



HHS Public Access

Author manuscript

Phys Life Rev. Author manuscript; available in PMC 2017 June 12.

Published in final edited form as:

Phys Life Rev. 2015 June ; 13: 155–185. doi:10.1016/j.pprev.2015.02.007.

Mapping complex traits as a dynamic system

Lidan Sun and Rongling Wu

Center for Statistical Genetics, Departments of Public Health Sciences and Statistics, The Pennsylvania State University, Hershey, PA 17033, USA

Abstract

Despite increasing emphasis on the genetic study of quantitative traits, we are still far from being able to chart a clear picture of their genetic architecture, given an inherent complexity involved in trait formation. A competing theory for studying such complex traits has emerged by viewing their phenotypic formation as a “system” in which a high-dimensional group of interconnected components act and interact across different levels of biological organization from molecules through cells to whole organisms. This system is initiated by a machinery of DNA sequences that regulate a cascade of biochemical pathways to synthesize endophenotypes and further assemble these endophenotypes toward the end-point phenotype in virtue of various developmental changes. This review focuses on a conceptual framework for genetic mapping of complex traits by which to delineate the underlying components, interactions and mechanisms that govern the system according to biological principles and understand how these components function synergistically under the control of quantitative trait loci (QTLs) to comprise a unified whole. This framework is built by a system of differential equations that quantifies how alterations of different components lead to the global change of trait development and function, and provides a quantitative and testable platform for assessing the multiscale interplay between QTLs and development. The method will enable geneticists to shed light on the genetic complexity of any biological system and predict, alter or engineer its physiological and pathological states.

Keywords

Complex trait; quantitative trait loci; genetic mapping; functional mapping; systems mapping; network mapping; statistical models; differential equations

1. Introduction

The past two decades have witnessed increasing applications of quantitative genetics to a wide spectrum of life sciences from plant and animal breeding to clinical medicine [16,101,118,142]. However, a considerable body of research has shown that quantitative traits are extremely difficult to study because their formation involves many unknown

Corresponding author: Rongling Wu, 500 University Dr., Penn State College of Medicine, Hershey, PA 17033, USA, Tel: 001-(717)-531-2048, rwu@phs.psu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

physiological mechanisms that guide or are guided by the underlying genetic factors that operate in a complicated way [7,100]. By regressing phenotypic values of traits directly on molecular markers from the genome, Lander and Botstein [78] have pioneered an approach for mapping and identifying specific genetic loci, known as quantitative trait loci (QTLs), that contribute to trait variation. Depending on the type of segregating populations used, this approach is called linkage mapping for controlled crosses or association mapping for natural populations [110]. With the increasing availability of inexpensive DNA sequencing and genotyping techniques, it has become a routine tool to dissect the genetic architecture of complex traits, providing unprecedented promises to construct the genotype-phenotype predictive map [62]. However, thousands of thousands of significant QTLs identified so far in a variety of species by this approach have gained little mechanistic insight because a majority of these loci have not been translated into genes and pathways [20]. The translation of functional QTLs requires knowledge of how they act and interact through a series of biochemical pathways toward the end-point phenotype.

To understand the genetic control of QTLs over the process of trait formation, a dynamic model, called functional mapping, has been developed [98,174] and recognized as an important approach for genetic mapping [69,131,141,188]. By integrating the dynamic pathways underlying phenotypic formation using mathematical equations, this model is renovated to identify QTLs involved in rate-limiting processes and to quantify the dynamic effect pattern of these genes across a time and space scale [58,89]. More recently, functional mapping has been extended to systems mapping by viewing a phenotype as a dynamic system [46,47,176]. The key insight of systems mapping is that the dynamics of a complex system depends on how its elements causally influence each other by means of QTLs. By identifying QTLs that determine information flows between different elements, systems mapping can reconstruct a genotype-phenotype map from developmental pathways [12].

Many existing genetic approaches are built on a direct genotype-phenotype association. Although this is a simple strategy easy to be used, it neglects the biology inside the “black box” that links genotype and phenotype through causal networks of interacting genes and pathways. Several authors have recognized the essentiality of incorporating transcript, protein and metabolite abundance into genotype-phenotype prediction models and constructing transcriptional and regulatory networks affecting high-order phenotypes [20,101]. With these established networks, the causative and downstream effects of DNA sequences on phenotypic variation can be clearly understood by perturbing gene expression, proteins and metabolites that play a critical role in the connectivity of DNA variation through endophenotypes to the end-point phenotype.

Figure 1 illustrates a big picture of the formation process of a complex trait from DNA to a final phenotype through a cascade of regulatory pathways. This picture presents a general system of information flow applicable for any trait or disease, but its implementation into a practical genetic study is extremely difficult, if not impossible. At the current level of biotechnology, however, it is feasible to dissolve the whole process of trait formation into multiple continuous smaller-scale systems, in each of which the comprising elements can be readily identified from prior knowledge, and further connect these systems in tandem as a functional whole. A system is defined by the elements that it constitutes, the interactions

between these elements, and the natural rules of the system [113,147]. The system rules operate through individual or subsystem elements, but are effective only at the entire system level. How the function of a system is recognized and emphasized depends on the investigator or user's perspective [5].

In this review, we present a general philosophy of mapping complex traits by broadly dissecting trait formation into its three interactive underlying systems, (1) a static system of genetic architecture composed of DNA-based variants, (2) a dynamic system of morphogenesis from early to adult phenotypes, and (3) a dynamic system of regulatory networks originated from DNA and ended at the molecules as precursors for synthesizing the phenotype. Phenotypic formation of one organism may not only be affected directly by its own genes, but also indirectly by genes of its conspecific in a community. We also discuss how a dynamic system of ecological interactions determines the phenotypic variation of a complex trait. We describe the basic principle of dissecting each of these systems and the computational framework of modeling the structure, organization and function of a system through differential equations. As an approach of genetic mapping, we pinpoint several commonly used types of mapping populations that accommodate to key characteristics of different species. To the end, this review provides a dynamic strategy for using the genotype of an organism to predict its phenotype across a range of environments. Our review does not simply provide literature documentation for results of genetic mapping, rather than attempts to provide a broad picture of mapping complex traits from a systematic standpoint, which is inevitably unbalanced in the coverage and references of the presentation. Despite their significant contributions to genetic mapping, we apologize in advance to those authors whose work is not cited in this review.

2. Modeling the genetic architecture of complex traits

There has been a wealth of seminal reviews on the status and opportunities of studies on genetic architecture by leading quantitative geneticists [7,61,63,99–101]. Hansen [55] discussed the pattern of how genetic architecture evolves and how it influences evolution. All the approaches for studying genetic architecture are to link individual DNA variants to the end-point phenotype (Fig. 1). Here, we define the genetic architecture of a complex trait from a systems perspective and review statistical methods for estimating the structure and organization of such a high-dimensional system.

2.1 A traditional strategy of mapping complex traits

The genetic theory for analyzing quantitative traits has well been established since the early twenty century [42]. This theory, equipped with statistical methods, such as the analysis of variance and path coefficients invented by Fisher [42] and Wright [170], enables geneticists to estimate the resemblance between relatives and partition phenotypic variation into its genetic (including additive, dominant and epistatic) and environmental components. By implementing the quantitative genetic theory into the design of experiment derived from various mating designs, Cockerham [21] translated the estimates of experimental variances into the estimates of additive and dominant variance components, largely expanding the practical application of the original theory to dissecting quantitative traits. Wu [173] further

integrated clonal replicates, which can be available for many species, such as forest trees, into Cockerham's [21] analytical model to extract part of the epistatic genetic variance.

Pure quantitative genetic analysis cannot reveal the "black box" behind quantitative genetic variation that is believed to involve many loci with unknown actions and interactions [33,97,101] until the revolutionary emergence of DNA recombinant technologies in the 1980s. Taking advantage of the rapid development of statistical and computational techniques, Lander and Botstein [78] pioneered an innovative approach for dissecting quantitative traits into their underlying genetic components, i.e., QTLs, at the DNA level. This so-called interval mapping has quickly demonstrated its widespread application in quantitative genetic studies for a wide range of species including microorganisms, plants, animals and humans (reviewed in [7,33,61,63,99–101]) and has also received extensive modifications and reformations for different purposes [15,68,71,102,112,129,186,187,191,194,198]. While genetic mapping is usually based on a genetic linkage map constructed from molecular markers, which may not have a detailed coverage of the genome, a more in-depth genetic approach, genome-wide association studies (GWAS), has emerged as a powerful tool since high-throughput single nucleotide polymorphism (SNP) genotyping techniques have become available [57,62,82]. GWAS attempts to search for all possible genes that control a complex trait or disease in a way a fisherman fishes from a big pool of water by using a high-density web.

As shown in Fig. 1, a conventional analytical strategy of genetic mapping and GWAS is based on a simple regression model that directly associates individual markers with a phenotype without considering the intermediate process from DNA to the end-point phenotype. This analysis is conducted on a controlled cross of segregating individuals, such as the backcross or F2, initiated with two inbred lines, with the model expressed as

$$y_i = \mu + \xi_i a_j + \zeta_i d_j + e_i \quad (1)$$

where y_i is the phenotypic value of individual i ; μ is the overall mean; a_j and d_j are the additive and dominant genetic effects of the j th QTL, respectively; ξ_i and ζ_i are the indicator variables that define the genotype of a QTL, expressed as

$$\xi_i = \begin{cases} 1 & \text{If the QTL is the homozygote } AA \\ 0 & \text{If the QTL is the homozygote } Aa \\ -1 & \text{If the QTL is the homozygote } aa \end{cases} \quad (2)$$

$$\zeta_i = \begin{cases} 1 & \text{If the QTL is the homozygote } Aa \\ 0 & \text{If the QTL is the homozygote } AA \text{ or } aa \end{cases} \quad (3)$$

and e_i is the residual error, assumed to be normally distributed with mean 0 and variance σ^2 .

Statistical methods, such as maximum likelihood or regression analysis, have been implemented to estimate the genetic effects and other model parameters [78]. After these parameters are estimated, hypothesis tests are performed to test whether the genetic effects are significant singly or jointly. In genetic mapping, permutations tests are usually used to determine the genome-wide critical threshold for claiming the significance of the QTLs detected. For GWAS, each SNP is subject to statistical analysis by the regression model (1), obtaining a p -value from the significance test. Put together, the p -values are plotted against the genome locations of all SNPs to produce a Manhattan plot, from which a final set of significant SNPs are determined after the adjust for multiple comparisons using Bonferroni correction, false discovery rate or other approaches.

As an example, Figure 2 shows the application of genetic mapping and GWAS in an F_1 family of interspecific hybrids in poplar. By scanning a linkage group based on the log-likelihood ratio (LR) test, a peak, beyond the critical threshold at the 5% significance level determined from permutation tests, was identified between markers AG/CGC-235R and GT/CAT-430RD, suggesting the existence of a significant QTL for stem diameter growth in this marker interval (Fig. 2A). The same hybrid population, genotyped for about 110,000 SNPs, was analyzed for the same trait by GWAS (R. Wu and M. Huang, unpublished data). From the Manhattan plot obtained, we identify several significant loci at the 5% significant level after Bonferroni correction (Fig. 2B). As like in genetic mapping, we need to estimate the proportion of total phenotypic variance explained by each significant locus. This proportion is called single-QTL heritability that contributes to the overall heritability of 11-year stem diameter.

Linkage mapping and association studies have provided a powerful tool to illustrate a trait's genetic architecture by identifying the numbers and genome locations of genes that affect the trait, the magnitude of their unique effects and pleiotropic effects, and the relative contributions of additive, dominant, and epistatic genetic effects. The traditional single-locus/single-phenotype analysis used, however, is not able to provide a precise picture of genetic architecture for complex traits because it is not implemented with the intrinsic complexity of genetic control. It has been widely recognized that complex traits are determined simultaneously by a number of genes, each acting singly or interacting with other genes and environmental factors to different degrees [33,97]. Such multifactorial nature of a complex trait has made the description of its underlying genetic architecture a difficult and challenging task.

2.2 Genetic actions and interactions

A gene exerts its effects on a complex trait in a complicated manner. It may act singly through main effects, or interact with other genes via epistasis, forming a network of actions and interactions. For a single gene, its main effect can be partitioned into additive and dominant components, whereas the epistasis of a given pair of genes includes additive \times additive, additive \times dominant, dominant \times additive and dominant \times dominant interactions [42]. By charting such a genetic network, we can comprehend the genetic control mechanisms of a complex trait and elucidate the rules for translating genetic variation to the phenotypic variation of the trait. We diagrammatize the genetic architecture of botanical

traits for a plant composed of all possible genes distributed on its total five chromosomes (Fig. 3). Some genes are collectively distributed on a small region (e.g., chromosome 2), whereas the others have a sporadic distribution throughout a wide region of the genome. Some interact with many other genes (e.g., gene 1), some only have a few interactions (e.g., gene 2), and the others have no interactions (e.g., genes 24 and 25). Some genes tend to interact with those within the same region (e.g., genes on chromosome 5), whereas others only interact with those from different chromosomes (e.g., genes on chromosome 1).

Pleiotropy, believed to play an important role in the genotype-phenotype relationship, is the other main element in genetic architecture. A gene is thought of being pleiotropic if it affects two or more traits at the same time [146]. Of the genes that affect an above-ground trait of the plant, some also pleiotropically influence its below-ground trait (e.g., genes 2, 3, 15 among others; Fig. 3). These genes are one of the causes for the genetic correlation between these two types of traits. Of course, the two traits are correlated partly because of the genetic linkage between two genes, each controlling a different trait (e.g., two adjacent genes 3 and 4). Although the below-ground trait is highly related to the above-ground trait, they obviously present two different modules because of distinct discrepancy in their underlying genetic basis. As shown in Fig. 3, a large portion of genes that affect the below-ground trait are not involved in variation in the above-ground trait. On the other hand, the genes for the above-ground trait are generally also responsible for the below-ground trait. This difference suggests that this particular below-ground trait is more autonomous than the above-ground trait assumed.

Given that the complexity of a phenotypic trait arises from the highly interactive relationships of its underlying components, it is reasonable to hypothesize that the metabolic pathways for trait formation involve multiple interacting gene products and regulatory loci [106]. Three-way interactions among different QTLs were believed to be involved in metabolic pathways toward a complex trait [134]. A mathematical approach was derived to describe multi-way genetic interactions to study the genetic structure of fitness landscapes for *Escherichia coli* [9]. A genome-scale knockout design was framed to detect high-order epistatic relationships between components of large metabolic networks [66]. Hansen and Wagner [56] showed that higher-order genetic interactions play a pivotal role in trait phenotype if the total genomic mutation rate is large and the interaction density among loci is not too low.

2.3 Genetic imprinting and epigenetic actions

Many recent studies have unraveled the importance of genomic imprinting and epigenetic modification as a mechanism for genetic variation [19,90,92,148,149]. Genomic imprinting is a genetic phenomenon by which certain genes are expressed or repressed depending on their parental origin [79]. Violating the classical Mendelian inheritance, imprinted genes are either expressed only from the allele inherited from the mother, such as H19 or CDKN1C, or from the allele inherited from the father, such as IGF-2. From a quantitative genetic perspective, genomic imprinting may provide organisms with evolutionary merits by contributing additional genetic variation and conferring a fitness benefit in changing environments [120]. Nowadays, different forms of genomic imprinting have been detected in

a variety of species and thought to play an important role in regulating crucial aspects of embryonic growth and development as well as pathogenesis. Recent bioinformatic analyses suggest that the number of imprinted genes may be higher than we thought previously but this remains to be demonstrated experimentally [67].

One of the forces causing genetic imprinting is systematic or stochastic changes in chromatin states, such as DNA methylation, chromatin remodeling, histone modification and RNA interference [17,157]. These epigenetic modifications have also been thought to provide an additional driving force for phenotypic variation in complex traits and diseases [31,39,40,70,122,123]. Different chromatin states, called epialleles, that occur in the same sequence allele cannot be captured by an analysis based on DNA sequence alone (Johannes et al. 2008). With the increasing availability of epigenome technologies, there has been an unprecedented opportunity to understand the role of epiallelic variants in maintaining and inducing functional variation for organisms to better buffer against environmental perturbations. Like genetic imprinting, epigenetic variants should also be considered for elucidating the genotype-phenotype map.

Genetic mapping has been used as an appealing approach to discovering imprinted QTLs (iQTLs) based on particular mating designs. Basically, these designs include outcrossed families [29,74], sex-specific linkage analysis [19,25], reciprocal crosses [90,148,149,192], multigenerational pedigrees [92], open-pollinated populations [138] and case-control studies [87]. From these designs, significant imprinted QTLs were detected for body composition and body weight in pigs, chickens, sheep and canine, endosperm traits in maize, and stem growth in hickory tree. Computational models have also been developed to detect and map epiQTLs, i.e, those QTLs at which epigenetic marks are differentially expressed in response to environmental changes [137]. Wang et al. [155] proposed a model to estimate the relative contribution of epiQTLs to overall genetic variation in a natural population.

2.4 Genotype-environment interactions

In order to adapt to the new or fluctuating environment, the organism will often change its phenotypic expression, a phenomenon called phenotypic plasticity [126,144]. Differences in the capacity of phenotypic plasticity among individuals cause genotype-environment interactions. This explains a vast body of observations of why the genetic effect of a QTL detected on a phenotypic trait depends on the environmental and developmental context, such as temperature, irradiation, nutrition, or age [101,154]. Obviously, genotype-environment interactions are a key factor that determines the pattern of the genetic architecture of a complex trait. The investigations of genotype-environment interactions will not only help to design an efficient breeding program as like a traditional consideration, but also are being integrated into conceptual model construction for evolutionary, climate change and clinical studies [8,32,108].

Special attention has been paid to study the genetic mechanisms of phenotypic plasticity [127]. More recently, El-Soda et al. [32] summarized five genetic models proposed to explain the genetic basis of phenotypic plasticity: (1) Overdominance: phenotypic plasticity is negatively correlated with the number of heterozygous loci; i.e., the more homozygous a genotype, the more plastic it is. (2) Allelic sensitivity or incomplete pleiotropy: the

magnitude of allelic effects at a gene varies among different environment; the gene has a pleiotropic but different effect on the same trait expressed in different environments. (3) Epistasis: loci causing plasticity regulate other genes through epistatic interactions to be switched on or off in a particular environment. (4) Linkage: alleles responsible for plasticity may be linked with alleles that produce either low or high fitness. (5) Epigenesis: an epiallele of a gene becomes sensitive to the environment due to chromatin modification and DNA methylation, while the standard allele is stable. All possible models implicate the mechanistic complexity of genotype-environment interactions, making it essential to chart genetic architecture from a systems perspective.

Most studies of genotype-environment interactions focus on genetic variation in phenotypic plasticity between two discrete environments [32]. Such studies are not able to unravel phenotypic response, or reaction norms, to a gradient of environmental conditions [145]. Wang et al. [154] implemented a dynamic model to study the genetic architecture of the gradient expression of a complex trait across a range of environments. This model was framed on mathematical aspects of phenotypic plasticity, equipped with a capacity to unravel the quantitative attribute of trait response to the environment. By testing the curve parameters that specify reaction norm trajectories, the model enables geneticists to test a series of fundamental hypotheses a quantitative way about the interplay between genes and environmental sensitivity. The model can also make the dynamic prediction of genetic control over phenotypic plasticity within the context of changing environments.

2.5 Direct and indirect genetic effects from genome-genome interactions

Increased recognition has been given to the role of genetic interactions between QTLs from different genomes in both direct and indirect ways [75,77,168,196]. For example, in flowering plants that undergo a double-fertilization process leading to the differentiation of the embryo and endosperm in the seed, seed traits are controlled simultaneously by three genomes, diploid offspring (embryo), triploid offspring (endosperm), and the maternal plant which carries the seed. While offspring genomes confer direct genetic effects on the seed, the maternal genome provides an indirect genetic effect. Seed development is thus the result of a mosaic of different gene expression programs occurring in parallel in different compartments. Wu and group have developed a series of statistical models that can characterize the genetic effects on seed traits derived from the maternal genome, the offspring genome and maternal-offspring interactions [22–24,26,150,178,180]. Likewise, given that a tumor is supplied with nutrients from normal tissues, its growth should depend on an interactive effect of genes directly from the normal tissue and indirectly the tumor tissue [88].

When different organisms grow in the same environment, the phenotype of an organism may be determined not only by its own genome, but also by the genome of the other organisms around it and genetic interactions between different genomes. In a gene expression study of plant-pathogen interactions, Zhu et al. [196] identified the genetic architecture of plant resistant traits composed of genes from the host and pathogen. Kolenbrander et al. [75] reviewed evidence for intimate interactions between genomes of the human and microorganisms in dental plaque communities. Lambrechts [77] argued that the phenotype

of viral infection does not merely result from additive effects of host and pathogen genotypes, but also from a specific interaction between the two genomes. Wolf et al. [169] integrated genetic mapping and structural equation models to detect a QTL for flowering time in focal plants that pleiotropically affects the expression of developmental traits in the neighbor plants. Also, there has been a body of other literature that reports the detection of QTLs for phenotypic traits related to interspecific interactions, such as immune response [80], tolerance to herbivory [160], mate recognition [130], or predator-defense traits [96], by using conventional mapping approaches.

3.6 High-dimensional model of inferring a complete picture of genetic architecture

Despite the complexity of the underlying process from genes to phenotype, the vast majority of genetic analysis is built around a very simple model of associating genotype with phenotype. To overcome this simplicity of trait analysis, a more sophisticated statistical model based on high-dimensional variable selection has been developed, allowing a large number of SNPs to be analyzed simultaneously [83]. This model involves a two-stage procedure for multi-SNP modeling and analysis in GWAS, by first producing a “preconditioned” response variable using a supervised principle component analysis and then formulating Bayesian LASSO to select a subset of significant SNPs. This model has proven to be particularly powerful for selecting the most relevant SNPs for GWAS where the number of predictors exceeds the number of observations.

More recently, high-dimensional variable selection has been extended to an ultrahigh-dimensional scale which enables the modeling and estimation of SNP-SNP interactions in GWAS [85]. The extended model incorporates a two-stage sure independence screening (TS-SIS) procedure [36]: first, to generate a pool of candidate SNPs and interactions served as predictors to explain and predict the phenotypes of a complex trait; and second, regularization regression methods, such as LASSO or smoothly clipped absolute deviation (SCAD) [34,35], are applied to further identify important genetic effects. A rates adjusted thresholding estimation (RATE) approach was used to control the false positive rate of the selected model by a general independent screening procedure.

Consider a GWAS in which a simple regression model (1) was extended to include all possible main effects (including the additive and dominant) and interaction effects, as well as all possible covariates. Thus, the phenotypic value of a particular individual i can be expressed as

$$y_i = \mu + \sum_{k=1}^q x_{ki} \alpha_k + \sum_{j=1}^p \xi_{ji} a_j + \sum_{j=1}^p \zeta_{ji} d_j + \sum_{j=1}^p \sum_{j'=1}^p \xi_{ji} \xi_{j'i} I_{jj'}^{aa} + \sum_{j=1}^p \sum_{j'=1}^p \xi_{ji} \zeta_{j'i} I_{jj'}^{ad} + \sum_{j=1}^p \sum_{j'=1}^p \zeta_{ji} \xi_{j'i} I_{jj'}^{da} + \sum_{j=1}^p \sum_{j'=1}^p \zeta_{ji} \zeta_{j'i} I_{jj'}^{dd} + \varepsilon_i, \quad i=1, \dots, n; j < j' \quad (4)$$

where μ is the overall mean, x_{ki} is the k th covariate for individual i , which could be either discrete or continuous, α_k is the effect of the k th covariate ($k = 1, \dots, q$), a_j and d_j are the additive effect and dominant effect of the j th SNP ($j = 1, \dots, p$), respectively, $I_{jj'}^{aa}$, $I_{jj'}^{ad}$, $I_{jj'}^{da}$

and $I_{jj'}^{dd}$ are the additive \times additive, additive \times dominant, dominant \times additive and dominant \times dominant epistatic effects between SNP j and j' ($j < j' = 1, \dots, p$), respectively, and ε_j is the residual error assumed to follow an $\mathcal{N}(0; \sigma^2)$ distribution. For individual i , ξ_{ji} and ζ_{ji} are the indicators of the additive and dominant effects of the j th SNP, respectively, as explained in equations (2) and (3).

Let D_a and D_d denote a set of additive and dominant effect indices, respectively. The first round SIS is conducted on $D_a^{(0)} = \{1, \dots, p\}$ and $D_d^{(0)} = \{1, \dots, p\}$ for all SNPs. After the first stage SIS, a subset of SNPs with potential nonzero additive ($D_a^{(1)}$) and dominant effects ($D_d^{(1)}$) is selected. Note that each SNP in any subset may either have true nonzero main effects, or serve as a root in a two-way interaction. Let $(D_{aa}^{(2)})$, $(D_{ad}^{(2)})$, $(D_{da}^{(2)})$ and $(D_{dd}^{(2)})$ denote a set of selected roots for additive \times additive, additive \times dominant, dominant \times additive and dominant \times dominant epistatic effects, respectively.

After two-stage sure independence screening, the dimensionality of GWAS model has greatly reduced. In order to precisely select important SNPs and epistatic interactions from a pool of candidate effects, penalized regressions widely used in main-effect analysis could be incorporated here. Specifically, we put penalties on the sizes of additive, dominant and all epistatic effects and minimize the following penalized least squares:

$$\begin{aligned} & \frac{1}{2n} \|y - Ey\|^2 + \sum_{j \in D_a^{(1)}} p_\lambda(|a_j|) + \sum_{j \in D_d^{(1)}} p_\lambda(|d_j|) \\ & + \sum_{j \in D_{aa}^{(2)}} p_\lambda(|I_{jj'}^{aa}|) + \sum_{j \in D_{ad}^{(2)}} p_\lambda(|I_{jj'}^{ad}|) + \sum_{j \in D_{da}^{(2)}} p_\lambda(|I_{jj'}^{da}|) + \sum_{j \in D_{dd}^{(2)}} p_\lambda(|I_{jj'}^{dd}|) \end{aligned} \quad (5)$$

where the penalty function $p_\lambda(\cdot)$ is implemented to shrink sufficiently small effects to zero and thus exclude the inactive predictors. The SCAD penalty least squares function, characterized by unbiasedness, continuity and sparsity properties [36], has been used to solve model (5), with the estimates of regression coefficients being estimated with the aid of local linear approximation. Note that the SCAD penalized least squares (5) can be easily minimized using L_1 penalized regression.

Results from simulation studies show that the TS-SIS procedure is computationally efficient and has an outstanding finite sample performance in selecting potential SNPs as well as gene-gene interactions. The procedure was used to analyze an ultra-high dimensional GWAS dataset from the Framingham Heart Study [28], leading to the detection of 32 significant SNPs including 42 epistatic interactions for body mass index (BMI) (Fig. 4). We summarized the results as follows: (1) Epistasis appears to be distributed randomly throughout the genome, although a few SNPs, such as ss66142093 on chr 3, ss66249128 and ss66468842 on chr 7 tend to interact with many other SNPs. (2) Active epistasis may not be due to interactions between two SNPs both of which display active marginal effects. Of the 42 selected pairs, there are five cases in which both SNPs have active marginal effects and there are 21 cases in which only one SNP has an active marginal effect whereas the

counterpart has not. There are as many as 16 pairs in which no SNP is active for its marginal effect. Notably, the DD interaction between SNP ss66249128 on chr 7 and SNP ss66451087 on chr 12 can explain 2.32% of the BMI variation, although the latter is marginally uncorrelated with BMI. In the presence of SNP ss66451087, the D effect of SNP ss66249128 is dramatically impacted.

Since this ultrahigh-dimensional GWAS model can analyze a large number of SNPs and interactions simultaneously, the resulting discoveries should be statistically powerful to chart the genetic architecture of a complex trait than those from traditional single-SNP approaches. Genetic loci detected by this approach show high more biological relevance. For example, SNP ss66142093 on chromosome 3, detected to explain 2.97% single-locus heritability, is near a candidate gene ANAPC13 involved in pathways for bone and cartilage development that affects human height and stature through cell cycle regulation and mitosis [158]. Also, significant loci detected on chromosomes 3, 4, 6, 9 and 12 are within the genes responsible for type 2 diabetes, a disease highly correlated with BMI [44]. These results demonstrate the capability of the TS-SIS model to resolve the complexity of genetic control mechanisms.

3. Developmental dissection of complex traits

The formation of every trait undergoes a series of developmental changes in an organism's ontogeny. For example, in the lifecycle of a soybean (Fig. 5), plant development includes a broad spectrum of processes, i.e., the formation of a complete embryo from a zygote, seed germination, the elaboration of a mature vegetative plant from the embryo, the formation of flowers, fruits, and seeds, and many of the plant's responses to its environment. Each of these processes is fundamental to determine the size, shape and production of a higher plant. Knowledge of the genetic basis of the variation in each process is important for understanding adaptive evolution and deriving elite domestic crop varieties. Traditional approaches map a phenotype by measuring it at a fixed time point (Fig. 1), view the process of trait formation as a "black box" and, thus, fail to capture the dynamic structure and pattern of the process. A conceptual functional mapping model has been derived to address this issue by considering biological mechanisms and processes of phenotypic formation ([98,174]; Fig. 1). By measuring a phenotypic trait repeatedly at a series of time points, a biologically meaningful curve, e.g., growth curve, can be used to fit to these observations. This procedure has been integrated into functional mapping which thus unifies the strengths of statistics, genetics, and developmental biology to facilitate the test of the interplay between genetic action and development and the characterization of physiological and developmental pathways involved in the function of genes [181]. Here we do not focus on the review of functional mapping because it has been described elsewhere [58,88].

3.1. From functional mapping to systems mapping

Although functional mapping is a big jump in exploring the developmental machineries of trait formation from Lander and Botstein's [78] internal mapping model, it is still too simple to chart a complete dynamic picture of trait formation. In general, a phenotypic trait, such as biomass, is formed through complex interactions and coordination of the underlying

components expressed at different organizational levels. Given that biology is multifactorial, any analytical tools used to study biology should be sophisticated sufficiently to capture such complexity. By viewing a complex phenotype as a dynamic system, Wu et al. [176] extended functional mapping to the level of systems mapping. The central principle of systems mapping is to dissect a trait into its underlying components, coordinate the interactions of different components in terms of biological laws through mathematical equations, and map specific genes that mediate each component and its connection with other components. As a bottom-top model, systems approach can identify specific QTLs that govern the developmental interactions of various components giving rise to the function and behavior of the system. By estimating and testing mathematical parameters that specify the system, systems mapping enables the prediction or alteration of the physiological status of a phenotype based on the underlying genetic control mechanisms.

Biomass partitioning—Here, we use an example of biomass partitioning to explain systems mapping. To best adapt to a particular environment, a plant tends to allocate its biomass into living parts in an optimal way that channels a maximum amount of resources to the target of harvest (leaves, stem, roots, or fruits) (Fig. 6; [50]). Given its such an ecological interest [105], theoretical modeling of biomass allocation pattern has received considerable attention [18,50]. Modeling work has focused on the understanding of how different organs of a plant coordinate and interact to optimize the capture of nutrients, light, water, and carbon dioxide in a manner that maximizes plant growth rate through a specific developmental program. Chen and Reynolds [18] developed coordination theory to model the dynamic allocation of carbon to different organs during growth in relation to carbon and water/nitrogen supply by a group of differential equations. To maintain the growth of a plant, its organs, the leaf, stem, branch, coarse root and fine root, should coordinate as a cohesive whole. Wu et al. [176] integrated the coordination and optimization model to study the pattern of biomass partitioning by incorporating the allometric scaling theory [161–163] into a system of ordinary differential equations (ODE), expressed as

$$\begin{aligned}\frac{dM_F}{dt} &= \alpha_F M^{\beta_F} - \lambda_F M_F \\ \frac{dM_S}{dt} &= \alpha_S M^{\beta_S} \\ \frac{dM_B}{dt} &= \alpha_B M^{\beta_B} \\ \frac{dM_R}{dt} &= \alpha_R M^{\beta_R} \\ \frac{dM_H}{dt} &= \alpha_H M^{\beta_H} - \lambda_H M_H\end{aligned}\quad (6)$$

where M_F , M_S , M_B , M_R , and M_H are the biomasses of the foliage (F), stem (S), branches (B), roots (R) and root hairs (H), respectively, with whole-plant biomass $M = M_F + M_S + M_B + M_R + M_H$; α and β are the constant and exponent power of an organ biomass scaling as whole-plant biomass; and λ is the rate of eliminating ageing leaves or root hairs [177].

The complex interactions between different parts of a plant that underlie design principles of plant biomass growth can be modeled and studied by estimating and testing the ODE parameters ($\alpha_F, \beta_F, \lambda_F; \alpha_S, \beta_S; \alpha_B, \beta_B; \alpha_R, \beta_R; \alpha_H, \beta_H, \lambda_H$). For example, plants are equipped with a capacity to optimize their fitness under low nutrient availability by shifting

the partitioning of carbohydrates to processes associated with nutrient uptake at a cost of carbon acquisition [59]. These parameters can be used to quantify and predict such regulation between different plant parts in response to environmental and developmental changes.

Model—Assume a plant's mapping population composed of n RILs, initiated with two inbred lines. This population is genotyped by a panel of molecular markers from which a genetic linkage map is constructed. All RILs are phenotyped for biomass traits, separately for the foliage (F), stem (S), branches (B), roots (R) and root hairs (H), at a multitude of T time points on the ontogeny of the plants. Let

$$\begin{aligned} \mathbf{y}_i &= (\mathbf{y}_i^F; \mathbf{y}_i^S; \mathbf{y}_i^B; \mathbf{y}_i^R; \mathbf{y}_i^H) \\ &= (\mathbf{y}_i^F(1), \dots, \mathbf{y}_i^F(T); \mathbf{y}_i^S(1), \dots, \mathbf{y}_i^S(T); \mathbf{y}_i^B(1), \dots, \mathbf{y}_i^B(T); \mathbf{y}_i^R(1), \dots, \mathbf{y}_i^R(T); \mathbf{y}_i^H(1), \dots, \mathbf{y}_i^H(T)) \end{aligned}$$

denote the vector of time-dependent biomass traits of RIL i measured for five organs.

Consider a QTL or a set of QTLs that affect the dynamic changes of five biomass traits at a time. Based on Lander and Botstein's [78] theory, we formulate a mixture-based likelihood expressed as

$$L(\mathbf{y}) = \sum_{i=1}^n \left[\omega_{1|i} f_1(\mathbf{y}_i^F; \mathbf{y}_i^S; \mathbf{y}_i^B; \mathbf{y}_i^R; \mathbf{y}_i^H) + \dots + \omega_{J|i} f_J(\mathbf{y}_i^F; \mathbf{y}_i^S; \mathbf{y}_i^B; \mathbf{y}_i^R; \mathbf{y}_i^H) \right] \quad (7)$$

where ω_{ji} ($j = 1, \dots, J$) is the conditional probability of the j th QTL genotype given the marker genotype of RIL i that are linked with the QTLs, and $f_j(\mathbf{y}_i^F; \mathbf{y}_i^S; \mathbf{y}_i^B; \mathbf{y}_i^R; \mathbf{y}_i^H)$ is a high-dimensional multivariate normal distribution density function of RIL i with mean vector for QTL genotype j expressed as

$$\begin{aligned} (\boldsymbol{\mu}_j^F; \boldsymbol{\mu}_j^S; \boldsymbol{\mu}_j^B; \boldsymbol{\mu}_j^R; \boldsymbol{\mu}_j^H) &= (\mu_j^F(1), \dots, \mu_j^F(T); \mu_j^S(1), \dots, \mu_j^S(T); \mu_j^B(1), \dots, \mu_j^B(T); \\ &\mu_j^R(1), \dots, \mu_j^R(T); \mu_j^H(1), \dots, \mu_j^H(T)) \end{aligned} \quad (8)$$

and covariance matrix, expressed as

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sum_F^F & \sum_F^S & \sum_F^B & \sum_F^R & \sum_F^H \\ \sum_S^F & \sum_S^S & \sum_S^B & \sum_S^R & \sum_S^H \\ \sum_B^F & \sum_B^S & \sum_B^B & \sum_B^R & \sum_B^H \\ \sum_R^F & \sum_R^S & \sum_R^B & \sum_R^R & \sum_R^H \\ \sum_H^F & \sum_H^S & \sum_H^B & \sum_H^R & \sum_H^H \end{pmatrix}. \quad (9)$$

Systems mapping uses ODE (6) to model the time-dependent mean vectors for each QTL genotype j using ODE parameters

$(\alpha_j^F, \beta_j^F, \lambda_j^F; \alpha_j^S, \beta_j^S, \lambda_j^S; \alpha_j^B, \beta_j^B, \lambda_j^B; \alpha_j^R, \beta_j^R, \lambda_j^R; \alpha_j^H, \beta_j^H, \lambda_j^H)$ ($j = 1, \dots, J$). A basic mathematical algorithm for ODE solving, the fourth Runge-Kutta, has been incorporated into mixture model to estimate these QTL genotype-specific ODE parameters [46,47,174]. Because the covariance matrix $\boldsymbol{\Sigma}$ contains longitudinal structure, it can be modeled by some

parsimonious approaches. Zhao et al.'s [195] multivariate structured antedependence (SAD) model can be implemented within systems mapping. The SAD is found to have an advantage in capturing the covariance structure with a few parameters.

The existence of significant QTLs involved in the ODE system can be detected by scanning the linkage group using a log-likelihood ratio profile, a method as shown in Fig. 2A. After a significant QTL is detected, systems mapping provide an elegant procedure to test mechanistic hypotheses of biological relevance [151]. They include:

1. **Size-shape relationship:** Given the ODE parameters for system (6), one can see how much biomass has been allocated to the leaves, stem, branches, roots and root hairs. It is possible that some plants have a dominant main stem, with less leaves, while some plants allocate more carbon to the roots than the leaves, stem and branches. Systems mapping can estimate specific effects of a QTL on a plant's size and form or shape and, furthermore, quantify how the QTL governs the dynamic relationship between size and shape.
2. **Structural-functional relationship:** The structure of a system can determine its function. On other hand, function may also lead to the change of structure. In general, a plant in drought soil would allocate more energy into its root system in order to increase its survival rate and fitness. Systems mapping enables geneticists to predict the dynamical change of such structural-functional relationships and gain new insight into the genetic mechanisms involved in balancing vegetative and reproductive growth.
3. **Cause-effect relationship:** All organs contribute to the whole-plant biomass through a complex web of direct and indirect relationships. The foliage performing photosynthesis may not only make a direct contribution to the whole-plant biomass by producing carbohydrates, but also affect the latter through an indirect relationship of other traits, such as roots and root hairs. Systems mapping can estimate the genetic effects on these direct and indirect relationships.
4. **Sink-source relationship:** In plants, leaves serve as main supply areas (sources) that produce and transport carbohydrates through the phloem to areas of growth or storage (sinks). Sinks can locate at many places including the stem and root system. The rate at which carbohydrates are transported is primarily dependent on the sink strength of plant organs. The specific QTLs that affect these sink-sources relationships can be identified from systems mapping.

Example—We used systems mapping to analyze the dynamic changes of above-ground biomass composed of the stem and leaves (including foliage and branches) in a soybean RIL mapping population derived from two cultivars, Kefeng No.1 and Nannong1138-2. Systems mapping detects two QTLs, one on linkage group 3, named *dynQ1*, and the second on linkage group 24, named *dynQ2* (Fig. 7). The two QTLs were found to determine the functional relationship between leaf and stem and whole-plant biomass. *dynQ1* triggers its effect on organ biomass growth immediately when seeds germinate into seedlings, whereas *dynQ2* is operational when the seedlings have grown into a certain size. At *dynQ1*, alleles

inherited from Kefeng No.1 contributes to the favorable allometric correlation of leaf and stem biomass with the whole-plant biomass. The alleles for *dynQ2* that favorably affect this correlation are derived from Nannong1138-2. These findings suggest that specific QTLs are involved in the control of developmental timing and the structural-functional relationship of biomass allocation, with the pattern of action being taken to cope with genetic background.

3.2. The application of systems mapping in pharmacology

In clinical pharmacology for HIV/AIDS, the response of a patient to an antiviral drug or multiple drugs can be viewed as a system in which the body interacts with the drug through biochemical and biophysical principles to control or eliminate the expansion of virus. Because this process is under the joint control of genes derived from the host and pathogen [3], an understanding of the genetic architecture of drug response is important for designing personalized medicines to effectively combat HIV/AIDS. Systems mapping is shown to be particularly powerful for tackling the complexity of the genetic control mechanisms of virus-host interactions [54,184]. The decline trend of viral load in the host after an antiviral therapy, followed by the possible emergence of drug resistant viruses in the therapy, can be described by a high-dimensional group of ordinary differential equations (ODE). By incorporating the difference of wild-type and mutant viruses, Bonhoeffer et al. [13] obtained the following ODE:

$$\begin{aligned}\frac{dx}{dt} &= \lambda - dx - \beta_1 xv_1 - \beta_2 xv_2 \\ \frac{dy_1}{dt} &= \beta_1(1 - \mu)xv_1 + \beta_2\mu xv_2 - ay_1 \\ \frac{dy_2}{dt} &= \beta_1\mu xv_1 + \beta_2(1 - \mu)xv_2 - ay_2 \\ \frac{dv_1}{dt} &= k_1y_1 - uv_1 \\ \frac{dv_2}{dt} &= k_2y_2 - uv_2\end{aligned}\quad (10)$$

where there are five variables: uninfected cells, x , cells infected by wild-type virus, y_1 , cells infected by mutant virus, y_2 , free wild-type virus, v_1 , and free mutant virus, v_2 . These five types of cells interact with each other to determine the pharmacodynamic changes of drug resistant virus in a host's body. The mutation rate between wild-type and mutant is given by μ (in both directions). For a small μ , the basic reproductive ratios of wild-type and mutant virus are $R_1 = \beta_1\lambda k_1/(adu)$ and $R_2 = \beta_2\lambda k_2/(adu)$. The system (7), defined by nine parameters ($\lambda, d, \beta_1, \beta_2, \mu, a, k_1, k_2, u$), shows that the expected pretreatment frequency of resistant mutant depends on the number of point mutations between wild-type and resistant mutant, the mutation rate of virus replication, and the relative replication rates of wild-type virus, resistant mutant, and all intermediate mutants [13].

Mathematical models of pharmacodynamics reactions (10) have proved very powerful for characterizing complex three-way interactions among the host, pathogen and drug to provide quantitative insights into the biochemical mechanisms that leads to drug sensitivity and resistance. Now, these models have been integrated into systems mapping, gleaning an additional dimension of how genes participate in drug response [54,184]. The basic condition of this integration is to genotype a panel of patients for molecular markers who are measured for those five variables that constitute ODE (7) repeatedly at a series of time. Founded on mixture (7), this integration is implemented with linkage disequilibria between

markers and QTLs since the mapping population is derived from a natural population. Guo et al. [54] gave a procedure for the pattern of genetic control over the dynamic change of uninfected cells, cells infected by wild-type virus, cells infected by mutant virus, free wild-type virus and free mutant virus after the patients are administered. This procedure can particularly test whether and how genetic pleiotropy impacts on the correlated change of these different cell and virus types. Such genetic information is helpful for clinical professionals to design and deliver the drug in an effective way based on a patient's genetic blueprint [152].

3.3. Systems mapping meets stochastic biology

There are fundamental physical reasons why biochemical processes might be subject to noise and stochastic fluctuations [125]. A stochastic modeling approach based on the probabilistic description of the transition rates of different stages has been shown to be useful for predicting the dynamic behavior of a system. It is thought to produce more accurate results than the deterministic approach. One interesting application of the stochastic model is to study the degree of increase of disease risk with age. For example, based on the mutation accumulation hypothesis stating that cancer arises through the sequential accumulation of mutations within cell lineages, Frank [43] proposed a system of differential equations to describe how the mutations increase with age to lead to cancer, expressed as

$$\begin{aligned} \frac{dy_0}{dt} &= -u_0 y_0 \\ \frac{dy_b}{dt} &= u_{b-1} y_{b-1} - u_b y_b, \quad b=1, \dots, M-1 \\ \frac{dy_M}{dt} &= u_{M-1} y_{M-1} \end{aligned} \quad (11)$$

where y_m ($m = 0, \dots, M$) is the number of cell lineages with m mutations at age t , and u_m is the rate at which lineages progress from having m mutations to having $m+1$ mutations. It is assumed that an individual has cancer once she has a single lineage with M mutations. The simplest deterministic model assumes that the u_m values are constant over age. But a realistic model is based on a system of stochastic differential equations (SDE), which views these values as changing with age. The advantage of SDEs is that they allow the decomposition of the noise affecting the system into a system noise term representing unknown or incorrectly specified dynamics and a measurement noise term accounting for uncorrelated errors. This advantage will help to increase the flexibility of the models for describing the system.

To reflect that cancer arises as uncontrolled clonal growth of cancerous cells following several mutations, Frank [43] implemented clonal expansion into the model (11). A new model is written as

$$\begin{aligned} \frac{dy_0}{dt} &= y_0 D_0(t, 0) \\ \frac{dy_b}{dt} &= \int_0^t u_{b-1} y_{b-1} D_b(t, s) ds, \quad b=1, \dots, M-1 \\ \frac{dy_M}{dt} &= \int_0^t u_{M-1} y_{M-1} ds \end{aligned} \quad (12)$$

where $u_{b-1}y_{b-1}$ is the influx of cells with $b-1$ mutations at time s ,

$D_b(t, s) = \exp\left(-\int_s^t u_b(Z) dZ\right)$ is the decay or outflux of these cells over the remaining period from s to t , with Z being the number of cells in the clone. Clonal expansion y follows a logistic equation, expressed as

$$Z_b = \frac{K_b e^{r_b t}}{K_b + e^{r_b t} - 1} \quad (13)$$

where K_b is the carrying capacity and r_b is the intrinsic rate of increase of the clone with b mutations [43]. It is possible that clones with different numbers of mutations have different carrying capacities and rates of increase. Let v denote the mutation rate per cell, which, multiplied by the clone size, is the total mutational capacity of a cell lineage. Thus, we have

$$D_b(t, s) = \exp\left(-\int_s^t v_b Z_b(\tau) d\tau\right) = \left(\frac{K_b}{K_b + e^{r_b(t-s)} - 1}\right)^{\frac{v_b K_b}{r_b}}.$$

The total transition rate from cells with b mutations to cells with $b+1$ mutations is the mutation rate per cell, v_b , multiplied by the average size of clones with b mutations, which leads ultimately to

$$U_b = v_b \int_0^t u_{b-1} y_{b-1} D_b(t, s) Z_b(t-s) ds / y_b. \quad (14)$$

Being more sophisticated than SDE (11), SDE (12) integrates mutation accumulation and clonal expansion to model age-specific accelerations of cancer via two types of parameters (v_b , r_b , K_b), allowing the estimation of any number of rounds of clonal expansion. If these two types of SDE are incorporated into systems mapping, we can construct a predictive model for the occurrence of cancer from an individual's genetic structure based on prior knowledge obtained from a genetic mapping study. Assume that such a study is launched, which contains a set of genotyped individuals whose number of cell lineages with different numbers of mutations is measured at different ages. By analyzing the marker and phenotypic data, systems mapping enables the characterization of specific QTLs involved in age-specific acceleration of cancer.

Suppose that, from the above-mentioned study, a significant QTL has been detected by implementing equations (12). For each of three possible genotypes of this QTL, a different set of SDE parameters was estimated, with which the curves of acceleration were drawn (Fig. 8). The patterns of age-specific acceleration were found to be very different, depending on QTL genotypes. QTL genotype qq , with three rounds of clonal expansion, display a pronounced peak in the acceleration curve of cancer at age 60 years, which is much higher and more delayed than the other genotype Qq and QQ with two and one round of clonal expansion. From this hypothesized example, it is shown that systems mapping has great power to predict the pattern and time of cancer occurrence based on patient's genotypes at significant loci.

3.4. Systems mapping of phenotypic landscapes

The development of complex traits is the consequence of interactions among a multitude of genetic and environmental factors that each trigger an impact on every step of trait development. This process is inherently complicated, but can be illustrated by a landscape of phenotype formed by genetic and environmental variables [121,168]. The surface of such a phenotype landscape defines the phenotype determined by a particular combination of underlying genetic (such as additive, dominant or epistatic) and environmental factors (such as temperature, light or moisture) that interact with each other through developmental pathways. The number of underlying factors contributing to phenotypic variation defines the number of dimensions of the landscape space. In theory, the number of underlying factors can be unlimited, implying that a landscape can exist in a very high dimensional space (i.e., *hyperspace*) [168]. By characterizing the topographical features of such a landscape, typically including slope, curvature, peak-valley and ridge [168], a fundamental question of how each underlying factor contributes to phenotypic expression individually or through an interactive web can be addressed.

The topography of a phenotypic landscape can be determined by the degree and pattern of phenotypic changes in response to different environmental factors. For example, photosynthesis as the primary process in plant growth is determined by many biotic or abiotic environmental factors, such as leaf age, CO₂ concentration, temperature, irradiance, nutrient, and water potential. The responsiveness of photosynthetic rate to each of these factors follows different physiological mechanisms during a multi-step of development, leading to a particular phenotypic landscape of photosynthesis. Wu et al. [171] integrated functional mapping to determine the genetic machinery of the reaction norms of photosynthetic rate simultaneously to these factors by using mechanistic or empirical models for describing the relationship of various biotic or abiotic factors, separately or jointly, with photosynthetic rate [139,172].

Systems mapping can play a central role in illustrating an overall picture of the genetic control of phenotypic sensitivity to multiple environmental factors and discerning the differences of genetic control over response to each environmental factor. This can be done by integrating a system of partial differential equations (PDE) that can deal with functions of multiple variables, beyond ordinary differential equations describing functions of a single variable. Consider plant height growth as a phenotypic landscape. It is determined by many factors, but two of which, temperature (f_1) and nutritional concentration (f_2), are considered. The PDE that specifies the topography of plant height growth (H) is expressed as

$$\frac{dH}{dt} = rH \left(1 - \frac{H}{K} \right) + D_1 \left(\frac{\partial^2 H}{\partial f_1^2} \right) + D_2 \left(\frac{\partial^2 H}{\partial f_2^2} \right) \quad (14)$$

where r is the plant's intrinsic rate of growth, K is the carrying capacity, and D_1 and D_2 are the diffusion coefficients that measures the rates of change per temperature and nutritional concentration, respectively. By estimating a set of PDE parameters (r , K , D_1 , D_2), we can visualize the phenotypic landscape of plant height growth.

Systems mapping allows QTL genotype-specific PDE parameter sets to be estimated and tested and, therefore, is empowered with a capacity to test how a specific QTL affects the slope, curvature, peak-valley and ridge of the growth landscape. This test enables geneticists to understand the mechanistic formation of the landscape. More interestingly, it can also test whether there is the common genetic basis for different environmental responses and how the QTL affects a web of interactions between different environmental responses toward plant height growth curves. Mathematically, SDE (14) is fairly flexible to accommodate to multiple developmentally related phenotypic traits, such as plant height and diameter growth. The integration of these high-dimensional SDE into systems mapping can enhance the level of addressing biological questions, such as physiological pleiotropy of different traits [93], ecological pleiotropy of phenotypic plasticity to different environmental factors [135], and developmental pleiotropy of the same trait varying over time [117]. Because SDE has not been popular to geneticists, a heavy involvement of applied mathematicians is essential for the implementation of SDE into systems mapping and the mechanistic underpinnings and interpretation of SDE models in a biologically intuitive way.

4. Ecological dissection of complex traits

There has been increasing recognition that the traits of an individual are not only influenced by its own genes but also strongly by the genotypes of its neighbors [75,77]. For example, the feather condition of a laying hen is controlled both by its own genes and the genes of its partners through some serotonin pathway [11]. Using integrated genetic mapping and structural equation models, a QTL was detected for flowering time in focal plants that pleiotropically affects the expression of developmental traits in the neighbor plants [169]. Given these findings, phenotypic formation can be viewed as a system of cooperative or competing interactions between biological entities. Game theory can be integrated to quantify the extent of individual-individual interactions [107,114] and reveal the genetic machineries for cooperation and competition that take place in a community.

4.1. Dynamic game theory

Game theory, originally developed to study conflict and cooperation between decision makers, has found its implications in biology because of the pioneering work of evolutionary game theory by John Maynard-Smith and George R. Price [104]. On a dynamic scale, the evolutionary game theory can be mathematically formulated to describe the growth of two individuals in the shared environment by using a coupled group of ODE, expressed as

$$\begin{aligned}\frac{dx}{dt} &= \alpha_x - \beta_x x - F(x, y; \gamma_{x \leftarrow y}) \\ \frac{dy}{dt} &= \alpha_y - \beta_y y - F(x, y; \gamma_{y \leftarrow x})\end{aligned}\quad (15)$$

which characterize the rate of change of sizes of the two individuals, x and y . The first two terms of equations (15) on the right-hand side describe neighbor-independent growth in terms of a Gompertz model, where α_x and α_y are thought of as an intrinsic growth rate due to the uptake of resources, and $\beta_x x$ and $\beta_y y$ are a metabolic loss proportional to the size of the individuals. The last term on the right-hand side couples the growth of the neighboring

individual by means of a kernel function $F(x,y)$, which in general depends at least on the size of the target individual and the size of the neighbor, with the extent characterized by $\gamma_{x\leftarrow y}$ (quantifying how individual y affects individual x) and $\gamma_{y\leftarrow x}$ (quantifying how individual x affects individual y). Many studies have discussed the choice of the kernel function although more work is needed to obtain an optimal function [128].

We can cast this formulation in a game theoretic context in the following way. Given equations (15), we may assume that a rest point for the ODE is an energy minimizing (payoff maximizing) point for the two individuals engaged in competition. In this case, we can write

$$\begin{aligned}\pi_x(x, y) &= \alpha_x x - \frac{1}{2} \beta_x x^2 + F(x, y; \gamma_{x\leftarrow y}) \\ \pi_y(x, y) &= \alpha_y y - \frac{1}{2} \beta_y y^2 + F(x, y; \gamma_{y\leftarrow x})\end{aligned}\quad (16)$$

as the two payoff functions (energy-cost functions) of the individual. A first order necessary condition for a Nash equilibrium is [111,124]:

$$\begin{aligned}\frac{\partial \pi_x}{\partial x} &= \alpha_x - \beta_x x + F(x, y; \gamma_{x\leftarrow y}) = 0 \\ \frac{\partial \pi_y}{\partial y} &= \alpha_y - \beta_y y + F(x, y; \gamma_{y\leftarrow x}) = 0\end{aligned}\quad (17)$$

when we assume that

$$\begin{aligned}\frac{\partial f(x, y; \gamma_{x\leftarrow y})}{\partial x} &= F(x, y; \gamma_{y\leftarrow x}) \\ \frac{\partial f(x, y; \gamma_{y\leftarrow x})}{\partial y} &= F(x, y; \gamma_{x\leftarrow y})\end{aligned}\quad (18)$$

Thus, Nash equilibrium in the two-player continuous strategy game implies a fixed point in the differential system (16). Following Antoniadis et al. [4], if we assume an evolutionary rule given by a better response function (i.e., Jacobi iteration) as opposed to the traditional replicator dynamics of evolutionary game theory [159], then we recover exactly the dynamics in equations (15) as a result of co-evolution.

After a proper kernel function is chosen, the test of parameters $\gamma_{x\leftarrow y}$, $\gamma_{y\leftarrow x}$ can provide information about how the growth of one individual is affected by the second individual. If $\gamma_{x\leftarrow y}$, $\gamma_{y\leftarrow x} = 0$, the two individuals are independent. If $\gamma_{x\leftarrow y}$, $\gamma_{y\leftarrow x} < 0$, this suggests that the two individuals are of conflict, while if $\gamma_{x\leftarrow y}$, $\gamma_{y\leftarrow x} > 0$ they are thought to be cooperative. Such mathematical formulation of conflict and cooperation allows game theory to be quantified and further used for the ecological dissection of trait expression. By embedding it into systems mapping, this theory helps to characterize the genetic components that contribute to the formation and progression of complex traits driven by ecological interactions.

4.2. Game theory meets system mapping

Are there specific genes that control the strengths and patterns of interactions between different species in a dynamic community and ecosystem? This question has been raised intensively because of the recent emergence of community and ecosystem genetics as a new discipline aimed to study how genetic variation in one species influences the composition of associated communities and the functioning of ecosystems [1,6,10,60,164–166]. The integration of systems mapping and evolutionary game theory can address this question, enabling the genetic mapping of genes involved in a network of interspecific interactions. It is likely that this integration can facilitate the movement of community genetics research from its first generation focusing on intra- and interspecific genetic variation to its second generation seeking for a mechanistic understanding of genes that structure communities and ecosystems [2].

While broad-scale empirical and theoretical efforts have addressed the relative importance of inter- and intra-specific effects and variation on community processes, it is now a time to initiate fine-scale genetic mechanistic approaches through systems mapping. The confluence of two traditionally separated areas, statistical genetics and game theory, helps to build a cohesive and predictive framework between community ecology, evolutionary biology, genetics and genomics.

5. Regulatory dissection of complex traits

The development of any trait can be modeled as a dynamic system consisting of various biological parts which coordinate to determine a final phenotype through genetic regulation (Fig. 1). In this section, we focus on regulatory pathways from DNA to cellular physiology through transcript, protein and metabolic expression. We do not review specific examples of how to construct genotype-phenotype maps with regulatory data (see [20]), rather than provide a general procedure of integrating this information flow into systems mapping.

5.1. Network mapping

To measure a cellular system, an understanding of how gene, protein, metabolic and physiological events are expressed in time course is crucial. Attempts to correlate gene expression data with proteomic and metabolomic data have been made by using classic correlation methods or multivariate statistics. However, they have often proved to be unsatisfactory in large part because the timescales of various biological control functions are either very different or simply unknown. Figure 9, modified from Nicholson et al. [113], shows such a problem, in which two pairs of gene-protein couples are expressed in time course. It can be seen that the levels of mRNA transcripts are not consistent with protein levels on timescales. If the action of the stimulus, such as environmental intervention, is at the genetic level, it will take a finite amount of time in a cell for the associated protein synthesis (or post-translational modification) to occur. Also, the duration of the gene events (t_g) and protein events (t_p) may be very different. All this will lead to the consequence that the observed covariance of the gene and protein events is highly dependent on sampling time point and frequency. In some instances, a single sampling point (say τ_3) might result in the incorrect assumption that gene 2 co-varied with protein 1. The times of maximal activity or

expression for protein may not be constant; for example, the t_p values and related turnover times are larger for protein 2 than protein 1. A problem also exists when attempting to correlate proteomic with metabolomic data. Currently proteomic data sets contain no information on the activity of specific proteins, which is dependent on their location in the cell and the presence of cofactors or inhibitors.

To understand true quantitative relationships between thousands of gene-protein couples or protein-metabolite couples in a cellular system, a series of high-dimensional systems of differential equations have been established by considering several critical factors, such as the time displacement of the genetic and protein synthetic and post-translational events, their different timescales and their half-lives [14,72,81,113]. These mathematical equations are integrated with systems mapping, leading to the birth of a new mapping model, called network mapping, which can map interactive eQTLs, pQTLs, and mQTLs underlying transcriptional, proteomic, and metabolomic profiles and interaction networks among these different profiles [155]. The model has power to test what are the most important pathways that cause final phenotypes and how QTLs control these pathways.

5.2. Dissecting rhythmic biology

The molecular bases of circadian rhythms have been clarified during the past decade by experimental advances, first in *Drosophila* and *Neurospora*, and more recently in cyanobacteria, plants and mammals [30,52,193]. In view of the large number of variables involved and of the complexity of feedback processes that generate oscillations, mathematical models are necessary to comprehend the transition from simple to complex oscillatory behavior and to delineate the conditions under which they arise [52].

Based on the negative control exerted by the PER protein on the expression of *per*, a molecular model governed by a set of five ordinary differential equations was derived [51]. In the model (Fig. 10a), the *per* gene is first expressed in the nucleus and transcribed into *per* mRNA. The latter is transported into the cytosol, where it is translated into the PER protein, P_0 , and degraded. The PER protein undergoes multiple phosphorylation, from P_0 into P_1 and from P_1 into P_2 . These modifications, catalyzed by a protein kinase, are reverted by a phosphatase. The fully phosphorylated form of the protein is marked up for degradation and transported into the nucleus in a reversible manner. The nuclear form of the protein (P_N) represses the transcription of the gene.

In the model, the temporal variation of the concentrations of mRNA (M) and of the various forms of the regulatory protein—cytosolic (P_0, P_1, P_2) or nuclear (P_N)—is governed by the following system of ODE:

$$\begin{aligned} \frac{dM}{dt} &= \frac{v_s K_I^n}{K_I^n + P_N^n} - \frac{v_m M}{k_m + M} \\ \frac{dP_0}{dt} &= k_s M - \frac{V_1 P_0}{K_1 + P_0} + \frac{V_2 P_1}{K_2 + P_1} \\ \frac{dP_1}{dt} &= \frac{V_1 P_0}{K_1 + P_0} - \frac{V_2 P_1}{K_2 + P_1} - \frac{V_3 P_1}{K_3 + P_1} + \frac{V_4 P_2}{K_4 + P_2} \\ \frac{dP_2}{dt} &= \frac{V_3 P_1}{K_3 + P_1} - \frac{V_4 P_2}{K_4 + P_2} - k_1 P_2 + k_2 P_N - \frac{v_d P_2}{K_d + P_2} \\ \frac{dP_N}{dt} &= k_1 P_2 - k_2 P_N \end{aligned} \quad (19)$$

where per mRNA (M) is synthesized in the nucleus and transfers to the cytosol (where it accumulates at a maximum rate v_d) and it is degraded by an enzyme of maximum rate v_m and Michaelis constant K_m , and the rate of synthesis of the PER protein, proportional to M , is characterized by an apparent first-order rate constant k_s . Parameters V_1, \dots, V_4 and K_1, \dots, K_4 denote the maximum rate and Michaelis constant of the kinase(s) and phosphatase(s) involved in the reversible phosphorylation of P_0 into P_1 and P_3 into P_2 , respectively. The fully phosphorylated form (P_2) is degraded by an enzyme of maximum rate v_d and Michaelis constant K_d , and transported into the nucleus at a rate characterized by the apparent first-order rate constant k_1 . Transport of the nuclear, bisphosphorylated form of PER (P_N) into the cytosol is characterized by the apparent first-order rate constant k_2 . The negative feedback exerted by nuclear PER on per transcription is described by an equation of the Hill type, in which n denotes the degree of cooperativity, and K_I denotes the threshold constant for repression. A more sophisticated set of differential equations can be derived when the pathways in Fig. 10b or 10c are considered.

We integrate five-variable ODE (19) into network mapping to identify QTLs for circadian rhythms and test how these QTLs have sustained oscillations of the system. Figure 11 illustrates the rhythmic curves of three genotypes at a QTL assumed to be detected by network mapping. When plotting the time evolution of per mRNA (M) as a function of the total amount of PER protein (P_{tot}), these oscillations evolve toward a limit cycle (i.e., a closed curve) (Fig. 11). It is interesting to see that the shapes and periods of rhythmic curves are very sensitive to different combinations of the parameters that determine the rhythms. A small change will lead to marked differences in rhythmic curves (Fig. 11), suggesting that testing genetic differences in parameter combinations is an effective way for understanding the genetic mechanisms for biological rhythms.

In several studies of drug response, the so-called clock genes were found to affect patients' circadian rhythms through clock-controlled transcription factors, holding a promise for the determination of an individualized optimal body time for drug administration based on a patient's genes. It has been suggested that drug administration at the appropriate body time can improve the outcome of pharmacotherapy by maximizing potency and minimizing the toxicity of the drug, whereas drug administration at an inappropriate body time can induce severe side effects [116]. In practice, body time-dependent therapy, termed "chronotherapy" [53,76], can be optimized by implementing the patient's genes that control expression levels of his/her physiological variables during the course of a day.

Network mapping has great power to unravel the genetic control of different aspects of biological rhythms and determine the best time for "chronotherapy" based on individual patient's genetic background. Also, some fundamental questions will be addressed from these tests: **(1)** How do QTLs determine periodic motions in a biological system? **(2)** How can the synchronization of different variables be achieved in a rhythmic system through genetic regulation? **(3)** How can mathematical equations be integrated into genetic mapping models for circadian rhythms related to clock-and-wavefront, reaction- diffusion, and cell-cycle processes?

6. Genetic and statistical considerations

In the preceding sections, we have described and reviewed basic principles of dissecting complex traits in terms of their underlying genetic, developmental and regulatory underpinnings using genetic mapping approaches. Genetic mapping relies on two issues, mapping population and statistical approach. Lander and Botstein's [78] original mapping population is based on an experimental cross. Other types of mapping populations can also be used, depending on the organism. We will describe those commonly used mapping populations. Although there have been numerous statistical approaches developed for genetic mapping, we will briefly review the mathematical and statistical methods that are critically used in systems mapping and network mapping.

6.1. Genetic designs

Experimental crosses—Advanced generations, like the backcross, F_2 or recombinant inbred lines (RILs), produced through continuous controlled crosses initiated with two inbred strain, form a segregating population suitable for genetic mapping. In many non-model outcrossing species, such as forest trees, the F_1 generation from two heterozygous parents, called the full-sib family, can produce the segregation of marker genes. However, unlike the advanced-generation populations, the full-sib family contains inconsistent patterns of marker segregation. The algorithm has been developed to simultaneously analyze all these different marker types [94,133,182]. Genetic mapping with experimental crosses is founded on linkage analysis of the recombination fraction between different loci.

Multigenerational families—In humans, neither adequate numbers of progeny can be generated from a single family nor can any controlled cross be made possible. A nuclear family with multiple successive generations in humans is often used in order to accumulate a sufficient number of progeny for genetic mapping. The recombination fraction and identical by descent (IBD) coefficient are the key determinants of genetic mapping with multigenerational families.

Unrelated individuals randomly sampled from natural populations—The genetic mapping of complex traits can be conducted by sampling a collection of unrelated individuals at random from a natural population. In a population, different loci are genetically associated, with the extent described by a parameter called the linkage disequilibrium (LD). By making use of recombinants accumulated over a long history of generations, LD-based mapping is meritorious in terms of high-resolution dissection of a target QTL into a narrow genomic region [132].

Unrelated families randomly sampled from natural populations—Although LD mapping has tremendous potential to fine map functional genes for a complex trait, it may provide a spurious estimate of LD in practice when the association between genetic loci is due to evolutionary forces, such as mutation, drift, selection, population structure, and admixture. A mapping strategy that samples unrelated families (composed of parents and offspring) from a natural population is helpful for overcoming the limitation of LD mapping by simultaneously estimating the linkage and linkage disequilibrium [64,84,175,179]. This

design can also estimate the genetic imprinting effect of QTLs by discerning the difference of maternal- and paternal-inherited alleles [138].

Natural populations with related families—Genetic studies of some particular complex traits require a mapping population to be derived from multiple related families. For such a genetic design, the recombination fraction, IBD, and linkage disequilibrium will be needed to be estimated at a time, tracing the co-transmission of alleles from parents to their offspring. This design is powerful for studying the evolution of genes that control human traits, such as drug response.

6.2. Mathematical and statistical solutions

The methodological challenge of systems mapping and network mapping are that we need to implement and develop rigorous mathematical and statistical approaches for resolving differential equations. The recent years have seen tremendous developments in addressing these challenges. Our review is focused on those which have been used or have a direct application to the problems described in this article. Interested readers are referred to the relevant mathematical or statistical literature [95,119,185].

How to solve high-dimensional ordinary differential equations (HiODE)—As the dimension of a system increases, it is crucial to better handle HiODE. Convergence rates and estimation consistency are important for the parameter estimation of HiODE. Mathematical analysis and computer simulation have been heavily used to investigate the mechanistic underpinnings of HiODE and the behavior of dynamic systems described by these equations [95].

How to solve delay differential equations (DDEs)—While modeling the oscillation behavior of mRNA and protein concentrations, we often encounter time delays of a system. Feng and Navaratna [41] and Verdugo and Rand [143] developed mathematical approaches for manipulating linear or nonlinear relationships between the dependent variables and time delays. When a DDE contains multiple time delays, the following questions should be addressed: (1) in which range of each of differential parameters, does Hopf bifurcation occur, i.e., there are sustained oscillations? (2) how do different time delays determine periodic motions in a biological system? (3) how can the synchronization of different variables be achieved in a rhythmic system? (4) how can mathematical equations be integrated into biological models for circadian rhythms, such as the clock-and-wavefront model, the reaction-diffusion model, and the cell-cycle model?

How to solving stochastic differential equations (SDE)—Numerical solution of SDE is a relatively young field. Almost all algorithms used for the solution of ODE may work poorly for SDEs. A few basic algorithms, such as Euler's method and It 's method, have been proposed for solving SDE (incorporating white noise) [73]. These algorithms can provide the condition under which a given SDE has a solution, and determine whether or not the solution is unique.

How to solve partial differential equations (PDE)—PDE and their solutions exhibit rich and complex structures. The finite element discretization method has been implemented for solving SDEs. This numerical approach can easily handle geometrically complicated domains and construct high-order interactions. No prior knowledge is needed for this method [140].

6.3. Modeling covariance structure

Robust modeling of longitudinal covariance structure (9) is a prerequisite for appropriate statistical inference of genetic effects on longitudinal traits. Several different approaches are available for covariance modeling, including parametric, nonparametric, and semiparametric.

Parametric modeling—The advantage of parametric approaches includes the existence of closed forms for the determinant and inverse of a longitudinal matrix. This will greatly help to enhance computational efficiency. The following is a list of existing parametric approaches used to model functional mapping: stationary parametric model – assuming the stationarity of variance and correlation (e.g., autoregressive (AR) model) [98], non-stationary parametric model – variance and correlation may not be stationary (e.g., structured antedependence (SAD) model) [195,197], and general parametric model (autoregressive moving average model ARMA(p,q)) [86]. For each model above, it is important to determine its optimal order for covariance structure modeling in the most parsimonious way by a model selection criterion.

Nonparametric modeling—In many cases, the covariance structure may not follow a parametric structure, which can be better modelled by a nonparametric approach. There is a nonparametric estimator derived to model the covariance structure without assuming stationarity using kernel-weighted local linear regression smoothing of sample variograms ordinates and of squared residuals [197]. Das et al. [27] implemented a B-spline approach for modeling time-varying covariances and covariances for longitudinal traits.

Semiparametric modeling—More recently, Fan et al. [38] proposed a semiparametric model for the covariance structure of irregular longitudinal data, in which they approached the time-dependent correlation by a parametric function and the time-dependent variance by a nonparametric kernel function. The Fan et al. model's advantage lies in the combination between the flexibility of nonparametric modeling and parsimony of parametric modeling. The establishment of a robust estimation procedure and asymptotic properties of the estimators will make this semiparametric model useful in the practical estimation of the covariance function [190].

In many longitudinal trials, data are often collected at irregularly spaced time points and with measurement schedules specific to different subjects. The efficient estimation of covariance structure in this situation will be a significant concern for detecting the QTLs that control dynamic traits. Although there are many challenges in modeling the covariance structure of subject-specific irregular longitudinal data, many workers have considered this issue using different approaches. These include nonparametric analysis derived from a two-

step estimation procedure [37,183], a semiparametric approach [35], a penalized likelihood approach [65], and functional data analysis [189].

7. Discussion and prospects

The genetic mapping of complex traits has developed to a point at which current sequencing technologies can allow trait phenotypes to be partitioned into their underlying components at the DNA, regulatory, cellular and developmental levels [20,48,115]. As shown in a diagram of genetic mapping that links DNA variants with final phenotypes at different levels (Fig. 1), genetic mapping can be performed by five approaches. First, genetic mapping, originally proposed by Lander and Botstein [78], is a fundamental approach for linking DNA polymorphisms with end-point phenotypes measured at a static point. Second, beyond the original mapping approach, functional mapping has been derived for mapping QTLs that govern the developmental process of trait development through mathematical equations [98,174]. Third, functional mapping has been extended to systems mapping in which a phenotype is mapped by treating it as a dynamic system composed of interacting components linked through a system of differential equations [12,176]. Fourth, to identify the pathway mediating the genetic effect on the physiological phenotype, network mapping has been developed [155]. This approach is grounded on the regulatory process of genes that govern mRNA and protein expression toward phenotypic formation through cellular metabolism, a process characterized as a dynamic system by a group of differential equations [81]. All these mapping approaches have afforded an unprecedented understanding of how the control of trait formation is triggered by the interplay between genetic, epigenetic, and environmental factors, facilitating the construction of a precise genotype-phenotype map. Especially, functional mapping can identify developmental genes, supporting evolutionary developmental biology (evo-devo) theory concerned with the discovery and understanding of the changes in developmental mechanisms and their role in the evolutionary origin of aspects of the phenotype [109].

The biological process of trait formation is extremely complex from an ecological view. One cause of this ecological complexity is, as Strauss et al. [136] articulated in their review article, "...species exhibit traits shaped by collections of other species that co-occur with them," leading to "... every effect of a species on a community will loop back to influence the species itself, either positively or negatively" [167]. Thus, to chart a complete picture of quantitative genetic architecture, we need to study how the traits of one organism are affected by the genes from its own genome and by those from the genome of its partner in the community. It is also possible that epistasis takes place not only among genetic loci in the same genome, but also across genomes [45]. All this growing recognition of the role of ecological interactions in shaping complex traits [1, 6,10,60, 136,164–166] appeals the emergence of the fifth approach for genetic mapping, ecosystem mapping of complex traits.

A biological system, even a cellular system, involved in phenotypic formation, can be dissected into countless continuous subsystems in tandem spanning from DNA variants through regulatory pathways to cellular and developmental processes toward the end-point phenotype. Each subsystem is affected in a complex, usually nonlinear, manner, by biotic and abiotic factors. Today's sequencing, genotyping and phenotyping technologies have

made available large stores of expression data from different levels of organization. The challenge will never be how these data are collected, rather than how the experiment is designed to maximize the use and interpretation of the data, from which biological rules for the formation and control of every part of a living system can be extracted. The use of genetic mapping from a systematic perspective can facilitate our understanding of how different parts are coordinated and organized into a whole system and what are the genetic roots of the function of these parts. To the end, by altering the pathways of one or more parts, the behavior and outcome of this system can be changed toward what we expect.

Acknowledgments

We thank Sisi Feng, Guifang Fu, Ningtao Wang, Zhong Wang and Xiaoxia Yang for their assistance in making the figures. This work is supported by grants NIH/NHLBI-1U10HL098115 and UL1 TR000127. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

References

1. Agrawal AA. Community genetics: new insights into community ecology by integrating population genetics. *Ecology*. 2003; 84:543–544.
2. Alonso-Blanco C, Aarts MG, Bentsink L, Keurentjes JJ, Reymond M, Vreugdenhil D, Koornneef M. What has natural variation taught us about plant development, physiology, and adaptation? *Plant Cell*. 2009; 21:1877–1896. [PubMed: 19574434]
3. An P, Winkler CA. Host genes associated with HIV/AIDS: advances in gene discovery. *Trend Genet*. 2010; 26:119–131.
4. Antoniadis P, Fdida S, Griffin C, Jin Y, Kesidis G. Distributed medium access control with conditionally altruistic users. *EURASIP J Wireless Comm Network*. 2013; 2013:202.
5. Baghalian K, Hajirezaei MR, Schreiber F. Plant metabolic modeling: Achieving new insight into metabolism and metabolic engineering. *Plant Cell*. 2014; 26:3847–3866. [PubMed: 25344492]
6. Bailey JK, Schweitzer JA, Ubeda F, Koricheva J, LeRoy CJ, Madritch MD, Rehill BJ, Bangert RK, Fischer DG, Allan GJ, Whitham TG. From genes to ecosystems: synthesizing the effects of plant genetic factors across systems. *Phil Trans R Soc B*. 2009; 364:1607–1616. [PubMed: 19414474]
7. Barton NH, Keightley PD. Multifactorial genetics: Understanding quantitative genetic variation. *Nat Rev Genet*. 2002; 3:11–21. [PubMed: 11823787]
8. Bateson P, Barker D, Clutton-Brock T, Deb D, D'Udine B, Foley RA, Gluckman P, Godfrey K, Kirkwood T, Lahr MM, McNamara J, Metcalfe NB, Monaghan P, Spencer HG, Sultan SE. Developmental plasticity and human health. *Nature*. 2004; 430:419–421. [PubMed: 15269759]
9. Beerenwinkel S, Pachter L, Sturmfels B, Elena SF, Lenski RE. Analysis of epistatic interactions and fitness landscapes using a new geometric approach. *BMC Evol Biol*. 2007; 2007(7):60.
10. Benfey PN, Mitchell-Olds T. From genotype to phenotype: systems biology meets natural variation. *Science*. 2008; 320:495–497. [PubMed: 18436781]
11. Biscarini F, Bovenhuis H, van der Poel J, Rodenburg TB, Jungerius AP, van Arendonk JA. (2010) Across-line SNP association study for direct and associative effects on feather damage in laying hens. *Behav Genet*. 2010; 40:715–727. [PubMed: 20496162]
12. Bo WH, Fu GF, Wang Z, Xu F, Shen Y, Xu JC, Huang ZW, Gai JY, Vallejos E, Wu R. Systems mapping: How to map genes for biomass allocation towards an ideotype. *Brief Bioinform*. 2014; 15:660–669. [PubMed: 23428353]
13. Bonhoeffer S, May RM, Shaw GM, Nowk MA. Virus dynamics and drug therapy. *Proc Natl Acad Sci U S A*. 1997; 94:6971–6976. [PubMed: 9192676]
14. Bonneau R, Facciotti MT, Reiss DJ, Schmid AK, Pan M, Kaur A, Thorsson V, Shannon P, Johnson MH, Bare JC, et al. A predictive model for transcriptional control of physiology in a free living cell. *Cell*. 2007; 131:1354–65. [PubMed: 18160043]

15. Broman KW, Speed TP. A model selection approach for the identification of quantitative trait loci in experimental crosses (with discussion). *J Roy Stat Soc B.* 2002; 64:641–656.
16. Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, et al. The genetic architecture of maize flowering time. *Science.* 2009; 325:714–718. [PubMed: 19661422]
17. Calarco JP, Borges F, Donoghue MT, Van Ex F, Jullien PE, Lopes T, Gardner R, Berger F, Feijó JA, Becker JD, Martienssen RA. Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell.* 2012; 151:194–205.
18. Chen JL, Reynolds JF. A coordination model of whole-plant carbon allocation in relation to water stress. *Ann Bot.* 1997; 80:45–55.
19. Cheverud JM, Hager R, Roseman C, Fawcett G, Wang B, Wolf JB. Genomic imprinting effects on adult body composition in mice. *Proc Natl Acad Sci U S A.* 2008; 105:4253–4258. [PubMed: 18337500]
20. Civelek M, Lusk AJ. Systems genetics approaches to understand complex traits. *Nat Rev Genet.* 2014; 15:34–48. [PubMed: 24296534]
21. Cockerham CC. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics.* 1954; 39:859–882. [PubMed: 17247525]
22. Cui YH, Wu R. Statistical model for characterizing epistatic control of triploid endosperm triggered by maternal and offspring QTLs. *Genet Res.* 2005; 86:65–75. [PubMed: 16181524]
23. Cui YH, Wu R. Mapping genome-genome epistasis: A multi-dimensional model. *Bioinformatics.* 2005; 21:2447–2455. [PubMed: 15728111]
24. Cui YH, Casella G, Wu R. Mapping quantitative trait locus interactions from the maternal and offspring genomes. *Genetics.* 2004; 167:1017–1026. [PubMed: 15238549]
25. Cui YH, Lu Q, Cheverud J, Littell RC, Wu R. Model for mapping imprinted quantitative trait loci in an inbred F₂ design. *Genomics.* 2006; 87:543–551. [PubMed: 16413163]
26. Cui YH, Wu JG, Shi CH, Zhu J, Littell RC, Wu R. Modeling epistatic effects of embryo and endosperm QTL on seed development. *Genet Res.* 2006; 87:61–71. [PubMed: 16545151]
27. Das K, Li JH, Fu GF, Wang Z, Li R, Wu R. Dynamic semiparametric Bayesian models for genetic mapping of complex trait with irregular longitudinal data. *Stat Med.* 2012; 32:509–523. [PubMed: 22903809]
28. Dawber TR, Meadors GF, Moore FE. Epidemiological approaches to heart disease: The Framingham study. *Am J Publ Health.* 1951; 41:279–286.
29. de Koning D-J, Bovenhuis H, Van Arendonk JAM. On the detection of imprinted quantitative trait loci in experimental crosses of outbred species. *Genetics.* 2002; 161:931–938. [PubMed: 12072486]
30. Dunlap JC. Molecular bases for circadian clocks. *Cell.* 1999; 96:271–290. [PubMed: 9988221]
31. Eichten SR, Swanson-Wagner RA, Schnable JC, Waters AJ, Hermanson PJ, et al. Heritable epigenetic variation among maize inbreds. *PLoS Genet.* 2011; 7(11):e1002372. [PubMed: 22125494]
32. El-Soda M, Malosetti M, Zwaan BJ, Koornneef M, Aarts MG. Genotype × environment interaction QTL mapping in plants: lessons from Arabidopsis. *Trends Plant Sci.* 2014; 19:390–398. [PubMed: 24491827]
33. Falconer, DS., Mackay, TFC. Introduction to quantitative genetics. 4th. Harlow, UK: Longman; 1996.
34. Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J Am Stat Assoc.* 2001; 96:1348–1360.
35. Fan J, Li R. New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *J Am Stat Assoc.* 2004; 99:710–723.
36. Fan J, Lv J. Sure independence screening for ultrahigh dimensional feature space (with discussion). *J Roy Stat Soc Ser B.* 2008; 70:849–911.
37. Fan J, Zhang J. Two-step estimation of functional linear models with applications to longitudinal data. *J Roy Stat Soc B.* 2000; 62:303–322.

38. Fan J, Huang T, Li R. Analysis of longitudinal data with semiparametric estimation of covariance function. *J Am Stat Assoc.* 2007; 35:632–641.
39. Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. *Nature.* 2007; 447:433–440. [PubMed: 17522677]
40. Feinberg AP, Irizarry RA. Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc Natl Acad Sci U S A.* 2010; 107(Suppl 1):1757–1764. [PubMed: 20080672]
41. Feng P, Navaratna M. Modelling periodic oscillations during somitogenesis. *Math Biosci Engin.* 2007; 4:661–673.
42. Fisher RA. The correlation between relatives on the supposition of Mendelian inheritance. *Trans Roy Soc (Edinburgh).* 1918; 52:399–433.
43. Frank SA. Age-specific acceleration of cancer. *Curr Biol.* 2004; 14:242–246. [PubMed: 14761658]
44. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science.* 2007; 316:889–894. [PubMed: 17434869]
45. Friesen ML, Sachs JL, Porter SS, von Wettberg EJ, Stark SC, Martinez-Romero E. Microbially mediated plant functional traits. *Ann Rev Ecol Evol Syst.* 2011; 42:23–36.
46. Fu GF, Luo J, Berg A, Wang Z, Li JH, Das K, Li R, Wu R. A dynamic model for functional mapping of biological rhythms. *J Biol Dyn.* 2010; 4:1–10. [PubMed: 22881066]
47. Fu GF, Wang Z, Li JH, Wu R. A mathematical framework for functional mapping of complex systems using delay differential equations. *J Theor Biol.* 2011; 289:206–216. [PubMed: 21871898]
48. Gagneur J, Stegle O, Zhu C, Jakob J, Tekkedil MM, Aiyar RS, Schuon AK, Pe'er D, Steinmetz LM. Genotype-environment interactions reveal causal pathways that mediate genetic effects on phenotype. *PLoS Genet.* 2013; 9(9):e1003803. [PubMed: 24068968]
49. Gallego M, Eide EJ, Woolf MF, Virshup DM, Forger DB. An opposite role for tau in circadian rhythms revealed by mathematical modeling. *Proc Natl Acad Sci U S A.* 2006; 103:10618–10623. [PubMed: 16818876]
50. Genard M, Dauzat J, Franck N, et al. Carbon allocation in fruit trees: from theory to modelling. *Trees Struc Func.* 2008; 22:269–282.
51. Goldbeter, A. *Biochemical oscillations and cellular rhythms: The molecular bases of periodic and chaotic behaviour.* Cambridge: Cambridge University Press; 1996.
52. Goldbeter A. Computational approaches to cellular rhythms. *Nature.* 2002; 420:238–245. [PubMed: 12432409]
53. Gorbacheva VY, Kondratov RV, Zhang R, Cherukuri S, Gudkov AV, Takahashi JS, Antoch MP. Circadian sensitivity to the chemotherapeutic agent cyclophosphamide depends on the functional status of the CLOCK/BMAL1 transactivation complex. *Proc Natl Acad Sci U S A.* 2005; 102:3407–3412. [PubMed: 15689397]
54. Guo YQ, Luo JT, Wang JX, Wang YQ, Wu R. How to compute which genes control drug resistance dynamics. *Drug Discov Today.* 2011; 16:339–334. [PubMed: 21315181]
55. Hansen TF. The evolution of genetic architecture. *Annu Rev Ecol Evol Syst.* 2006; 37:123–57.
56. Hansen TF, Wagner GP. Epistasis and the mutation load: A measurement- theoretical approach. *Genetics.* 2001; 158:477–485. [PubMed: 11333254]
57. 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 Genet.* 2010; 6(9):e1001139. [PubMed: 20927186]
58. He QL, Berg A, Li Y, Vallejos CE, Wu R. Modeling genes for plant structure, development and evolution: Functional mapping meets plant ontology. *Trends Genet.* 2010; 26:39–46. [PubMed: 19945189]
59. Hermans C, Hammond JP, White PJ, Verbruggen N. How do plants respond to nutrient shortage by biomass allocation? *Trends Plant Sci.* 2006; 11:610–617. [PubMed: 17092760]
60. Hersch-Green EI, Turley NE, Johnson MTJ. Community genetics: what have we accomplished and where should we be going? *Phil Trans Roy Soc B.* 2011; 366:1453–1460. [PubMed: 21444318]

61. Hill WG. Understanding and using quantitative genetic variation. *Philos Trans R Soc Lond B Biol Sci.* 2010; 365:73–85. [PubMed: 20008387]
62. Hindorff, LA., MacArthur, J (European Bioinformatics Institute), Morales, J (European Bioinformatics Institute), Junkins, HA., Hall, PN., Klemm, AK., Manolio, TA. A Catalog of Published Genome-Wide Association Studies. Available at: www.genome.gov/gwastudies. Accessed, December 27, 2014
63. Holland JB. Genetic architecture of complex traits in plants. *Curr Opin Plant Biol.* 2007; 10:156–161. [PubMed: 17291822]
64. Hou W, Liu T, Li Y, Li Q, Li J, Das K, Berg A, Wu R. Multilocus genomics of outcrossing plant populations. *Theor Pop Biol.* 1009; 76:68–76.
65. Huang JZ, Liu N, Pourahmadi M, Liu L. Covariance selection and estimation via penalized normal likelihood. *Biometrika.* 2006; 93:85–98.
66. Imielinski M, Belta C. Exploiting the pathway structure of metabolism to reveal high-order epistasis. *BMC Syst Biol.* 2008; 2:40. [PubMed: 18447928]
67. Imumorin IG, Kim EH, Lee YM, De Koning DJ, van Arendonk JA, De Donato M, Taylor JF, Kim JJ. Genome scan for parent-of-origin QTL effects on bovine growth and carcass traits. *Front Genet.* 2011; 2:44. [PubMed: 22303340]
68. Jin C, Fine JP, Yandell BS. A unified semiparametric framework for QTL analyses, with application to spike phenotypes. *J Am Stat Assoc.* 2007; 102:56–67.
69. Johannes F, Colot V, Jansen RC. Epigenome dynamics: a quantitative genetics perspective. *Nat Rev Genet.* 2008; 9:883–890. [PubMed: 18927581]
70. Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, Simon M, Agier N, Bulski A, Albuissou J, Heredia F, Bouchez D, Dillmann C, Guerche P, Hospital F, Colot V. Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet.* 2009; 5:e1000530. [PubMed: 19557164]
71. Kao CH, Zeng ZB, Teasdale RD. Multiple interval mapping for quantitative trait loci. *Genetics.* 1999; 152:1203–1216. [PubMed: 10388834]
72. Karlebach G, Shamir R. Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol.* 2008; 9:770–80. [PubMed: 18797474]
73. Kloeden, PE., Platen, E. Numerical solution of stochastic differential equations. New York: Springer-Verlag; 1999.
74. Knott SA, Marklund L, Haley CS, Andersson K, Davies W, Ellegren H, Fredholm M, Hansson I, Hoyheim B, Lundström K, Moller M, Andersson L. Multiple marker mapping of quantitative trait loci in a cross between outbred wild boar and large white pigs. *Genetics.* 1998; 149:1069–1080. [PubMed: 9611214]
75. Kolenbrander PE, Eglund PG, Diaz PI, Palmer RJ Jr. Genome-genome interactions: bacterial communities in initial dental plaque. *Trend Microbiol.* 2005; 13:11–15.
76. Labrecque G, Belanger PM. Biological rhythms in the absorption, distribution, metabolism and excretion of drugs. *Pharmacol Thera.* 1991; 52:95–107.
77. Lambrechts L. Dissecting the genetic architecture of host–pathogen specificity. *PLoS Pathog.* 2010; 6(8):e1001019. [PubMed: 20700450]
78. Lander ES, Botstein D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics.* 1989; 121:185–199. [PubMed: 2563713]
79. Lawson HA, Cheverud JM, Wolf JB. Genomic imprinting and parent-of-origin effects on complex traits. *Nat Rev Genet.* 2013; 14:609–617. [PubMed: 23917626]
80. Leach RJ, O'Neill RG, Fitzpatrick JL, Williams JL, Glass EJ. Quantitative trait loci associated with the immune response to a bovine respiratory syncytial virus vaccine. *PLoS ONE.* 2012; 7(3):e33526. [PubMed: 22438944]
81. Legewie S, Herzel H, Westerhoff HV, Blüthgen N. Recurrent design patterns in the feedback regulation of the mammalian signalling network. *Mol Syst Biol.* 2008; 4:190. [PubMed: 18463614]
82. Lettre G, Jackson AU, Gieger C, Schumacher FR, Berndt SI, Sanna S, Eyheramendy S, Voight BF, Butler JL, Guiducci C, et al. Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat Genet.* 2008; 40:584–591. [PubMed: 18391950]

83. Li JH, Das K, Fu G, Li R, Wu R. The Bayesian lasso for genome-wide association studies. *Bioinformatics*. 2012; 27:516–523.
84. Li JH, Li Q, Hou W, Han K, Li Y, Wu S, Li YC, Wu R. An algorithmic model for constructing a linkage and linkage disequilibrium map in open-pollinated progeny populations. *Genet Res*. 2009; 91:9–21.
85. Li JH, Zhong W, Li R, Wu R. A fast algorithm for detecting gene-gene interactions in genome-wide association studies. *Ann Appl Stat*. 2014; 8:2292–2318. [PubMed: 26457126]
86. Li N, McMurry T, Berg A, Wang Z, Berceli SA, Wu R. Functional clustering of periodic transcriptional profiles through ARMA(p,q). *PLoS ONE*. 2010; 5(4):e9894. [PubMed: 20419127]
87. Li X, Sui YH, Liu T, Wang J, Li Y, Lin Z, Hegarty J, Koltun WA, Wang Z, Wu R. A model for family-based case-control studies of genetic imprinting and epistasis. *Brief Bioinform*. 2014; 15:1069–1079. [PubMed: 23887693]
88. Li Y, Wu R. Modeling host-cancer genetic interactions with multilocus sequence data. *J Comp Sci Syst Biol*. 2009; 2:24–43.
89. Li Y, Wu R. Functional mapping of growth and development. *Biol Rev*. 2010; 85:207–216. [PubMed: 19930171]
90. Li YC, Coelho CM, Liu T, Wu S, Zeng Y, Li Y, Hunter B, Dante R, Larkins B, Wu R. A statistical strategy to estimate maternal-zygotic interactions and parent-of-origin effects of QTLs for seed development. *PLoS ONE*. 2008; 3:e3131. [PubMed: 18769549]
91. Liu GD, Kong L, Wang Z, Wu R. Systems mapping of metabolic genes through control theory. *Adv Drug Deliv Rev*. 2013; 65:918–928. [PubMed: 23603209]
92. Liu T, Todhunter RJ, Wu S, Hou W, Mateescu R, Zhang Z, Burton-Wurster NI, Acland GM, Lust G, Wu R. A random model for mapping imprinted quantitative trait loci in a structured pedigree: An implication for mapping canine hip dysplasia. *Genomics*. 2007; 90:276–284. [PubMed: 17531439]
93. Lovell JT, Juenger TE, Michaels SD, Lasky JR, Platt A, Richards JH, Yu X, Easlson HM, Sen S, McKay JK. Pleiotropy of FRIGIDA enhances the potential for multivariate adaptation. *Proc Biol Sci*. 2013; 280:20131043. [PubMed: 23698015]
94. Lu Q, Cui Y, Wu R. A multilocus likelihood approach to joint modeling of linkage, parental diplotype and gene order in a full-sib family. *BMC Genet*. 2004; 5:20. [PubMed: 15274749]
95. Lu T, Liang H, Li H, Wu H. High dimensional ODEs coupled with mixed-effects modeling techniques for dynamic gene regulatory network identification. *J Am Stat Assoc*. 2011; 106:1242–1258. [PubMed: 23204614]
96. Lucek K, Haesler MP, Sivasundar A. When phenotypes do not match genotypes-unexpected phenotypic diversity and potential environmental constraints in Icelandic stickleback. *J Hered*. 2012; 103:579–584. [PubMed: 22563124]
97. Lynch, M., Walsh, B. *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer Associates; 1998.
98. Ma CX, Casella G, Wu R. Functional mapping of quantitative trait loci underlying the character process: A theoretical framework. *Genetics*. 2002; 161:1751–1762. [PubMed: 12196415]
99. Mackay TFC. The genetic architecture of quantitative traits. *Annu Rev Genet*. 2001; 35:303–339. [PubMed: 11700286]
100. Mackay TFC. Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nat Rev Genet*. 2014; 15:22–33. [PubMed: 24296533]
101. Mackay TFC, Stone EA, Ayroles JF. The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet*. 2009; 10:565–577. [PubMed: 19584810]
102. Manichaikul A, Moon JY, Sen S, Yandell BS, Broman KW. A model selection approach for the identification of quantitative trait loci in experimental crosses. *Genetics*. 2009; 181:1077–1086. [PubMed: 19104078]
103. Mather, K., Jinks, JL. *Biometrical genetics, the study of continuous variation*. third. London: Chapman and Hall; 1982.
104. Maynard-Smith J, Price GR. The logic of animal conflict. *Nature*. 1973; 246:15.

105. McConnaughay KDM, Coleman JS. Biomass allocation in three plants: ontogeny or optimality? A test along three resource gradients *Ecology*. 1999; 80:2581–2593.
106. McMullen MD, Byrne PF, Snook ME, Wiseman BR, Lee EA, Wildstrom NW, Coe EH. Quantitative trait loci and metabolic pathways. *Proc Natl Acad Sci U S A*. 1998; 95:1996–2000. [PubMed: 9482823]
107. McNamara JM. Towards a richer evolutionary game theory. *J R Soc Interface*. 2013; 10:20130544. [PubMed: 23966616]
108. Merilä J, Hendry AP. Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. *Evol Appl*. 2014; 7:1–14. [PubMed: 24454544]
109. Muller GB. Evo–devo: extending the evolutionary synthesis. *Nat Rev Genet*. 2007; 8:943–949. [PubMed: 17984972]
110. Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES. Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell*. 2009; 21:2194–2202. [PubMed: 19654263]
111. Nash J. Equilibrium points in n-person games. *Proc Natl Acad Sci U S A*. 1950; 36:48–49. [PubMed: 16588946]
112. Nettleton D, Doerge RW. Accounting for variability in the use of permutation testing to detect quantitative trait loci. *Biometrics*. 2000; 56:52–58. [PubMed: 10783776]
113. Nicholson JK, Holmes E, Lindon J, Wilson ID. The challenges of modeling mammalian biocomplexity. *Nat Biotech*. 2004; 22:1268–1274.
114. Nowak MA, Sigmund K. Evolutionary dynamics of biological games. *Science*. 2004; 303:793–799. [PubMed: 14764867]
115. Ober U, Ayroles JF, Stone EA, Richards S, Zhu D, Gibbs RA, Stricker C, Gianola D, Schlather M, Mackay TF, Simianer H. Using whole-genome sequence data to predict quantitative trait phenotypes in *Drosophila melanogaster*. *PLoS Genet*. 2012; 8(5):e1002685. [PubMed: 22570636]
116. Ohdo S, Koyanagi S, Suyama H, Higuchi S, Aramaki H. Changing the dosing schedule minimizes the disruptive effects of interferon on clock function. *Nat Med*. 2001; 7:356–360. [PubMed: 11231636]
117. Pavlicev M, Wagner GP. A model of developmental evolution: selection, pleiotropy and compensation. *Trend Ecol Evol*. 2012; 27:316–322.
118. Plomin R, Haworth CM, Davis OS. Common disorders are quantitative traits. *Nat Rev Genet*. 2009; 10:872–878. [PubMed: 19859063]
119. Ramsay JO, Hooker G, Campbell D, Cao J. Parameter estimation for differential equations: a generalized smoothing approach” (with discussion). *J Roy Stat Soc Ser B*. 2007; 69:741–796.
120. Reik W, Walter J. Genomic imprinting: parental influence on the genome. *Nat Rev Genet*. 2001; 2:21–32. [PubMed: 11253064]
121. Rice SH. A general population genetic theory for the evolution of developmental interactions. *Proc Natl Acad Sci U S A*. 2002; 99:15518–15523. [PubMed: 12438697]
122. Richards EJ. Natural epigenetic variation in plant species: A view from the field. *Curr Opin Plant Biol*. 2011; 14:204–209. [PubMed: 21478048]
123. Richards CL, Bossdorf O, Pigliucci M. What role does heritable epigenetic variation play in phenotypic evolution? *BioScience*. 2010; 60:232–237.
124. Rosen JB. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica*. 1965; 33:520–534.
125. Samoilov MS, Price G, Arkin AP. From fluctuations to phenotypes: the physiology of noise. *Sci STKE*. 2006; 2006(366):re17. [PubMed: 17179490]
126. Scheiner SM. Genetics and evolution of phenotypic plasticity. *Annu Rev Ecol Syst*. 1993; 24:35–68.
127. Schlichting CD, Smith H. Phenotypic plasticity: linking molecular mechanisms with evolutionary outcomes. *Evol Ecol*. 2002; 16:189–211.
128. Schneider MK, Law R, Illian JB. Quantification of neighbourhood-dependent plant growth by Bayesian hierarchical modeling. *J Ecol*. 2006; 94:310–321.

129. Sen S, Churchill GA. A statistical framework for quantitative trait mapping. *Genetics*. 2001; 159:371–387. [PubMed: 11560912]
130. Shaw KL, Parsons YM, Lesnick SC. QTL analysis of a rapidly evolving speciation phenotype in the Hawaiian cricket *Laupala*. *Mol Ecol*. 2007; 16:2879–2892. [PubMed: 17614904]
131. Sillanpää MJ, Pikkukookana P, Abrahamsson S, Knürr T, Fries A, Lerceteau E, Waldmann P, García-Gil MR. Simultaneous estimation of multiple quantitative trait loci and growth curve parameters through hierarchical Bayesian modeling. *Heredity*. 2012; 108:134–146. [PubMed: 21792229]
132. Slatkin M. Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nat Rev Genet*. 2008; 9:477–485. [PubMed: 18427557]
133. Stam P. Construction of integrated genetic linkage maps by means of a new computer package: Join Map. *Plant J*. 2003; 3:739–744.
134. Stich B, Yu J, Melchinger AE, Piepho H-P, Utz HF, Maurer HP, Buckler ES. Power to detect higher-order epistatic interactions in a metabolic pathway using a new mapping strategy. *Genetics*. 2007; 176:563–570. [PubMed: 17194777]
135. Strauss SY, Irwin RE. Ecological and evolutionary consequences of multispecies plant-animal interactions. *Ann Rev Ecol Evol Syst*. 2004; 35:435–466.
136. Strauss SY, Sahli H, Conner JK. Toward a more trait-centered approach to diffuse (co)evolution. *New Phytol*. 2005; 165:81–90. [PubMed: 15720623]
137. Sui Y, Wu W, Wang Z, Wang J, Wang Z, Wu R. A case-control design for testing and estimating epigenetic effects on complex diseases. *Brief Bioinform*. 2014; 15:319–326. [PubMed: 23334587]
138. Sun L, Zhu X, Bo W, Xu F, Cheng T, Zhang Q, Wu R. An open-pollinated design for mapping imprinting genes in natural populations. *Brief Bioinform*. 2014
139. Thornley, JHM., Johnson, IR. *Plant and crop modelling: A mathematical approach to plant and crop physiology*. Oxford: Clarendon Press; 1990.
140. Tveito, A., Winther, R. *Introduction to partial differential equations: A computational approach*. New York: Springer-Verlag; 1999.
141. van Eeuwijk FA, Bink MC, Chenu K, Chapman SC. Detection and use of QTL for complex traits in multiple environments. *Curr Opin Plant Biol*. 2010; 13:193–205. [PubMed: 20137999]
142. Varshney RK, Kudapa H, Pazhamala L, Chitkineni A, Thudi M, Bohra M, Gaur PM, Janila P, Fikreb A, Kimurto P, Ellis N. Translational genomics in agriculture: some examples in grain legumes. *Crit Rev Plant Sci*. 2015; 34:169–194.
143. Verdugo A, Rand R. Hopf bifurcation in a DDE model of gene expression. *Commun Nonlin Sci Num Simul*. 2008; 13:235–242.
144. Via S, Lande R. Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution*. 1985; 39:505–522. [PubMed: 28561964]
145. Via S, Gomulkiewicz R, De Jong G, Scheiner SM, Schlichting CD, Van Tienderen PH. Adaptive phenotypic plasticity: consensus and controversy. *Trends Ecol Evol*. 1995; 10:212–217. [PubMed: 21237012]
146. Wagner GP, Zhang J. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex adaptations. *Nat Rev Genet*. 2011; 12:204–213. [PubMed: 21331091]
147. Walpole J, Papin JA, Peirce SM. Multiscale computational models of complex biological systems. *Annu Rev Biomed Eng*. 2013; 15:137–154. [PubMed: 23642247]
148. Wang CG, Wang Z, Luo JT, Li Q, Li Y, Ahn K, Prows DR, Wu R. A model for transgenerational imprinting variation in complex traits. *PLoS ONE*. 2010; 5(7):e11396. [PubMed: 20644725]
149. Wang CG, Wang Z, Prows DR, Wu R. A computational framework for the inheritance of genomic imprinting for complex traits. *Brief Bioinform*. 2012; 13:34–45. [PubMed: 21565936]
150. Wang XF, Xu CW, Wu R, Larkins BA. Genetic dissection of complex endosperm traits. *Trend Plant Sci*. 2009; 14:391–398.
151. Wang YQ, Wang NT, Wang Z, Han H, Wu R. Genetic dissection of complex traits: From functional mapping to systems mapping. *J Biomet Biostat*. 2012; 3:e119.

152. Wang YQ, Wang NT, Wang JX, Wang Z, Wu R. Delivering systems pharmacogenomics towards precision medicine through mathematics. *Adv Drug Deliv Rev.* 2013; 65:905–911. [PubMed: 23523629]
153. Wang Z, Wang ZH, Wang JX, Sui YH, Zhang J, Liao DP, Wu R. A quantitative genetic and epigenetic model of complex traits. *BMC Bioinform.* 2012; 13:274.
154. Wang Z, Pang XM, Lv YF, Xu F, Zhou T, Li X, Feng SS, Li JH, Li ZK, Wu R. A dynamic framework for quantifying the genetic architecture of phenotypic plasticity. *Brief Bioinform.* 2013; 14:82–95. [PubMed: 22396460]
155. Wang Z, Wang YQ, Wang NT, Wang JX, Wang ZH, Wu R. Towards a comprehensive picture of the genetic landscape of complex traits. *Brief Bioinform.* 2014; 15:30–42. [PubMed: 22930650]
156. Wang ZH, Wang Z, Fu GF, Luo JT, Wu R. Stochastic modeling of systems mapping for drug response. *Adv Drug Deliv Rev.* 2013; 65:912–917. [PubMed: 23528445]
157. Waters AJ, Makarevitch I, Eichten SR. Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. *Plant Cell.* 2011; 23:4221–4233. [PubMed: 22198147]
158. Weedon MN, Frayling TM. Reaching new heights: Insights into the genetics of human stature. *Trend Genet.* 2008; 24:595–603.
159. Weibull, JW. Evolutionary game theory. MIT Press; 1997.
160. Weinig C, Stinchcombe JR, Schmitt J. QTL architecture of resistance and tolerance traits in *Arabidopsis thaliana* in natural environments. *Mol Ecol.* 2003; 12:1153–1163. [PubMed: 12694279]
161. West GB, Brown JH, Enquist BJ. A general model for the origin of allometric scaling laws in biology. *Science.* 1997; 276:122–126. [PubMed: 9082983]
162. West GB, Brown JH, Enquist BJ. A general model for the structure and allometry of plant vascular systems. *Nature.* 1999; 400:664–667.
163. West GB, Brown JH, Enquist BJ. The fourth dimension of life: fractal geometry and allometric scaling of organisms. *Science.* 1999; 284:1677–1679. [PubMed: 10356399]
164. Whitham TG, Bailey JK, Schweitzer JA, Shuster SM, Bangert RK, LeRoy CJ, Lonsdorf E, Allan GJ, DiFazio SP, Potts BM, et al. A framework for community and ecosystem genetics: from genes to ecosystems. *Nat Rev Genet.* 2006; 7:510–523. [PubMed: 16778835]
165. Whitham TG, DiFazio SP, Schweitzer JA, Shuster SM, Allan GJ, Bailey JK, Woolbright SA. Extending genomics to natural communities and ecosystems. *Science.* 2008; 320:492–495. [PubMed: 18436780]
166. Whitham TG, Martinsen GD, Young W, Gehring CA, Schweitzer JA, Shuster SM, Wimp GM, Bailey JK, Fischer RD, Lindroth R, Keim P. Community and ecosystem genetics: a consequence of the extended phenotype. *Ecology.* 2003; 84:559–573.
167. Wilson DS. Evolution on the level of communities. *Science.* 1976; 192:1358–1360. [PubMed: 1273598]
168. Wolf JB. The geometry of phenotypic evolution in developmental hyperspace. *Proc Natl Acad Sci U S A.* 2002; 99:15849–15851. [PubMed: 12461185]
169. Wolf JB, Mutic JJ, Kover PX. Functional genetics of intraspecific ecological interactions in *Arabidopsis thaliana*. *Phil Trans R Soc B.* 2011; 366:1358–1367. [PubMed: 21444310]
170. Wright S. Systems of mating. Parts I–V. *Genetics.* 1921; 6:111–178. [PubMed: 17245958]
171. Wu JS, Zhu J, Zeng YR, Wu R. Functional mapping of norm reactions to environmental signals. *Genet Res.* 2007; 89:27–38. [PubMed: 17517157]
172. Wu R. Simulated optimal structure of a photosynthetic system: Implications for the breeding of forest crop ideotype. *Can J For Res.* 1993; 23:1631–1638.
173. Wu R. Detecting epistatic genetic variance with a clonally replicated design: Models for low- vs high-order nonallelic interaction. *Theor Appl Genet.* 1996; 93:102–109. [PubMed: 24162206]
174. Wu R, Lin M. Functional mapping – how to map and study the genetic architecture of dynamic complex traits. *Nat Rev Genet.* 2006; 7:229–237. [PubMed: 16485021]
175. Wu R, Zeng Z-B. Joint linkage and linkage disequilibrium mapping in natural populations. *Genetics.* 2001; 157:899–909. [PubMed: 11157006]

176. 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 Syst Biol.* 2011; 5:84. [PubMed: 21615967]
177. Wu R, Grisson JE, McKe SE, O'Malley DM. Phenotypic plasticity of fine roots and increased plant growth in pines. *BMC Ecol.* 2004; 4:14. [PubMed: 15353004]
178. Wu R, Lou X-Y, Ma C-X, Wang XL, Larkins BA, Casella G. An improved genetic model generates high-resolution mapping of QTL for protein quality in maize endosperm. *Proc Natl Acad Sci USA.* 2002; 99:11281–11286. [PubMed: 12177407]
179. Wu R, Ma C-X, Casella G. Joint linkage and linkage disequilibrium mapping of quantitative trait loci in natural populations. *Genetics.* 2002; 160:779–792. [PubMed: 11861578]
180. Wu R, Ma C-X, Gallo-Meagher M, Littell RC, Casella G. Statistical methods for dissecting triploid endosperm traits using molecular markers: An autogamous model. *Genetics.* 2002; 162:875–892. [PubMed: 12399397]
181. Wu R, Ma C-X, Lin M, Casella G. A general framework for analyzing the genetic architecture of developmental characteristics. *Genetics.* 2004; 166:1541–1551. [PubMed: 15082567]
182. Wu R, Ma C-X, Painter I, Zeng Z-B. Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. *Theor Pop Biol.* 2001; 61:349–363.
183. Wu WB, Pourahmadi M. Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika.* 2003; 90:831–844.
184. Wu WM, Feng SS, Wang Y, Wang N, Hao H, Wu R. Systems mapping of genes controlling chemotherapeutic drug efficiency for cancer stem cells. *Drug Discover Today.* 2014; 19:1125–1130.
185. Xun X, Cao J, Mallick B, Maity A, Carroll RJ. Parameter estimation of partial differential equation models. *J Am Stat Assoc.* 2013; 108:1099–1020.
186. Xu S. Estimating polygenic effects using markers of the entire genome. *Genetics.* 2003; 163:789–801. [PubMed: 12618414]
187. Xu S, Atchley WR. Mapping quantitative trait loci for complex binary diseases using line crosses. *Genetics.* 1996; 143:1417–1424. [PubMed: 8807312]
188. Yang R, Xu S. Bayesian shrinkage analysis of quantitative trait loci for dynamic traits. *Genetics.* 2007; 176:1169–1185. [PubMed: 17435239]
189. Yao F, Müller HG, Wang J-L. Functional data analysis for sparse longitudinal data. *J Am Stat Assoc.* 2005; 100:577–590.
190. Yap J, Fan J, Wu R. Nonparametric modeling of covariance structure in functional mapping of quantitative trait loci. *Biometrics.* 2009; 65:1068–1077. [PubMed: 19302406]
191. Yi N, Shriner D, Banerjee S, Mehta T, Pomp D, Yandell BS. Efficient strategies for Bayesian mapping of genome-wide interacting QTL. *Genetics.* 2007; 176:1865–1877. [PubMed: 17483424]
192. Yin DN, Zhu XL, Jiang LB, Zhang J, Zeng YR, Wu R. A reciprocal cross design to map the genetic architecture of complex traits in apomictic plants. *New Phytol.* 2014 accepted.
193. Young MW, Kay SA. Time zones: A comparative genetics of circadian clocks. *Nat Rev Genet.* 2001; 2:702–715. [PubMed: 11533719]
194. Zeng Z-B. Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci. *Proc Natl Acad Sci U S A.* 1993; 90:10972–10976. [PubMed: 8248199]
195. Zhao W, Hou W, Littell RC, Wu R. Structured antedependence models for functional mapping of multivariate longitudinal quantitative traits. *Stat Methods Mol Genet Biol.* 2005; 4(Issue 1)
196. Zhu S, Wang Z, Wang JX, Wang YQ, Wang NT, Xu M, Xu XH, Huang MR, Wang MX, Zhang SG, Wu R. A quantitative model of transcriptional differentiation driving host-pathogen interactions. *Brief Bioinform.* 2013; 14:713–723. [PubMed: 22962337]
197. Zimmerman DL, Nunez-Anton V. Parametric modeling of growth curve data: An overview (with discussion) *Test.* 2001; 10:1–73.
198. Zou F, Nie L, Wright FA, Sen PK. A robust QTL mapping procedure. *J Stat Plan Infer.* 2009; 139:978–989.

Highlights

- Phenotypic formation is dissolved into several parts, each modeled as a system.
- We review a conceptual framework for genetic mapping of each of these systems.
- This framework is equipped by differential equations to map a system's structure.
- Systems mapping identifies QTLs useful to construct genotype-phenotype map.

Genotype-Phenotype Map

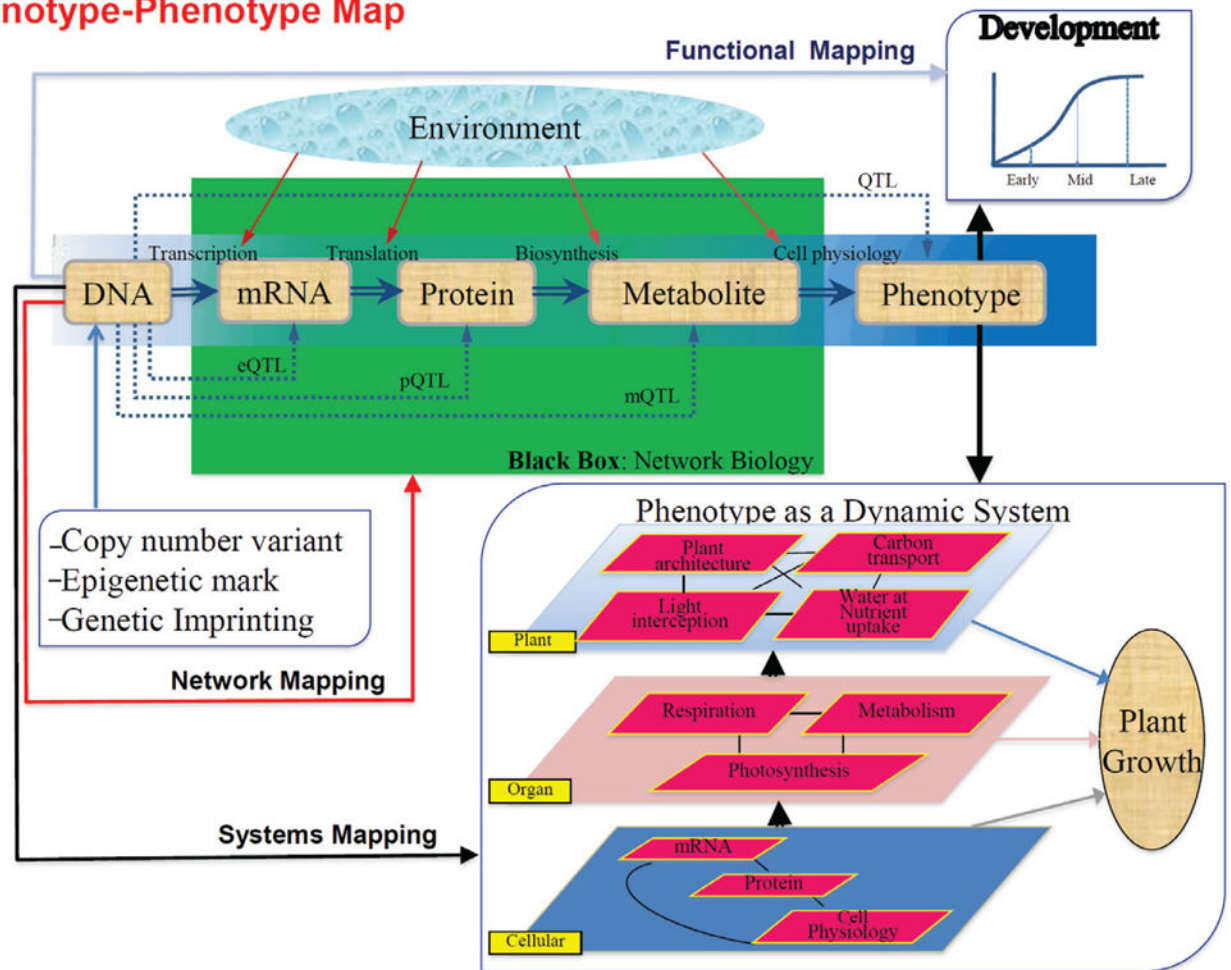


Figure 1.

Genetic mapping of complex phenotypes toward constructing a genotype-phenotype map of plant growth. Traditional direct DNA-phenotype associations to find the underlying QTLs are reformed by functional mapping that considers the developmental process of phenotypic traits, systems mapping that treats phenotypic formation as a dynamic system composed of many interactive, functional elements at different levels of organization, and network mapping that integrates the regulatory pathways from DNA sequences to endophenotypes providing essential materials for the physiological function of cells.

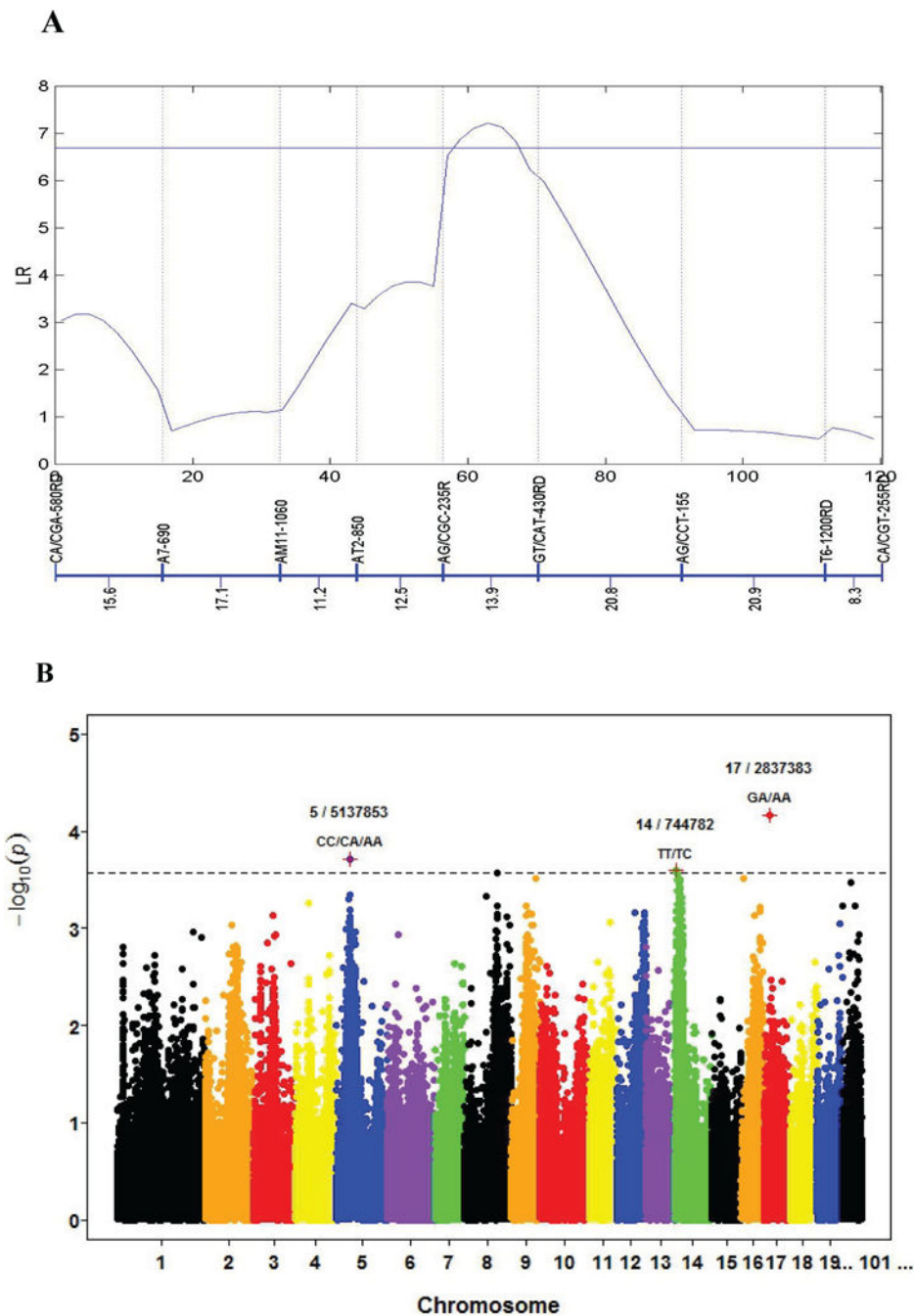


Figure 2. Graphic presentation of the results by linkage mapping and GWAS for 11-year stem diameter growth in a full-sib family of interspecific hybridization between different *Populus* species. **(A)** Plot of log-likelihood ratio (LR) test statistics for testing the existence of a significant QTL on a linkage group, with the names and genetic distances (in cM) of molecular markers given beneath. The horizontal line denotes the genome-wide threshold at the 5% significance level determined from permutation tests. **(B)** Manhattan plot of minus $\log_{10}(p)$ values against the genome locations of SNPs distributed throughout the genome. The

horizontal dash line indicates the 5% significance level after the adjustment for multiple testing from Bonferroni correction. Several significant loci are shown. Data was supplied by Prof. Minren Huang at Nanjing Forestry University with permission.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

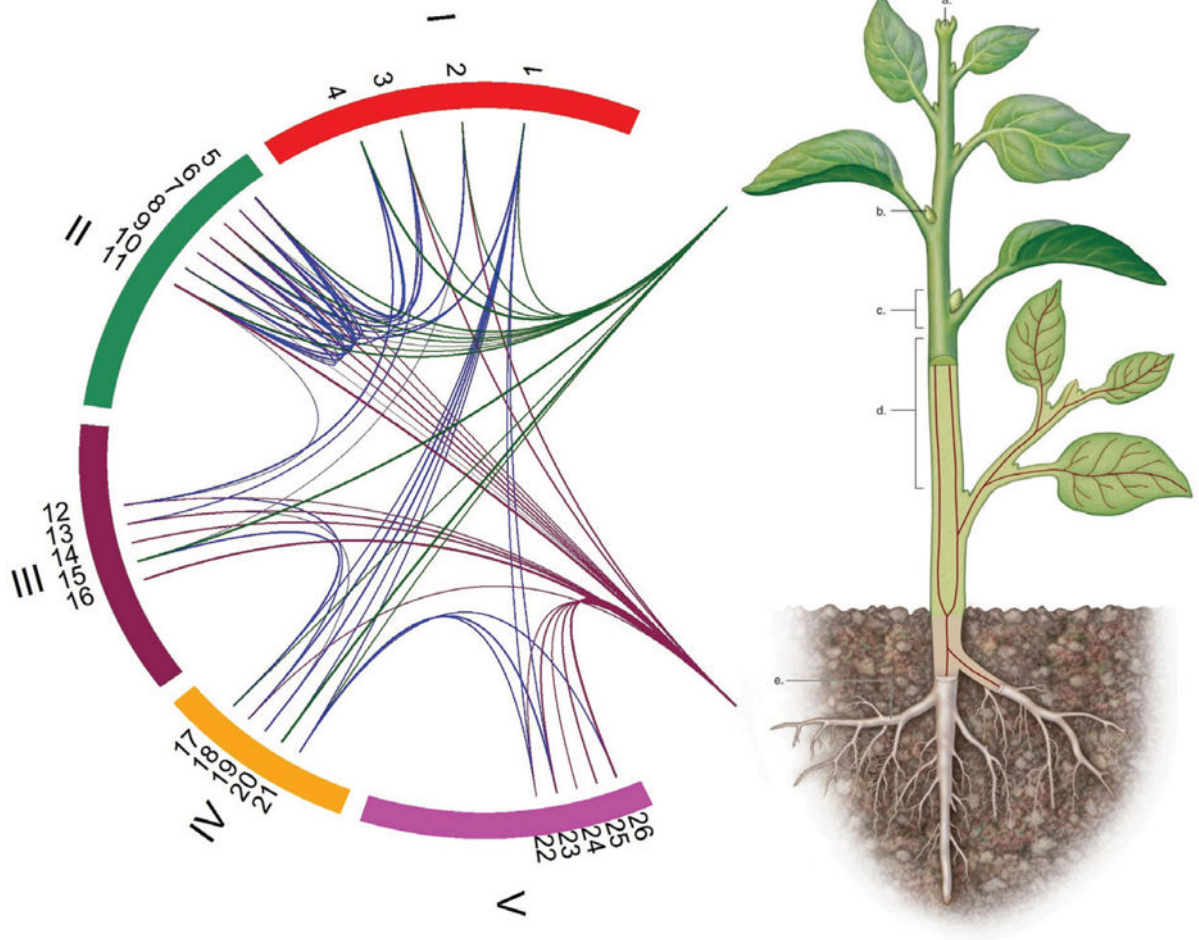


Figure 3.

A network of genetic actions and interactions among genes distributed on different chromosomes that affect leaf trait and root traits of a plant. The plant in the diagram contains five chromosomes (I – V) on which the locations of genes, labelled by 1 – 26, are indicated. The direct effects of genes are indicated by green lines for leaf traits and brown lines for root traits. The epistatic interactions of different genes are indicated by blue lines.

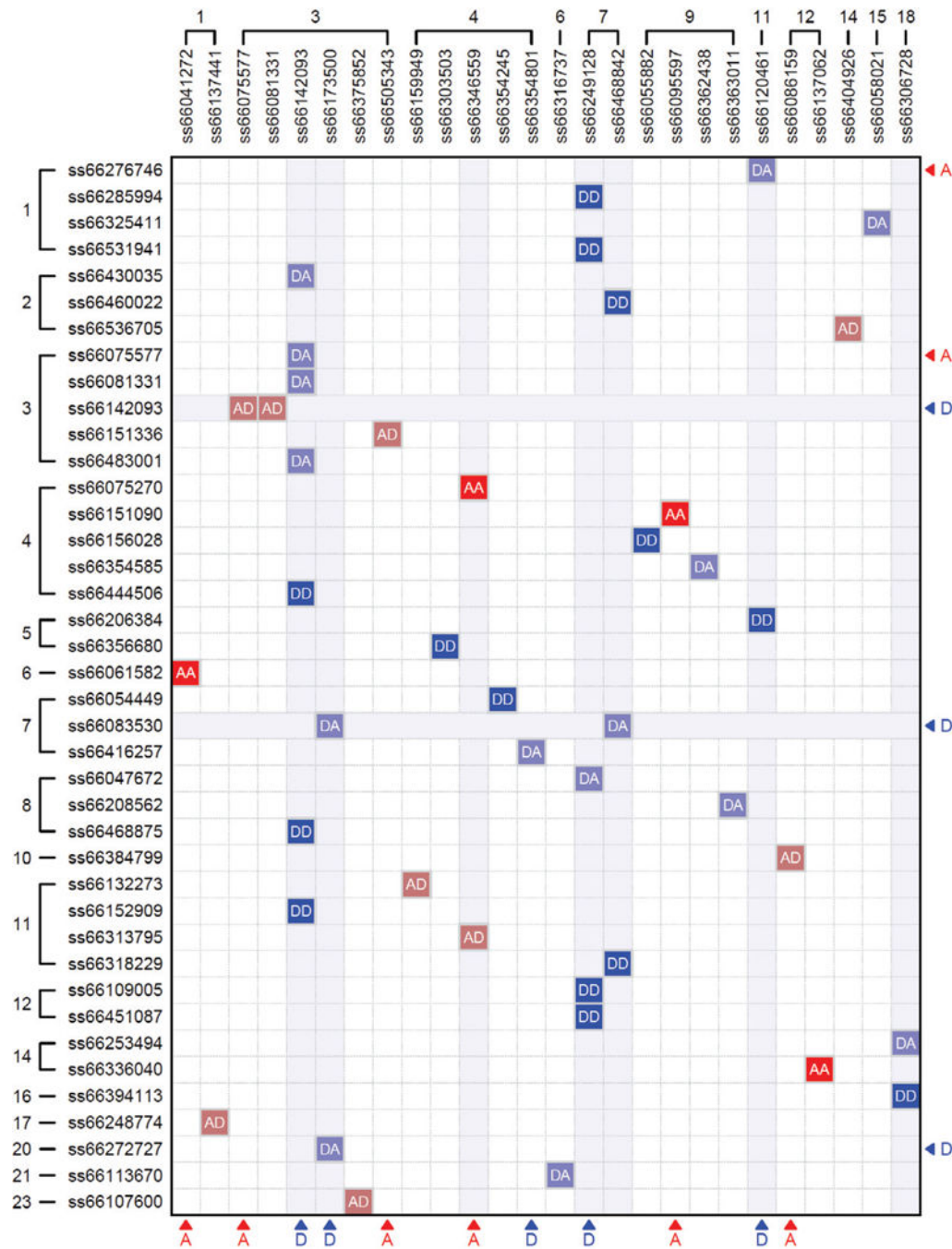


Figure 4. A picture of significant genetic interactions between different SNPs that determine BMI in the Framingham Heart Study. The numbers beside SNPs are chromosome numbers. The SNPs that display significant main additive (A) and dominant (D) genetic effects are indicated by arrows. AA, AD, DA and DD represent four corresponding types of epistasis. Modified from Li et al. [85].

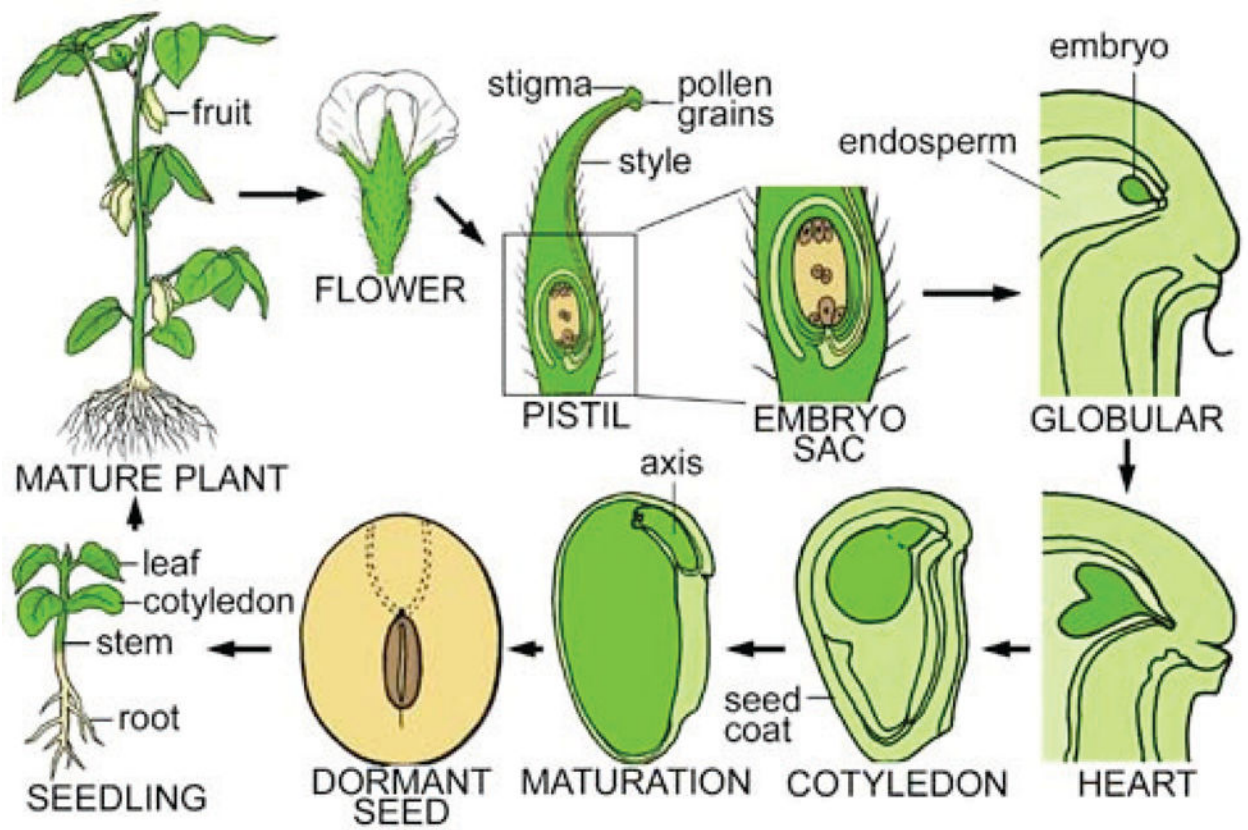


Figure 5. Lifecycle of a typical flowering plant – soybean, which experiences many distinct stages of vegetative and reproductive growth. Key organs of the plants are shown.

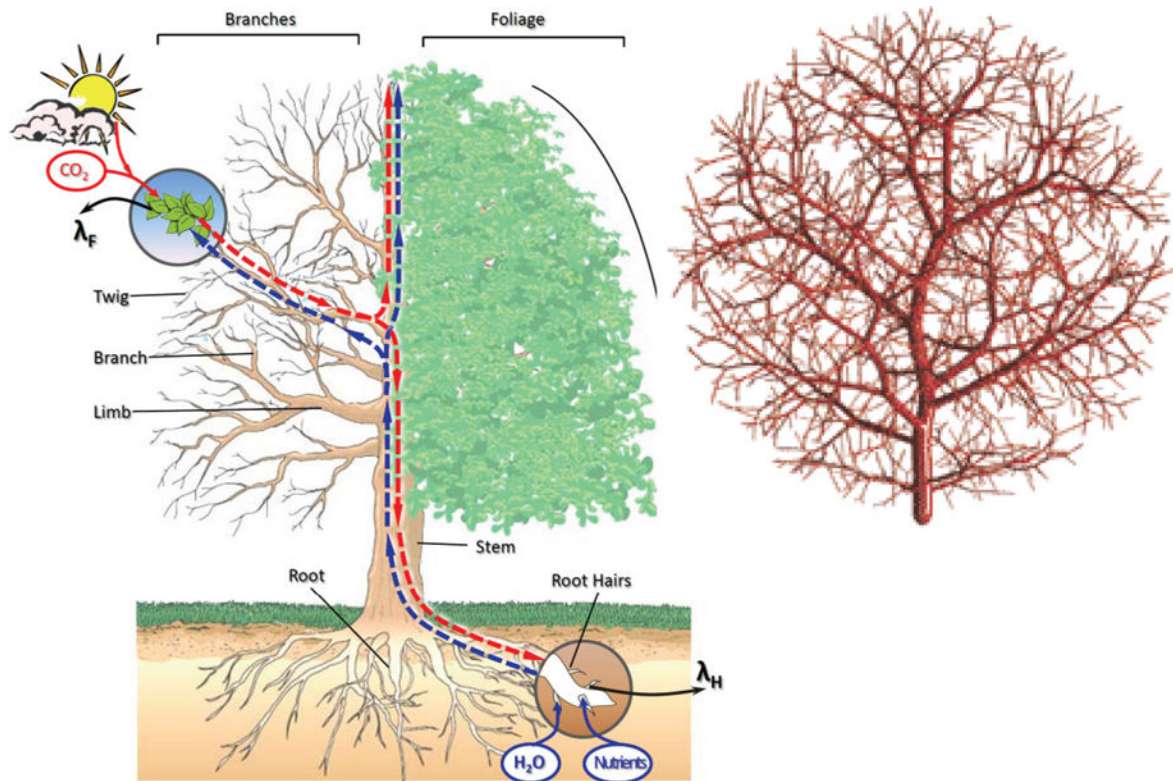


Figure 6. Biological design principles that build the system of ordinary differential equations (6). **Left:** Botanical architecture of a living plant and the pathways of absorbing water and nutrients from root hairs and transporting carbohydrates from leaves that perform photosynthesis with CO_2 and sunlight. **Right:** A fractal model that is used to explain universal allometric scaling laws that constrain the relationship of physiology and body size in living systems.

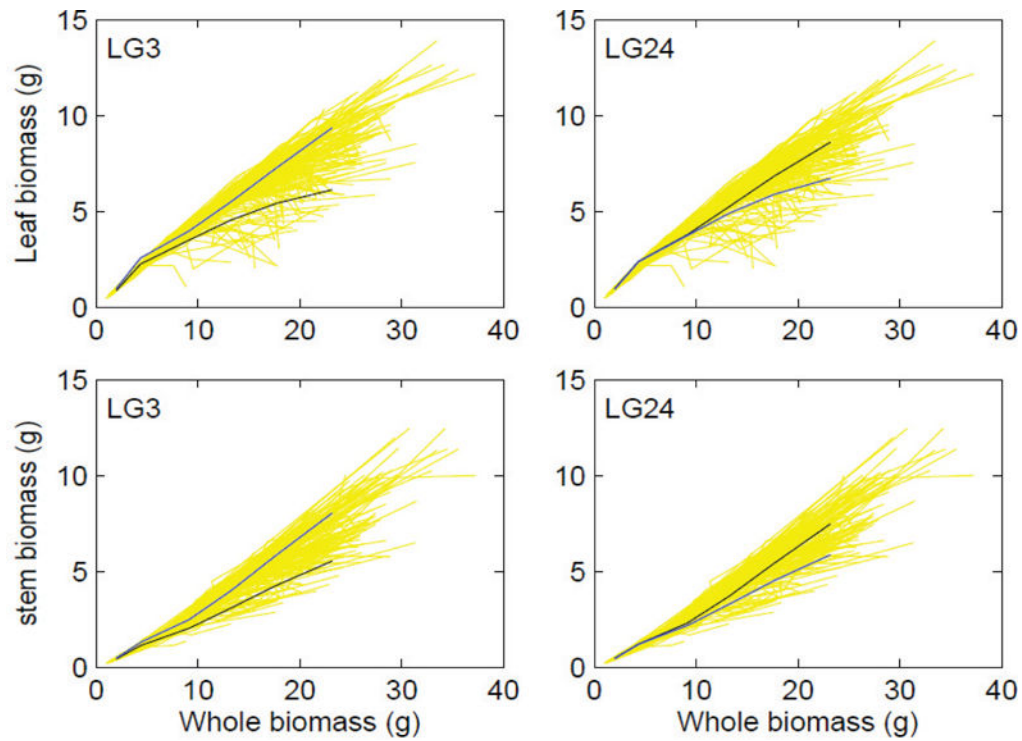


Figure 7. Dynamic pattern of biomass allocated to the leaves and stem for two different genotypes (dark curves) at each of the two QTLs, *dynQ1* and *dynQ2*, detected on LG3 and LG24, respectively, detected from a recombinant inbred line population of soybean. Curves in yellow each represent the biomass growth trajectory of a progeny in soybean. The system of ODEs (1) is reduced to a two-dimensional system, expressed as

$$\begin{aligned}\dot{M}_F &= \alpha_F W^{\beta_F} - \lambda_F M_F \\ \dot{M}_S &= \alpha_S W^{\beta_S}\end{aligned}$$

where $W = M_L + M_S$ is the sum of the foliage (M) and stem biomass (S). Data was supplied by Prof. Junyi Gai at Nanjing Agricultural University with permission.

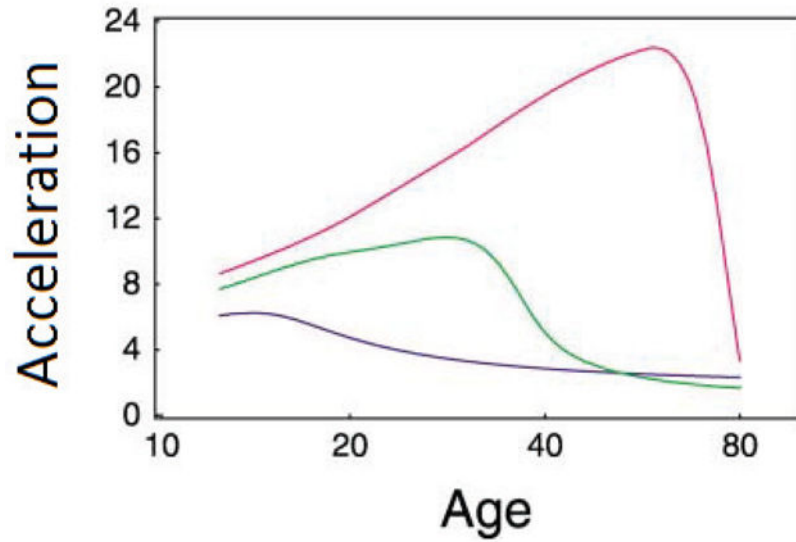


Figure 8.

Patterns of age-specific acceleration for different genotypes at a hypothesized QTL. Three genotypes show different forms of acceleration curves, indicating the role of the QTL in cancer pathogenesis. The first genotype QQ is denoted by a blue curve, for which clonal expansion occurs only in the last round before cancer, so $K_{M-2} = K_{M-3} = 1$; the second genotype is denoted by a green curve, whose clonal expansion occurs in the last two rounds before cancer, with $K_{M-2} = 10^6$ and $K_{M-3} = 1$; and the third genotype is denoted by a red curve, where clonal expansion occurs in the last three rounds before cancer, with $K_{M-2} = K_{M-3} = 10^6$. The mutation rates for these three genotypes are $v_m = 5.8 \times 10^{-4}$, 9.3×10^{-5} , and 1.55×10^{-6} for m mutations to keep the total incidence of cancer at 10%, respectively. Basic parameters are $M = 4$, $x_0(0) = 10^4$, $r_m = 0.5$ for all m , $K_0 = 1$, and $K_{M-1} = 10^6$. This example was made up from Frank's [43] analysis.

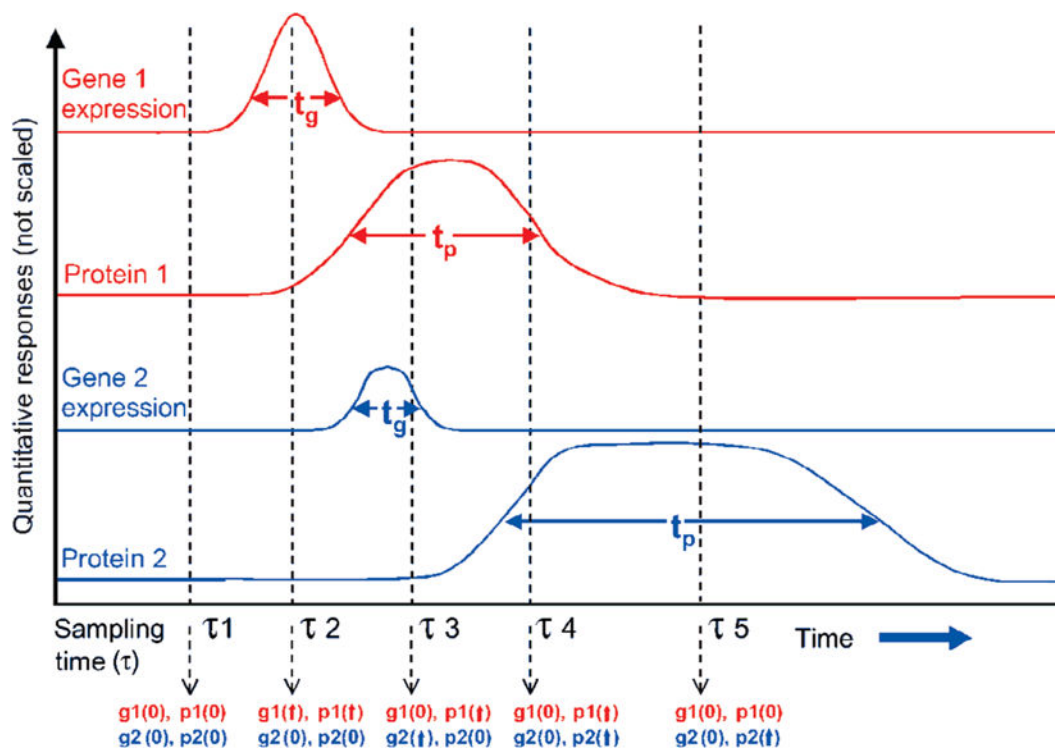
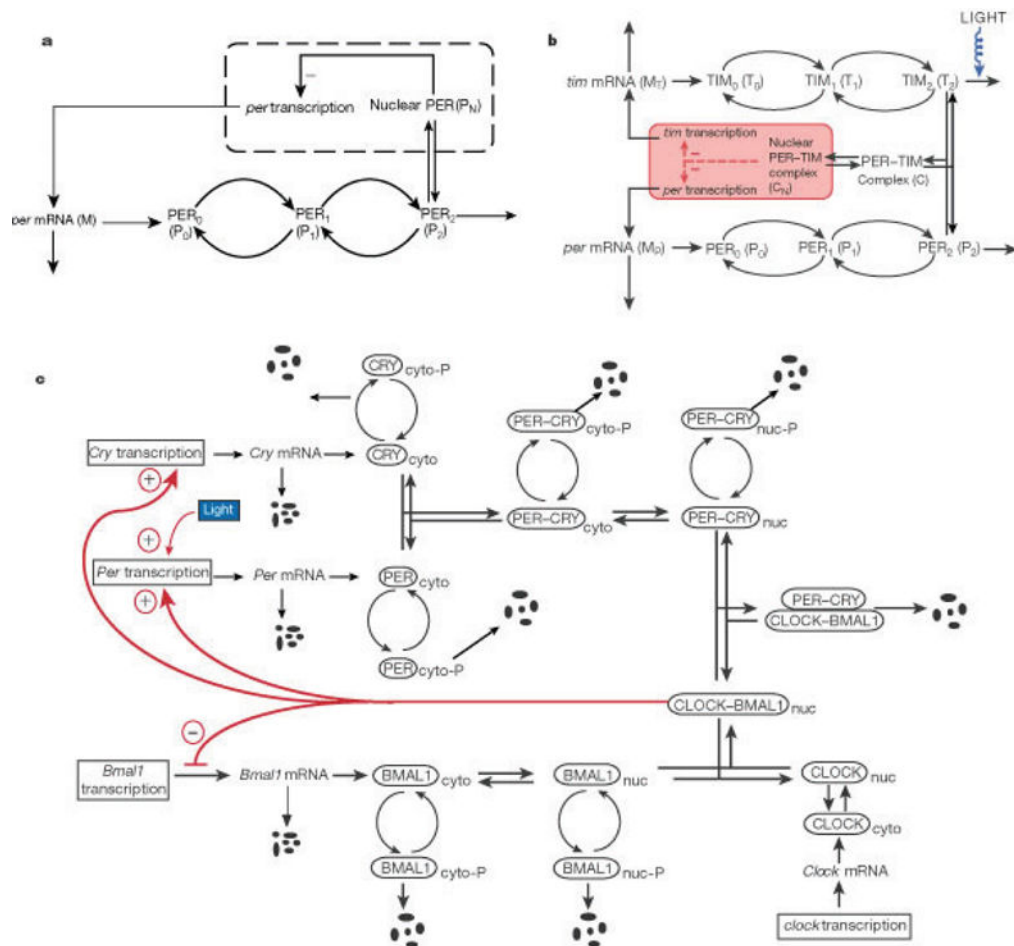


Figure 9.

Dynamic changes of two hypothesized gene-protein couples (relating transcription activity to protein level) that are expressed in response to an environmental stimuli at time zero. $g_1(0)$, $g_2(\uparrow)$; $p_1(0)$, $p_2(\uparrow)$ describe the relative condition for the expression of each gene and protein at a given time-point (e.g. τ_1 , τ_2 , etc.), where (0) indicates baseline level and (\uparrow) indicates expression. It can be seen that the post intervention observation of relative state is dependent on sampling time-point, which can be considered by differential equations. Adapted from Nicholson et al. [113].

**Figure 10.**

Molecular models for explaining circadian oscillations. (a) Model for circadian oscillations in *Drosophila* based on negative autoregulation of the *per* gene by its protein product PER. (b) Model for circadian oscillations in *Drosophila* incorporating the formation of a complex between the PER and TIM proteins. (c) Model for circadian oscillations in mammals incorporating indirect, negative autoregulation of the *Per* and *Cry* genes through binding of the PER-CRY dimer to the complex formed between the two activating proteins CLOCK and BMAL1. Adapted from Goldbeter [52].

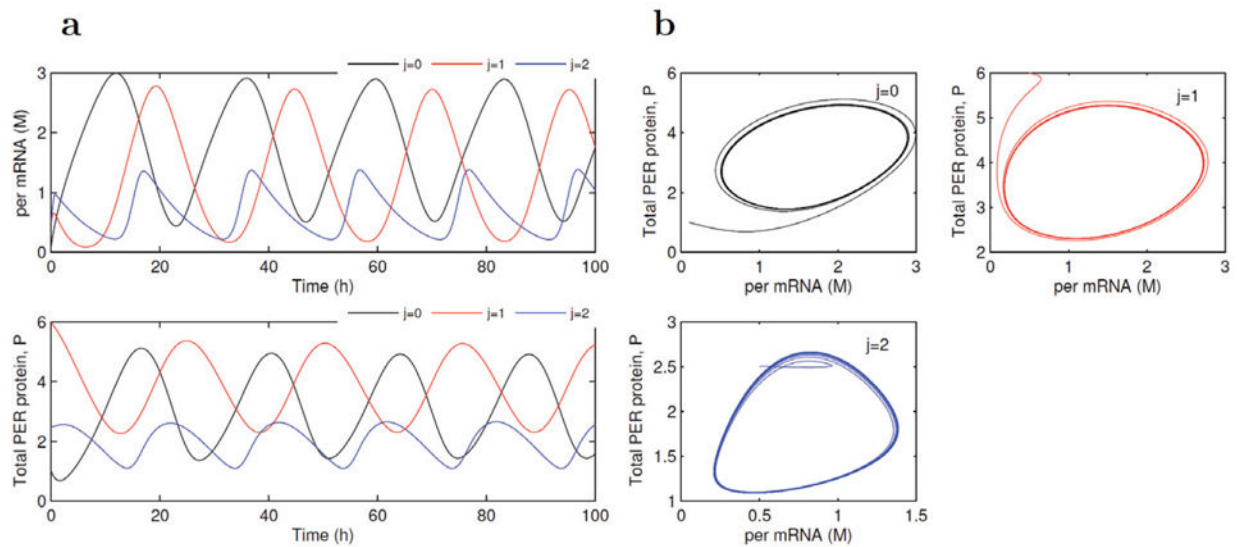


Figure 11.

Sustained oscillations of per mRNA and the total amount of PER protein and limit cycle generated by ODE (19). Rhythmic curves for three genotypes. $j = 0$ for qq , 1 for Qq , and 2 for QQ , at a hypothesized QTL are determined by a combination of parameters, i.e.,

$(v_s, v_m, K_s, v_d, k_1, k_2, K_b, K_d, K_1, K_2, K_3, K_4, n, v_1, v_2, v_3, v_4, K_m) = (0.76,$

$0.65, 0.38, 0.95, 1.9, 1.3, 1.0, 2, 2, 2, 2, 4, 3.2, 1.58, 5, 2.5, 0.5)$ for curve genotype $j = 0$,

$(0.82, 0.5, 0.28, 0.6, 2.3, 1.1, 1.6, 0.5, 2.2, 2.2, 2.2, 2.2, 2.2, 6, 2.2, 1.38, 8, 1.5, 0.7)$ for curve genotype $j =$

1, and $(0.30, 0.48, 0.75, 1.5, 0.8, 0.6, 0.7, 2.5, 2.5, 2.5, 2.5, 9.9, 4.2, 1.8, 1.2, 9, 0.5, 1.5)$ curve genotype $j = 2$.