

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Functional Mapping of Plant Growth in *Arabidopsis thaliana*

Kaiyue Liu, Wenhao Bo, Lina Wang,
Rongling Wu and Libo Jiang

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.74424>

Abstract

Most traits important to agriculture, biology, and biomedicine are complex traits, determined by both genetic and environmental factors. The complex traits that change their phenotypes over different stages of development are called dynamic traits. Traditional quantitative trait loci (QTLs) mapping approaches ignore the dynamic changes of complex traits. Functional mapping, as a powerful statistical tool, can not only map QTLs that control the developmental pattern and process of complex traits, but also describe the dynamic changes of complex traits. In this study, we used functional mapping to identify those QTLs that affect height growth in 10th generation recombinant inbred lines derived from two different *Arabidopsis thaliana* accessions. Functional mapping identified 48 QTLs that are related to height traits. The growth curves of different genotypes can be drawn for each significant locus. By GO gene function annotations, we found that these QTLs detected are associated with the synthesis of biological macromolecules and the regulation of biological functions. Our findings provide unique insights into the genetic control of height growth of *A. thaliana* and will provide a theoretical basis for the study of complex traits.

Keywords: complex traits, QTL, functional mapping, growth, *Arabidopsis thaliana*

1. Introduction

Complex traits are genetic traits controlled by multiple genes. They are sensitive to environmental changes and easily affected by the environment [1]. The phenotypic expression of complex traits in individuals within a population displays a continuous variation and generally a normal distribution. Most important biochemical, medical, and agronomic traits

and the majority of human diseases are complex traits that are controlled by interconnected genetic networks and environmental factors. In these gene networks, the effect of each gene is small [2]. The characteristics also change at different developmental stages and demonstrate the dynamic features of complex traits. The study of these complex traits is an important topic in modern biology.

Genetic mapping aimed at mapping underlying genes to genomic locations is a powerful tool for dissecting the genetic architecture of complex traits. Lander and Botstein have proposed an approach for mapping quantitative trait loci (QTLs) based on a sparse-density linkage map of molecular markers. This so-called interval mapping method can overcome the confounding problem of marker-QTL recombination [3]. Composite interval mapping includes other markers as covariates to control the overall genetic background, which displays increased power in QTL detection [4]. Considering QTL-QTL epistatic interactions in a linkage map, Kao et al. proposed using multiple marker intervals to map QTLs [5]. Currently, statistical methodologies for QTL mapping include regression analysis, maximum likelihood, and the Bayesian approach. With the development of high-throughput single nucleotide polymorphism (SNP) genotyping techniques, genome-wide association studies (GWAS) have provided a powerful means of mapping a complete set of genes underlying complex traits [6–8]. The genetic structure of a trait is explained by GWAS, which identify the numbers and chromosomal locations of each gene, the size of each gene's unique and pleiotropic effects, and the relative contributions of additive, dominant, and epistatic genetic effects. This provides an unprecedented tool for preparing a genotype-phenotype map [9]. So far, GWAS have detected many genetic variants for a wide range of complex traits, including those pertaining to agriculture, forestry, and human disease [9, 10].

It should be noted, however, that traits such as height and weight vary with time or other independent environmental stimuli. The traditional QTL mapping method directly associates a single marker with a single phenotype at a time point, which ignores the dynamic characteristics of organisms at different developmental stages, and cannot exactly reflect the whole genetic architecture of complex traits. Although thousands of QTLs were detected in many individuals, only a small number of QTLs were cloned and separated [11]. The reason for this problem is that the QTLs that have undergone rigorous statistical testing are divorced from biological relevance, which limited the projections of the genetic structure of traits. Genetic analysis of dynamic traits presents a serious statistical challenge. To solve these problems, Ma et al. [13] proposed a QTL mapping method based on a logistic-mixture model [12]. The QTL effect on developmental traits during ontogeny is considered as a function of time, and a series of growth formulas can be derived from the logistic curve describing plant height, size, and weight [13], arriving at a model which is expected to be improved in parameter estimation and statistical inference over previous models. Ma et al. [13] developed a maximum likelihood statistical framework based on a logistic-mixture model for the characteristics of function-valued traits, which change as a function of a specific variable. This QTL mapping strategy is called functional mapping.

Functional mapping combines mathematical functions that describe biological processes and assembles mathematical formulas into the statistical framework of QTL mapping to study the interactions between genes and phenotypic traits of organisms during growth and development. We estimate the parameters of a specific genotype that determines the development of

a trait by statistical hypothesis testing and parameter estimation, rather than directly estimating the genetic effects of the gene at all-time points. Because of the combination of biological laws (defined by mathematical models) and the reduced number of genetic parameters that need to be estimated, functional mapping increased the potency of detecting significant QTLs [14]. Statistical merits of functional mapping will be exemplified in which case data are recorded irregularly, bringing on data sparsity. Such sparse longitudinal data cannot be well analyzed by traditional mapping for two reasons. First, because the problem of missing data existing at a given time point, traditional method is unable to use all of individuals, thus leading to a biased parameter estimation and reduced power of QTL detection. Second, individuals are measured at a few number of time points, limiting the fit of growth equation. Functional mapping is robust for handling longitudinal sparse data in which no single time point has the phenotypic data for all individuals, facilitating the QTL mapping to study the genetic architecture of hard-to-measure traits.

A set of tree data is used to assess the statistical validity of functional mapping. Several QTLs affecting the developmental trajectories of poplar were detected with the QTL mapping method based on the logistic-mixture model [15], and these QTLs were located on a genetic linkage map constructed by polymorphic markers. Studies have shown that functional mapping is useful in establishing gene-phenotype relationships and predicting QTL phenotypes of individual organisms based on the control of a trait. Functional mapping combines the principles of Mendelian genetics with statistical and developmental mechanisms, and is superior to traditional QTL mapping methods that combine the principles of statistics and genetics. To date, functional mapping has been used for mapping dynamic QTLs in poplars [16], jujubes [17], soybeans [18], rice [19], maize [20], yeast [21], oysters [22], mice [23], humans [24], and drug responses [25, 26].

Arabidopsis thaliana is a small, annual or winter annual, rosette plant. *A. thaliana* is a central genetic model and universal reference organism in plant and crop science. The successful integration of different fields of research in the study of *A. thaliana* has made a large contribution to our molecular understanding of key concepts in biology. The *Arabidopsis* reference genome sequence was the first published nuclear genome of a flowering plant in 2000 (<http://www.arabidopsis.org>) [27, 28]. The ease and speed with which experiments can be conducted on *A. thaliana* has allowed enormous fundamental progress in our knowledge of the molecular principles of plant development, cell biology, metabolism, physiology, genetics, and epigenetics [29]. The uses of *Arabidopsis* as the universal reference plant continue to expand, particularly in the field of systems biology [30, 31]. Important work has been done to investigate the molecular networks that mediate environmentally controlled developmental switches in *A. thaliana*. Examples include the transition from vegetative to reproductive development, also termed flowering time control [32–34], seed dormancy and germination control [35, 36], and the light-regulated development of seedlings [37, 38]. *A. thaliana* has also served as a model research organism for exploring many areas of fundamental biology, including photobiology, the circadian clock, DNA methylation, DNA repair, RNA silencing, protein degradation, and G-protein signaling, many of which have direct application in human health [39–42]. Insights into the functions of a multitude of individual genes, as well as the elucidation of biosynthetic pathways and regulatory networks, in *A. thaliana* have proven invaluable for identifying the genetic basis of agronomically important traits in crops such as plant height and flowering time [43, 44]. Classical molecular genetics approaches have been used in

A. thaliana to dissect the patterning and development of flowers [45], embryos [46], leaves, and roots [47]. In this study, we describe the implementation of functional mapping to identify and map QTLs for height trajectories in a population of *A. thaliana*. The mapping population is composed of 144 recombinant inbred lines (RILs) derived from a Landsberg erecta (LER) cultivar and a Shadara (SHA) cultivar. Functional mapping identified 48 QTLs that determine the height growth of *A. thaliana*. The identification of these QTLs will help us address fundamental questions about the genetic mechanisms of height growth.

2. Materials and methods

2.1. Materials

In this study, the mapping population included 114 RILs derived from two cultivars of *Arabidopsis*. The main phenotype of the LER cultivar includes round leaves, short petiole, short pedicel, flowers clustered at inflorescence tips, short pod width, sharp tip, inflorescence compact, and a plant height of 10–25 cm. The general traits of the SHA cultivar are described in The *Arabidopsis* Information Resource database as a slightly narrow leaf with a height of about 30 cm. To ensure adequate seeds for the sustainability of the material, each RIL carried out a generation of expansion. For the sterilization of *Arabidopsis* seeds, the seeds were shaken in a centrifuge tube with 10% sodium hypochlorite for 6 min; 1 ml of 95% alcohol was then added, and the tubes were shaken for 5 min. The seeds were rinsed 5–6 times to thoroughly wash away the remaining sodium hypochlorite solution on the seed surface. The alcohol was poured out, and the tubes were dried at a super clean bench to obtain sterile *Arabidopsis* seeds.

For vernalization of *A. thaliana* seeds, distilled water was added to the seeds and the seeds were placed in a 4°C refrigerator for 3 days of dark treatment before sowing. To sow the seeds after vernalization, vermiculite, peat, and limestone were mixed evenly in a 1:1:1 ratio and used as the soil. An 18 × 18 cm plastic grid with 90 squares was placed over the plant pot. The seeds were placed into each mesh with a toothpick. The planting diagram of the experiment is shown in **Figure 1**. Circles and triangles represent the two different lines of *A. thaliana*, planted in the same growing space. After planting, the plant pot was covered with plastic wrap and then transferred to the long day (16 h) and short day (8 h) artificial climate room (22°C). When the *Arabidopsis* sprouts displayed green shoots, the plastic wrap was removed.

2.2. Phenotypic data acquisition

Plant height was measured after 1 week of planting. Each line of *A. thaliana* was randomly selected for 10 strains. In total, 1160 strains of *Arabidopsis* were selected. The height of *Arabidopsis* plants were measured once per week until the end of the plant's life cycle by manual measurement with a 1 mm ruler (accurate to 1 mm).

The plant height phenotype was measured for 1160 *A. thaliana* strains. The height of 1–60 lines of *Arabidopsis* was measured seven times, and the height of 60–116 lines of *Arabidopsis* was measured seven times. Because of environmental or genotypic reasons, nine lineages of *Arabidopsis* did not complete a life cycle (numbers 1–4, 22, 28–30, and 60).

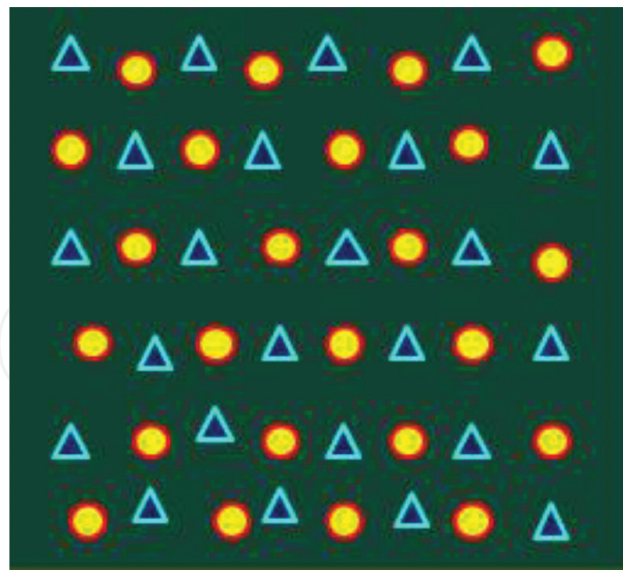


Figure 1. Schematic diagram of *Arabidopsis* planting.

The growth of *Arabidopsis* strains can be seen in **Figure 2a–c**, which indicates the growth of the strains for 1, 3, and 6 weeks, respectively.

2.3. Acquisition of SNP data

A. thaliana DNA was extracted with a DNA extraction kit (TIANGEN DP305) according to the manufacturer's instructions. The plant material used was 100 mg of the young leaf. After DNA was detected in the samples, the DNA sequence was determined by an outside company. The resulting sequence data was subjected to SNP calling.

2.3.1. DNA quality inspection

The degree of degradation of the DNA was analyzed by agarose gel electrophoresis and imaged with an ultraviolet gel imager. The DNA sample was determined to be without degradation if

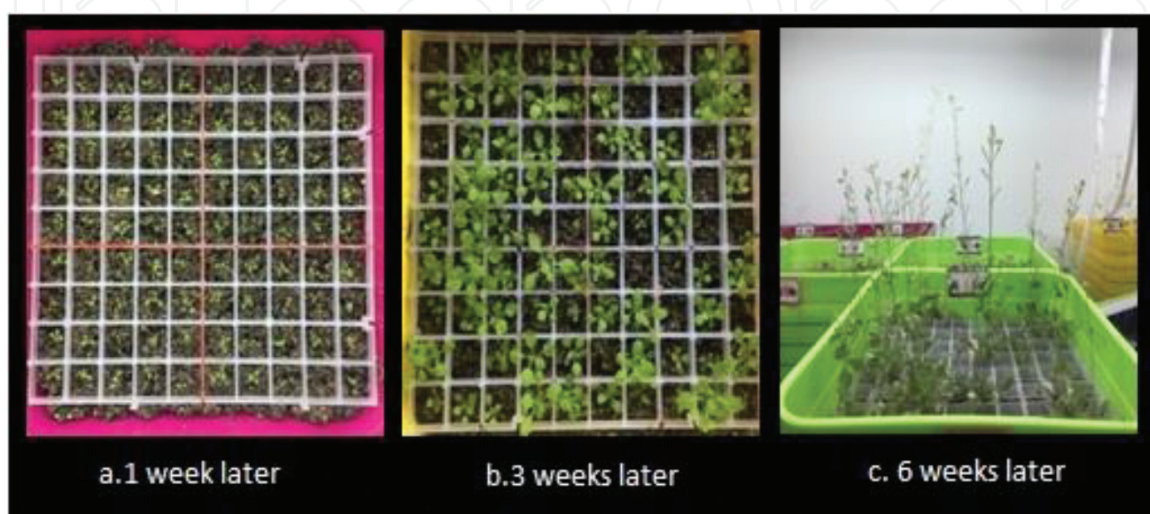


Figure 2. *Arabidopsis* growth conditions at 1 week (a), 3 weeks (b), and 6 weeks (c).

the sample did exhibit dispersion phenomena. All of the samples were determined not to be degraded and were sequenced.

We measured the absorbance of the DNA samples at wavelengths of 260 and 280 nm and calculated the value of the optical density (OD) 260/OD280. Samples were determined to be of sufficient quality if this value was greater than 1.8 and less than 2.2. After testing, all samples were found to meet the requirements for sequencing.

2.3.2. Sequencing and mapping

DNA from the sample was sequenced on the Illumina HiSeq 2000 platform to generate 125 bp paired-end reads at greater depth. All these RILs sequencing were performed at Total Genomics Solution (TGS) Institute. After sequencing quantity control, the data of 107 samples reached 117.68 G, with an average of 1.10 G per sample. The percentage of Q30 bases was more than 90%, the percentage of Q20 bases was more than 95%, and the distribution of GC was normal. Thus, the quantity and quality of the sequencing met the requirements for subsequent analysis.

The per-base coverage depth across all contigs was calculated by mapping raw reads from each RIL against reference genomes. The results showed that the average mapping ratio and sequencing coverage of the samples were higher than 95 and 92%, respectively. The sequence depth of the samples was 9.91×. In addition to the line 110, the other lines are better for subsequent variation detection and analysis.

2.4. SNP calling

SNP detection was performed with the widely accepted mutation detection software GATK. SNP. The screening criteria were as follows: The depth of sequencing for each sample was greater than or equal to 4, otherwise the sample was marked as missing. Additionally, the quality value of the comparison must be greater than or equal to 20, and the variation detection quality value must be greater than or equal to 50. If the allele frequency was less than 5% or the absence rate of the sample was greater than 50%, the site was filtered out.

The RIL population contains 105 progeny and two parents; the parent number is "A518_LER" and "A518_SHA." The reads of each sample were compared with the reference genome. The average ratio of the samples was above 95%, and the coverage was greater than 92%. The average sequencing depth was 9.19×. By processing, we obtained 107 samples and 1,023,325 whole genome SNP markers. According to the situation of this population, individuals with heterozygous genotype ratios over 25% and markers indicating that parents possessed heterozygous genotypes were eliminated. Finally, 609,427 SNPs and 80 individuals met the model requirements and could be used to detect the QTLs that affect the growth height of *A. thaliana*.

3. Functional mapping

3.1. Statistical model

Let $y_i = c(y_i(1), \dots, y_i(T))$ denote the vector of trait values for RIL, i measured at T time-points. Consider a SNP with two alleles Q and q , generating two genotypes: QQ with n_1 RILs and qq with n_2 RILs. In this study, the likelihood for height growth data of *A. thaliana* is expressed as

$$L(\Phi | y) = \prod_{i=1}^{n_1} f_1(y_i) \prod_{i=1}^{n_2} f_2(y_i) \quad (1)$$

where Φ indicates the unknown parameters including the time-dependent effects of different QTL genotypes and the time-dependent residual variance and correlations. $f_1(y_i)$ and $f_2(y_i)$ are a multivariate normal distribution with a time-dependent mean vector for genotype QQ and qq ,

$$\begin{cases} \mu_1 = (\mu_1(1), \dots, \mu_1(T)) \text{ for } QQ, \\ \mu_2 = (\mu_2(1), \dots, \mu_2(T)) \text{ for } qq. \end{cases} \quad (2)$$

$(T \times T)$ -dimensional longitudinal covariance matrix is expressed as Σ , which can be modeled by using a statistical approach such as the first-order autoregressive [AR(1)] model or an autoregressive moving-average process (ARMA). The maximum-likelihood estimates (MLEs) of the unknown parameters are implemented with the simplex algorithm in R software [48].

3.2. Modeling the mean vector

One of the most important equations for capturing time-specific change in growth is the logistic curve [49, 50], which we used to describe height growth of the QTL genotype according to the following expression:

$$g(t) = \frac{a}{1 + be^{-rt}} \quad (3)$$

where $g(t)$ represents the trait value at time point t , a indicates the asymptotic value of g when $t \rightarrow \infty$, b is a parameter to position the curve on the time axis, and r indicates the relative growth rate. Consequently, any specific growth characteristics described by the logistic growth equation can be captured by parameter a , b , and r , and these can be used to determine the coordinates of biologically important benchmarks along the growth trajectory. The mean vector for the QTL genotypes QQ and qq from time 1 to T in the multivariate normal density function is expressed as:

$$\begin{cases} \mu_1 = \left(\frac{a_1}{1+b_1 e^{-r_1}}, \dots, \frac{a_1}{1+b_1 e^{-r_1 T}} \right) \text{for } QQ, \\ \mu_2 = \left(\frac{a_2}{1+b_2 e^{-r_2}}, \dots, \frac{a_2}{1+b_2 e^{-r_2 T}} \right) \text{for } qq. \end{cases} \quad (4)$$

3.3. Modeling the covariance structure

The first-order autoregressive [AR(1)] model has been successfully applied to model the structure of the within-subject covariance matrix for functional mapping. The AR(1) model includes two basic assumptions: (1) variance stationarity (the residual variance, σ^2) is unchanged over time points and (2) covariance stationarity (the correlation between different time points) decreases proportionally (in ρ) with increased time interval. The AR(1) is described as:

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & L & \rho^{T-1} \\ \rho & 1 & L & \rho^{T-2} \\ L & L & L & L \\ \rho^{T-1} & \rho^{T-2} & L & 1 \end{bmatrix} \quad (5)$$

where $0 < \rho < 1$ is the proportion parameter with which the correlation decays with a time point.

3.4. Hypothesis tests

After the MLEs of the parameters are obtained, the hypothesis concerning the existence of a QTL affecting overall growth can be written as:

$$\begin{cases} H_0: a_1 = a_2, b_1 = b_2, r_1 = r_2 \\ H_1: \text{at least one of the equalities above does not hold} \end{cases} \quad (6)$$

where the null hypothesis H_0 corresponds to the reduced model and the alternative hypothesis H_1 corresponds to the full model. The test statistics for testing the hypotheses is the log-likelihood ratio (LR) of the full over reduced model. The critical threshold is determined from permutation tests.

3.5. Candidate gene function annotation

Gene ontology (GO) annotation analysis was performed using Blast2GO [51]. Finally, R script language programming was used to translate the GO annotation results into charts.

4. Results

4.1. QTL detection

By plotting the total growth against growth week, it was observed that each of the 116 mapped genotypes followed the logistic growth curve. **Figure 3** illustrates the growth trajectories of

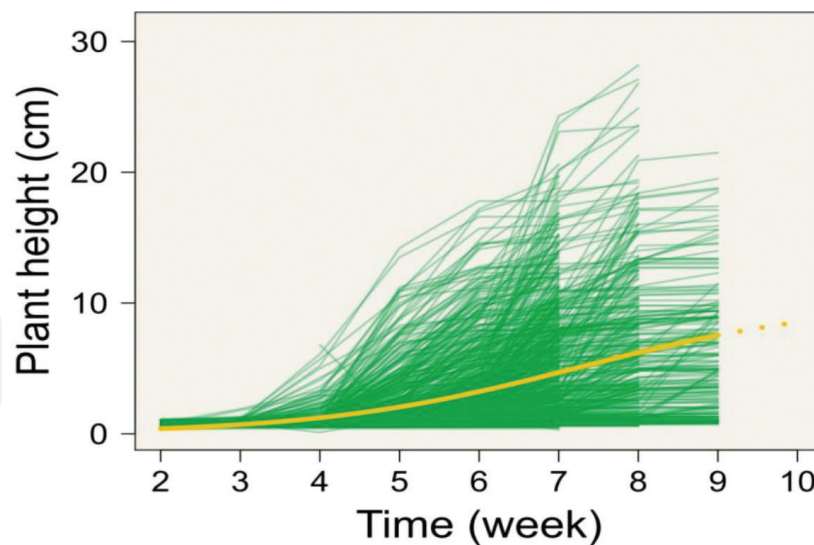


Figure 3. The growth trajectories of height for individual recombinant inbred lines (RILs) are shown in green, whereas yellow lines are the mean growth trajectories of all RILs fitted to a curve. The x- and y-axes of the plot denote time (in weeks) and height (in cm).

height for each RIL over 9 weeks. A least-squares approach was used to fit height growth with the logistic curve (Eq. (3)) for each RIL. Based on statistical tests, all genotypes can be fit by a logistic curve containing parameters a , b , and r ($P < 0.01$, **Figure 3**).

According to the fit results, the relative growth rate of *A. thaliana* was $r = 0.61$, and the asymptotic value of plant height was calculated to be $a = 10.08$. We can see from **Figure 3** that the plant height of *A. thaliana* increases exponentially over a period of time, and as growth time increases, the trend of the curve begins to flatten, until the plant height does not change anymore. In addition, different genotypes showed different growth curves, suggesting the possibility of genetic control over the growth trajectories. The statistical model built upon the logistic growth curve model is used to map QTLs responsible for growth trajectories in plant height.

Functional mapping was implemented to analyze the mapping population. A Manhattan plot of LR values against the genome locations of SNPs distributed throughout the genome is shown in **Figure 4**. We found QTLs that affect the growth trajectory of plant height on the fourth chromosome of the RIL population (**Figure 4**). The genome-wide empirical estimate of the critical value is obtained from permutation tests. The profile of the LR value of the full and reduced model across the length of chromosome 4 has a clear peak from 7.9 to 22.5 kb. A total of 48 significant QTLs responsible for growth trajectories of plant heights were identified. The LR value at this peak is 437.78, which is beyond the empirical critical threshold at the significance level $p = 0.05$.

Functional mapping can also be used to observe the dynamic expression of a QTL over time such as when a QTL starts to affect a growth process. The parameter combination of the QQ genotype was $a_1 = 9.07$, $b_1 = 370.08$, and $r_1 = 0.92$. The parameter combination of the qq genotype was $a_2 = 39.62$, $b_2 = 114.61$, and $r_2 = 0.28$. Different QTL genotypes corresponded to different parameter combinations, which indicate that the QTL controls the developmental

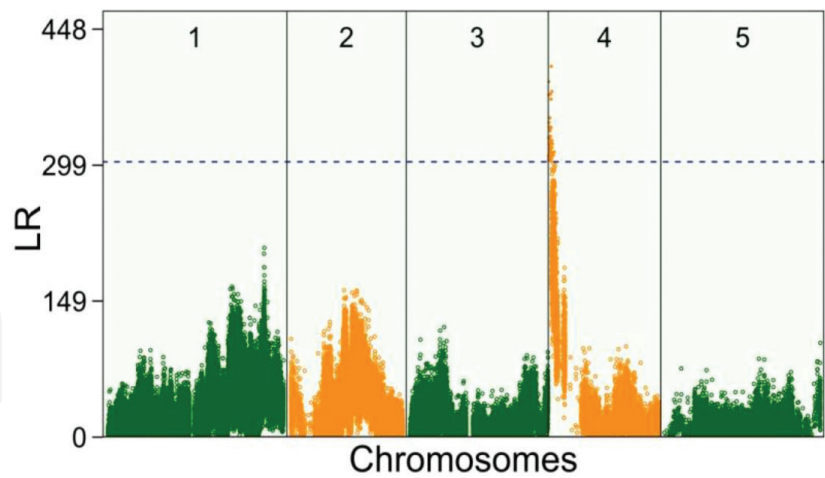


Figure 4. The log-likelihood ratio (LR) of height change of *Arabidopsis thaliana*. The profiles of the LR between the full model (there is a quantitative trait locus [QTL]) and reduced model (there is no QTL) for height growth trajectories throughout the *A. thaliana* genome. The critical thresholds for claiming the genome-wide existence of a QTL are obtained from permutation tests.

trajectory of the plant height. Using the estimates from the growth curves, we drew two different curves, each corresponding to a genotype at each of the detected QTLs on the fourth chromosome (**Figure 5**).

On the basis of the hypothesis test (6), this QTL is detected to be inactive until *A. thaliana* grew to 4 weeks, and its effect on height growth increased with time. At 9 weeks, the genotype *qq* exhibited height growth more than its alternative *QQ*. The effect of *qq* on plant height was more significant than that of *QQ*. If different genotypes at a given QTL correspond to different trajectories, the QTL must affect the differentiation of this trait. Apparently, this mapped QTL interacts significantly with time to affect the height growth of *A. thaliana*.

4.2. Candidate gene function annotation

GO classification is widely used for gene classification and functional annotation, and GO provides three types of semantic terms, including cell component, molecular function, and biological

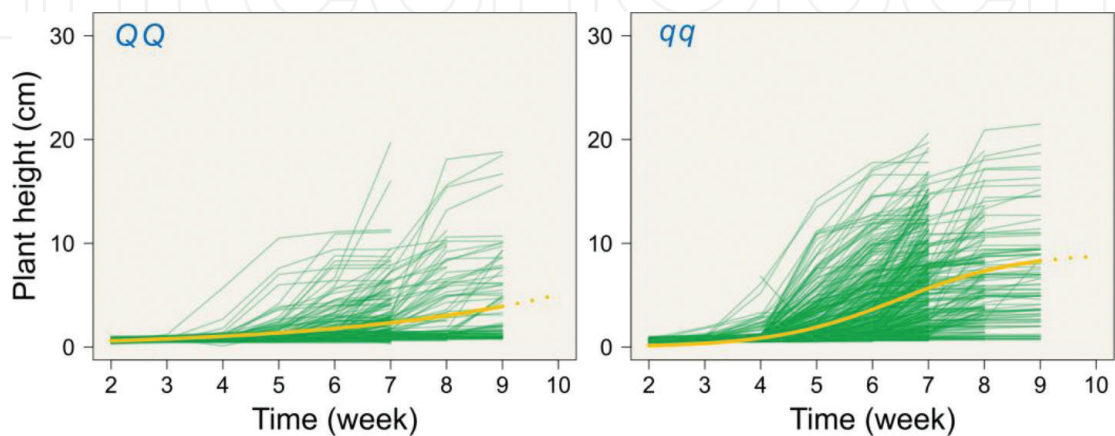


Figure 5. Growth trajectory for different genotypes at the QTLs detected on groups.

process, to describe the characteristics of genes and gene products [52, 53]. To further summarize the data of GO classification, the cytological components include cell structure, tissue, protein complex, extracellular structure, and cell process. Biological processes include developmental processes, physiological processes, regulatory processes, and the processes of responding to stress. Molecular functions include binding, catalysis, activation, structural molecules, and transcriptional regulatory functions (**Figure 6**).

Functional prediction and classification analysis of 48 loci were screened with the National Center for Biotechnology and the Joint Genome Institute (JGI) databases. There are 20 gene families in these loci, which include the F-box and calcium-binding EF-hand protein families. Pathway cluster analysis was used to compare the 48 genes with the known protein sequences in the JGI database.

It was found that 7 of the 48 genes corresponded to functions in the JGI database. They were divided into nine groups, including carbon fixation pathways in prokaryotes, biosynthesis of

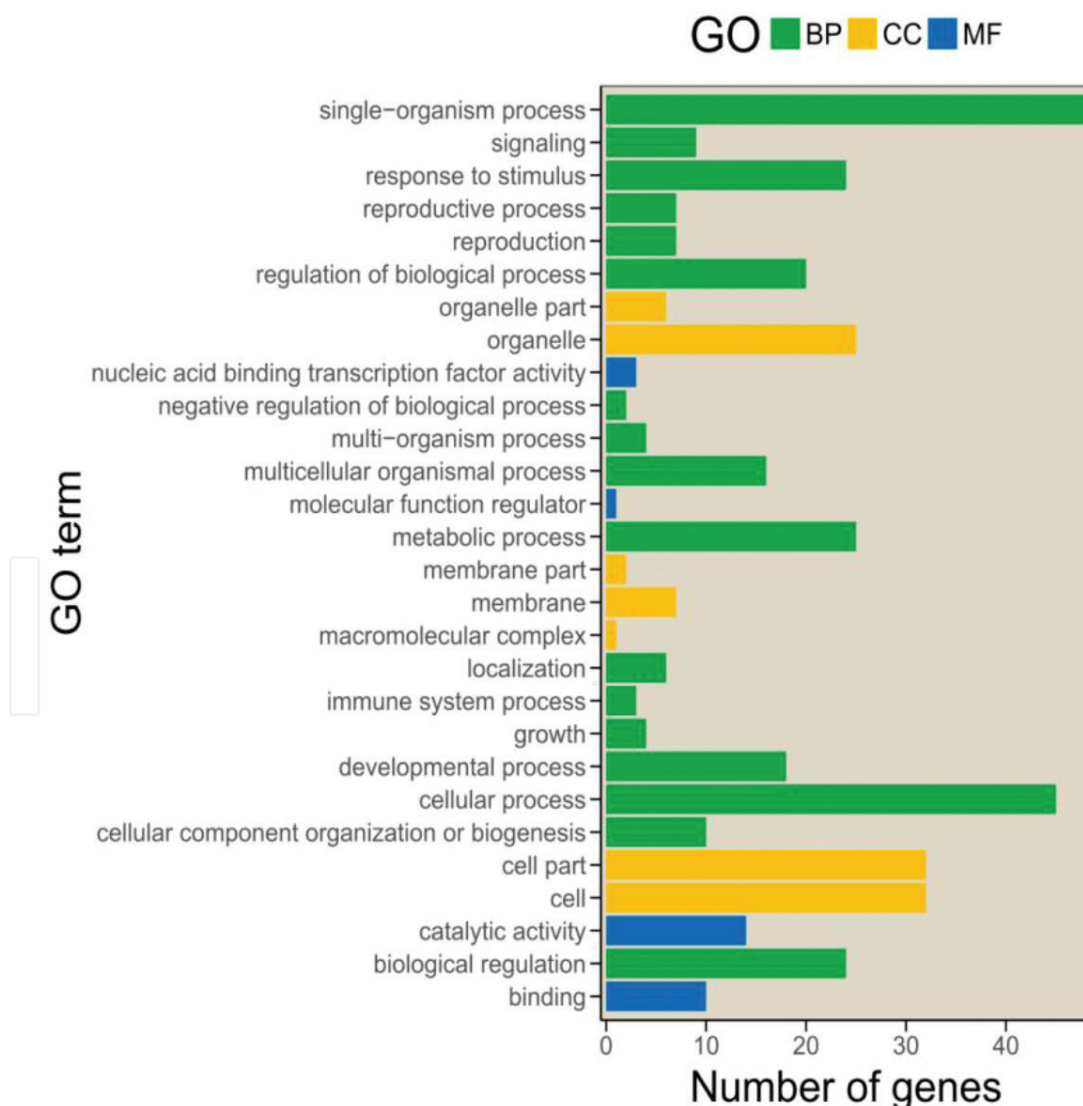


Figure 6. Gene ontology annotations of significant sites.

antibiotics, one-carbon pool by folate, starch, and sucrose metabolism, glycerolipid metabolism (synthase), glycerolipid metabolism (lipase), polyketide sugar unit biosynthesis, streptomycin biosynthesis, and pentose and glucuronate interconversions. We found that these nine pathways are mostly associated with carbohydrate, energy, and amino acid metabolism. These biological processes play an important role in the development of height in *Arabidopsis*.

We also found that AT4G00160.1 encodes an F-box protein in the signal transduction pathway. It has been shown that F-box is an auxin receptor [54]. Plant height development is regulated by gibberellin (GA) and auxin (indole-3-acetic acid [IAA]) [55], and GA20ox and GA30ox are encoded by multiple genes, and mutations in these loci can result in dwarfing of the plant in the later stage of GA biosynthesis [56]. The most significant site, AT4G01150.1, is related to protein curvature thylakoid chloroplastic. Protein curvature thylakoid chloroplastic tends to be located in leaves and stems, and plays an important role in plant photosynthesis, which affects plant growth [57].

5. Discussion

Gene mapping has been shown to be a powerful approach for the study of the genetic architecture of complex traits. It has been instrumental for the characterization of QTLs that control quantitative traits of interest to agriculture, biology, and human disease [58, 59]. However, traditional mapping strategies do not provide much insight into the genetic control mechanisms for phenotypic variations if some statistical and biological issues related to the approach are not resolved. Ma et al. (2002) integrated some fundamental biological principles into the mapping framework, aimed at generating more biologically meaningful discoveries related to trait formation and development, further proposing so-called functional mapping [13]. Functional mapping attempts to combine strong statistical and molecular genetics with the developmental mechanisms of biological features, and to elucidate the genetic mechanisms of complex traits. Since functional mapping combines different mathematical functions with biological significance, it possesses three advantages over traditional mapping methods in QTL mapping: (1) because the underlying biological mechanism is considered, the results of functional mapping are closer to biological reality; (2) a smaller sample size can be used to achieve sufficient accuracy for QTL detection because multiple measurements of the same individual improve mapping accuracy; and (3) by treating the growth process as a smooth curve, a large number of variables can be analyzed simultaneously, and the estimation of a small number of parameters can improve the accuracy of the parameter estimation and flexibility of the model.

With the development of high-throughput sequencing technology and the reduction of sequencing cost, GWAS have become an important tool for studying complex traits and have been widely used in genetic studies of complex traits in humans, animals, and plants [60]. Most GWAS only use single phenotypic data to perform regression analysis with each SNP such as with Plink software [61]. In addition, some GWAS have been developed to solve the false positive loci of population structure and genetic relationship [62–64]. The successes and potential of GWAS have not been explored when complex phenotypes arise as a curve. In any regard, a curve is more informative than a point in describing the biological features of a trait. To apply functional mapping to GWAS by integrating GWAS and functional aspects of

dynamic traits, a new analytical model for genome-wide association analysis of dynamic data, called functional GWAS (fGWAS), has been derived [65]. There are two advantages to fGWAS over GWAS: (1) fGWAS is able to identify genes that determine the final form of the trait and (2) it provides the ability to study the temporal pattern of genetic control over a time course.

The regulation network of plant height traits has been studied intensively in molecular biology. We already know that the development of plant height traits is regulated by growth hormones such as GA and IAA. Using functional mapping, we found 48 growth QTLs in *A. thaliana*. Through the GO annotation of QTLs, we found that there are many genes among the significant loci identified in this study that are related to the pathways for synthesis and conduction of growth hormones, such as AT4G00160.1, which encodes an F-box protein in the signal transduction pathway. It has been shown that the F-box is an auxin receptor. Thus, we can see that the QTLs identified in this study may not only be applicable to *A. thaliana*, but also to other plants. These results show that functional mapping can reveal more intricate details of dynamic traits such as height growth and other phenotypes.

Functional mapping is far from enough to fully study complex traits, and there are still many limitations in describing the developmental pathways leading to the final phenotype and revealing the underlying genetic mechanisms for the formation and development of these traits. It is too simple to draw a complete dynamic diagram of complex traits. Wu [66] extends functional mapping to system mapping. By identifying the dynamic formation process of complex traits as a system and decomposing it into several parts, the QTL that controls the interaction of each component during the development of complex characters is identified. From the point of view of ecology, the process of character formation is extremely complex. To draw a complete quantitative genetic structure, we need to study the characteristics of an organism affected by its own genes as well as the influence from the community partner genome. In nature, most organisms live in groups, and individuals compete with each other. Wu combined game theory with QTL mapping, which opens up new opportunities for improving the accuracy and resolution of complex phenotype QTL recognition [67].

Acknowledgements

This work is supported by National Natural Science Foundation of China (grant 31700576) and grant 201404102 from the State Administration of Forestry of China.

Author details

Kaiyue Liu¹, Wenhao Bo¹, Lina Wang¹, Rongling Wu^{1,2} and Libo Jiang^{1*}

*Address all correspondence to: libojiang@bjfu.edu.cn

1 Center for Computational Biology, College of Biological Sciences and Technology, Beijing Forestry University, Beijing, China

2 Center for Statistical Genetics, Pennsylvania State University, Hershey, PA, USA

References

- [1] Zhong WH, Xin JX. Genetic analysis method of developmental characters QTL. Journal of Henan Institute of Science and Technology (Nature Science Edition). 2008;**36**(04):7-10
- [2] Cheverud JM. Genetics and analysis of quantitative traits. American Journal of Physical Anthropology. 1999;**108**(03):548-549
- [3] Lander ES, Botstein D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics. 1989;**121**(01):185-199
- [4] Zeng ZB. Precision mapping of quantitative trait loci. Genetics. 1994;**136**(04):1457-1468
- [5] Kao CH, Zeng ZB, Teasdale RD. Multiple interval mapping for quantitative trait loci. Genetics. 1999;**136**(04):1457-1468
- [6] Hayes BJ, Pryce J, Chamberlain AJ, Bowman PJ, Goddard ME. Genetic architecture of complex traits and accuracy of genomic prediction: Coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. PLoS Genetics. 2010;**6**(09):e1001139
- [7] Hindorff LA, MacArthur J, Morales J, Junkins HA, Hall PN, Klemm AK, Manolio TA. A Catalog of Published Genome-Wide Association Studies. 2011. www.genome.gov/gwastudies/ [Accessed May 27, 2011]
- [8] Lettre G, Jackson AU, Gieger C, Schumacher FR, Berndt SI, Sanna S, Eyheramendy S, Voight BF, Butler JL, Guiducci C. Identification of ten loci associated with height highlights new biological pathways in human growth. Nature Genetics. 2008;**40**:584-591
- [9] Sun L, Wu R. Mapping complex traits as a dynamic system. Physics of Life Reviews. 2015;**13**:155
- [10] Shilin Y, Yi W, Zu ZZ, Jun XW, Nan C, Guang YX, Jian C. Genome-wide association study on susceptibility genes associated with Yang-deficiency constitution: A small sample case-control study. Chinese Journal of Integrative Medicine. 2015;**21**(08):601-609
- [11] Jiao Y, Zhao H, Ren L, Song W, Zeng B, Guo J, Wang B, Liu Z, Chen J, Li W, Lai J, et al. Genome-wide genetic changes during modern breeding of maize. Nature Genetics. 2012;**44**(7):813
- [12] Leal SM. Genetics and analysis of quantitative traits. American Journal of Human Genetics. 2001;**68**(2):548
- [13] Ma CX, Casella G, Wu R. Functional mapping of quantitative trait loci underlying the character process: A theoretical framework. Genetics. 2002;**161**(04):1751-1762
- [14] Rice SH. The analysis of ontogenetic trajectories: When a change in size or shape is not Heterochrony. Proceedings of the National Academy of Sciences of the United States of America. 1997;**94**(03):907-912
- [15] Zhong WH. Genetic analysis and functional mapping of complex quantitative traits. Henan Agricultural Science. 2006;**35**(11):19-22
- [16] Wu R, Ma CX, Lou XY, Casella G. Molecular dissection of allometry, ontogeny, and plasticity: A genomic view of developmental biology. AIBS Bulletin. 2003;**53**(11):1041-1047

- [17] Zhang B, Tong C, Yin T. Detection of quantitative trait loci influencing growth trajectories of adventitious roots in *Populus* using functional mapping. *Tree Genetics and Genomes*. 2009;**5**:539-552
- [18] Huang Z, Tong C, Bo W, Pang X, Wang Z, Xu J, Wu R, et al. An allometric model for mapping seed development in plants. *Briefings in Bioinformatics*. 2013;**15**(4):562-570
- [19] Li H, Huang W, Wu S. A conceptual framework to mapping quantitative trait loci governing ontogenetic allometry. *PLoS One*. 2007;**2**(08):1245
- [20] Cui Y, Zhu J, Wu RL. Functional mapping for genetic control of programmed cell death. *Physiological Genomics*. 2006;**25**:458-469
- [21] Coelho WS, Li Y. Identification of quantitative trait loci that affect endo reduplication in *Zea mays* endosperm. *Theoretical and Applied Genetics*. 2007;**115**:1147-1162
- [22] Li N, McMurphy T, Berg A. Functional clustering of periodic transcriptional profiles through ARMA. *PLoS One*. 2010;**5**(04):e9894
- [23] Liu T, Wu R. A Bayesian algorithm for: Functional mapping of dynamic traits. *Algorithms*. 2009;**2**:667-691
- [24] Li N, Das K, Wu R. Functional mapping of human growth trajectories. *Journal of Theoretical Biology*. 2009;**261**:33-42
- [25] Lin M, Aquilante C, Johnson J. Sequencing drug response with HapMap. *Pharmacogenomics*. 2005;**5**:149-156
- [26] Lin M, Hou W, Li H. Modeling: Sequence-sequence interactions for drug response. *Bioinformatics*. 2007;**23**:1251-1257
- [27] Somerville C, Koornneef M. A fortunate choice: The history of *Arabidopsis* as a model plant. *Nature Reviews Genetics*. 2002;**3**:883-889
- [28] Koornneef M, Meinke D. The development of *Arabidopsis* as a model plant. *The Plant Journal*. 2010;**61**:909-921
- [29] Krämer U. Planting molecular functions in an ecological context with *Arabidopsis thaliana*. *Elife*. 2015;**4**:1-13
- [30] Brady SM, Orlando DA, Lee JY, Wang JY, Koch J, Dinneny JR, Mace D, Ohler U, Benfey PN. A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science*. 2007;**318**:801-806
- [31] Endo M, Shimizu H, Nohales MA, Araki T, Kay SA. Tissue-specific clocks in *Arabidopsis* show asymmetric coupling. *Nature*. 2014;**515**:419-422
- [32] Simpson GG, Dean C. *Arabidopsis*, the Rosetta stone of flowering time. *Science*. 2002;**296**:285-289
- [33] Valverde F, Mouradov A, Soppe W, Ravenscroft D, Samach A, Coupland G. Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. 2004

- [34] Andres F, Coupland G. The genetic basis of flowering responses to seasonal cues. *Nature Reviews Genetics*. 2012;**13**:627-639
- [35] Bentsink L, Jowett J, Hanhart CJ, Koornneef M. Cloning of DOG1, a quantitative trait locus controlling seed dormancy in *Arabidopsis*. *Proceedings of the National Academy of Sciences of USA*. 2006;**103**:17042-17047
- [36] Graeber K, Nakabayashi K, Miatton E, Leubner Metzger G, Soppe WJ. Molecular mechanisms of seed dormancy. *Plant, Cell & Environment*. 2012;**35**(10):1769-1786
- [37] Deng XW, Matsui M, Wei N, Wagner D, Chu AM, Feldmann KA, Quail PH. COP1, an *Arabidopsis* regulatory gene, encodes a protein with both a zincbinding motif and a G beta homologous domain. *Cell*. 1992;**71**:791-801
- [38] Nemhauser JL. Dawning of a new era: Photomorphogenesis as an integrated molecular. *Current Opinion in Plant Biology*. 2008;**11**:4-8
- [39] Hays JB. *Arabidopsis thaliana*, a versatile model system for study of eukaryotic genome-maintenance functions. *DNA Repair*. 2002;**1**:579-600
- [40] Thomma BP. Disease induction by human microbial pathogens in plant-model systems. *Drug Discovery Today*. 2007;**12**:167-173
- [41] Jones AM, Chory J, Dangl JL, Estelle M, Jacobsen SE, Meyerowitz EM, Nordborg M, Weigel D. The impact of *Arabidopsis* on human health: Diversifying our portfolio. *Cell*. 2008;**133**:939-943
- [42] Schlaich NL. *Arabidopsis thaliana* – The model plant to study host-pathogen interactions. *Current Cancer Drug Targets*. 2011;**12**:955-966
- [43] Doebley JF, Gaut BS, Smith BD. The molecular genetics of crop domestication. *Cell*. 2006;**127**:1309-1321
- [44] Olsen KM, Wendel JF. A bountiful harvest: Genomic insights into crop domestication phenotypes. *Annual Review of Plant Biology*. 2013;**64**:47-70
- [45] Meyerowitz EM, Bowman JL, Brockman LL, Drews GN, Jack T, Sieburth LE, Weigel D. A genetic and molecular model for flower development in *Arabidopsis thaliana*. *Development*. 1991;**1**:157-167
- [46] Mayer U, Torres Ruiz RA, Berleth T, Misera S, Jurgens G. Mutations affecting body organization in the *Arabidopsis* embryo. *Nature*. 1991;**353**:402-407
- [47] Di Laurenzio L, Wysocka-Diller J, Malamy JE, Pysh L, Helariutta Y, Freshour G, Hahn MG, Feldmann KA, Benfey PN. The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root. *Cell*. 1996;**86**:423-433
- [48] Ortutay C, Ortutay Z. *Molecular Data Analysis Using R*. Hoboken, New Jersey: Wiley Blackwell

- [49] Niklas KJ. Plant Allometry: The Scaling of Form and Process. Chicago: University of Chicago Press. 1994
- [50] West GB, Brown JH, Enquist BJ. A general model for ontogenetic growth. *Nature*. 2001;**413**:628-631
- [51] Huang ZX, Ke CH, Chen J. Bioinformatics process of large scale GO annotation. *Journal of Xiamen University (Natural Science Edition)*. 2012;**51**(1):139-143
- [52] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Epping JT, Harrise MA, Hill DP, Issel-Tarver L, Kasarsk A, Lew S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: Tool for the unification of biology. *Gene Ontology Consortium Nature Genetics*. 2000;**25**:25-29
- [53] Berardini TZ, Mundodi S, Reiser L, Huala E, Garacia-Hernandze M, Zhang P, Mueller Yoon J, Doyle A, Lander G, Moseyko N, Yoo D, Xu I, Zoeckley B, Montoya M, Miller N, Weems D, Rhee SY. Functional annotation of the *Arabidopsis* genome using controlled vocabularies. *Plant Physiology*. 2004;**135**:745-755
- [54] Sakamoto TK, Miura H, Itoh T, Tatsumi M. An overview of gibberellin metabolism enzyme genes and their related mutants in rice. *Plant Physiology*. 2014;**134**:1642-1653
- [55] Hedden P, Phillips AL. Gibberellin metabolism: New insights revealed by the genes. *Trends in Plant Science*. 2000;**5**(12):523-530
- [56] Itoh H, Tatsumi T, Sakamoto T, Otomo K. A rice semi-dwarf gene, Tan-Ginbozu (D35), encodes the gibberellin biosynthesis enzyme, ent-kaurene oxidase. *Plant Molecular Biology*. 2004;**54**:533-547
- [57] Li P, Liu Y, Wang H, He Y, Wang X, Lv F, Chen H, Pang X, Liu M, Shi T, et al. PubAngioGen: A database and knowledge for angiogenesis and related diseases. *Nucleic Acids Research*. 2015;**43**:D963-D967
- [58] Niedziela A, Bednarek PT, Labudda M, Mańkowski DR, Anioł A. Genetic mapping of a 7R al tolerance QTL in triticale (x *Triticosecale* Wittmack). *Journal of Applied Genetics*. 2014;**55**(01):1-14
- [59] Recla JM, Robledo RF, Gatti DM, Bult CJ, Churchill GA, Chesler EJ. Precise genetic mapping and integrative bioinformatics in diversity outbred mice reveals Hydin as a novel pain gene. *Mammalian Genome*. 2014;**25**(5-6):211-222
- [60] Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics*. 2005;**6**(2):95-108
- [61] Christopher CC, Carson CC, Laurent CT, Shashaank V, Shaun MP, James JL. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience*. 4(1):7
- [62] Yu J, Pressoir G, Briggs WH. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*. 2006;**38**(02):203-208

- [63] Klasen JR, Barbez E, Meier L. A multi-marker association method for genome-wide association studies without the need for population structure correction. *Nature Communications*. 2016;**7**:13299
- [64] Stephan J, Stegle O, Beyer A. A random forest approach to capture genetic effects in the presence of population structure. *Nature Communications*. 2015;**6**:7432
- [65] Stefan E, Lorenzo BJ. Handbook of Adaptive Designs in Pharmaceutical and Clinical Development. A. Pong and S. C. Chow (2010). Boca Raton: Chapman & Hall/CRC Press. *Biometrical Journal*. 2011;**53**(4):708-709
- [66] Wu R, Cao JG, Huang ZW, Wang Z, Gai JY, Vallejos E. Systems mapping: How to improve the genetic mapping of complex traits through design principles of biological systems. *BMC Systems Biology*. 2011;**5**:84
- [67] Zhu X, Jiang L, Ye M. Integrating evolutionary game theory into mechanistic genotype-phenotype mapping. *Trends in Genetics*. 2016;**32**(05):256-268