j}(t_{T_3}); \mu_{4j}(t_4), \ldots, \mu_{4j}(t_{T_4})] \]

and \((T_1+T_2+T_3+T_4) \times (T_1+T_2+T_3+T_4)\) covariance matrix \(\Sigma\).

The biological merit of functional mapping is to model genotype-specific growth curves based on a biologically meaningful mathematical equation, rather than estimate all values in the genotypic mean vectors. Below, we will formulate functional mapping for different stages of plant growth.

(i) **The seed formation model.** For the seed in a growing plant, a sigmoid equation can be used to capture its growth trajectory [27, 28]. For a joint genotype \(j\), this is expressed as

\[ \mu_{1j} = \left( \frac{\alpha_j}{1 + \beta_j e^{-\gamma_j t / \alpha_j}} \right), \]

where \(\alpha_j\) is the asymptotic value of seed growth, \(\beta_j\) is the initial value of seed growth and \(\gamma_j\) is the relative growth rate of seed biomass.

(ii) **The seed germination model.** Seed germination is the first part of plant vegetative growth. In this stage, the seedling has not yet reached its maximum size. A growth model that takes into account this feature was proposed by Jens and Bayley [29]. For joint genotype \(j\), this model is expressed as

\[ \mu_{2j} = (A_j + B_j t_{1j} - c_j / D_j t_{1j}, \ldots, A_j + B_j t_{T_2} - c_j / D_j t_{T_2}), \]

where \(A_j\) is the initial growth, \(B_j\) is the velocity of growth, and \(C_j\) and \(D_j\) describe rapid early growth or acceleration.

(iii) **The vegetative growth model.** Biomass growth after the emergence of the first photosynthetic leaf is complex, which includes exponential and stationary phases. Pearl and Reed [30] proposed a logistic curve that can reflect this complexity. For joint genotype \(j\), this model is expressed as

\[ \mu_{3j} = \left( \frac{k_j}{1 + m_j e^{b_{j1} t_{1j} + b_{j2} t_{2j} + \ldots + b_{jR} t_{T_Rj}}}, \ldots, \frac{k_j}{1 + m_j e^{b_{j1} t_{1j} + b_{j2} t_{2j} + \ldots + b_{jR} t_{T_Rj}}} \right), \]

where \(k_j\) is the asymptotic value, \(m_j\) is the initial growth and \(b_{j1}, b_{j2}, \ldots, b_{jR}\) are the \(R\)-order polynomial coefficients that describe the growth rate of plant biomass.

(iv) **The reproductive growth model.** Reproductive organs emerge from vegetative tissues. Thus, the growth of flowers relies on vegetative growth, which can be described by an allometric equation [31]. Considering the point in ontogeny where flower growth begins relative to vegetative growth, the allometric equation for joint genotype \(j\) is revised as

\[ \mu_{4j} = \left( a_j (\mu_{3j}(t_{4j}) - c_j)^{r_j}, \ldots, a_j (\mu_{3j}(t_{T_4j}) - c_j)^{r_j} \right), \]

where \(\mu_{3j}(t_{4j}), \ldots, \mu_{3j}(t_{T_4j})\) are the vegetative growth at time points \(t_{4j}, \ldots, t_{T_4j}\), respectively. \(aj, cj, rj\) is a constant, \(rj\) is the scaling exponent and \(c_j\) is a parameter that describes the timing of flowering.

Growth Equations (8–11) establish the biological foundation of functional mapping by which one can estimate the temporal pattern of biomass growth at different stages of plant development. There is also a statistical strength of functional mapping which lies in the parsimonious and flexible modeling of the covariance structure \(\Sigma\) [32, 33].

**Estimation and test**

The unknown parameters \((\Theta)\) contained in the likelihood (6) include three parts:

(i) Genotype-specific curve parameters \((\Theta_1)\) that specify growth trajectories of biomass at different stages of plant development, i.e. \((\alpha_j, \beta_j, \gamma_j)\) for seed formation, \((A_j, B_j, C_j, D_j)\) for seed germination, \((k_j, m_j, b_{j1}, b_{j2}, \ldots, b_{jR})\) for vegetative growth and \((a_j, c_j, r_j)\) for reproductive growth.

(ii) Covariate effect parameters \((\Theta_2)\) that describe the contributions of covariates to phenotypic values.

(iii) The structural parameters \((\Theta_3)\) that model the covariance matrix. The first autoregressive [AR(1)] model assuming stationary variance and covariance is one of the most popular approaches for modeling the covariance in functional mapping [9] because of its elegant mathematical properties that facilitate computational implementation. A transform–both–sides (TBS) model has been incorporated to mitigate the heteroscedastic problem of the residual variance, making the AR(1) model more useful.
stationarity assumption may still be violated even after the TBS transformation. In this case, we can use nonstationary covariance models. One representative model is the speckled antedependence (SAD) approach, which can specify time-varying variance and correlation in the analysis of longitudinal traits [34]. The SAD model of different orders has been used in functional mapping [35]. In Yap et al. [32, 33], several general approaches are discussed for modeling the covariance by parametric or nonparametric approaches.

Because of nonlinear complexities of Equations (811), no analytical estimators can be derived for the parameters underlying the equations. We will implement a hybrid of the Nelder–Mead Simplex and least-squares methods to obtain the maximum likelihood estimates (MLEs) of the unknown parameters in the likelihood (6). The Simplex method searches for the optimal estimates of curve parameters (θ1) and covariance-structuring parameters (θ3), while the least-squares method provides an analytic solution for the covariate effects (θ2). This integrative algorithm has been packed into computer software CycFunMap (which will be online available right after this manuscript is accepted).

To test whether the gene considered is significant, we calculate the ratio of the log-likelihoods under the following hypotheses:

\[ H_0: (\alpha_j, \beta_j, \gamma_j) \equiv (\alpha, \beta, \gamma); \quad (A_j, B_j, C_j, D_j) \equiv (A, B, C, D); \quad (k_j, m_j, b_{1j}, \ldots, b_{Rj}) \equiv (k, m, b_1, \ldots, b_R); \quad (a_j, c_j, r_j) \equiv (a, c, r) \quad (j = 1, \ldots, S) \]

\[ H_1: \text{Not all equalities in the } H_0 \text{ hold.} \]

Although the log-likelihood ratio (LR) calculated from the above hypotheses is thought of being asymptotically \( \chi^2 \)-distributed, the critical threshold can be determined using an alternative way based on permutation tests. This alternative does not rely on the distribution of the LR under null hypothesis and can be used to determine the threshold for any data type and sample size.

**DEVELOPMENTAL PATTERN OF GENETIC ACTION**

**Genome-genome interactions**

Seed development in flowering plants is triggered by a double-fertilization process that leads to the differentiation of the embryo, endosperm and seed coat that are the major regions of the seed and essential for seed viability and plant reproduction. Many different developmental and physiological events occur within each seed region during development, which are programmed, in part, by the activity of different genes. Seed development, therefore, is the result of a mosaic of different gene expression programs occurring in parallel in different seed compartments. Functional mapping can integrate genes from different regions and models how they are organized into unique regulatory circuits within the plant genome to ‘make a seed.’

Consider the gene A that forms five possible joint genotypes across the maternal, embryo and endosperm tissues (Table 1). After curve parameters for Equations (8–11) are estimated, we can test how this gene triggers its effect through genome–genome interactions on different parts of plant growth. For seed biomass growth, we use the estimated curve parameters \((a_j, b_j, c_j)\) to calculate time-varying genotypic means \(\mu_{ij}\), which are applied to estimate the temporal pattern of various genetic effects based on Equations (1–4). These time-dependent effects include the additive effect \(a_{1(t_1)}\), maternal–embryo interaction effect \(d_{11(t_1)}\), maternal–endosperm interaction effect \(d_{12(t_1)}\) and embryo–endosperm interaction effect \(d_{13(t_1)}\), expressed as

\[ a_{1(t_1)} = \frac{1}{6} [\mu_1(t_1) - \mu_2(t_1)] \]

\[ d_{11(t_1)} = \frac{1}{3} [2\mu_1(t_1) - 9\mu_2(t_1)] + 9\mu_3(t_1) - 3\mu_4(t_1) + \mu_5(t_1)] \]

\[ d_{12(t_1)} = \frac{1}{3} [-\mu_1(t_1) + 6\mu_2(t_1) - 6\mu_3(t_1)] + \mu_5(t_1)] \]

\[ d_{13(t_1)} = \frac{1}{3} [\mu_1(t_1) - \mu_2(t_1) - \mu_3(t_1) + \mu_4(t_1)] \]

A series of hypotheses tests can be generated to investigate whether each of these effects is significant on the growth trajectory of seed biomass, with null hypotheses expressed as \( H_{01} : a_{1(t_1)} = 0 \), \( H_{02} : d_{11(t_1)} = 0 \), \( H_{03} : d_{12(t_1)} = 0 \) and \( H_{04} : d_{13(t_1)} = 0 \), respectively. The LR for each of these hypotheses can be calculated and compared with a \( \chi^2 \)-distribution critical value.
Using a similar procedure, we can employ curve parameters \((A_j, B_j, C_j, D_j)\) to test whether and how the gene controls biomass growth at the cotyledon stage. Each type of time-varying genetic effects, expressed as \(a_2(t_{2\ell}), d_{21}(t_{2\ell}), d_{22}(t_{2\ell})\) and \(d_{23}(t_{2\ell})\), is estimated using equations analogous to Equations (12–15). Also, we use curve parameters \((k_j, m_j, b_{j1}, \ldots, b_{jR})\) and \((a_j, c_j, r_j)\) to test whether and how the gene controls biomass growth at the vegetative and reproductive stages by estimating genetic effects \(a_3(t_{3\ell}), d_{31}(t_{3\ell}), d_{32}(t_{3\ell})\) and \(d_{33}(t_{3\ell})\) and \(a_4(t_{4\ell}), d_{41}(t_{4\ell}), d_{42}(t_{4\ell})\) and \(d_{43}(t_{4\ell})\), respectively. The above hypotheses enable geneticists to address fundamental genetic questions, e.g. how do genes from different genomes interact and coordinate in a cohesive way to affect plant growth and development?

### Pleiotropic control

The lifecycle of a plant involves a sequence of developmental events; each of these events may be controlled by the same or different sets of genes. Given the importance of pleiotropic control in plant development, we formulate a design to test whether a pleiotropic gene affects seed growth in a growing plant, seedling growth at the cotyledon stage and plant growth at the post-cotyledon stage. The null hypothesis for so doing is expressed as

\[
H_0 : (\alpha_j, \beta_j, \gamma_j) \equiv (\alpha, \beta, \gamma), \quad j = 1, \ldots, 5
\]  

(16)

\[
H_0 : (A_j, B_j, C_j, D_j) \equiv (A, B, C, D), \quad j = 1, \ldots, 5
\]  

(17)

\[
H_0 : (k_j, m_j, b_{j1}, \ldots, b_{jR}) (k, m, b_1, \ldots, b_R), \quad j = 1, \ldots, 5
\]  

(18)

\[
H_0 : (a_j, c_j, r_j) \equiv (a, c, r), \quad j = 1, \ldots, 5.
\]  

(19)

If any two of the above null hypotheses are rejected at the same time, it suggests that this gene triggers a pleiotropic effect on different stages of plant development. Thus, this testing procedure allows the identification of pleiotropic control for any pair of developmental processes. The LR for each of these hypotheses is calculated, which is asymptotically \(\chi^2\)-distributed.

### Genetic control over developmental transitions

As mentioned above, plant development includes several sequential distinct phases, seed development from a complete embryo, vegetative growth after the germination of the seed, flower growth, pollination and seed formation. Since development within each phase is a quantitative change over time, mathematical models can well be used to describe the pattern of development. However, the connection of any two adjacent phases through mathematical models is challenging given that they are qualitatively different from each other. Statistical models for changing points [36] can be incorporated into functional mapping, which can detect specific genes that regulate the transition of development from one phase to next.

### MODEL DEMONSTRATION

We test and validate the model of functional mapping to estimate the temporal pattern of genetic control at different stages of plant development. A gene of interest was simulated from a plant genome that controls seed development and plant growth. We assume that there are five time points for the seed formation stage, six time points for the seed germination stage and 10 time points for the vegetative growth stage, all equally spaced. To avoid computing burden, we did not include reproductive growth in the simulation. Parameter values for each curve at different developmental stages were chosen for maize [16] for a hypothesized backcross design. To demonstrate the model’s power and robustness, we simulated the data with different levels of noise (reflected by a heritability value; see below) and different sample sizes, 200, 400 and 600, which are allocated to five groups of joint maternal–embryo–endosperm genotypes (Table 1).

Time-varying phenotypic values were simulated as the summation of individual genotypic means and normally distributed residual errors. Since this study aims to provide a new idea for genetic mapping during the entire ontogeny, we assumed a simple AR(1) structure (specified two parameters, variance \(\sigma^2\) and correlation \(\rho\)) for the covariance matrix. More complex patterns of the covariance matrix with structure or without structure have been studied in the previous literature [32, 33, 35] which can be integrated into the current model with no technical difficulty. To facilitate computing, we assume no correlation between different stages. The values of the variance were determined by scaling the heritabilities \(H^2 = 0.05\) and \(0.10\) at middle stages of each developmental process.

Table 2 gives the estimates of genotype-specific curve parameters and their standard errors from...
1000 simulation replicate when a sample size (200) and small heritability ($H^2 = 0.05$) were used. In general, this sample size can provide reasonably good estimates of the parameters even under a small heritability. When the sample size is doubled or tripled and/or heritability increases, the precision of parameter estimates increases strikingly (results not shown).

Using the simulation scenario of Table 2, we illustrate the estimated and true curves for plant development at three different stages (Fig. 4), demonstrating that the model provides reasonably accurate estimates of response curves. The estimates of four time-varying genetic effects in three stages estimated from Equations (12–15) are shown in Figure 5, in comparison with the true curves.

Additional simulation studies were performed to compare the power for functional mapping with different sample sizes. We focused on power analysis for detecting genome–genome interactions on seed development and a pleiotropic additive effect on plant growth trajectories at cotyledon and post-cotyledon stages. It was found that there is reasonably high power to detect effects of maternal–embryo, maternal–endosperm and embryo–endosperm interactions on seed development when both sample size and heritability are modest. Also, the power for detecting pleiotropic control for two different stages of plant development is high, increasing dramatically when the sample size and/or heritability increase. By simulating and analyzing phenotypic data based on a

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Joint-maternal-embryo-endosperm genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Seed development $\alpha_j$</td>
<td>0.703 (0.032)</td>
</tr>
<tr>
<td>$\beta_j$</td>
<td>705.7 (226.9)</td>
</tr>
<tr>
<td>$\gamma_j$</td>
<td>2.182 (0.128)</td>
</tr>
<tr>
<td>Seed germination $A_j$</td>
<td>90.938 (10.193)</td>
</tr>
<tr>
<td>$B_j$</td>
<td>3.173 (1.691)</td>
</tr>
<tr>
<td>$C_j$</td>
<td>5.541 (0.092)</td>
</tr>
<tr>
<td>$D_j$</td>
<td>1.196 (0.163)</td>
</tr>
<tr>
<td>Seedling growth $k_j$</td>
<td>A12345</td>
</tr>
<tr>
<td>$b_{1j}$</td>
<td>0.019 (0.016)</td>
</tr>
<tr>
<td>$b_{2j}$</td>
<td>0.035 (0.021)</td>
</tr>
</tbody>
</table>

Figure 4: True (solid) and estimated (dot) curves for time-varying genotypic values for five joint genotypes (Table 1) in three different stages, seed development (A), seed germination (B) and seedling growth (C).
single curve of plant development, we examined the false positive rates (FPR) of the model. Even with a modest sample size (200) and heritability (0.05), the FPR is below 0.05.

FUTURE DIRECTIONS

Genetic mapping that attributes a phenotypic trait to its underlying QTLs has proven powerful for studying the genetic architecture of the trait [37]. However, traditional approaches for mapping QTLs with phenotypes separately for different times fail to capture the dynamic structure and pattern of the process, which greatly limits the scope of inference about its genetic architecture. To overcome this limitation, a new statistical method, called functional mapping, has been developed [8–14]. Functional mapping integrates biological mechanisms and processes of the trait into the genetic mapping framework through mathematical and computational models. Functional mapping unifies the strengths of statistics, genetics and developmental biology, thus facilitating the test of the interplay between genetic action and development. To make functional mapping applicable in practice, a package of software has been developed [38, 39] and given at our webpage http://statgen.psu.edu/software.

In this article, we argue that functional mapping can be used to map genes that govern different stages of development in plants. Any organisms including plants undergo several distinct stages during ontogeny, each characterized by different types of developmental events and landmarks [16, 17]. Transition from the vegetative to reproductive phase is distinguished by the production of flowers, each of which contains a series of continuous or discrete alternations forming multiple sub-phases. Comparing the patterns of genetic control for different developmental stages can not only help us understand the genetic regulation and mechanisms underlying each stage, but also provide a way to draw a complete view of developmental processes in an organism’s lifetime [40–42]. Such knowledge is essential for understanding the evolution of plant development, explaining why plants are so successful in adapting themselves to changing environments, and selecting elite genotypes of economic and social value in agriculture and forestry.

Because of its significance, geneticists have developed molecular tools to identify the machineries of phase transitions that are programmed into the entire process of ontogeny. Several transcriptional factors targeted by microRNAs, such as SQUAMOSA PROMOTER BINDING-LIKE (SPL), SERRATE and EARLY FLOWERING IN SHORT DAYS, have been found to affect the transitions from a dormant embryonic phase to a germinating seedling and from the adult vegetative to reproductive phases [43–48]. Functional mapping, complementing molecular genetic tools, can be integrated with these molecular discoveries, facilitating the quantitative prediction of developmental pattern and process. Although the statistical behavior of functional mapping has been studied extensively in previous work, a simulation study was performed here to investigate how this approach can be used to detect genes for the entire process of plant development. The results from the simulation validated the usefulness and utilization of...
functional mapping to address several fundamental developmental questions, such as how genome–
genome interactions control seed development in a growing plant, whether there is a pleiotropic mechanism for different stages of development, and how genes control the transition of development from one stage to next.

This is a first modeling framework that can dictate a comprehensive picture of the genetic control and regulation for dynamic development in plants. There is a plenty of room for its improvement. First, He et al. [13] recognized that the integration of functional mapping with developmental ontology could gain new insight into the genetic mechanisms for plant structure, development and evolution. He et al.’s viewpoint can be realized if a detailed sequence of vegetative and reproductive growth processes is integrated into functional mapping. Second, we used an annual plant as an example to demonstrate the model. The same principle can be extended to study the genetic control of ontogeny in long-lived trees, in which complex vegetative phase changes also occur from year to year [41, 48]. Third, we outlined a testing procedure for the genetic control of plant development by assuming one single gene. This simplified assumption helps to describe and define the interaction effects among genes from different genomes. However, it is not realistic given the genetic complexity of development. It is crucial to incorporate multiple genes and their interactions within and between genomes.

With recent rapid development of network biology, there is a good reason to integrate ‘omics’ information into our functional mapping to explore the regulatory mechanisms of phenotype formation [49]. Taken together, when genome-wide association studies are coming of age in plants, we are in an excellent position to integrate genetics and development through functional mapping.

**FUNDING**

This research is supported by the Changjiang Scholars Award and ‘One-thousand Person’ Award.

**References**