Mapping genes for plant structure, development and evolution: functional mapping meets ontology

Qiuling He¹, Arthur Berg², Yao Li³, C. Eduardo Vallejos⁴ and Rongling Wu⁵,²,⁶,¹

¹The Key Laboratory of Forest Genetics and Gene Engineering, College of Forest Resources and Environment, Nanjing Forestry University, Nanjing, JS 210037, China
²Department of Public Health Sciences, Pennsylvania State University, Hershey, PA 17033, USA
³Department of Statistics, West Virginia University, Morgantown, WV 26506, USA
⁴Department of Horticultural Sciences, University of Florida, Gainesville, FL 32606, USA
⁵The State Key Laboratory of Tree Breeding, College of Life Sciences and Biotechnology, Beijing Forestry University, Beijing 100083, China
⁶Department of Statistics, Pennsylvania State University, University Park, PA 16802, USA

One of the fundamental tasks in biology is the identification of genes that control the structure and developmental pattern of complex traits and their responses to the environment during trait development. Functional mapping provides a statistical means for detecting quantitative trait loci (QTLs) that underlie developmental traits, such as growth trajectories, and for testing the interplay between gene action and development. Here we describe how functional mapping and studies of plant ontology can be integrated so as to elucidate the expression mechanisms of QTLs that control plant growth, morphology, development, and adaptation to changing environments. This approach can also be used to construct an evo-devo framework for inferring the evolution of developmental traits.

Functional mapping for developmental traits

Plants and animals are complex organisms that have evolved into a diverse array of forms through the accumulation of mutations in genes that control the cascades of developmental processes leading to mature forms [1–3]. An increasing number of evolutionary studies have identified the genetic changes that must take place to produce a phenotype that is divergent from its ancestral form [4–6]. It has been demonstrated that changes in the size and shape of organisms can be the result of heterochrony: the change in the rate and/or timing of developmental events [1]. Thus, heterochrony can lead to the evolution of new species from ancestral forms.

Genetic mapping with molecular markers has proved to be a powerful tool for dissecting complex phenotypes into their underlying quantitative trait loci (QTLs) [7,8]. This methodology has been used extensively to map a diverse array of complex phenotypes, but in the vast majority of cases the endpoint phenotypes have been used as the target trait. For example, as one of the most interesting endpoint phenotypes in plant breeding, plant height at maturity has been extensively studied by traditional genetic mapping, but this approach does not take into account the genetic mechanisms underlying the developmental trajectory of this phenotype. To illustrate this point, consider a group of segregants from a population displaying exactly the same endpoint phenotype. If these seemingly identical phenotypes develop over significantly different periods of time, then differences can be attributed to variation in the genes that control the developmental trajectory of the phenotype. Functional mapping (Box 1) is a QTL mapping approach that directly targets the genes that control growth and development by treating these biological processes as dynamic traits rather than static endpoint phenotypes [9–11].

The power of functional mapping rests on its parsimonious modeling of temporal trends and the covariance structure of the model. For these reasons it is statistically more advantageous in parameter estimation, gene detection power, and computational efficiency and feasibility than traditional QTL mapping methods of complex traits at individual time points. From a developmental biology perspective, results from functional mapping can be visualized as dynamic curves of gene expression in time [12–15] and, in principle, offers an opportunity to identify and map genes that play a pivotal role in regulating the timing of major developmental milestones [16].

In this article we discuss the versatility of functional mapping for unraveling the genetic and molecular mechanisms that control developmental trajectories and milestones, and examine pleiotropic effects exerted by some genes on different developmental landmarks and biological processes. Functional mapping can play a strengthening role in the emerging discipline of biology–ontology [17–23]. The marriage between functional mapping and ontology is opening up a new gateway for characterizing the genetic architecture of ontogenetic trajectories from embryo to senescence in developmental time and space. The results of this ‘marriage’ are likely to provide impetus for the construction of the evo-devo framework of organismic structure and development so as to address fundamental questions related to how developmental processes evolved.
Although the framework applies to growth trajectories in other species, we have articulated our argument using plants as an example.

Plant ontology

Biological ontology addresses questions concerning the components of biological systems, the types of relationships between these elements, and how this information can be organized to facilitate retrieval and utilization. The increasing availability of data from several genome sequencing projects has allowed the implementation of this concept. Examples include the Gene Ontology Consortium site (www.geneontology.org), that covers model organisms from all Life Domains, and the Plant Ontology Consortium (POC: http://www.plantontology.org) that focuses on model plant species. These and other sites were launched to develop standard vocabularies for the generic description of molecular function, biological processes, and cellular and higher order structures [17,21,22]. The standard vocabularies allow a uniform annotation of datasets across diverse taxonomic groups [23]. This approach makes ontology a powerful tool for investigating genes and their functional and sequence homologs that are involved in growth and development. The Open Biomedical Ontology (www.obofoundry.org) was recently created in an effort to standardize vocabularies among ontologies developed for diverse organisms [24]. Standardization will greatly facilitate the phylogenetic and evolutionary comparisons of morphological structure and developmental patterns.

A biological ontology has a hierarchical structure similar to that of a tree. Biological ontologies addressing different levels of complexity have been developed. Let us consider some plant ontologies to illustrate their usefulness. The Plant Structure Ontology depicted in Figure 1 has three hierarchies and several nodes that are interconnected unidirectionally [21]. In contrast to a normal tree, some nodes have connectors originating in more than one higher-level node, as seen for the vascular tissues. A higher level of complexity is represented by the Whole-Plant Growth Stages Ontology that covers the entire lifecycle of a plant [20]. For example, Figure 2 describes the maize lifecycle that is partitioned into several developmental stages from germination to maturity and senescence [25]. More recently, the Plant Structure Development Stages Ontology has been
proposed to describe the developmental stages of individual plant structures of a generic flowering plant (i.e., the flower, leaf, fruit, inflorescence, root, and seed) [22]. These ontologies can be integrated and operated as sub-domains, as long as the links that connect their nodes are established. The gene ontology could play an important role in connecting the nodes of these different ontologies that describe morphological structures and developmental patterns in plants.

Ontologies are dynamic by design because new relationships can be discovered and others can be deemed obsolete. Thus the genetic analysis of dynamic traits, in particular those associated with growth and development, can help establish connections between nodes in an ontological sub-domain, but also between sub-domains.

**Growth gene ontology**

*Why use functional mapping?*

The treatment of a phenotypic trait as a dynamic process permits functional mapping to identify genes that control different aspects of plant growth and development. To explain this, we consider growth data measured at different time points. As one of the most prevalent laws observed in biological systems [26], growth can be described by the Richards growth equation [27]:

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

(Eqn 1)

where \( g(t) \) is the plant height at age \( t \), \( a \) is the asymptotic value when \( t \) tends to be infinite, \( b \) is the parameter that...
describes the initial growth of plant height, \( r \) is the relative growth rate, and \( k \) is the power at which the metabolic rate of an organ scales as the \( k \)th power of its weight [27].

Consider two distinct genotypes (\( P \) and \( Q \)) for which the dynamic phenotype of a growth trait has been characterized using the Richards growth function (Eqn 1) with parameters in Table 1. Figure 3a illustrates the difference in growth curves between these two genotypes. Segregating progeny obtained from \( P \) and \( Q \) genotypes would have a diverse array of trajectories, and functional mapping can detect and locate the genes controlling these differences in growth trajectories by testing the effects of each of the growth parameters, \( a, b, r, \) and \( k \), in Eqn 1.

Assuming that four independent sets of genes control these parameters, we can examine the effect of replacing one set at a time in the parental genotypes. The replacements for genotypes \( P \) and \( Q \) are shown in Figures 3b and c, respectively, in which the subscripts of \( P \) and \( Q \) indicate the replacement of the genotype’s original growth curve by a specific parameter. In all cases the trajectory was halted when it reached 99.9% of the final value (the \( a \) parameter). These plots show that genes controlling the initial size (the \( b \) parameter) have a continuously increasing or decreasing effect until the final size is reached (compare \( P \) with \( P_{b} \), or \( Q \) with \( Q_{b} \)), but consequently have a modest effect on the time it takes to reach maximum size. Nevertheless, these genes can have an impact on other organs. If the trait considered is leaf area, then having a greater leaf area throughout the growth period might result in an increase in carbon assimilation with consequent physiological and ecological advantages. These plots also show that genes that control the rate (\( r \)) will increase the differences in the timing of the inflection point at which the relative rate of growth reaches a maximum (compare \( P \) with \( P_{r} \), or \( Q \) with \( Q_{r} \)). The most significant consequence of this genetic effect is on the length of the growing period. Cytosolic invertases provide an example of the genes that control growth rate in Arabidopsis [28]. The replacement of power \( k \) will lead to delay (compare \( P \) with \( P_{k} \)) or early occurrence (compare \( Q \) with \( Q_{k} \)) of asymptotic growth. By far, the genes that control the final size (\( a \)) appear to have the greatest effect (compare \( P \) with \( P_{a} \) or \( Q \) with \( Q_{a} \)). Growth results from the combined effects of cell division, cell expansion, and the spatial and temporal coordination of these activities. Several genes that control organ size have been cloned and characterized. For instance, \( DA1 \) [29] and \( KLUH \) [30] are genes that control the timing of cell proliferation, whereas \( AINTEGUMENTA \) [31] promotes cell division but it is controlled by the \( FO \) gene [32]. However, genes such as \( BIGPETAL \) [33] and that encoding Aquaporin [34] modulate or control cell expansion. This simplistic analysis illustrates the advantage of using functional

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( P )</th>
<th>( Q )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a )</td>
<td>25</td>
<td>15</td>
</tr>
<tr>
<td>( b )</td>
<td>80</td>
<td>41</td>
</tr>
<tr>
<td>( r )</td>
<td>0.65</td>
<td>0.35</td>
</tr>
<tr>
<td>( k )</td>
<td>2.0</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Table 1. Four parameters used to define a growth equation [1] for two contrasting genotypes \( P \) and \( Q \).
mapping in the dissection of genetic mechanisms that control growth and development at the organ or organismal level.

The catalog of growth genes
Different allelic combinations of different genes – specifically those that regulate growth rates and those that control their timing and duration of the expression – can lead to different growth trajectories, as depicted in Figure 4. Based on the pattern of genetic control for growth parameters we can broadly classify the growth genes into four categories:

(i) Faster–slower genes (Figure 4a) control the relative growth rate, $r$, and exert an effect on the duration of the exponential phase of growth (Figure 4a). Of two genotypes that share the same endpoint phenotype, the one with the ‘faster’ alleles will reach the endpoint earlier than the one with the ‘slower’ alleles. Examples for faster–slower genes include those that lead to the differences between $P$ and $P_r$ (Figure 3b) as well as $Q$ and $Q_r$ (Figure 3c).

(ii) Bigger–smaller genes (Figure 4b) control asymptotic growth, $a$, and plants with the ‘bigger’ genotype are consistently better in growth during lifetime than those with the ‘smaller’ genotype. The bigger and smaller genotypes are parallel and do not cross over in the entire progress of growth. Examples include those that are responsible for the differences between $P$ and $P_a$ (Figure 3b) as well as $Q$ and $Q_a$ (Figure 3c).

(iii) Earlier–later genes (Figure 4c) control the amount of initial growth (described by parameter $b$) and scaling power $k$. For this type of gene, an ‘earlier’ genotype grows faster in the early stage because it can exploit more resources than can a ‘later’ genotype. Thus, it is likely that an earlier–later gene functions depending on the expression of other genes associated with the capacity to exploit resources rapidly. For example, an ‘earlier’ genotype in poplar trees exhibits faster early growth from a budded cutting than a ‘later’ genotype, probably because the former carries favorable genes for increased rooting capacity. Examples of earlier–later genes include those leading to differences between $P_b$ and $P_k$ (Figure 3b) as well as $Q_b$ and $Q_k$ (Figure 3c).

(iv) Age-dependent bigger–smaller genes (Figure 4d) determine the variation in growth curves by altering

Figure 4. Four categories of QTLs that each trigger their effects on growth in time. For each QTL category there are two genotypes each displayed by a growth curve. $T$ denotes the time when the extent of growth reaches a maximal value and $t$ denotes the time at which growth rate is maximal. Times $T$ and $t$, subscripted by a QTL genotype (1 or 2), can be used to describe the shape of a growth curve related to the developmental characteristics of an organism.
their direction of effect as a function of time. The genotype that is bigger at the final size does not fully display its growth potential in the early stage because of a lower scaling power $k$, whereas the counterpart genotype has a larger $k$ value (and therefore a better early growth) but does not have great potential. This category of genes produces a maximum degree of variation in growth trajectories. Examples for age-dependent bigger–smaller genes include those that are responsible for differences between $P_k$ and $P_a$ (Figure 3b) as well as $Q_k$ and $Q_a$ (Figure 3c).

In practice it is possible that growth is simultaneously controlled by all four categories of genes, so increasing the extent of growth curve variation. In addition, the expression of a typical gene can be age-specific. For example, the bigger–smaller gene might be activated after a particular time point, whereas the earlier–later gene is active only before a particular time point prior to adult age.

**Growth and reproductive genes**

A flowering plant undergoes several distinct developmental stages including germination, rosette growth, flowering, and senescence, that cover vegetative and reproductive lifecycles [20]. In maize the vegetative (V) stage includes emergence (V$n$) and tasseling where $n$ is the number of collars present on the plant, whereas the reproductive (R) stage includes silking, blister, milk, dough, dent, and physiological maturity (Figure 2). Each of these sub-stages for the V and R lifecycles determines grain yield solely by influencing the number of mature grains per plant and the average weight of the individual grains.

Functional mapping can be used to integrate different growth, physiological and developmental changes during the lifetime of an organism (Figure 2). Functional mapping permits geneticists to test whether there is a shared genetic control mechanism for vegetative and reproductive processes through the pleiotropic effect of the same gene or linkage disequilibria between different genes [35]. If the answer is yes, functional mapping can further quantify the extent to which the genetic basis is shared between these two different types of biological processes. Figure 5 illustrates the pleiotropic control of the QTL on the vegetative growth process and a reproductive event. If the pleiotropic effect is important for these two types of traits, functional mapping can determine its direction—in other words whether it is positive (as shown in Figure 5a) or negative (Figure 5b).

With appropriate extensions, functional mapping can address several central questions in plant biology:

(i) How does a QTL jointly affect the growth stages of multiple different structures of a plant, such as the flower, leaf, fruit, inflorescence, root, and seed? If different QTLs exist to determine each trait, then how do they coordinate to affect the whole-plant during development?

(ii) Are there specific QTLs for developmental trajectories and for responses to temperature and day-length in plants? If yes, how do these QTLs affect the patterns of development and the norms of reaction? Because covariance structure matters in the modeling
of longitudinal data, more robust statistical approaches are needed when considering the dependence of longitudinal responses to different agents.

(iii) What is the impact of developmental stage and the environment on the expression of these QTLs? Do they interact to form a web of epistatic network that regulates plant growth and development at the whole plant level?

(iv) How can the developmental pattern of a morphological trait be genetically coupled with gene expression profiling, proteomic profiling, and metabolic profiling at different stages of growth? Can we use these profile data as biomarkers to predict developmental trajectories based on a plant’s genetic makeup?

Concluding remarks

Plant systems are complex, both structurally and functionally, and the genes that direct the growth and development of a plant will also have complex patterns of gene expression in time and space. Functional mapping provides a powerful statistical tool for identifying the interplay between genes and development and for determining how the roles of specific genes evolve as a function of time. In contrast to traditional QTL mapping approaches that target endpoint phenotypes, functional mapping models address discrete growth stages that span the entire life cycle of the plant, thereby maximizing our ability to detect subtle changes that affect only a limited aspect of development.

With the completion of genome-sequencing projects for several plants and animals [36–44], the next major challenge will be the large-scale determination of gene function in diverse plant species. However, complete understanding of a gene’s function is only achieved once it is associated with a phenotype of interest at the organismal level [4]. The next decade will witness a surge of interest in accumulating a vast amount of phenotypic data comprising anatomical, morphological, and developmental characteristics of plants. In this regard we must consider that the formation and progression of dynamic phenotypes require temporal and spatial coordination of the expression of different gene networks. Functional mapping allows dynamic phenotypes to be linked with their respective underlying genotypes throughout growth and development, and can be used to assist plant ontology in capturing and interpreting phenomena associated with these processes. On a wider view, the marriage of functional mapping and ontology can shed some light on how genes controlling developmental trajectories play an important role in heterochrony and the evolution of form and structure in plants.

As our understanding of the gene networks and underlying molecular details of developmental pathways improves, we are faced with the challenge of constructing a genetic regulatory network with data generated from high-throughput genotyping, microarrays, proteomics and transcriptomics. Functional mapping, plant ontology, gene and ‘omics’ ontologies will together identify problems and gaps in knowledge related to gene annotation in different plants species in which their developmental and evolutionary relationships are not yet entirely clear.

Acknowledgements

This work is partially supported by NSF/NIH Joint grant DMS/NIGMS-0540745, the Changjiang Scholarship Award at Beijing Forestry University, and the Innovation Fund of the Institutes of Food and Agricultural Sciences at the University of Florida.

References

19 Bard, J. et al. (2005) An ontology for cell types. Genome Biol. 6, R21
29 Li, Y.H. et al. (2008) Control of final seed and organ size by the DA1 gene family in Arabidopsis thaliana. Genes Dev. 22, 1331–1336
33 Szecsi, J. et al. (2006) BIGPETALp, a bHLH transcription factor is involved in the control of Arabidopsis petal size. EMBO J. 25, 3912–3920
34 Ma, N. et al. (2008) Rh-PIP2;1, a rose aquaporin gene, is involved in ethylene-regulated petal expansion. Plant Phys. 148, 894–907
54 von Sengbusch, P. (2003) Botany Online – The Internet Hypertextbook (http://www.biologie.uni-hamburg.de/b-online/e00/default.htm)