A Model Framework for Identifying Genes that Guide the Evolution of Heterochrony

Lidan Sun,†1 Meixia Ye,†2 Han Hao,‡3 Ningtao Wang,§3 Yaqun Wang,§3 Tangren Cheng,∥1 Qixiang Zhang,*1 and Rongling Wu,*2,3

1Beijing Key Laboratory of Ornamental Germplasm Innovation and Molecular Breeding, National Engineering Research Center for Floriculture, College of Landscape Architecture, Beijing Forestry University, Beijing, China
2Center for Computational Biology, College of Biological Sciences and Technology, Beijing Forestry University, Beijing, China
3Center for Statistical Genetics, The Pennsylvania State University

†These authors contributed equally to this work.
*Corresponding author: E-mail: rwu@phs.psu.edu, zqx@bjfu.edu.cn.

Associate editor: Shizhong Xu

Abstract

Heterochrony, the phylogenetic change in the time of developmental events or rate of development, has been thought to play an important role in producing phenotypic novelty during evolution. Increasing evidence suggests that specific genes are implicated in heterochrony, guiding the process of developmental divergence, but no quantitative models have been instrumented to map such heterochrony genes. Here, we present a computational framework for genetic mapping by which to characterize and locate quantitative trait loci (QTLs) that govern heterochrony described by four parameters, the timing of the inflection point, the timing of maximum acceleration of growth, the timing of maximum deceleration of growth, and the length of linear growth. The framework was developed from functional mapping, a dynamic model derived to map QTLs for the overall process and pattern of development. By integrating an optimality algorithm, the framework allows the so-called heterochrony QTLs (hQTLs) to be tested and quantified. Specific pipelines are given for testing how hQTLs control the onset and offset of developmental events, the rate of development, and duration of a particular developmental stage. Computer simulation was performed to examine the statistical properties of the model and demonstrate its utility to characterize the effect of hQTLs on population diversification due to heterochrony. By analyzing a genetic mapping data in rice, the framework identified an hQTL that controls the timing of maximum growth rate and duration of linear growth stage in plant height growth. The framework provides a tool to study how genetic variation translates into phenotypic innovation, leading a lineage to evolve, through heterochrony.

Key words: developmental timing, growth rate, heterochrony, functional mapping, quantitative trait loci, phylogeny.

Background

Changes in developmental timing and rate, named as heterochrony, have long been believed to be a major force in the evolution of phenotype (Gould 1977; Wilson et al. 1987; Magwene 2001; Drake 2011). It has been observed that relatively few genetic changes in heterochrony through the endocrine regulation of metamorphosis can cause profound morphological consequences (Moss 2007). For example, although humans and chimpanzees are closely related, their skull shape and brain growth are different dramatically during early development (Rice 2002; King 2004; Mitteroecker et al. 2004). A recent phylogenetic investigation revealed that changes in developmental timing are a crucial step for birds to evolve from dinosaurs (Bhullar et al. 2012). The consequence of these changes leads birds to take months to reach sexual maturity, allowing them to retain the physical characteristics of baby stages characterized by dinosaurs that take years to mature. One question that naturally arises from these evolutionary divergences is what mechanisms are implicated for heterochrony and the change of biological clock.

Many studies have pursued to identify the molecular control of developmental timing. Using Caenorhabditis elegans as an example, some of these studies have identified heterochronic genes that orchestrate the timing of cell divisions and fates during development regulated by microRNAs and their targets (Ambros 2000; Rougvie 2001; Pasquinelli and Ruvkun 2002; Banerjee and Slack 2002; Moss 2007). By comparing gene expression profiles of Saccharomyces cerevisiae, S. paradoxus, and their sterile hybrid at multiple meiotic stages during sporulation, Lenz et al. (2014) found that although there are no differences in expression levels across meiosis between the hybrid and parents, the hybrid meiotic program occurs earlier than either parent. Such heterochronic changes in the meiotic expression program are thought to explain regulatory differences between species and misexpression in interspecific hybrids. Focusing on particular pathways causing heterochronic changes, none of these studies has provided an
Mapping Genes for the Evolution of Heterochrony  ·  doi:10.1093/molbev/msu156

Ontogenetic Trajectories

The ontogenetic trajectory of a particular organism is the path it takes through a “developmental space” that describes changes in form. Such trajectories differ between taxonomic groups, thereby used as a systematic parameter. Differences in ontogenetic trajectories among related taxa are composed of those in various heterochronic processes (Alberch et al. 1979; McKinney and McNamara 1991; Rice 1997). However, the issues of how heterochronic phenomena can be quantified and how they contribute to trajectories differences remain unsolved (Reilly et al. 1997; Magwene 2001). Mathematical equations have been widely used to model ontogenetic trajectories of many traits, such as plant height and flower size, from which heterochronic phenomena can be discerned in a quantitative way.

It has been recognized that the growth of organisms at any organizational level from cell to organ to organism should be S-shaped according to fundamental principles of biophysics and biochemistry (West et al. 2001). One mathematical form of the S-shaped curve is expressed as

\[ g(t) = a(1 + b e^{-rt})^{1/k} \]  

where \( g(t) \) is the trait value at time \( t \), \( a \) is the asymptotic value of the trait, \( b \) is a parameter to position the curve on the time axis, \( r \) is the growth rate that determines the spread of the curve along the time axis, and \( k \) is the shape parameter of the curve. Thus, overall and specific characteristics of growth can be captured by estimating the set of parameters \( (a, b, r, k) \).

The growth equation (1) has several features that can fully describe the pattern and form of developmental process. There are three physiologically important points, with coordinates denoted as \( P_a, P_t \), and \( P_d \), respectively, on the growth curve. The point \( P_d \) is the inflection point at which growth rate reaches its maximum. By calculating the second derivative of the growth equation, we obtain the coordinates of \( P_t \) as

\[ (t,g(t)) = \left( \frac{1}{r} \ln \left[ \frac{b}{k-1} \right], ak^{1/r} \right) \]  

Because of \( P_t \), the growth curve is divided into two phases, the exponential growth (from time \( t = 0 \) to \( P_t \)) and the asymptotic growth (from \( P_t \) to the infinite time).

The points \( P_a \) and \( P_d \) present the timing of maximum acceleration and maximum deceleration of growth, which are the first and second inflection points of growth rate curve, respectively. These two points partition the growth curve into three phases, the exponential growth (from time \( t = 0 \) to \( P_a \)), the linear growth (from \( P_a \) to \( P_d \)), and the ageing (from \( P_d \) to the infinite). By calculating the third derivatives of the growth equation (1) with respect to time, the coordinates of \( P_a \) and \( P_d \) can be obtained as, respectively,

\[ (t_a,g_a) = \left( t + \frac{1}{r} \ln \left[ \frac{(k-1)+1}{1-\eta_1} \right], \eta_1 \right) \]  

\[ (t_d,g_d) = \left( t + \frac{1}{r} \ln \left[ \frac{(k-1)+1}{1-\eta_2} \right], \eta_2 \right) \]  

where

\[ \eta_1 = \left[ \frac{(k+1) - (k-1) \sqrt{k(k+4)}}{2(k-1)} \right]^{\frac{1}{r}} \]  

\[ \eta_2 = \left[ \frac{(k+1) + (k-1) \sqrt{k(k+4)}}{2(k-1)} \right]^{\frac{1}{r}} \]

The duration of linear growth can be calculated as

\[ T = t_d - t_a. \]

This parameter, along with the timing of the inflection points in equations (2–4), can help to better understand the shape and process of trait growth.

Functional Mapping

Statistical Model

Consider \( n \) phylogenetically related taxa whose genomes have shared synteny, that is, the co-localization of genes within a chromosomal region, derived from the same ancestor. Suppose there is a mapping population of size \( n_l \) generated for each taxon \( l \). The same set of molecular markers is genotyped throughout the genomes, from which a genetic linkage map that covers the entire genome has been constructed for each taxon. All individuals from each mapping
population are phenotyped for the same trait at a series of $P$ time points for $m$ taxa. Assume that all taxa are measured by the same schedule, that is, at the same time point, with the same time interval, although the model allows the measurement schedule to vary among taxa.

If specific QTLs exist to affect the dynamic change of the trait for all taxa, the growth parameters $\Theta$ that specify the change should be different among QTL genotypes. Genetic mapping uses a mixture model-based likelihood to estimate QTL genotype-specific parameters (Lander and Botstein 1989). For taxon $l$, such a likelihood is expressed as

$$L_l(\Phi_l | y_l) = \prod_{i=1}^{m} \left[ \omega_{jl}(y_l) + \cdots + \omega_{jl}(y_l) \right]$$

(6)

where $\Phi_l$ is the unknown parameters including the QTL position, time-dependent effects, and time-dependent residual variances and correlations; $y_l = (y_l(1), \ldots, y_l(P))$ is the phenotypic vector of individual $l$ from taxon $l$ measured at $P$ time points; $\omega_{jl}$ is the conditional probability of QTL genotype $j$ ($j = 1, \ldots, J$) given the marker genotype of individual $l$ (assuming that all taxa have constructed a syntenic linkage map); $f_{jl}(y_l)$ is a multivariate normal distribution with expected mean vector for genotype $j$ from taxon $l$,

$$\mu_j = [\mu_j(1), \ldots, \mu_j(P)]$$

(7)

and $(P \times P)$-dimensional longitudinal covariance matrix $\Sigma_j$ expressed as

$$\Sigma_j = \begin{pmatrix} \sigma^2_1 & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma^2_p \end{pmatrix}$$

(8)

Assuming that phenotypic measurements are independent among taxa, a joint likelihood that combines all mapping populations can be formulated as

$$L(y) = \prod_{l=1}^{m} L_l(y_l),$$

(9)

from which taxa-dependent parameters are estimated according to the following procedure.

The likelihood (eq. 6) contains the conditional probabilities of QTL genotypes given the genotypes of two flanking markers that bracket the QTL in an experimental cross population (Wu et al. 2007). Let $r_v$, $r_p$, and $r_M$ denote the recombination fractions between the left marker and the QTL, between the QTL and the right marker, and between the two markers, respectively. Such conditional probabilities are derived as a function of $r_v$, $r_p$, and $r_M$ (see Wu et al. 2007; Xu 2012 for the detail of the derivation). By maximizing the joint likelihood (eq. 9), a grid approach was used to search for the position of the QTL throughout the genome.

Rather than estimating all elements in the mean vector (eq. 7) and covariance matrix (eq. 8), functional mapping maximizes the likelihood (eq. 9) by incorporating a mathematical equation to model the time-dependent mean values and a stochastic process to model the autocorrelation structure of the longitudinal matrix (Ma et al. 2002). The two tasks can not only enhance the biological relevance of functional mapping but also increase its parsimony and statistical precision. For the growth trait, the mean vector (eq. 7) can be modeled by growth equation (1), that is,

$$\mu_j = [\mu_j(1), \ldots, \mu_j(P)]$$

$$= [a_j(1 + b_j e^{-r_j})^{-1/2}, \ldots, a_j(1 + b_j e^{-r_j})^{-1/2}].$$

Each QTL genotype $j$ has a set of growth parameters $(a_j, b_j, r_j, k_j)$ for taxon $l$. The covariance matrix (eq. 8) follows an autoregressive structure which can be modeled by many statistical processes, such as parametric Brownian motion, autoregressive, structured antedependence and autoregressive moving average, and nonparametric and semiparametric (see Zhao et al. 2005; Yap et al. 2009 for the discussion about covariance structure). The EM algorithm implemented with optimality techniques have been developed to estimate the parameters that model the mean vector and the parameters that model the covariance structure (Zhao et al. 2004). The existence of significant QTLs for growth trajectories can be tested using a general procedure outlined in Ma et al. (2002).

Hypothesis Tests

Functional mapping can not only test how the QTL affects the pattern and process of growth but also test and estimate the effect of the QTLs detected on heterochrony. As mentioned above, the heterochrony of growth can be described by several parameters as follows: 1) the timing of inflection point, $t_v$; 2) the timing of maximum acceleration of growth, $t_w$; 3) the timing of maximum deceleration of growth, $t_d$; 4) growth rate, $r$, and 5) the duration of linear growth, $\Delta T$ (Rice 2002; Smith 2008). How QTLs affects these heterochronic parameters can be tested in the following aspects.

The QTLs Affect a Heterochronic Parameter within Each Taxon

For example, for a specific taxon, $l$ ($l = 1, \ldots, m$), to test whether a QTL segregating in a mapping population (e.g., backcross, doubled haploids, or recombinant inbred lines) affects the inflection point, we can formulate the null hypothesis as follows:

$$H_0 : \frac{1}{r_{1l}} \ln \frac{b_{1l}}{k_{1l} - 1} = \frac{1}{r_{2l}} \ln \frac{b_{2l}}{k_{2l} - 1} \sqrt{a^2 + b^2}$$

(10)

$$H_0 : a_j k_{jy}^{-1} = a_j k_{jy}^{-1}$$

(11)

where the backcross has two genotypes 1 and 2 at a QTL considered. The rejection of the null hypotheses (eq. 10) and (eq. 11) indicates that the QTL has a significant effect on the timing and growth at the inflection point, respectively. The hypotheses about the genetic control of the other heterochronic parameters can be tested in a similar way.
The QTL Interacts with Taxon to Determine a Heterochronic Parameter

Assuming the inflection point, for a pair of taxa, $l_1$ and $l_2$ ($l_1 < l_2 = 1, \ldots, m$), QTL x taxon interaction can be tested by the null hypothesis

$$H_0 : \ln \left[ \frac{b_{11}^l}{k_{11}^l - 1} \right] = \ln \left[ \frac{b_{21}^l}{k_{21}^l - 1} \right] = \ln \left[ \frac{b_{12}^l}{k_{12}^l - 1} \right] = \ln \left[ \frac{b_{22}^l}{k_{22}^l - 1} \right]$$

(12)

The rejection of equation (12) indicates that the genetic effect of the QTL on heterochrony is taxon dependent. The reformation of hypothesis (eq. 12) leads to a new expression as

$$H_0 : \ln \left[ \frac{b_{11}^l}{k_{11}^l - 1} \right] = \ln \left[ \frac{b_{12}^l}{k_{12}^l - 1} \right] = \ln \left[ \frac{b_{21}^l}{k_{21}^l - 1} \right] = \ln \left[ \frac{b_{22}^l}{k_{22}^l - 1} \right]$$

(13)

It is interesting to see that each side of equation (13) is the difference of the inflection point between the same genotype of two different taxa, reflecting an evolutionary change of gene expression across taxa. Thus, the test based on hypothesis (eq. 12) is equivalent to testing how a QTL detected drives the evolution of development.

The QTL Governs the Evolution of Multiple Developmental Features through Pleiotropy

Hypotheses (eq. 11) and (eq. 12) can be extended to test the pleiotropic effect of the same QTL on two or more than two heterochronic parameters. For a specific taxon, $l$, whether the QTL pleiotropically affects the timing of the inflection point and the duration of linear growth can be tested by the null hypotheses,

$$H_0 : \ln \left[ \frac{b_{11}^l}{k_{11}^l - 1} \right] = \ln \left[ \frac{b_{12}^l}{k_{12}^l - 1} \right] = \ln \left[ \frac{b_{21}^l}{k_{21}^l - 1} \right] = \ln \left[ \frac{b_{22}^l}{k_{22}^l - 1} \right]$$

(14)

$$H_0 : \Delta T_1 = \Delta T_2.$$  

(15)

If both null hypotheses are rejected, this suggests that the QTL has a pleiotropic effect on the two heterochronic parameters. The question of how pleiotropy governs the evolution of development is of fundamental interest to evolutionary biologists. This can be addressed by functional mapping. For example, the pleiotropy of the QTL on the timing of the inflection point and the duration of linear growth can be tested by formulating a pair of null hypotheses, expressed as

$$H_0 : \ln \left[ \frac{b_{11}^l}{k_{11}^l - 1} \right] = \ln \left[ \frac{b_{21}^l}{k_{21}^l - 1} \right] = \ln \left[ \frac{b_{12}^l}{k_{12}^l - 1} \right] = \ln \left[ \frac{b_{22}^l}{k_{22}^l - 1} \right]$$

(16)

$$H_0 : H_0 : T_{11}^l - T_{12}^l = T_{21}^l - T_{22}^l.$$  

(17)

for two taxa $l_1$ and $l_2$. The rejection of both null hypotheses indicates the significant pleiotropy of the QTL on the evolution of development.

Note that all hypothesis tests like equations (10–17) are made on the basis of a log-likelihood ratio (LR) derived from the likelihoods under the null and alternative hypotheses. The LR values estimated does not violate regularity assumption, rather than can be viewed to follow a $\chi^2$ distribution, because all these tests are based on the position of each significant QTL estimated from functional mapping, in which case the QTL position is fixed under both null and alternative hypotheses.

Computer Simulation

The utility and usefulness of our hQTL mapping model were assessed through computer simulation. We simulated multiple taxa to compare differences in the genetic control of hQTL on heterochrony, but for the simplicity of presentation, we reported results from two taxa. In each taxon, a doubled haploid population is simulated, in which a QTL affects growth trajectories. The QTL was assumed to be at 37 cM from the first marker on a linkage group of length 200 cM evenly spaced by 20 markers. Two sets of growth parameters ($a_1, b_1, r_1, k_1$) and ($a_2, b_2, r_2, k_2$) for each QTL genotype, QQ or qq, were chosen from a space of parameters observed in an annual plant height grows (Zhao et al. 2004). The residual covariance matrix for each population was assumed to follow a first-order autoregressive (AR(1)) structure by two parameters, variance ($\sigma^2$) and correlation ($\rho$). The phenotypic values of growth trajectories were simulated by the summing of genotypic values of the QTL plus residual errors following a multivariate normal distribution with mean vector 0 and (co)variances, depending on the size of heritability explained by the QTL. Different heritabilities, $H^2 = 0.05, 0.10$, and $0.20$, were used. The two mapping populations assumed the same sample size, which is $n = 200, 400$, and 600 for each population, but only the results from $n = 200$ and 400 were reported.

The hQTL mapping model was used to analyze the simulation data, with results given in table 1. The model can provide reasonably precise estimation of the QTL location, even with a small sample (200) and under a modest heritability (0.05). In general, growth parameters for each QTL genotype can be reasonably precisely estimated, although a high sample size (400) is recommended under a modest heritability or a low sample size (200) can be used unless the heritability is high (0.20). The heritability can increase through minimizing experimental errors and phenotyping errors.

The focus of this study is to test how an hQTL affects developmental processes and population diversification and evolution. Based on the given values of growth parameters, QTL genotype-specific growth trajectories for each taxon were drawn, illustrated in figures 1 and 2, in which the characteristics of four key heterochronic parameters, that is, the timing of the inflection point ($t_i$), the timing of maximum acceleration of growth ($t_a$), the timing of maximum deceleration of growth ($t_d$), and the length of linear growth ($\Delta T$), are shown. The estimated values of these parameters using equations (2–5) were tabulated in table 2. It can be seen that all these heterochronic parameters were very well estimated under different sample sizes and heritabilities. This result suggests that developmentally meaningful heterochronic parameters can be estimated as ordinary statistical variables within the framework of functional mapping.
Between the two taxa simulated, the effects of the hQTL on the heterochronic parameters vary, displaying larger values for \( t_B \) and \( t_c \) in taxon 2 (fig. 1B) than taxon 1 (fig. 1A), and taxon-dependent directions for \( \Delta T \). Different influences of the same hQTL on developmental processes of different taxa imply that this QTL is potential an evolutionary driver for population diversification. Figure 2 describes a taxon-specific diversification in growth trajectories for each QTL genotype. Overall, such a diversification is more pronounced for geno-

Table 1. MLEs of the Position of a Putative QTL and Growth Parameters that Define Organ Growth Trajectories for Different QTL Genotypes Segregating in Different Simulated Mapping Populations from Two Taxa under Different Sample Sizes and Heritabilities.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>True Value</th>
<th>( n = 200 )</th>
<th>( n = 400 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H^2 = 0.05 )</td>
<td>( H^2 = 0.1 )</td>
<td>( H^2 = 0.2 )</td>
<td>( H^2 = 0.05 )</td>
</tr>
<tr>
<td>QTL position</td>
<td>37</td>
<td>37.4 (2.386)</td>
<td>37.3 (1.579)</td>
</tr>
<tr>
<td>Taxon 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QQ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( a_1 )</td>
<td>9.8</td>
<td>9.799 (0.137)</td>
<td>9.801 (0.079)</td>
</tr>
<tr>
<td>( b_1 )</td>
<td>60</td>
<td>66.07 (38.87)</td>
<td>60.36 (23.39)</td>
</tr>
<tr>
<td>( r_1 )</td>
<td>0.62</td>
<td>0.618 (0.037)</td>
<td>0.617 (0.027)</td>
</tr>
<tr>
<td>( k_1 )</td>
<td>1.8</td>
<td>1.777 (0.158)</td>
<td>1.782 (0.111)</td>
</tr>
<tr>
<td>Covariance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \rho )</td>
<td>0.8</td>
<td>0.798 (0.011)</td>
<td>0.798 (0.012)</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>1.217 (0.069)</td>
<td>0.578 (0.033)</td>
<td>0.260 (0.038)</td>
</tr>
<tr>
<td>Taxon 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QQ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( a_1 )</td>
<td>10</td>
<td>9.972 (0.148)</td>
<td>9.930 (0.098)</td>
</tr>
<tr>
<td>( b_1 )</td>
<td>100</td>
<td>146.86 (81.78)</td>
<td>186.96 (63.69)</td>
</tr>
<tr>
<td>( r_1 )</td>
<td>0.65</td>
<td>0.674 (0.046)</td>
<td>0.700 (0.031)</td>
</tr>
<tr>
<td>( k_1 )</td>
<td>2.4</td>
<td>2.513 (0.215)</td>
<td>2.657 (0.138)</td>
</tr>
<tr>
<td>Covariance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \rho )</td>
<td>0.8</td>
<td>0.806 (0.011)</td>
<td>0.809 (0.012)</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>2.001 (0.110)</td>
<td>0.964 (0.060)</td>
<td>0.457 (0.034)</td>
</tr>
</tbody>
</table>

Note.—The standard errors of the estimates were given in parentheses from 200 simulation replicates.

Using hypothesis tests like equations (12) and (13), we can test how an hQTL interacts with taxon to determine development processes. We performed a simulation study to assess the power of detecting hQTL × taxon interactions for the timing of inflection point and linear growth length (table 4). In general, such power is very high, suggesting that the model has a great capacity to detect the genetic machinery of developmental divergence that act as an evolutionary force making one taxon different from the other.

**Worked Example**

The new model can be validated perfectly by analyzing an example of genetic mapping for multiple taxa. Unfortunately, we currently have no multiple mapping populations from...
different taxa. Although it is not ideal, we tested our model for mapping QTLs that control heterochronic parameters by using a single mapping population derived from two inbred lines in rice. A doubled-haploid (DH) population of 123 lines derived from semidwarf IR64 and tall Azucena (Huang et al. 1997) was used to construct a genetic linkage map with 135 RFLP and 40 isozyme and RAPD markers that cover 12 chromosomes. This map is 2,005-cM long with an average distance of 11.5 cM between a pair of adjacent markers. The DH population, analytically identical to a backcross population, has two genotypes at a QTL, QQ (coded as 1) and qq (coded as 2). The DH population was cloned and different clonal replicates of the same genotype were grown in a randomized complete design with two replicates. After 10 days of transplanting into the field trial, plant heights were measured every 10 days until all lines had headed (Yan et al. 1998).

Zhao et al. (2004) have used this population to map growth QTLs for plant height based on functional mapping. In this study, we focus on testing and estimating the effects of QTLs detected on heterochronic processes. By implementing growth equation (1) into functional mapping, we obtained the maximum likelihood estimates (MLEs) for the set of growth parameters, \((a_j, b_j, r_j, k_j)\), for different QTL genotypes \(j (j = 1, 2)\) using the EM algorithm. We used the AR(1) model to structure the covariance matrix. The likelihood ratio test was used to test the existence of any possible QTLs for plant height growth with the null hypothesis formulated as

\[
H_0 : \begin{pmatrix} a_1, b_1, r_1, k_1 \end{pmatrix} = \begin{pmatrix} a_2, b_2, r_2, k_2 \end{pmatrix}
\] (18)

To the end, three significant QTLs were detected on chromosome 7, 8, and 12 using testing procedure of Ma et al. (2002) As an example, we chose the QTL located between RG574 and RZ816 on chromosome 12 to further analyze how it affects the pattern and process of growth.

Using the MLEs of \((a_j, b_j, r_j, k_j)\), we drew genotype-dependent curves for plant height growth in rice and further calculated the coordinates of three key inflection points, \(P_1\) (point of maximum growth rate), \(P_A\) (point of maximum growth acceleration), and \(P_d\) (point of maximum growth deceleration) (fig. 3). We estimated the genetic effects of the QTL on these two coordinates (table 5) and further tested

**Fig. 1.** Genetic variation in developmental timing caused by a putative heterochronic QTL from simulated data. The QTL simulated exerts a larger effect on developmental timing in taxon 2 (B) than 1 (A), showing a role of gene in taxon diversification. Three landmarks \(P_a, P_I\) and \(P_d\) are the timing of maximum acceleration of growth, maximum rate of growth, and maximum deceleration of growth, respectively, and time interval \(T\) is the duration of linear growth, all parameters subscribed by the notation of QTL genotypes QQ (1) and qq (2). The effects of the QTL on the three landmarks and linear growth length are denoted as \(\Delta t_a, \Delta t_I, \Delta t_d\) and \(\Delta T\), respectively. Relative to taxon 1 (A), taxon 2 (B) is denoted by a prime.

**Fig. 2.** Taxon-dependent diversification in developmental timing for the same genotype at a putative heterochronic QTL from simulated data. Larger differences in developmental timing are observed for genotype qq (B) than QQ (A), suggesting that the former is a greater force causing taxa to diversify than the latter. Three landmarks \(P_a, P_I, P_d\) are the timing of maximum acceleration of growth, maximum rate of growth, and maximum deceleration of growth, respectively, and time interval \(T\) is the length of linear growth, all parameters subscribed by the notation of QTL genotypes QQ (1) and qq (2). Relative to taxon 1, taxon 2 is denoted by a prime. Taxon-dependent differences in the three landmarks and linear growth length are denoted as \(\Delta t_a, \Delta t_I, \Delta t_d\) and \(\Delta T\) for genotype QQ (A), and \(\Delta t_{aB}, \Delta t_{IB}, \Delta t_{dB}\) and \(\Delta T_B\) for genotype qq (B), respectively.
Two Taxa under Different Sample Sizes and Heritabilities.

Table 2. MLEs of Four Heterochronic Parameters at Each of Two QTL Genotypes Segregating in Different Simulated Mapping Populations from Two Taxa under Different Sample Sizes and Heritabilities.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>True Value</th>
<th>n = 200</th>
<th>n = 400</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H^2 = 0.05</td>
<td>H^2 = 0.1</td>
<td>H^2 = 0.2</td>
</tr>
<tr>
<td>Taxon 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QQ</td>
<td>t_{11}</td>
<td>6.964</td>
<td>6.957 (0.125)</td>
</tr>
<tr>
<td></td>
<td>t_{x1}</td>
<td>4.936</td>
<td>4.934 (0.153)</td>
</tr>
<tr>
<td></td>
<td>t_{d1}</td>
<td>8.991</td>
<td>8.818 (0.132)</td>
</tr>
<tr>
<td></td>
<td>T_1</td>
<td>4.055</td>
<td>4.048 (0.138)</td>
</tr>
<tr>
<td>QQ</td>
<td>t_{12}</td>
<td>6.584</td>
<td>6.586 (0.145)</td>
</tr>
<tr>
<td></td>
<td>t_{x2}</td>
<td>4.504</td>
<td>4.507 (0.189)</td>
</tr>
<tr>
<td></td>
<td>t_{d2}</td>
<td>8.664</td>
<td>8.664 (0.146)</td>
</tr>
<tr>
<td></td>
<td>T_2</td>
<td>4.160</td>
<td>4.156 (0.172)</td>
</tr>
<tr>
<td>Taxon 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QQ</td>
<td>t_{11}</td>
<td>6.657</td>
<td>6.632 (0.133)</td>
</tr>
<tr>
<td></td>
<td>t_{x1}</td>
<td>4.374</td>
<td>4.495 (0.179)</td>
</tr>
<tr>
<td></td>
<td>t_{d1}</td>
<td>8.760</td>
<td>8.768 (0.141)</td>
</tr>
<tr>
<td></td>
<td>t_1</td>
<td>4.386</td>
<td>4.273 (0.182)</td>
</tr>
<tr>
<td>QQ</td>
<td>t_{12}</td>
<td>4.796</td>
<td>4.807 (0.135)</td>
</tr>
<tr>
<td></td>
<td>t_{x2}</td>
<td>2.702</td>
<td>2.718 (0.168)</td>
</tr>
<tr>
<td></td>
<td>t_{d2}</td>
<td>6.890</td>
<td>6.896 (0.166)</td>
</tr>
<tr>
<td></td>
<td>T_2</td>
<td>4.188</td>
<td>4.177 (0.198)</td>
</tr>
</tbody>
</table>

Note.—The standard errors of the estimates were given in parentheses from 200 simulation replicates.

Whether these effects are significant. Whether the QTL affects the timing and growth of P_1 was tested using the null hypotheses (eq. 10) and (eq. 11). We further test whether the QTL affects the timing and growth of P_2 based on the null hypotheses expressed as

\[ H_0 : t_{11} + \frac{1}{r_1} \ln \left( \frac{(k_1 - 1)\eta_{11}^{k_1 - 1}}{1 - \eta_{11}^{k_1 - 1}} \right) = t_{12} + \frac{1}{r_2} \ln \left( \frac{(k_2 - 1)\eta_{21}^{k_2 - 1}}{1 - \eta_{21}^{k_2 - 1}} \right) \]

(19)

\[ H_0 : a_1 \eta_{11} = a_2 \eta_{21} \]

(20)

where \( t_j = \frac{1}{\eta_j} \ln \left( \frac{b_j}{k_j - 1} \right) \), \( \eta_j = \left( \frac{k_j(k_j + 1) + (k_j - 1)\sqrt{k_j(k_j + 1)}}{2k_j(2k_j - 1)} \right)^{1/2} \), \( j = 1 \) (QQ), 2 (qq).

Similarly, whether the QTL affects the timing and growth of P_3 can be tested using the null hypotheses:

\[ H_0 : t_{11} + \frac{1}{r_1} \ln \left( \frac{(k_1 - 1)\eta_{12}^{k_1 - 1}}{1 - \eta_{12}^{k_1 - 1}} \right) = t_{12} + \frac{1}{r_2} \ln \left( \frac{(k_2 - 1)\eta_{22}^{k_2 - 1}}{1 - \eta_{22}^{k_2 - 1}} \right) \]

(21)

\[ H_0 : a_1 \eta_{12} = a_2 \eta_{22} \]

(22)

where \( \eta_{j2} = \left( \frac{k_j(k_j + 1) + (k_j - 1)\sqrt{k_j(k_j + 1)}}{2k_j(2k_j - 1)} \right)^{1/2} \), \( j = 1 \) (QQ), 2 (qq).

By calculating the LRs for each case and comparing them against the \( \chi^2 \) distribution, we find that the two QTL genotypes have marginally significant differences in the timing of the inflection point P_1 (table 5). The timing of P_2 is not significantly different between two genotypes, but there is a significant difference in the timing of P_3. The amount of growth at all these three inflection points involves a significant genetic effect.

We also calculated the duration of linear growth for each QTL genotype based on equation (5). Statistical test shows that these genotypes are basically significantly different in the duration of linear growth (\( P = 0.065 \)). It can be seen from figure 3 that the two genotypes display an increasing difference in the stage of exponential growth (\( 0 - P_3 \)), which becomes more dramatic after entering the stage of linear growth (\( P_3 - P_3 \)), leading to a pronounced difference in plant height at the adult stage. The QTL detected has a significant effect on the growth amount of linear growth period (table 5). Only using functional mapping can one identify the developmental basis for phenotypic variation in final plant height.

**Discussion**

Although many studies of developmental timing have used comparative and experimental approaches for revealing the nature of developmental time from individual cells to whole
organisms, a quantitative model that can map individual QTLs for heterochrony has not yet been developed. As a ubiquitous phenomenon, heterochrony, defined as a shift in the timing of events, can occur in any stage of development at any level of organization from cells to organs to organisms (Rougvie 2001; Rice 2002; Mitteroecker et al. 2004; Moss 2007; Lenz et al. 2014). In this article, focusing on organ growth, we have discovered and renovated the key feature of functional mapping, allowing it to map hQTL, and more importantly, embedded it within the phylogenetic context to study the genetic origin that causes the evolution of development. The hQTLs prompt the identification of a set of interacting regulators involved in the developmental timing mechanisms where they are not yet known to exist (Rougvie 2001; Lenz et al. 2014).

The major innovation of hQTL mapping lies in its capacity to characterize the genetic mechanisms for the timing of development and identify specific loci that govern developmental divergence along a phylogeny. Although the integration of development and evolution, that is, evo-devo has emerged as a promising concept for revolutionizing the understanding of evolution (Arthur 2002; Müller 2007; Sommer 2009), there is a significant lack in the systematic detection of genes for evo-devo processes and phylogenetic diversification. Without the discovery and analysis of such genes, reciprocal questions of fundamental importance in evolutionary biology remain unanswered: How species with shared developmental genetic toolkits can still produce a diversity of life forms versus how similar forms can derive from different toolkits (Canestro et al. 2007). hQTL mapping bridges this gap by providing a powerful means of detecting evo-devo genes and implementing them into a predictive model of evolution and speciation.

Since in the 1980s, a series of genes, called Hox genes, were discovered to set the identity of segments of insect bodies from head to tail, a tremendous interest has been directed toward identifying regulatory genes that guide development by controlling the expression and function of coding genes (Kikuta et al. 2007; Woolfe and Elgar 2007; Vaquerizas et al. 2009; Vidal et al. 2011). However, this study is the first that is equipped with a power to characterize the quantitative effects of regulator genes on the process and pattern of development and link genotype to phenotype for quantitative traits. Although the current model was developed on linkage mapping using a genetic linkage map, it can be readily extended to embrace genome-wide association studies (GWAS), allowing the comprehensive detection of hQTLs throughout the genome. The integration of hQTL mapping with GWAS through high-dimensional statistical models (Li et al. 2012; Liu et al. 2014) will provide an unprecedented opportunity to increase our understanding of phylogenomics, phylogenetic inference on a genome-wide scale (Canestro et al. 2007).

The new model requires mapping data from multiple syntonic taxa. Extensive simulation studies have been performed to demonstrate its statistical usefulness and utilization in identifying heterochronic parameters and their genetic dissection. Although such data are not available yet to the authors, an increasingly widespread application of new cost-effective biotechnologies, such as next-generation sequencing, has changed this situation to favor and materialize the use of the new model. As a demonstration, we used a mapping population of rice to detect QTLs for developmental events in plant height growth during ontogeny. One QTL detected to affect plant height growth trajectories was analyzed by our hQTL mapping model, which finds that the final plant height growth trajectory is determined by

<table>
<thead>
<tr>
<th>Taxon 1</th>
<th>n = 200</th>
<th>H$^2$ = 0.05</th>
<th>H$^2$ = 0.1</th>
<th>H$^2$ = 0.2</th>
<th>n = 400</th>
<th>H$^2$ = 0.05</th>
<th>H$^2$ = 0.1</th>
<th>H$^2$ = 0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_l$</td>
<td>0.55</td>
<td>0.85</td>
<td>0.98</td>
<td>0.81</td>
<td>0.98</td>
<td>0.81</td>
<td>0.98</td>
<td>1</td>
</tr>
<tr>
<td>$t_u$</td>
<td>0.41</td>
<td>0.79</td>
<td>0.96</td>
<td>0.7</td>
<td>0.99</td>
<td>0.7</td>
<td>0.99</td>
<td>1</td>
</tr>
<tr>
<td>$t_d$</td>
<td>0.51</td>
<td>0.77</td>
<td>0.98</td>
<td>0.81</td>
<td>0.99</td>
<td>0.81</td>
<td>0.99</td>
<td>1</td>
</tr>
<tr>
<td>$T$</td>
<td>0.11</td>
<td>0.13</td>
<td>0.28</td>
<td>0.15</td>
<td>0.16</td>
<td>0.15</td>
<td>0.16</td>
<td>0.19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Taxon 2</th>
<th>n = 200</th>
<th>H$^2$ = 0.05</th>
<th>H$^2$ = 0.1</th>
<th>H$^2$ = 0.2</th>
<th>n = 400</th>
<th>H$^2$ = 0.05</th>
<th>H$^2$ = 0.1</th>
<th>H$^2$ = 0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_l$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$t_u$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$t_d$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$T$</td>
<td>0.14</td>
<td>0.125</td>
<td>0.15</td>
<td>0.17</td>
<td>0.16</td>
<td>0.16</td>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Empirical Power of Detecting Genetic Effects of a QTL on Four Heterochronic Parameters in Different Simulated Mapping Populations from Two Taxa under Different Sample Sizes and Heritabilities.

Table 4. Empirical Power of Detecting the Taxon-Specific Difference of Genetic Effects on Four Heterochronic Parameters under Different Sample Sizes and Heritabilities.

Fig. 3. Genetic variation in developmental timing caused by an hQTL detected in a DH population of rice. Three landmarks $P_a$, $P_b$, and $P_d$ are the timing of maximum acceleration of growth, maximum rate of growth, and maximum deceleration of growth, respectively, and time interval $T$ is the length of linear growth, all parameters subscribed by the notation of QTL genotypes QQ (1) and qq (2). The effects of the QTL on the three landmarks and linear growth length are denoted as $\Delta t_a$, $\Delta t_b$, $\Delta t_d$ and $\Delta T$, respectively.
may be explained by genetic variation in linear growth length. This discovery enables geneticists to predict and select plant height according to its underlying mechanistic basis. Similar analyses can also be conducted in other species including animals.

The hypothesis test procedure in hQTL mapping allows evolutionary geneticists to map genes that govern evolutionary rates over a series of taxa, thereby facilitating the comparison of the patterns in evolutionary biology from a phylogenetic perspective (Adams 2013). By mapping trait development for a series of syntetic taxa, hQTL mapping can be equipped to construct a phylogenetic tree across the taxa based on developmental processes of each genotype at a QTL detected (Felsenstein 2004; Broman et al. 2012; Zhang et al. 2012), compare genotype-dependent differences in tree structure, and infer the impact of the QTL on phylogenetic diversity and evolutionary transformations. Incorporating the approach for hQTL mapping with phylogenetic studies would therefore provide researchers with a powerful means to testing evolutionary hypotheses of character change and the accumulation of phenotypic diversity within and across lineages.

Acknowledgments

The authors thank the associate editor and the two anonymous reviewers for their constructive comments on this manuscript. This work was supported by Special Fund for Forest Scientific Research in the Public Welfare (201404102) from the State Forestry Administration of China, grants 2011AA100207 and 2013AA102607 from the Ministry of Science and Technology of China, and the Arvalis Award.

References


