Published in final edited form as:

Adv Drug Deliv Rev. 2013 June 30; 65(7): 918–928. doi:10.1016/j.addr.2013.04.007.

Systems mapping of metabolic genes through control theory*

Guodong Liu^a, Lan Kong^a, Zhong Wang^a, Chenguang Wang^{b,*}, and Rongling Wu^{a,**}

^aDivision of Biostatistics and Bioinformatics, Pennsylvania State University, Hershey, PA 17033, USA

^bThe Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD 21287, USA

Abstract

The formation of any complex phenotype involves a web of metabolic pathways in which one chemical is transformed through the catalysis of enzymes into another. Traditional approaches for mapping quantitative trait loci (QTLs) are based on a direct association analysis between DNA marker genotypes and end-point phenotypes, neglecting the mechanistic processes of how a phenotype is formed biochemically. Here, we propose a new dynamic framework for mapping metabolic QTLs (*m*QTLs) responsible for phenotypic formation. By treating metabolic pathways as a biological system, robust differential equations have proven to be a powerful means of studying and predicting the dynamic behavior of biochemical reactions that cause a high-order phenotype. The new framework integrates these differential equations into a statistical mixture model for QTL mapping. Since the mathematical parameters that define the emergent properties of the metabolic system can be estimated and tested for different *m*QTL genotypes, the framework allows the dynamic pattern of genetic effects to be quantified on metabolic capacity and efficacy across a time-space scale. Based on a recent study of glycolysis in *Saccharomyces cerevisiae*, we design and perform a series of simulation studies to investigate the statistical properties of the framework and validate its usefulness and utilization in practice. This framework can be generalized to mapping QTLs for any other dynamic systems and may stimulate pharmacogenetic research toward personalized drug and treatment intervention.

Keywords

Systems mapping; QTL; Metabolic system; Differential equations; Mixture model

1. Introduction

The formation of any complex traits or diseases occurs in a series of building blocks of cellular processes, involving metabolic pathways or protein–protein interaction networks [1]. Metabolomics, an emerging discipline of studying small-molecule metabolites, may be

*This review is part of the *Advanced Drug Delivery Reviews* theme issue on “Mathematical modeling of systems pharmacogenomics towards personalized drug delivery”.

© 2013 Elsevier B.V. All rights reserved.

*Corresponding author. Tel.: +1 410 614 4961; fax: +1 410 955 0859. cwang68@jhmi.edu (C. Wang). **Corresponding author. Tel.: +1 717 531 2037; fax: +1 717 531 0480. rwu@phs.psu.edu (R. Wu).

particularly useful for understanding the physiology of dynamic cellular processes and for diagnosis of disease [2,3]. For example, by studying granular metabolic pathways, which reflect biological responses to exogenous and endogenous factors, the molecular mechanisms of cardiovascular diseases can be better uncovered [4]. It has been recognized that metabolic profiles of the neonate, in terms of levels of amino acid, organic acid and fatty acid oxidation metabolites, are associated with obesity, diabetes and cardiovascular disease in adulthood through metabolic programming [2].

The phenotypes of metabolites and their associations with disease phenotypes may be controlled by genes [2,5,6]. Therefore, the elucidation of the underlying genetic architecture of metabolic traits predisposing to diseases can enhance our understanding of the genetic regulation of complex metabolic networks. Genetic mapping by linking molecular markers and trait phenotypes in segregating populations has been instrumental for identifying the genetic determinants of metabolic profiles, known as metabolic quantitative trait loci (*mQTL*), in a relation to particular traits or diseases [7–9]. Traditional strategies for *mQTL* mapping are based on a direct association analysis between marker genotypes and end-point phenotypes. It is likely that these strategies affect our statistical inference about the dynamic regulation of *mQTL*, given that metabolic responses and fluxes respond to various environmental or physiological stimuli in a dynamic matter [10,11].

More recently, a more powerful mapping strategy has been developed for genetic mapping of complex traits by dissolving the phenotype of a trait into its developmental, physiological, anatomical or biochemical components and further mapping QTLs involved in coordination and organization among a web of interactive components [12]. This strategy, called systems mapping, capitalizes on a system of differential equations to specify the dynamic behavior of phenotypic formation as a biological system. Compared to traditional static mapping for a steady state, systems mapping possesses several unique advantages: first, it maps and studies the genetic architecture of complex traits from their underlying mechanistic processes and pathways, with results being biologically more meaningful; second, it makes full use of mathematical equations to simplify the complexity of trait formation into key tractable elements, enhancing the statistical power of QTL mapping; and third, it allows various hypotheses about the interplay between genes and development to be formulated and tested in a quantitative way. Systems mapping has been employed for genetic mapping in a variety of biological phenomena from biomass allocation [13] to circadian rhythm [14,15] to viral dynamics [16].

In this article, we develop a new dynamic framework for mapping *mQTLs* that control metabolic pathways within the context of systems mapping. For a particular metabolic pathway, one chemical is transformed through the catalysis of enzymes into the other which may further form a new chemical. We use a system of ordinary differential equations (ODEs) to describe and quantify this process, aimed to investigate how different metabolites function synergistically to produce a desirable final phenotype. We implement ODE aspects of metabolic pathways into a statistical mixture model in which a mixture component is represented by an *mQTL* genotype. Since the mathematical parameters that define the emergent properties of the metabolic system can be estimated and tested for different *mQTL*

genotypes, the framework allows the dynamic pattern of genetic effects to be quantified on metabolic capacity and efficacy across a time-space scale.

The way of how genotype-specific ODE parameters for systems mapping are estimated represents a statistical challenge. The last five years have witnessed the development of various statistical methods for estimating ODE parameters based on state variables measured at multiple time points [17–20]. The estimation of ODE parameters within systems mapping is even more complex because the QTL genotypes that define the dynamic system are unobserved. Although conventional mathematical approaches for ODE solution have been successfully integrated into systems mapping [12,14,15], the performance may deteriorate quickly when the number of parameters increases. In order to effectively and efficiently solve ODEs for *m*QTL mapping, we propose to adopt the well-established results in control theory by representing the system of ODEs as a feedback control system. Control theory has proven to be useful in the study of biological and biochemical systems where the issues of regulation and control are central [21]. Characterization of general control rules that underpin metabolic dynamics is an important part of systems analysis in biology. It has been long recognized that many biological regulatory mechanisms have evolved to optimize cellular adaptation in response to external stimuli, concordant with the design principle of control theory. We have performed extensive simulation studies to investigate the statistical properties of systems mapping equipped with control theory. The results from computer simulation have validated the utility and usefulness of the new mapping model for *m*QTL identification in practice.

2. Modeling metabolic reactions as a system

We consider metabolic dynamics as a feedback control system and apply the concept of control theory to facilitate the optimization problem in *m*QTL mapping. To describe our mapping model, we use glycolytic oscillations, a well-studied metabolic reaction, as a case study. Glycolysis is a metabolic process that provides central energy in the form of ATP (adenosine triphosphate) for a living cell. As the most studied control system, many mathematical models have been developed to describe glycolytic oscillations, in which the concentrations of metabolites fluctuate. To identify the function and underlying mechanisms of glycolytic oscillations, Chandra et al. [22] have recently proposed a minimal three-reaction system with specific mechanisms both necessary and sufficient for oscillations in *Saccharomyces cerevisiae*, which is specified by a system of ODEs, expressed as

$$\begin{aligned} \dot{x} &= \frac{2y^a}{1+y^{2h}} - \frac{2kx}{1+y^{2g}} \\ \dot{y} &= -q \frac{2y^a}{1+y^{2h}} + (q+1) \frac{2kx}{1+y^{2g}} - (1+\delta) \end{aligned}$$

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 \\ -q \end{bmatrix}}_{\text{PFK}} \underbrace{\frac{2y^a}{1+y^{2h}}}_{\text{PK}} + \underbrace{\begin{bmatrix} -1 \\ (q+1) \end{bmatrix}}_{\text{PK}} \underbrace{\frac{2kx}{1+y^{2g}}}_{\text{Consumption}} + \underbrace{\begin{bmatrix} 0 \\ -(1+\delta) \end{bmatrix}}_{\text{Consumption}} \quad (1)$$

where functions $x = x(t)$ and $y = y(t)$ represent the time-varying concentrations of the lumped variable of intermediate metabolites and output ATP, respectively. Parameter a is the cooperativity of ATP binding to phosphofructokinase (PFK); h is the feedback strength of

ATP on PFK; k is the intermediate reaction rate; g is the feedback strength of ATP on PK; q is the autocatalytic stoichiometry; and δ indicates the external perturbation in ATP consumption. Five parameters (a, q, g, h, k), each with a biological interpretation, determine the properties of glycolysis. As being modeled by system of ODEs (1), q ATP molecules are used up by PFK to produce a pool of intermediate metabolites, which are then fed into PK to generate $q + 1$ ATP molecules. So during each round of reactions, a net one ATP molecule is produced by PFK + PK reactions and then consumed at the third reaction that models the cell's use of ATP. ATP inhibits both PFK and PK reactions.

Chandra et al. [22] used control theory to illustrate how the tradeoffs between efficiency and robustness arise from individual parameters in this model, including the interplay of feedback control with autocatalysis of network products necessary to power and catalyze intermediate reactions. In addition, they explicitly derived analytic equations for hard tradeoffs with oscillations as an inevitable side effect. We integrate ODEs (1) into a systems mapping framework to devise a computational model for identifying specific *m*QTLs that control glycolytic oscillations and further explore how the change of QTL effect parameters leads to the alteration of system behavior, structure and robustness.

3. Design and model

3.1. Segregating population

There are a number of segregating populations that can be used for genetic mapping [23]. For simplicity, we consider a backcross population of size n , initiated with two homozygous lines, in which there are only two possible genotypes, denoted by 0 or 1, at a genetic marker. All backcross progeny are genotyped for molecular markers throughout the genome, which leads to the construction of a genetic linkage map covering the entire genome. Focusing on glycolytic oscillations, the backcross progeny are measured for the lumped variable of intermediate metabolites (x) and ATP concentration (y) at a series of T time points. Our hypothesis is that there exist some particular *m*QTLs that control the emergent properties of glycolytic oscillations.

Let first consider such an *m*QTL that is flanked by two adjacent markers, M_1 and M_2 . Let r , r_1 and r_2 denote the recombination rates between markers M_1 and M_2 , M_1 and *m*QTL, and *m*QTL and M_2 , respectively. Although QTL genotypes cannot be observed, they can be inferred from the genotypes of the two markers that bracket the *m*QTL. Assuming that the crossovers at adjacent intervals are independent, i.e., $r = r_1 + r_2 - 2r_1r_2$, the conditional probabilities of QTL genotypes, conditional upon the interval genotypes, can be derived and shown in Table 1. For an arbitrary backcross progeny whose marker genotypes are known, the prior probability with which it carries a particular QTL genotype can be calculated from the table. If there is no double crossover, we have $r \approx r_1 + r_2$ and define $\theta = r_1/r$ as a parameter to specify the location of the QTL within the marker interval.

Different from traditional static QTLs, the *m*QTL of interest to glycolytic oscillations affects a dynamic behavior of the system specified by a system of ODEs (1). Thus, if the *m*QTL plays a role in regulating glycolytic metabolism, a set of ODE parameters (a, q, g, h, k) should differ between the genotypes of *m*QTL. By testing and estimating QTL genotype-

specific differences in parameters (a, q, g, h, k), we can infer the pattern of whether and how the m QTL governs the emergent properties of glycolytic oscillations. Again, since QTL genotypes are unobservable, our model will be formulated within a mixture setting in which each mixture component is represented by a set of ODE parameters.

3.2. Likelihood function

For a backcross progeny i , the dynamic process of glycolysis is characterized by two traits z_{1i} and z_{2i} , corresponding to the concentrations of the lumped variable of intermediate metabolites and ATP, respectively. These two traits are measured for each progeny at the same set of time points $t = 1, \dots, T$, expressed as random vectors $Z_{1i} = \{z_{1i}(1), \dots, z_{1i}(T)\}$ and $Z_{2i} = \{z_{2i}(1), \dots, z_{2i}(T)\}$. We consider the bivariate longitudinal trait $Z_i = \{Z_{1i}, Z_{2i}\}$ arising from a mixture of multivariate normal distributions:

$$Z_i \sim f(z_i | \Omega, \theta) = \sum_{j=1}^2 \omega_{j|i} f_j(Z_i | \Omega_{uj}, \Omega_v) \quad (2)$$

where $\omega_{j|i}$ is the mixture proportion of a QTL genotype j ($j = 1, 2$ for the backcross), $f_j(Z_i | \Omega_{uj}, \Omega_v)$ is a QTL genotype-specific multivariate normal density function with $2T \times 1$ mean vector, $\mu_j = \{\mu_{1j}, \mu_{2j}\}$ and $2T \times 2T$ common covariance matrix, Σ . Parameter vectors Ω_{uj} and Ω_v represent the set of parameters to define μ_j and Σ , respectively, whereas θ is the QTL location parameter contained within $\omega_{j|i}$.

Given the bivariate trait data $Z = \{Z_i; i = 1, \dots, n\}$ and marker data denoted by M for n progenies, we formulate the likelihood function of $\Omega = (\Omega_{u1}, \Omega_{u2}, \Omega_v)$ and by

$$L(\Omega, \theta | Z, M) = \prod_{i=1}^n \sum_{j=1}^2 \omega_{j|i} f_j(Z_i | \Omega_{uj}, \Omega_v). \quad (3)$$

The mean vector of multivariate normal distribution, μ_j , describes the average trait trajectory of those backcross progeny which carry QTL genotype j . In systems mapping, this vector is modeled by mathematical functions with biological meanings. Here, we adopt Chandra et al.'s ODEs (1) to model the mean functions of the concentrations of intermediate metabolites and ATP in glycolysis, which suggests that the QTL affects the glycolytic dynamic system through the mathematical parameters in the system of ODEs. Then, the solutions, say $\mu_{1j}(t)$ and $\mu_{2j}(t)$, to the ODEs (1) with parameters specified for QTL genotype j , form the mean vector $\mu_j = \{\mu_{1j}(1), \dots, \mu_{1j}(T), \mu_{2j}(1), \dots, \mu_{2j}(T)\}$.

With respect to the choice of covariance matrix Σ for the bivariate longitudinal traits, we use the structured antedependent (SAD) model which has shown many favorable properties in several studies [24]. The SAD model is more flexible than the autoregressive model by allowing non-stationary variances and correlations. The s th order SAD model assumes that an observation at time t is independent of all observations before $t - s$. Zhao et al. [25] incorporated a first-order SAD model to QTL functional mapping for bivariate longitudinal traits. They derived a closed form for the determinant and inverse of the structured covariance matrix, which significantly enhances the computing efficiency. Assuming a first-

order SAD (i.e. SAD (1)) model for residual errors, as shown in [25], we express the relationship between two traits $z_1(t)$ and $z_2(t)$ at time t for progeny i by

$$\begin{aligned} z_{1i}(t) &= \sum_{j=1}^2 \xi_{ij} \mu_{1j}(t) + e_{1i}(t) = \sum_{j=1}^2 \xi_{ij} \mu_{1j}(t) + \phi_1 e_{1i}(t-1) + \psi_1 e_{2i}(t-1) + \varepsilon_{1i}(t) \\ z_{2i}(t) &= \sum_{j=1}^2 \xi_{ij} \mu_{2j}(t) + e_{2i}(t) = \sum_{j=1}^2 \xi_{ij} \mu_{2j}(t) + \phi_2 e_{2i}(t-1) + \psi_2 e_{1i}(t-1) + \varepsilon_{2i}(t) \end{aligned} \quad (4)$$

where ξ_{ij} is the indicator variable denoted as 1 if progeny i carries QTL genotype j and 0 otherwise; $\mu_{1j}(t)$ and $\mu_{2j}(t)$ are the mean phenotypic values of QTL genotype j for two traits at time t ; $e_{1i}(t)$ and $e_{2i}(t)$ are the residuals of progeny i for the two traits including the polygenic and random error effects; ϕ_1 , ϕ_2 and ψ_1 , ψ_2 are the antedependence parameters induced by one trait itself or by the other trait, respectively; and $\varepsilon_{1i}(t)$ and $\varepsilon_{2i}(t)$ are the “innovation” errors assumed to be bivariate normally distributed with mean zero and variance matrix:

$$\Sigma_e(t) = \begin{pmatrix} \gamma_1^2(t) & \gamma_1(t)\gamma_2(t)\rho(t) \\ \gamma_1(t)\gamma_2(t)\rho(t) & \gamma_2^2(t) \end{pmatrix} \quad (5)$$

As shown in the Appendix A of Zhao et al. [25], both the determinant and inverse of covariance matrix Σ for bivariate traits Z_i have closed analytic forms, leading to the much easier computation of likelihood function (2). We further assume that $\Sigma_e(t)$ is time invariant so that the parameters of covariance matrix can be represented by $\Omega_c = (\phi_1, \phi_2, \psi_1, \psi_2, \gamma_1, \gamma_2, \rho)$.

3.3. Computational algorithm

The EM algorithm has been powerful for estimating a mixture model-based likelihood in traditional QTL mapping. However, there are two layers of difficulties in applying classic EM algorithm to our systems mapping of metabolic genes. First, the nonlinear ODEs that characterize the metabolic process usually have no analytical solution, implying that the mean vector of multivariate normal distribution cannot be represented in a closed form. Secondly, the number of parameters involved in the mean and covariance structures of longitudinal processes is fairly large, requiring more delicate handling of parameter initialization to ensure convergence of the EM algorithm at a reasonable rate. We approximate the solutions to the system of ODEs by Runge–Kutta methods [26]. To improve the initial estimation of ODE parameters, we treat glycolysis as a feedback control system and apply initial-/steady-/transient-state analysis from control theory to establish the constraints over the mathematical parameters.

In the EM algorithm, the objective function to be maximized is the conditional expectation of complete-data log likelihood function. Let u_{ij} be the unobserved indicator of carrying a particular QTL genotype j for progeny i . Then the complete data including both the observed and unobserved parts take the form of (Z, u, M) where $Z = \{Z_i, i = 1, \dots, n\}$, $u = \{u_{ij}, i = 1, \dots, n; j = 1, 2\}$ and M represents the genotypes of flanking markers for all progeny. For the parameters $\Omega = (\Omega_{u1}, \Omega_{u2}, \Omega_v)$ and θ in the observed data likelihood function (2), the

conditional expectation of complete-data log likelihood function given the estimates of Ω and θ at the τ th iteration is

$$\begin{aligned} Q(\Omega, \theta | \Omega^{(\tau)}, \theta^{(\tau)}) &= E \left[\log L(Z, u; \Omega, \theta | Z; \Omega^{(\tau)}, \theta^{(\tau)}) \right] \\ &= \sum_{i=1}^n \sum_{j=1}^2 E \left[u_{ij} \left\{ \log \omega_{j|i} + \log f_j(Z_i; \Omega_{uj}, \Omega_v) \right\} | Z_i; \Omega^{(\tau)}, \theta^{(\tau)} \right] \quad (6) \\ &= \sum_{i=1}^n \sum_{j=1}^2 \hat{u}_{ij} \log \omega_{j|i} + \sum_{i=1}^n \sum_{j=1}^2 \hat{u}_{ij} \log f_j(Z_i; \Omega_{uj}, \Omega_v) \end{aligned}$$

where

$$\hat{u}_{ij} = E \left[u_{ij} | Z_i; \Omega^{(\tau)}, \theta^{(\tau)} \right] = \frac{\omega_{j|i}(\theta^{(\tau)}) f_j(Z_i; \Omega_{uj}^{(\tau)}, \Omega_v^{(\tau)})}{\sum_{j'=1}^2 \omega_{j'|i}(\theta^{(\tau)}) f_{j'}(Z_i; \Omega_{uj'}^{(\tau)}, \Omega_v^{(\tau)})}.$$

In the E step, we calculate the posterior probability, $\hat{u}_{ij}$, with which a backcross progeny i carries a QTL genotype j given the observed data and current estimates of parameters; while at the M step, we maximize the function $Q(\Omega, \theta | \Omega^{(\tau)}, \theta^{(\tau)})$ to obtain the updated estimates of Ω and θ . We iterate between the E and M steps until the algorithm converges.

In the M step, we can maximize the function $Q(\Omega, \theta | \Omega^{(\tau)}, \theta^{(\tau)})$ by differentiating it with respect to each unknown parameter, setting the derivatives equal to zero and solving the resultant system of equations. Since the parameter θ is only involved in the first term of function $Q(\cdot)$ where the distributional parameter vector Ω is not present, we can estimate it as a function of posterior probabilities. Due to the fact that there exist closed-forms for the determinant and the inverse of covariance matrix Σ under the SAD (1) model, the estimating equations derived from $Q(\cdot)$ function for the covariance parameter vector, Ω_v , also has a closed form. Therefore, we can solve these equations to obtain the estimate of Ω_v by fixing the mean parameters at their current estimates. Because of the use of nonlinear ODEs for mean functions, there is no close form for the estimating equations associated with the QTL genotype-specific mean parameter vector Q_{mj} . We apply a search-based method such as Nelder–Mead algorithm [27] to estimate Q_{mj} by fixing the covariance parameter vector Ω_v . Our optimization strategy in the M step shares the same spirit as that of expectation conditional maximization algorithm [28] in which a sequence of conditional maximization steps are taken in the M step so that each parameter can be estimated individually, conditionally on the other parameters remaining fixed.

Systems mapping raises many computational challenges. First, the mixture log-likelihood function typically has many local maximums as in the traditional QTL mapping. Secondly, the incorporation of nonlinear ODEs to the multivariate normal distribution further complicates the estimation procedure. Moreover, like most greedy heuristic searching method, the Nelder–Mead method is prone to being trapped at the local extremes. Therefore, good initialization of unknown parameters is critical for the stability and convergence of the EM algorithm. We apply techniques such as initial-/steady-/transient-state analysis from control theory to derive more accurate initial values for the unknown mean parameters and

impose constraints among them to speed up the convergence rate of the Nelder–Mead optimization. In the following, we describe how each of the mean parameters that define the system of glycolysis is initialized on the basis of the results from state analyses of a control system. More details on the state analyses of glycolysis are provided in the Appendix A.

In the system of ODEs (1) proposed for glycolysis, five parameters (a, q, g, h, k) are used to characterize the trade-off between the efficiency and robustness of glycolytic oscillations. We may write the mean parameter vector $Q_{mj} = (a_j, q_j, g_j, h_j, k_j)$ for QTL genotype j , where $j = 1, 2$ for a backcross population. Due to the fact that, in glycolysis, two ATP (y) molecules are consumed upstream and four are produced downstream, we can normalize the parameter $q_j = 1$ (i.e., each ATP produces two downstream) with kinetic exponent $a_j = 1$.

1. **Initialization of k :** Recall that the solutions to the ODEs (1), $x(t)$ and $y(t)$, describe the mean functions of the lumped variable of intermediate metabolites and ATP concentration, respectively. Let x_0 and y_0 be the initial/undisturbed values of x and y at time $t = 0$. It follows the initial-state analysis of glycolysis that $k = y_0/x_0$. This result implies that, we may obtain a good initial value for k once we have good estimates of y_0 and x_0 for each QTL genotype. If we can find reasonable estimates of $x(t)$ and $y(t)$ based on the observed bivariate longitudinal trait data ($Z_1, \dots, Z_n$) from n progeny, we will use the resultant estimates of $x(0)$ and $y(0)$ to calculate the initial value for k . In practice, we obtain the crude estimate of $x(t)$ and $y(t)$ for a particular QTL genotype j by the weighted average of bivariate trait trajectories across all progeny, i.e.,

$$\bar{Z}_j = \frac{\sum_{i=1}^n \omega_{j|i} Z_i}{\sum_{i=1}^n \omega_{j|i}} \quad (7)$$

where $\omega_{j|i}$ is the conditional probability of a particular QTL genotype j for progeny i based on its flanking marker information (Table 1).

2. **Initialization of h :** A steady state of a feedback control system is the state when the signal essentially stabilizes at a steady value with little fluctuation even a small disturbance persists. Steady-state error (SSE) is a measure to quantify the difference between the steady-state value after transients died out and the initial value before any perturbation is added to the system. SSE divided by the perturbation magnitude (δ), namely the SSE ratio, is used to describe how good a control system is to buffer against the perturbation from outside of the system. Steady-state analysis of the glycolysis model (1) results in the following relationship

$$\left| \Delta \bar{y}_{\delta} \right| = \left| \frac{1}{h-a} \right| \quad (8)$$

where $\bar{y}$ is the average of SSE and δ is the average perturbation when the system approaches a steady equilibrium. This relationship can help set the initial value of parameter h once we estimate the SSE ratio from the crude estimate of mean

function $y(t)$, as described in (1). In practice, it is not difficult to identify a rough time point, λ , when a process is essentially at the steady state. For each QTL genotype, we can estimate $\bar{y}$ by averaging the estimated $y(t)$ from (1) over $t \rightarrow \lambda$. Usually, the external perturbation δ is known or can be well estimated. So we may have a good initial value for h when we set $a = 1$ by normalization.

3. **Initialization of g :** Once other ODE parameters are initialized, we apply a search-based optimization method to find the value of g that minimizes the least-squares errors between the mean function defined by ODEs (1) and the crude estimate of mean function.
4. **Initialization of Ω_v :** We initialize the covariance parameter vector Ω_v using the approach similar to that for the M step of EM algorithm. We substitute the initial values of ODE parameters and ω_{ji} into the $Q(\cdot)$ function (6), derive closed forms for the derivatives of $Q(\cdot)$ with respect to each unknown covariance parameter, and set them equal to zero. The solution to the resultant system of equations provides the initial values of Ω_v .

3.4. Hypothesis testing

We will test whether an m QTL exists between any two flanking markers to affect glycolytic oscillation. In so doing, we can formulate the null and alternative hypothesis in terms of the mathematical parameters in ODEs as follows:

H_0 The ODE parameters are the same for different QTL genotypes in any interval

H_1 At least one of the ODE parameters is different between QTL genotypes

Basically, the null hypothesis corresponds to the case when the bivariate trait data arise from a single glycolysis system, while the alternative hypothesis assumes that there exist two distinct glycolysis systems determined by the QTL genotypes. We test the hypothesis by using the log-likelihood ratio (LR) test statistic defined by

$$LR = -2 \ln \left[\frac{L(\hat{\Omega}_m, \hat{\Omega}_v)}{L(\hat{\Omega}_{m_1}, \hat{\Omega}_{m_2}, \hat{\Omega}_v)} \right] \quad (9)$$

where the estimates in the denominator are the MLEs under no restriction and those in the numerator are the MLEs under the null hypothesis H_0 . We calculate the LRs for all marker intervals and we would reject H_0 if any of the likelihood ratios is significantly higher than a threshold. Because it is difficult to analytically obtain the distribution of the test statistics, the critical threshold can be determined empirically from permutation tests [29]. We permute the trait values, breaking association between genotypes and phenotypes to mimic a null distribution. For each permutation data set, we calculate the maximum LR score. We construct the histogram of maximum LR scores and choose the upper 5th percentile as the cutoff threshold for a test with significance level at 0.05.

We use the so-called *grid search* method to obtain an optimal estimate of the QTL location. To assume that the QTL position, bracketed by two flanking markers, moves from one end

to the other by a fixed step size, we can calculate the LR test statistic for each assumed QTL position while treating the parameter θ as a fixed constant in the EM algorithm. If the test statistic at a region exceeds a predefined critical threshold, a significant QTL is indicated at the peak of the LR profile. In this way, we can systematically search for a significant QTL throughout a given linkage group or even the entire genome.

Systems mapping can also allow us to test how a significant *m*QTL govern the dynamic behavior of glycolytic oscillations. Five parameters (a, q, g, h, k) in ODEs (1) each determine different aspects of glycolysis as a dynamic system. By testing the QTL genotype-specific difference in these parameters individually or through a combination, we can find the pattern of the *m*QTL affecting the structure, organization and dynamics of glycolytic oscillations.

4. Computer simulation

We designed and conducted a series of simulation studies to investigate the statistical properties of our *m*QTL mapping framework and the performance of our proposed computation algorithm. A simulated backcross includes bivariate longitudinal trait values simulated per the glycolysis model defined by the system of ODEs (1) and the SAD (1) model for residual errors. The data simulation and parameter estimation were conducted in Matlab and Simulink. To demonstrate the usefulness of the model, we used three different experiments for *m*QTL identification by our systems mapping.

Experiment I: We simulated a linkage group on which a putative *m*QTL is located somewhere. Using the joint and conditional probabilities in Table 1, the marker genotypes and the genotype of QTL for a backcross of 200 progeny were simulated. For each progeny carrying a particular QTL genotype, we generated bivariate longitudinal traits, $Z(t)$, at 13 time points ($t = 1, \dots, 13$). In Table 2, the true values of ODE parameters that are chosen from the literature are given for different QTL genotypes, Qq and qq , with which mean trajectories $\mu_{1j}(t)$ and $\mu_{2j}(t)$ were drawn, respectively (Fig. 1). From these mean trajectories, plus the SAD(1) parameters, we simulated individual phenotypic trajectories. These two genotypes perform very differently, with Qq associated with a system having better balance between efficiency and complexity than qq .

The perturbation δ in the system of ODEs is assumed to be a step function with a magnitude of 0.05. We estimated the model parameters for each of 200 replicated data sets and summarize the results in Table 2. All the parameters can be well estimated. The averaged estimates over 200 replicates are well consistent to the true values. The empirical standard errors of parameter estimates are reasonably low, indicating the high precision of the estimates. Our computational algorithm has shown to be very effective, taking about 3 iterations to converge in general. We also note that the initial values obtained through the state analysis of a control system can be very close to the true values. The limit cycles of two metabolic variables that vary between two QTL genotypes can be also reasonably estimated (Fig. 2).

Experiment II: Here, we designed a simulation scenario with a smaller-effect QTL by setting up ODE parameters that are closer between two genotypes (Fig. 3), as compared to those in Experiment I, while letting the other parameters same. It can be seen that our method still produce very accurate and precise estimates for all the parameters (Table 3; Fig. 4), even if an *m*QTL is thought of being small-sized. The computational efficiency of our algorithm remains comparable to Experiment I.

Experiment III: We attempt to investigate how our framework performs when there exist multiple QTLs that jointly affect quantitative traits in glycolysis. We simulated the genotype data for 40 markers, located on four linkage groups each of 10 equally-spaced markers. We make the following assumptions: (1) there is a single QTL on each of the first three linkage groups, with QTL effect sizes varying from the strongest on the first group to the weakest on the third group; (2) there is no QTL on the fourth linkage group; and (3) the joint effect of three QTLs on glycolysis dynamics is additive, and is reflected by the change of parameters in the system of ODEs.

Following the previous notation, we denote $\mu_{1j}(t)$ and $\mu_{2j}(t)$ as the mean functions of two longitudinal traits associated with the glycolysis system. Based on the system of ODEs (1) and assuming that parameters a and q are normalized to be 1, we can express the mean trajectories, $\mu_j(t) = \{\mu_{1j}(t), \mu_{2j}(t)\}$, as functions of three parameters g , h and k ; i.e., $\mu_j(t) = B(t; g_j, h_j, k_j)$. To incorporate the QTL effects on (g, h, k) , we modeled these parameters by

$$g = g_0 + \Delta g_{lj}, h = h_0 + \Delta h_{lj}, \text{ and } k = k_0 + \Delta k_{lj},$$

where g_{lj} , h_{lj} and k_{lj} are the effects associated with genotype j ($j = 1, 2$) of the l th QTL ($l = 1, 2, 3$). We set $(g_0, h_0, k_0) = (1.3, 4.1, 1.98)$ and the additive effects as shown in Table 4. The covariance parameters in the SAD (1) model remain the same as those in the first two experiments.

We applied the grid-search method as aforementioned to estimating the locations of QTLs and test the existence of QTL using the likelihood ratio (LR) statistics. In general, with a proper sample size, our *m*QTL mapping algorithm can reasonably precisely estimate the locations of *m*QTLs of different effect sizes, defined as 1.00 (large), 0.84 (moderate) and 0.18 (small), respectively (Fig. 5). In particular, there is sufficient power to accurately determine the existence and estimate the exact locations of the QTLs of large and moderate effect sizes using a modest sample size. However, a modest sample size is difficult in finding an *m*QTL of a small effect size. To detect such a small *m*QTL, a sample size 300 – 400 is needed.

In practice, false positives may occur when indicating QTLs in a region where no QTL is present. Type I error is used to measure the rate of false positives. In QTL mapping, it is important to keep the type I error rate below 5%. We estimated the type I error rate of likelihood ratio test statistic (9) under the null hypothesis (i.e. No QTL is present) at 0.05 level in a setup similar to experiment I and II. Through 100 replicates, the type I error rate was estimated to be 0.04. Our strategies of using a grid search for QTLs along with permutations to estimate the threshold of the test statistic have been previously successfully

used in general QTLs mapping with robust performance in controlling the type I error rate below a commonly acceptable level, i.e. 0.05.

5. Discussion

A quantitative understanding of the complexity of cellular metabolism integrated with tissue, organ, and whole-body processes can shed light on the etiology of complex human diseases [2,3]. Through computer simulations and validation with experimental data, sophisticated mathematical models have been increasingly used to study the patterns of how cellular metabolic reactions are coordinated and organized and how a large number of chemical species are transported. ODE are particularly powerful for handling highly nonlinear phenomena of spatially lumped and/or distributed systems, allowing quantitative evaluation of metabolic pathways and regulatory mechanisms under normal and abnormal conditions [22].

Here, we argue that the integration of these physiologically-based mathematical models with a powerful dynamic mapping model, systems mapping [12,14,15], provides unprecedented power to quantify the genetic architecture (by mapping metabolic QTLs, *mQTLs*) and predict responses of metabolic systems based on an individual's genetic makeup. While traditional mapping of *mQTLs* is to simply associate molecular markers with end-point phenotypes, such integration incorporates a dynamic mechanism of metabolic pathways. To the end, it can provide a key step toward simulating the integrated effects of altering enzyme activities or substrate concentrations with personalized pharmacological agents.

Previous modeling of systems mapping using the system of ODEs has focused on the simulation and analysis of the behavior of state variables for a dynamic system. In this article, we consider metabolic dynamics as a feedback control system and apply the concepts of control theory to facilitate the optimization problem in robust *mQTL* mapping. In particular, by incorporating techniques like initial, steady, and transient state analyses from control theory into our *mQTL* mapping pipeline, we are able to tightly constraint all the ODE parameters to close local neighborhoods of the true values of the parameters and even make fairly accurate initial estimates of certain parameters. This success in obtaining initial parameter estimates would ensure the EM algorithm typically implemented for *mQTL* mapping to quickly converge, resulting in accurate estimation of the model parameters.

We have proposed a dynamic framework for mapping metabolic QTLs in which a systems mapping approach is shown to effectively model metabolic pathway interactions and constraints. This fundamental framework could be extended to modeling a variety of more complex biological phenomena and processes, such as genotype–environment interactions, epistatic interactions, epigenetic marks, and phenotypic plasticity. Current epidemiological studies have developed to a point at which metabolic phenotypes of individual subjects are associated with disease phenotypes in a large-scale trial. These so-called metabolome-wide association studies [30] can be favorably stimulated by our systems mapping toward personalized drug and treatment intervention.

Acknowledgments

This publication was supported by NSF/IOS-0923975 and UL1 TR000127 from the National Center for Advancing Translational Sciences (NCATS). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

References

1. Goonewardena SN, Prevette LE, Desai AA. Metabolomics and atherosclerosis. *Curr Atheroscler Rep.* 2010; 12:267–272. [PubMed: 20464531]
2. Srinivasan M, Patel MS. Metabolic programming in the immediate postnatal period. *Trends Endocrinol Metab.* 2008; 19:146–152. [PubMed: 18329279]
3. Davidovic SP, Nikolic SV, Curic NJ, Latinovic SL, Draskovic DO, Cabarkapa VS, Stosic ZZ. Changes of serum VEGF concentration after intravitreal injection of Avastin in treatment of diabetic retinopathy. *Eur J Ophthalmol.* 2012; 22:792–798. [PubMed: 22344470]
4. Shah SH, Hauser ER, Bain JR, Muehlbauer MJ, Haynes C, Stevens RD, Wenner BR, Dowdy ZE, Granger CB, Ginsburg GS, Newgard CB, Kraus WE. High heritability of metabolomic profiles in families burdened with premature cardiovascular disease. *Mol Syst Biol.* 2009; 5:258. [PubMed: 19357637]
5. Alul FY, Cook DE, Shchelochkov OA, Fleener LG, Berberich SL, Murray JC, Ryckman KK. The heritability of metabolic profiles in newborn twins. *Heredity.* 2012; 110:253–258. [PubMed: 23149456]
6. Nicholson G, Rantalainen M, Maher AD, Li JV, Malmmodin D, Ahmadi KR, Faber JH, Hallgrimsdottir IB, Barrett A, Toft H, Krestyaninova M, Viksna J, Neogi SG, Dumas M-E, Sarkans U, The Mol PC, Silverman BW, Donnelly P, Nicholson JK, Allen M, Zondervan KT, Lindon JC, Spector TD, McCarthy MI, Holmes E, Baunsgaard D, Holmes CC. Human metabolic profiles are stably controlled by genetic and environmental variation. *Mol Syst Biol.* 2011; 7:525. [PubMed: 21878913]
7. Dumas ME, Wilder SP, Bihoreau MT, Barton RH, Fearnside JF, Argoud K, D'Amato L, Wallis RH, Blancher C, Keun HC, Baunsgaard D, Scott J, Sidelmann UG, Nicholson JK, Gauguier D. Direct quantitative trait locus mapping of mammalian metabolic phenotypes in diabetic and normoglycemic rat models. *Nat Genet.* 2007; 39:666–672. [PubMed: 17435758]
8. Nicholson G, Rantalainen M, Li JV, Maher AD, Malmmodin D, Ahmadi KR, Faber JH, Barrett A, Min JL, Rayner NW, Toft H, Krestyaninova M, Viksna J, Neogi SG, Dumas M, Sarkans U, Donnelly P, Illig T, Adamski J, Suhre K, Allen M, Zondervan KT, Spector TD, Nicholson JK, Lindon JC, Baunsgaard D, Holmes E, McCarthy MI, Holmes CC. A genome-wide metabolic QTL analysis in Europeans implicates two loci shaped by recent positive selection. *PLoS Genet.* 2011; 7:e1002270. [PubMed: 21931564]
9. Cazier JB, Kaisaki PJ, Argoud K, Blaise BJ, Veselkov K, Ebbels TM, Tsang T, Wang Y, Bihoreau MT, Mitchell SC, Holmes EC, Lindon JC, Scott J, Nicholson JK, Dumas ME, Gauguier D. Untargeted metabolome quantitative trait locus mapping associates variation in urine glycerate to mutant glycerate kinase. *J Proteom Res.* 2012; 11:631–642.
10. Dash RK, Li Y, Kim J, Beard DA, Saidel GM, Cabrera ME. Metabolic dynamics in skeletal muscle during acute reduction in blood flow and oxygen supply to mitochondria: in-silico studies using a multi-scale, top-down integrated model. *PLoS One.* 2008; 3:e3168. [PubMed: 18779864]
11. Li Y, Solomon TP, Haus JM, Saidel GM, Cabrera ME, Kirwan JP. Computational model of cellular metabolic dynamics: effect of insulin on glucose disposal in human skeletal muscle. *Am J Physiol Endocrinol Metab.* 2010; 298:E1198–E1209. [PubMed: 20332360]
12. Wu R, Cao J, Huang Z, Wang Z, Gai J, 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]
13. Bo WH, Fu G, Wang Z, Xu F, Shen Y, Xu J, Vallejos CE, Wu RL. Systems mapping: how to map genes for biomass allocation towards an ideotype. *Brief Bioinform.* 2013 (in press).
14. Fu G, Luo J, Berg A, Wang Z, Li J, Das K, Li R, Wu R. A dynamic model for functional mapping of biological rhythms. *J Biol Dyn.* 2011; 5:84–101. [PubMed: 21278847]

15. Fu G, Wang Z, Li J, Wu R. A mathematical framework for functional mapping of complex phenotypes using delay differential equations. *J Theor Biol.* 2011; 389:206–216. [PubMed: 21871898]
16. Luo J, Hager WW, Wu R. A differential equation model for functional mapping of a virus-cell dynamic system. *J Math Biol.* 2010; 61:1–15. [PubMed: 19685244]
17. Ramsay JO, Hooker G, Campbell D, Cao J. Parameter estimation for differential equations: a generalized smoothing approach. *J R Stat Soc Ser B.* 2007; 69:741–796.
18. Liang H, Miao H, Wu H. Estimation of constant and time-varying dynamic parameters of HIV infection in a nonlinear differential equation model. *Ann Appl Stat.* 2010; 4:460–483. [PubMed: 20556240]
19. Miao H, Xia X, Perelson AS, Wu H. On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Rev.* 2011; 53:3–39.
20. Calderhead B, Girolami M. Statistical analysis of nonlinear dynamical systems using differential geometric sampling methods. *Interface Focus.* 2011; 1:821–835. [PubMed: 23226584]
21. Iglesias, PA.; Ingalls, BP. *Control Theory and Systems Biology.* MIT Press; Cambridge, Mass: 2010.
22. Chandra FA, Buzi G, Doyle JC. Glycolytic oscillations and limits on robust efficiency. *Science.* 2011; 333:187–192. [PubMed: 21737735]
23. Wu, R.; Ma, C.; Casella, G. *Statistical Genetics of Quantitative Traits: Linkage, Maps and QTL.* Springer-Verlag; New York: 2007.
24. Zimmerman D, Núñez-Antón V, Gregoire T, Schabenberger O, Hart J, Kenward M, Molenberghs G, Verbeke G, Pourahmadi M, Vieu P, Zimmerman D. Parametric modelling of growth curve data: an overview. *TEST.* 2001; 10:1–73.
25. Zhao W, Hou W, Littell RC, Wu R. Structured antedependence models for functional mapping of multiple longitudinal traits. *Stat Appl Genet Mol Biol.* 2005; 4 (Article 33).
26. Butcher, JC. *Numerical Methods for Ordinary Differential Equations.* J. Wiley; Chichester, West Sussex, England: 2003.
27. Nelder JA, Mead R. A simplex method for function minimization. *Comput J.* 1965; 7:308–313.
28. Meng XL, Rubin DB. Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika.* 1993; 80:267–278.
29. Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. *Genetics.* 1994; 138:963–971. [PubMed: 7851788]
30. Bictash M, Ebbels T, Chan Q, Loo R, Yap I, Brown I, De Iorio M, Daviglus M, Holmes E, Stamler J, Nicholson J, Elliott P. Opening up the “black box”: metabolic phenotyping and metabolome-wide association studies in epidemiology. *J Clin Epidemiol.* 2010; 63:970–979. [PubMed: 20056386]

Appendix A

As we present our computational algorithm for QTL mapping of glycolysis system, we describe how the initialization of parameters in the system of ODEs (1) can take advantage of state analysis results of a feedback control system that corresponds to the three-interaction glycolysis model. To help the readers understand how concepts and principles of control theory can be effectively applied to modeling complex biological systems/reactions such as glycolysis, we provide a step-by-step derivation of these results using classic initial-/steady-/transient-state analysis methodologies in control theory. Much of our mathematical derivation follows the work of Chandra et al. [22].

A.1. Initial-state analysis of glycolysis

The glycolysis model based on the ODEs (1) describes how intermediate metabolites (lumped into one variable x), and ATP (y) evolve and their interactions over time. We refer to $x(t)$ and $y(t)$, the solutions to ODEs, as signals in the context of a feedback control system. Let x_0 and y_0 be the initial/undisturbed concentration of x and y at time $t = 0$, i.e., $x_0 = x(0)$ and $y_0 = y(0)$. Assuming the signal velocity (rate of concentration change) at an initial state is zero, i.e. $\dot{x}_0 = 0$ and $\dot{y}_0 = 0$. Plugging it into the ODEs (1), we have

$$\begin{aligned} \frac{2y_0^a}{1+y_0^{2h}} - \frac{2kx_0}{1+y_0^{2g}} &= 0 \\ -q \frac{2y_0^a}{1+y_0^{2h}} + (q+1) \frac{2kx_0}{1+y_0^{2g}} - 1 &= 0 \end{aligned}$$

Since parameters a and q have been normalized to one in this glycolysis model, we can derive parameter k as follows:

$$k = \frac{y_0}{x_0}$$

A.2. Steady-state analysis of glycolysis

Linearization is a method for assessing a nonlinear system local stability at an equilibrium point from the stability properties of the linearized system. Linearization approximates extremely well to a non-linear function in a small neighborhood of a given point but significantly reduce its complexity. Chandra et al. [22] derived a linearization around an equilibrium point from the two-state glycolysis model as follows:

$$\begin{aligned} \begin{bmatrix} \Delta \dot{x} \\ \Delta \dot{y} \end{bmatrix} &= \begin{bmatrix} -k & a+g \\ (q+1)k & -qa-g(q+1) \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \end{bmatrix} \delta + \begin{bmatrix} -1 \\ q \end{bmatrix} h\Delta y \\ \begin{bmatrix} \Delta \dot{x}_1 \\ \Delta \dot{x}_2 \end{bmatrix} &= A \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} + B\delta = \begin{bmatrix} -k & (a-h)+g \\ (q+1)k & -q(a-h)-(q+1)g \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \end{bmatrix} \delta \\ \Delta y &= C \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} \end{aligned}$$

The closed-loop transfer function (also called the weighted sensitivity function) between the response y and ATP perturbation δ is as follows:

$$\begin{aligned} WS &= \frac{Y(s)}{\delta(s)} = C(sI - A)^{-1}B \\ &= \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} s+k & -(a-h)-g \\ -(q+1)k & s+q(q-h)+(q+1)g \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ -1 \end{bmatrix} \\ &= \frac{s+k}{s^2 + (k+g+aq-hq+gq)s + hk - ak} \end{aligned}$$

Let $N(s)$ be the numerator and $D(s)$ be the denominator of a transfer function. The roots of $N(s) = 0$ are called zeros; the roots of $D(s) = 0$ are called poles. The polynomial equation $D(s) = 0$ is also called the characteristic equation.

The number of zeros and poles and their individual values are critical to system stability as well as other important system properties. This is particularly true in the case of poles. For example, poles in the right half plane correspond to exponentially growing amplitudes, i.e. an unstable state; on the other hand, poles in the left half plane correspond to exponentially decreasing amplitudes, i.e. a stable state. In other words, all poles must be to the left of the imaginary axis for any stable control system. Effects of zeros are subtler than those of poles. Generally speaking, the closer to a pole means higher gain/amplitude while closer to a zero indicates lower gain/amplitude. Furthermore, in any physically realizable control systems, the number of poles must be greater than or equal to the number of zeros. This provides us with a qualitative understanding of what the system does at various frequencies.

To make it even simpler, we can determine if a system is stable by just looking at the coefficients of the characteristic equations using the Routh–Hurwitz stability criterion, which demands that coefficients of the polynomial of characteristic equation do not change signs for any stable control system. In our example of glycolysis, since the first coefficient of the characteristic equation is +1, both the second and third coefficients need to be positive. So this leads to

$$-k(a-h) > 0$$

Since $k > 0$, we have

$$h-a > 0$$

Furthermore, since

$$k+g+aq-hq+gq > 0$$

We then have

$$h-a < g + \frac{g+k}{q}$$

We can further derive more system parameter constraints of glycolysis using steady-state analysis, which studies system SSE ratio, among other steady-state properties. At an equilibrium point of glycolysis, we have

$$\begin{aligned}
 \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} &= \begin{bmatrix} -k & (a-h)+g \\ (q+1)k & -q(a-h)-(q+1)g \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \end{bmatrix} \delta = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
 \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} &= \begin{bmatrix} -k & (a-h)+g \\ (q+1)k & -q(a-h)-(q+1)g \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \delta \\
 \begin{bmatrix} \Delta x/\delta \\ \Delta y/\delta \end{bmatrix} &= \begin{bmatrix} -k & (q-h)+g \\ (q+1)k & -q(a-h)-(q+1)g \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
 \begin{bmatrix} \Delta x/\delta \\ \Delta y/\delta \end{bmatrix} &= \frac{1}{k(a-h)} \begin{bmatrix} (h-a)q-(q+1)g & h-a-g \\ k(1+q) & k \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
 \frac{\Delta y}{\delta} &= \frac{k}{k(a-h)} = \frac{1}{a-h}
 \end{aligned}$$

If we ignore the direction of the perturbation, we can get an absolute SSE ratio as follows:

$$\left| \frac{\Delta y}{\delta} \right| = \left| \frac{1}{h-a} \right|$$

Through a few rounds of mathematical manipulations we have derived constraints among parameters a , h and signals y and δ . In practice, we used the average SSE $\bar{y}$ and average disturbance δ when glycolysis approaches a steady equilibrium to approximate the true SSE y and disturbance δ .

Fig. 1.

Experiment I: Trajectories of two longitudinal traits (y : ATP, x : lumped variable of intermediate metabolites) associated with a big QTL having genotypes Qq (blue) and qq (red). The true and estimated trajectories are presented by solid thick and dashed thick curves, respectively. Thin curves with faded colors are individual noisy trajectories.

Fig. 2.

Experiment I: Limit cycles between the trajectories of two longitudinal traits (x : lumped variable of intermediate metabolites, y : ATP) associated with a big QTL having genotypes Qq (blue) and qq (red). The true and estimated trajectories are presented by solid thick and dashed thick curves, respectively.

Fig. 3.

Trajectories of two longitudinal traits (y : ATP, x : lumped variable of intermediate metabolites) associated with a small QTL having genotypes Qq (blue) and qq (red). The true and estimated trajectories are presented by solid thick and dashed thick curves, respectively. Thin curves with faded colors are individual noisy trajectories.

Fig. 4.

Experiment II: Limit cycles between the trajectories of two longitudinal traits (x : lumped variable of intermediate metabolites, y : ATP) associated with a big QTL having genotypes Qq (blue) and qq (red). The true and estimated trajectories are presented by solid thick and dashed thick curves, respectively.

Fig. 5.

The profile of log-likelihood ratio (LR). The red horizontal line indicates the critical threshold for declaring the existence of QTL on the basis of 100 permutation tests. The ticks on the x-axis indicate the positions of molecular markers. Vertical dashed lines indicate the identified QTL locations of QTLs of statistical significance above the critical threshold and red arrows point to the true QTL locations on the chromosomes.

Joint and conditional probabilities with interval mapping in a backcross genetic design. r is the recombination factor between the two flanking markers; r_1 is the recombination factor between the QTL and the left flanking marker; r_2 is the recombination factor between the QTL and the right flanking marker.

Table 1

Marker interval (M_1-M_2)		M_1 -QTL- M_2		QTL given two-marker information	
Genotype	Prob	Genotype	Joint prob	Conditional prob	Approx. cond. prob*
11	$(1-r)/2$	111	$(1-r_1)(1-r_2)/2$	$(1-r_1)(1-r_2)/(1-r)$	1
		101	$r_1r_2/2$	$r_1r_2/(1-r)$	0
10	$r/2$	110	$(1-r_1)r_2/2$	$(1-r_1)r_2/r$	$1-\theta = r_2/r$
		100	$r_1(1-r_2)/2$	$r_1(1-r_2)/r$	$\theta = r_1/r$
01	$r/2$	011	$r_1(1-r_2)/2$	$r_1(1-r_2)/r$	θ
		001	$(1-r_1)r_2/2$	$(1-r_1)r_2/r$	$1-\theta$
00	$(1-r)/2$	010	$r_1r_2/2$	$r_1r_2/(1-r)$	0
		000	$(1-r_1)(1-r_2)/2$	$(1-r_1)(1-r_2)/(1-r)$	1

Table 2

Parameter estimates and their standard errors obtained from 200 simulation replicates in *m*QTL mapping for Experiment I.

	Parameter	True value	Estimate mean estimate	SE
QTL Genotype qq	g_1	1.1	1.1013	0.0094
	h_1	4.5	4.5005	0.0162
	k_1	1.98	1.9799	7.7887e-4
QTL Genotype Qq	g_2	1.5	1.4992	0.0261
	h_2	3.9	3.8989	0.0288
	k_2	1.95	1.9501	0.0016
QTL Location	θ	0.3	0.3008	0.0518
SAD Parameters	γ_1	0.02	0.0199	2.6259e-4
	γ_2	0.0018	0.0018	2.4853e-5
	ϕ_1	0.1	0.0962	0.0196
	ϕ_2	0.2	0.1997	0.0049
	ψ_1	0.2	0.2062	0.0519
	ψ_2	0.4	0.3999	0.0018
	ρ	0.4	0.3991	0.0161

Table 3

Parameter estimates and their standard errors obtained from 200 simulation replicates in *m*QTL mapping for experiment II.

	Parameter	True value	Estimate	SE
QTL Genotype qq	g_1	1.1	1.0993	0.0108
	h_1	4.5	4.4982	0.0184
	k_1	2	2.0001	8.3736e-4
QTL Genotype Qq	g_2	1.3	1.3016	0.0179
	h_2	4.3	4.3029	0.0277
	k_2	2	1.9999	0.0011
QTL location	θ	0.3	0.3021	0.0531
SAD parameters	γ_1	0.02	0.0199	2.6755e-4
	γ_2	0.0018	0.0018	2.6477e-5
	ϕ_1	0.1	0.1006	0.0223
	ϕ_2	0.2	0.1992	0.0045
	ψ_1	0.2	0.1934	0.0517
	ψ_2	0.4	0.4001	0.0018
	ρ	0.4	0.3999	0.0159

Table 4

Additive effects of three QTLs on glycolysis system in experiment III.

QTL	Genotype (<i>j</i>)	<i>g</i>	<i>h</i>	<i>k</i>
1 (strong effect)	1	−0.15	0.01	0.01
	2	0.15	−0.01	−0.01
2 (moderate effect)	1	−0.1	−0.005	−0.005
	2	0.1	0.005	0.005
3 (weak effect)	1	−0.05	0	0
	2	0.05	0	0