

A differential equation model for functional mapping of a virus-cell dynamic system

Jiangtao Luo · William W. Hager · Rongling Wu

Received: 23 January 2009 / Revised: 13 July 2009
© Springer-Verlag 2009

Abstract The dynamic pattern of viral load in a patient's body critically depends on the host's genes. For this reason, the identification of those genes responsible for virus dynamics, although difficult, is of fundamental importance to design an optimal drug therapy based on patients' genetic makeup. Here, we present a differential equation (DE) model for characterizing specific genes or quantitative trait loci (QTLs) that affect viral load trajectories within the framework of a dynamic system. The model is formulated with the principle of functional mapping, originally derived to map dynamic QTLs, and implemented with a Markov chain process. The DE-integrated model enhances the mathematical robustness of functional mapping, its quantitative prediction about the temporal pattern of genetic expression, and therefore its practical utilization and effectiveness for gene discovery in clinical settings. The model was used to analyze simulated data for viral dynamics, aimed to investigate its statistical properties and validate its usefulness. With an increasing availability of genetic polymorphic data, the model will have great implications for probing the molecular genetic mechanism of virus dynamics and disease progression.

Keywords Differential equation · Functional mapping · Quantitative trait loci · Viral dynamics

Mathematics Subject Classification (2000) 62 Statistics

J. Luo · W. W. Hager
Department of Mathematics, University of Florida, Gainesville, FL 32611, USA

R. Wu (✉)
Departments of Public Health Sciences and Statistics,
Pennsylvania State University College of Medicine, Hershey, PA 17033, USA
e-mail: rww@hes.hmc.psu.edu

1 Introduction

Several serious human diseases, such as AIDS, hepatitis B, influenza, and rabies, are caused by viruses. To control these diseases, antiviral drugs have been developed to prevent infection of new viral cells or stop already-infected cells from producing infectious virus particles by inhibiting specific viral enzymes. This process constitutes a complex dynamic system, in which different types of viral cells, including uninfected cells, infected cells, and free virus particles, interact with each other to determine the pattern of viral change in response to drugs (Ho et al. 1999; Wei et al. 2004; Perelson et al. 1997, 1996; Sedaghat et al. 2008). A major challenge that faces drug development and delivery for controlling viral diseases is to develop a quantitative model for analyzing and predicting the dynamics of decline in virus load during drug therapy and further providing estimates of the rate of emergence of resistant virus.

The development of such a model can now be made possible with recent advances in two seemingly unrelated areas. First, the combination between novel instruments and an increasing understanding of molecular genetics has led to the birth of high-throughput genotyping assays for single nucleotide polymorphisms (SNPs). Through the construction of a haplotype map (HapMap) with SNP data (The International HapMap Consortium 2003), we are able to characterize concrete nucleotides or their combinations that encode a complex phenotype, and ultimately document, map and understand the structure and patterns of the human genome linked to drug response. Second, the past two decades have witnessed a tremendous growth of interest in deriving sophisticated mathematical models for characterizing virus dynamics from molecular and cellular mechanisms of interactions between virus and drug (Ho et al. 1999; Wei et al. 2004; Perelson et al. 1997, 1996; Sedaghat et al. 2008; Bonhoeffer et al. 1997; Bonhoeffer et al. 1999; Nowak and May 2000). These models mostly built with differential equations (DE) have been instrumental for studying the function of virus and the origins and properties of virus dynamics.

These two advances can be integrated to identify specific genes or quantitative trait loci (QTLs) that regulate a dynamic system of viral infection through a new statistical model called functional mapping (Ma et al. 2002; Wang and Wu 2004; Wu et al. 2004a,b,c; Wu and Lin 2006). The basic idea of functional mapping is to map dynamic QTL for the pattern of developmental changes in time course. The purpose of this article is to propose a statistical strategy for implementing a system of DE into the functional mapping framework, ultimately to map QTLs from the host genome that determine the dynamic pattern of virus load in patients' bodies. The new strategy is founded on a set of random samples drawn from a natural population at Hardy–Weinberg equilibrium (HWE). We integrate the Markov chain properties of dynamic data into the model to facilitate the estimation of parameters that define virus dynamics. Simulation studies were performed to investigate statistical properties of the model and validate its usefulness and utilization.

2 Dynamic models of virus load

2.1 Differential equations

A basic model for describing short-term virus dynamics was provided by many researchers (Bonhoeffer et al. 1997; Bonhoeffer et al. 1999; Nowak and May 2000). This model includes three variables: uninfected cells, x , infected cells, y , and free virus particles, v . These three types of cells interact with each other to determine the dynamic changes of virus in a host's body, which can be described by a system of ordinary differential equations (ODE):

$$\begin{aligned}\frac{dx}{dt} &= \lambda - dx - \beta xv \\ \frac{dy}{dt} &= \beta xv - ay \\ \frac{dv}{dt} &= ky - uv,\end{aligned}\tag{1}$$

where uninfected cells are yielded at a constant rate, λ , and die at the rate dx ; free virus infects uninfected cells to yield infected cells at rate βxv ; infected cells die at rate ay ; and new virus is yielded from infected cells at rate ky and dies at rate uv (Bonhoeffer et al. 1999). The system (1) is defined by six parameters $\{\lambda, d, \beta, a, k, u\}$ and the initial conditions for x , y , and v . The dynamic pattern of this system can be determined and predicted by the change of these parameters and the initial conditions of x , y , and v . There are some practical problems in the real application. First, we can only observe the data for x , y , and v at discrete time points, and it is difficult to get the continuous $\frac{dx}{dt}$, $\frac{dy}{dt}$, and $\frac{dv}{dt}$ terms. Second, any biological development is related to genes, but the model does not involve any genetic components. Third, the dynamic change of the virus is accompanied by noise which cannot be neglected in the dynamic modeling. It should be noted that the model (1) used to explain our idea in this article is a basic sculpture of real virus infection as it ignores the dynamics of immune responses and virus mutations.

Let $0 = t_0 < t_1 < \dots < t_N = T$ denote a mesh on the time interval $[0, T]$ and define $\Delta t_k = t_{k+1} - t_k$. The Euler approximation to the continuous differential equations (1) is

$$\begin{aligned}\frac{x(t_{k+1}) - x(t_k)}{\Delta t_k} &= \lambda - dx(t_k) - \beta x(t_k)v(t_k) \\ \frac{y(t_{k+1}) - y(t_k)}{\Delta t_k} &= \beta x(t_k)v(t_k) - ay(t_k) \\ \frac{v(t_{k+1}) - v(t_k)}{\Delta t_k} &= ky(t_k) - uv(t_k),\end{aligned}\tag{2}$$

or equivalently,

$$\begin{aligned}
 x(t_{k+1}) &= x(t_k) + \lambda \Delta t_k - dx(t_k) \Delta t_k - \beta x(t_k)v(t_k) \Delta t_k \\
 y(t_{k+1}) &= y(t_k) + \beta x(t_k)v(t_k) \Delta t_k - ay(t_k) \Delta t_k \\
 v(t_{k+1}) &= v(t_k) + ky(t_k) \Delta t_k - uv(t_k) \Delta t_k.
 \end{aligned}
 \tag{3}$$

2.2 Markov properties

Suppose there is a random sample with n patients from a population carrying a certain virus. Each patient is measured for uninfected cells, x , infected cells, y , and free virus particles, v , at a series of time points, (t_0, t_1, \dots, t_N) . Thus, three sets of serial measurements are expressed as $\mathbf{x}_i = [x_i(t_0), \dots, x_i(t_N)]$, $\mathbf{y}_i = [y_i(t_0), \dots, y_i(t_N)]$, and $\mathbf{v}_i = [v_i(t_0), \dots, v_i(t_N)]$, where the subscript i corresponds to the patient and t_j , $0 \leq j \leq N$, are the measurement times.

A transitional Markov model is used to describe the random process of the system by

$$\begin{aligned}
 x_i(t_{k+1}) &= x_i(t_k) + \lambda \Delta t_k - dx_i(t_k) \Delta t_k - \beta x_i(t_k)v_i(t_k) \Delta t_k + \epsilon_{x_i}(t_k) \\
 y_i(t_{k+1}) &= y_i(t_k) + \beta x_i(t_k)v_i(t_k) \Delta t_k - ay_i(t_k) \Delta t_k + \epsilon_{y_i}(t_k) \\
 v_i(t_{k+1}) &= v_i(t_k) + ky_i(t_k) \Delta t_k - uv_i(t_k) \Delta t_k + \epsilon_{v_i}(t_k),
 \end{aligned}
 \tag{4}$$

where $\epsilon_{x_i}(t_k) \sim N(0, \sigma_x^2)$, $\epsilon_{y_i}(t_k) \sim N(0, \sigma_y^2)$, and $\epsilon_{v_i}(t_k) \sim N(0, \sigma_v^2)$ are the innovation errors for three variables, x , y , and v , respectively, each of which is assumed to be iid and time-independent. To simplify our line of analysis, we assume that these three variables are independent of each other, although this assumption can be relaxed.

For simplicity, we use x_{ik} , y_{ik} , and v_{ik} to stand for $x_i(t_k)$, $y_i(t_k)$, and $v_i(t_k)$, respectively. For a conditional density function, $f(\cdot|\cdot)$, we derive the Markov properties of the dynamic system (1) as follows:

Theorem 1 *All the future values of uninfected cells, infected cells, and free virus particles depend statistically only on their present values. That is,*

$$\begin{aligned}
 &f(x_{ik+1}, y_{ik+1}, v_{ik+1} | (x_{i1}, y_{i1}, v_{i1}), \dots, (x_{ik}, y_{ik}, v_{ik})) \\
 &= f(x_{ik+1}, y_{ik+1}, v_{ik+1} | (x_{ik}, y_{ik}, v_{ik})), \\
 &f(x_{ik+1} | (x_{i1}, y_{i1}, v_{i1}), \dots, (x_{ik}, y_{ik}, v_{ik})) = f(x_{ik+1} | (x_{ik}, y_{ik}, v_{ik})), \\
 &f(y_{ik+1} | (x_{i1}, y_{i1}, v_{i1}), \dots, (x_{ik}, y_{ik}, v_{ik})) = f(y_{ik+1} | (x_{ik}, y_{ik}, v_{ik})), \\
 &f(v_{ik+1} | (x_{i1}, y_{i1}, v_{i1}), \dots, (x_{ik}, y_{ik}, v_{ik})) = f(v_{ik+1} | (x_{ik}, y_{ik}, v_{ik})).
 \end{aligned}$$

The proof follows directly from (4) and the definitions of $\epsilon_{x_i}(t_k)$, $\epsilon_{y_i}(t_k)$, and $\epsilon_{z_i}(t_k)$. From this theorem, we have the following results.

Corollary 2.1 *Conditional on (x_{ik}, y_{ik}, v_{ik}) , $(x_{ik-1}, y_{ik-1}, v_{ik-1})$ and $(x_{ik+1}, y_{ik+1}, v_{ik+1})$ are statistically independent.*

Corollary 2.2 *Conditional on (x_{ik}, y_{ik}, v_{ik}) , x_{ik-1} and x_{ik+1} are statistically independent.*

Corollary 2.3 *Conditional on (x_{ik}, y_{ik}, v_{ik}) , y_{ik-1} and y_{ik+1} are statistically independent.*

Corollary 2.4 *Conditional on (x_{ik}, y_{ik}, v_{ik}) , v_{ik-1} and v_{ik+1} are statistically independent.*

Since

$$\begin{aligned} & ((x_{ik+1}, y_{ik+1}, v_{ik+1})|(x_{ik}, y_{ik}, v_{ik}), (x_{ik-1}, y_{ik-1}, v_{ik-1})) \\ & = f((x_{ik+1}, y_{ik+1}, v_{ik+1})|(x_{ik}, y_{ik}, v_{ik})), \end{aligned}$$

conditional on (x_{ik}, y_{ik}, v_{ik}) , $(x_{ik-1}, y_{ik-1}, v_{ik-1})$ and $(x_{ik+1}, y_{ik+1}, v_{ik+1})$ are statistically independent (Bremaud 1999). Hence, Corollary 2.1 holds. The proofs of Corollaries 2.2–2.4 can be made in a similar way.

Now, we get the following theorems:

Corollary 2.5 *Conditional on (x_{ik}, y_{ik}, v_{ik}) , (x_{ij}, y_{ij}, v_{ij}) for $j = 0, 1, \dots, k - 1$ and $(x_{ik+1}, y_{ik+1}, v_{ik+1})$ are statistically independent.*

Corollary 2.6 *Conditional on (x_{ik}, y_{ik}, v_{ik}) , $\{x_{i1}, \dots, x_{ik-1}\}$, and x_{ik+1} are statistically independent.*

Corollary 2.7 *Conditional on (x_{ik}, y_{ik}, v_{ik}) , $\{y_{i1}, \dots, y_{ik-1}\}$, and y_{ik+1} are statistically independent.*

Corollary 2.8 *Conditional on (x_{ik}, y_{ik}, v_{ik}) , $\{v_{i1}, \dots, v_{ik-1}\}$, and v_{ik+1} are statistically independent.*

The Theorem and all these corollaries will be used to derive computing algorithms for solving a system of differential equations (4) embedded in functional mapping.

3 Functional mapping

3.1 Genetic design

Genetic mapping of QTLs can be based on linkage analysis for a pedigree (Lander and Bostein 1989) or linkage disequilibrium analysis for a natural population (Wang and Wu 2004). In this article, we assume that the population used to map human QTLs for viral load trajectories is composed of n patients randomly sampled from a natural population at HWE. A panel of SNP markers are genotyped for all patients, aimed at the identification of QTLs affecting virus dynamics. Suppose there is a functional QTL of alleles A and a for virus dynamics. Let q and $1 - q$ denote the allele frequencies of A and a . The QTL forms three possible genotypes, AA , Aa , and aa . We assume that this QTL is associated with a SNP marker of alleles M (in a frequency of p) and m

(in a frequency of $1 - p$). The detection of significant linkage disequilibrium between the marker and QTL implies that the QTL may be linked with and, therefore, can be genetically manipulated by the marker.

The four haplotypes for the marker and QTL are MA , Ma , mA and ma , with respective frequencies expressed as $p_{11} = pq + D$, $p_{10} = p(1 - q) - D$, $p_{01} = (1 - p)q - D$, and $p_{00} = (1 - p)(1 - q) + D$, where D is the linkage disequilibrium between the marker and QTL. Thus, the population genetic parameters (p , q , and D) can be estimated by solving a group of regular equations if we can estimate the four haplotype frequencies $\Phi = (p_{11}, p_{10}, p_{01}, p_{00})$. Joint marker-QTL diplotype frequencies can be expressed as a product of the corresponding haplotype frequencies under the HWE assumption, from which joint marker-QTL genotype frequencies are derived. Because the marker is observed, an unknown genotype of the QTL can be inferred from the conditional probability of the QTL genotype given a marker genotype.

Each sampled patient is measured for three different traits, uninfected cells, x , infected cells, y , and free virus particles, v , at a series of time points, $(t_{i1}, \dots, t_{iT_i})$.

3.2 Likelihood

For a given QTL genotype j ($j = 2$ for AA , 1 for Aa , or 0 for aa), the parameters describing virus dynamics are denoted by $\Theta_j = \{\lambda_j, d_j, \beta_j, a_j, k_j, u_j\}$. The comparisons of these parameters between the three different QTL genotypes can determine whether and how this QTL affects the pattern of virus dynamics.

The likelihood of longitudinal viral data $(\mathbf{x}_i, \mathbf{y}_i, \mathbf{v}_i) = \{x_i(t_k), y_i(t_k), v_i(t_k)\}_{k=0}^N$ and marker information \mathbf{M}_i for patient i is formulated by the mixture transitional Markov model, expressed as

$$L(\mathbf{x}, \mathbf{y}, \mathbf{v}; \mathbf{M}) = \prod_{i=1}^n \left[\sum_{j=0}^2 \omega_{ji} f_j(\mathbf{x}_i, \mathbf{y}_i, \mathbf{v}_i; \Theta_j, \Psi) \right], \tag{5}$$

where ω_{ji} is a mixture proportion, that is, the conditional probability of QTL genotype j given the marker genotype of subject i , which can be expressed as a function of haplotype frequencies (Table 1), and $f_j(\mathbf{x}_i, \mathbf{y}_i, \mathbf{v}_i; \Theta_j, \Psi)$ is a multivariate normal distribution with QTL genotype-specific mean vector specified by ODE parameters (Θ_j) and covariance matrix specified by parametric, non-parametric, or semiparametric models (Ψ) (Ma et al. 2002; Wu and Lin 2006).

Based on the Corollaries given above, the multivariate distribution can be specified by the following transition model

$$\begin{aligned} & f_j(\mathbf{x}_i, \mathbf{y}_i, \mathbf{v}_i; \Theta_j, \Psi) \\ &= f_j(x_{i1}, y_{i1}, v_{i1} | \Theta_j, \Psi) \prod_{k=0}^{N-1} f_j(x_{ik+1}, y_{ik+1}, v_{ik+1} | x_{ik}, y_{ik}, v_{ik}; \Theta_j, \Psi) \end{aligned} \tag{6}$$

Table 1 Joint genotype frequencies at the marker and QTL in terms of gametic haplotype frequencies, from which the conditional probabilities of QTL genotypes given marker genotypes can be calculated according to Bayes’ theorem

Genotype	Diplotype	AA	Aa	aa	Observations
		A A	A a + a A	a a	
MM	M M	p_{11}^2	$2p_{11}p_{10}$	p_{10}^2	N_1
Mm	M m	$2p_{11}p_{01}$	$2p_{11}p_{00} + 2p_{10}p_{01}$	$2p_{10}p_{00}$	N_2
mm	m m	p_{01}^2	$2p_{01}p_{00}$	p_{00}^2	N_3

where

$$\begin{aligned}
 & f_j(x_{ik+1}, y_{ik+1}, v_{ik+1} | x_{ik}, y_{ik}, v_{ik}; \Theta_j, \Psi) \\
 &= f_j(x_{ik+1} | x_{ik}, y_{ik}, v_{ik}; \Theta_j, \Psi) f_j(y_{ik+1} | x_{ik}, y_{ik}, v_{ik}; \Theta_j, \Psi) \\
 &\quad \times f_j(v_{ik+1} | x_{ik}, y_{ik}, v_{ik}; \Theta_j, \Psi), \\
 & f_j(x_{ik+1} | x_{ik}, y_{ik}, v_{ik}; \Theta_j, \sigma_x^2) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left[-\frac{1}{2\sigma_x^2} (x_{ik+1} - g_j(x_{ik+1}))^2\right], \\
 & f_j(y_{ik+1} | x_{ik}, y_{ik}, v_{ik}; \Theta_j, \sigma_y^2) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left[-\frac{1}{2\sigma_y^2} (y_{ik+1} - h_j(y_{ik+1}))^2\right], \\
 & f_j(v_{ik+1} | x_{ik}, y_{ik}, v_{ik}; \Theta_j, \sigma_v^2) = \frac{1}{\sqrt{2\pi\sigma_v^2}} \exp\left[-\frac{1}{2\sigma_v^2} (v_{ik+1} - l_j(v_{ik+1}))^2\right],
 \end{aligned}$$

with $\Psi = (\sigma_x^2, \sigma_y^2, \sigma_v^2)$, and

$$\begin{aligned}
 g_j(x_{ik+1}) &= x_{ik} + \lambda_j \Delta t_k - d_j x_{ik} \Delta t_k - \beta_j x_{ik} v_{ik} \Delta t_k \\
 h_j(y_{ik+1}) &= y_{ik} + \beta_j x_{ik} v_{ik} \Delta t_k - a_j y_{ik} \Delta t_k \\
 l_j(v_{ik+1}) &= v_{ik} + k_j y_{ik} \Delta t_k - u_j v_{ik} \Delta t_k.
 \end{aligned} \tag{7}$$

3.3 Estimation and algorithm

The EM algorithm (Dempster et al. 1977; Little and Rubin 2002) is implemented to get the maximum likelihood estimates (MLE) of all unknown parameters. The gradient of the log-likelihood function

$$\log L(\mathbf{x}, \mathbf{y}, \mathbf{v}; \mathbf{M}) = \sum_{i=1}^n \log \left[\sum_{j=0}^2 \omega_{j|i} f_j(\mathbf{x}_i, \mathbf{y}_i, \mathbf{v}_i; \Theta_j, \Psi) \right], \tag{8}$$

is given by

$$\begin{aligned} &\nabla_{\Theta_j} \log L(\mathbf{x}, \mathbf{y}, \mathbf{v}; \mathbf{M}) \\ &= \sum_{i=1}^n \sum_{j=0}^2 \frac{\omega_{j|i} f_j(\mathbf{x}_i, \mathbf{y}_i, \mathbf{v}_i; \Theta_j, \Psi)}{\sum_{j'=0}^2 \omega_{j'|i} f_{j'}(\mathbf{x}_i, \mathbf{y}_i, \mathbf{v}_i; \Theta_j, \Psi)} \nabla_{\Theta_j} \log f_j(\mathbf{x}_i, \mathbf{y}_i, \mathbf{v}_i; \Theta_j, \Psi), \end{aligned}$$

$$\begin{aligned} &\nabla_{\Psi} \log L(\mathbf{x}, \mathbf{y}, \mathbf{v}; \mathbf{M}) \\ &= \sum_{i=1}^n \sum_{j=0}^2 \frac{\omega_{j|i} f_j(\mathbf{x}_i, \mathbf{y}_i, \mathbf{v}_i; \Theta_j, \Psi)}{\sum_{j'=0}^2 \omega_{j'|i} f_{j'}(\mathbf{x}_i, \mathbf{y}_i, \mathbf{v}_i; \Theta_j, \Psi)} \nabla_{\Psi} \log f_j(\mathbf{x}_i, \mathbf{y}_i, \mathbf{v}_i; \Theta_j, \Psi), \end{aligned}$$

and

$$\begin{aligned} &\nabla_{\omega_{j|i}} \log L(\mathbf{x}, \mathbf{y}, \mathbf{v}; \mathbf{M}) \\ &= \sum_{i=1}^n \sum_{j=0}^2 \frac{\omega_{j|i} f_j(\mathbf{x}_i, \mathbf{y}_i, \mathbf{v}_i; \Theta_j, \Psi)}{\sum_{j'=0}^2 \omega_{j'|i} f_{j'}(\mathbf{x}_i, \mathbf{y}_i, \mathbf{v}_i; \Theta_j, \Psi)} \nabla_{\omega_{j|i}} \log(\omega_{j|i}). \end{aligned}$$

An iterative loop for the EM algorithm is formulated as follows. In the E step, the posterior probability with which a patient i carries a specific QTL genotype j based on the marker and phenotypic data is calculated by

$$\Omega_{j|i} = \frac{\omega_{j|i} f_j(\mathbf{x}_i, \mathbf{y}_i, \mathbf{v}_i; \Theta_j, \Psi)}{\sum_{j'=0}^2 \omega_{j'|i} f_{j'}(\mathbf{x}_i, \mathbf{y}_i, \mathbf{v}_i; \Theta_j, \Psi)}. \tag{9}$$

In the M step, the parameters are estimated by solving the following log-likelihood equations:

$$\nabla_{\Theta_j} \log L(\mathbf{x}, \mathbf{y}, \mathbf{v}; \mathbf{M}) = 0, \tag{10}$$

$$\nabla_{\Psi} \log L(\mathbf{x}, \mathbf{y}, \mathbf{v}; \mathbf{M}) = 0, \tag{11}$$

$$\nabla_{\omega_{j|i}} \log L(\mathbf{x}, \mathbf{y}, \mathbf{v}; \mathbf{M}) = 0. \tag{12}$$

Wang and Wu (2004) derived a closed algorithmic form to obtain the MLEs of haplotype frequencies p_{11} , p_{10} , p_{01} and p_{00} and, therefore, allele frequencies of the marker (p) and QTL (q) and their linkage disequilibrium (D). Genotype-specific mathematical parameters for viral dynamics and variances for the three types of viruses are calculated by implementing the Newton algorithm with the Armijo search (Bertsekas 2003).

4 Hypothesis testing

4.1 The significance of QTL

Whether there is a specific QTL responsible for viral dynamics described by a system of differential equations 1 can be tested by using the following hypotheses:

$$\begin{aligned} H_0 : \Theta_j &\equiv \Theta, \quad (j = 2, 1, 0) \\ H_1 : &\text{At least one of the equalities above does not hold,} \end{aligned} \quad (13)$$

The likelihoods under the null (L_0) and alternative hypotheses (L_1) are calculated, from which a log-likelihood ratio test statistic is computed by

$$\text{LR} = -2 \left[\log L_0(\tilde{\Theta}, \tilde{\Psi} | \mathbf{x}, \mathbf{y}, \mathbf{z}) - \log L_1(\hat{\Phi}, \hat{\Theta}_j, \hat{\Psi} | \mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{M}) \right],$$

where the tildes and hats present the MLE under the null and alternative hypotheses, respectively. Because of violation of the regularity assumption, the LR may not asymptotically follow a χ^2 -distribution with the degrees of freedom equal to the difference of parameter numbers between the two hypotheses (13). For this reason, the threshold for claiming the existence of a significant QTL is determined from empirical permutation tests (Churchill and Doerge 1994) because this approach does not rely on the distribution of LR values.

After a significant QTL is claimed, its significant association with the marker considered can be tested by the following hypotheses:

$$H_0 : D = 0 \text{ vs. } H_1 : D \neq 0, \quad (14)$$

whose log-likelihood ratio test statistic asymptotically follow the χ^2 -distribution with one degree of freedom.

4.2 Genetic mechanisms

The model allows the test of whether the QTL triggers a pleiotropic effect on three different types of cells. To do so, three null hypotheses for uninfected cells, infected cells, and free virus particles are formulated as follows:

$$H_0 : (\lambda_j, d_j, \beta_j) \equiv (\lambda, d, \beta), \quad (15)$$

$$H_0 : (\beta_j, a_j) \equiv (\beta, a), \quad (16)$$

$$H_0 : (k_j, u_j) \equiv (k, u), \quad (17)$$

for $j = 2, 1, 0$. If all the null hypotheses are rejected, then this means that the QTL pleiotropically affect these three different aspects of viral dynamics. The pleiotropic effect of the QTL on any pair of three types of cells can also be tested accordingly. An empirical approach for determining the critical threshold is based on simulation studies.

4.3 Physiological control of QTL

Several physiological important parameters define the dynamic system (1) (Bonhoeffer et al. 1999), including

1. The average life-times, $1/d$, $1/a$, and $1/u$, of uninfected cells, infected cells, and free virus, respectively,
2. The average number of virus particles or the burst size, k/a , yielded over the lifetime of a single infected cell,
3. Basic reproductive ratio, $R_0 = \beta\lambda k/(adu)$, i.e., the average number of newly infected cells that arise from any one infected cell when almost all cells are uninfected.

How a QTL affects these physiological aspects of viral dynamics separately or jointly can be tested.

5 Application to simulated data

Monte Carlo simulation was performed to examine the statistical properties of the model for genetic mapping of viral dynamics. Also, the use of the model to analyze simulated data will validate its practical usefulness and utilization. We randomly choose 100 subjects from an HWE population. Consider one of the markers genotyped for all subjects. This marker of two alleles M and m is used to infer a QTL of two alleles A and a for viral dynamics based on the non-random association between the marker and QTL. The allele frequencies are assumed as $p = 0.6$ for allele M 0.4 for allele m as well as $q = 0.6$ for allele A and 0.4 for allele a . A positive value of linkage disequilibrium ($D = 0.08$) between alleles M and A is assumed, suggesting that these two more common alleles are in coupling phase.

The three QTL genotypes, AA , Aa , and aa , are each hypothesized to have different response systems for uninfected cells, x , infected cells, y , and free virus particles, v , constructed by Eqs. (1). Six curve parameters $\{\lambda_j, d_j, \beta_j, a_j, k_j, u_j\}$ that define QTL genotype-specific systems were chosen from their spaces of biological relevance (Bonhoeffer et al. 1999). The phenotypic values of these three variables are expressed as the sum of the genotype-specific means and innovation errors assumed to follow a multivariate normal distribution. The phenotypic data were simulated for a practically reasonable number of equally spaced time points (say 22) under two different levels of heritability, low (0.1) and high (0.4). The genetic variance due to the QTL for virus response at a middle measurement point was used to define the heritability. The residual variances for each of the three virus traits were then calculated under different heritabilities. To assure the homoscedasticity of variances, the transforms-both-sides (TBS) model was used to simulate innovation errors. The TBS model can preserve biological means of parameters in original DE and also avoid negative phenotypic values (Wu et al. 2004b).

The DE-incorporated functional mapping model was used to analyze the simulated data, with the results suggesting that the QTL responsible for the dynamic system of viral infection can be detected using a molecular marker in association with the

QTL. As expected, population genetic parameters about QTL segregation in a population can well be estimated with a closed form of the EM algorithm derived in (Wang and Wu 2004). The curve parameters for virus responses of each QTL genotype can be estimated accurately and precisely with a modest sample size (100) even for a low heritability of viral loads (Tables 2, 3). The precision of all parameters can increase with increasing heritability level. By drawing the curves of viral trajectories with six parameters, the dynamic behavior of the system can be visualized. Figure 1 illustrates QTL genotype-specific curves of uninfected cells, infected cells, and free virus particles in a dynamic system from a random run of simulation. It is found that the shapes of the estimated curves are broadly consistent with the those of the true curves, suggesting that the system can be reasonably estimated with the new model.

Simulation studies showed that the new model displays reasonably high power, 0.75 for a modest heritability (0.1) and 0.99 for a high heritability (0.4), to detect a significant QTL responsible for a dynamic system of viral infection. Hypothesis tests described in Sects. 4.2 and 4.3 provide a general platform for addressing the genetic control machinery of viral dynamics. For a given set of simulation data, it appears that these tests can be reasonably made. For example, the power for detecting a pleiotropic QTL for three types of viral cell dynamics is adequately high (≥ 0.7) for a modest sample size and heritability level. On the other hand, under this circumstance, type I error rates for detecting a significant QTL despite its absence is reasonably low

Table 2 The MLEs of parameters that define virus-host dynamics for three different QTL genotypes, and the association between the marker and QTL in a natural population, assuming that the heritability of the simulated QTL is $H^2 = 0.1$

	AA		Aa		aa	
	Given	MLE	Given	MLE	Given	MLE
Virus-host parameters						
λ	11.00	11.9556 (0.0137)	10.00	010.99626 (0.0115)	12.20	012.6963 (0.0142)
d	0.4500	0.4622 (0.0017)	0.090	0.1042 (0.0014)	0.3800	0.4004 (0.0017)
β	0.1100	0.1196 (0.0001)	0.1200	0.1296 (0.0001)	0.1300	0.1435 (0.0031)
a	0.2000	0.1971 (0.0008)	0.2500	0.2477 (0.0006)	0.2500	0.2654 (0.0008)
k	0.3000	0.3089 (0.0010)	0.2000	0.2121 (0.0008)	0.2100	0.2187 (0.0012)
u	0.9800	1.0034 (0.0025)	0.6400	0.6621 (0.0023)	0.6800	0.7043 (0.0033)
	Given	MLE				
Genetic parameters and variances						
p	0.6	0.6010 (0.0312)				
q	0.6	0.5703 (0.0680)				
D	0.08	0.07701 (0.0108)				
σ_x^2	0.2538	0.2548 (0.0107)				
σ_y^2	0.2583	0.2978 (0.0105)				
σ_v^2	0.5976	0.5968 (0.0239)				

The numbers in the parentheses are the square roots of the mean square errors of the MLEs

Table 3 The MLEs of parameters that define virus-host dynamics for three different QTL genotypes, and the association between the marker and QTL in a natural population, assuming that the heritability of the simulated QTL is $H^2 = 0.4$

	<i>AA</i>		<i>Aa</i>		<i>aa</i>	
	Given	MLE	Given	MLE	Given	MLE
Rhythmic parameters						
λ	11.000	11.9507 (0.0069)	10.00	10.9599 (0.0057)	12.00	12.4235 (0.0100)
d	0.4500	0.4629 (0.0009)	0.090	0.1046 (0.0007)	0.380	0.3971 (0.0012)
β	0.1100	0.1198 (0.0000)	0.120	0.1278 (0.0000)	0.130	0.1386 (0.0000)
a	0.2000	0.1976 (0.0003)	0.250	0.2480 (0.0003)	0.250	0.2642 (0.0005)
k	0.3000	0.3061 (0.0004)	0.200	0.2106 (0.0003)	0.210	0.2192 (0.0006)
u	0.9800	1.0013 (0.0011)	0.640	0.6626 (0.0009)	0.680	0.7025 (0.0017)
	Given	MLE				
Genetic parameters						
p	0.6	0.5976 (0.0336)				
q	0.6	0.6030 (0.0407)				
D	0.08	0.0793 (0.0047)				
σ_x^2	0.0423	0.0456 (0.0020)				
σ_y^2	0.0430	0.0436 (0.0031)				
σ_v^2	0.0996	0.1000 (0.0045)				

The numbers in the parentheses are the square roots of the mean square errors of the MLEs

(≤ 0.1). These results suggest that our model will be practically useful in statistical analysis of the genetic control of viral dynamics.

6 Discussion

A combination of functional mapping (Ma et al. 2002; Wang and Wu 2004; Wu et al. 2004a,b,c; Wu and Lin 2006) and mathematical models (Bonhoeffer et al. 1997; Bonhoeffer et al. 1999; Nowak and May 2000) provides new insights into the genetic control of virus population dynamics. In this article, we have proposed a statistical model for mapping QTLs that affect the dynamic pattern of viral infection. One of the meritorious advantages of the new model, as compared to existing functional mapping models, lies in the organization of multiple correlated aspects of viral infection into a dynamic system through a group of ODE and the implementation of such a viral dynamic system into the framework of functional mapping. To our best knowledge, the work presented here is a first model of genetic mapping which treats multiple complex traits as a complex system.

The current model is not a simple extension of functional mapping for multiple traits (Zhao et al. 2005). The previous multi-trait models do not take into account the relationships of genotypic values of different traits, although they model across-trait correlations in residual errors. The new model views multiple traits as a whole in which different traits coordinate each other to determine the dynamic behavior of the

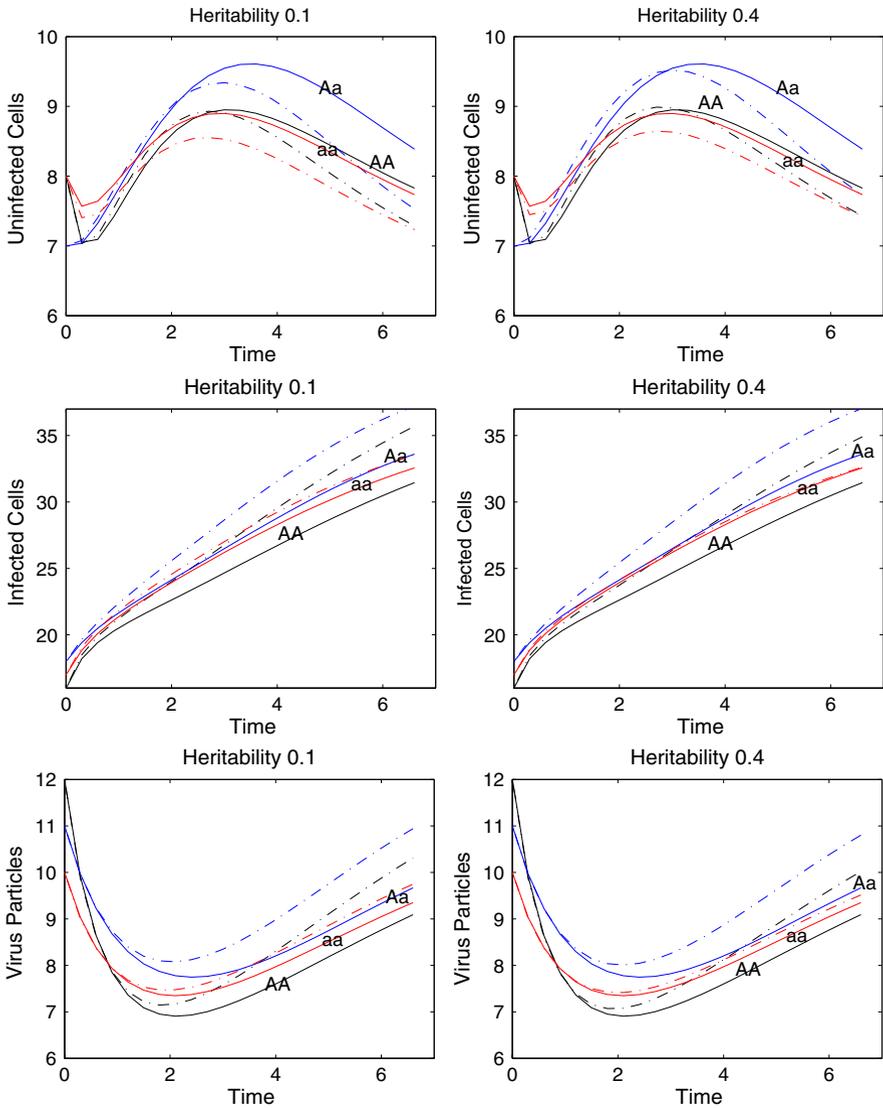


Fig. 1 Estimated and true curves for a system of viral infection including uninfected cells, x , infected cells, y , and free virus particles, v for three genotypes at a simulated QTL, AA , Aa , and aa , under different heritability levels, 0.1 (right panel) and 0.4 (left panel). The broad consistency between the estimated and true curves suggests that the model can provide a reasonably good estimate of the dynamic system

system. Thus, by altering one variable or trait, other variables will change, leading to the change of the entire system. The genetic mapping of genes for a dynamic system will provide a powerful means for understanding the genetic architecture of a biological process.

The mathematical strength of the new model is the deployment of a system of DE in a genetic mapping context. The solution of multiple DE, especially high-dimensional

ones, is computationally challenging. In this article, we apply a Newton algorithm with the EM setting to provide numerical estimates of the parameters that define the dynamic system. With the corollaries derived from several assumptions of independence, the algorithm is shown from simulation studies to be computationally efficient and provides precise estimates of the parameters, even when the sample size used is modest (Tables 2, 3).

As a demonstration of the new model, we assume that a dynamic system is controlled by a single QTL, although this assumption is too simple in real world. The genome-wide modeling of multiple QTLs throughout the genome can be incorporated into the current model setting, allowing the characterization of epistatic interactions among different QTLs (Wang et al. 2005; Wu et al. 2006). A multi-locus linkage disequilibrium model has been available to specify high-order non-random associations among multiple loci in a natural population (Li and Wu 2009). Although more parameters are involved in a multi-locus model, the closed forms derived for the EM algorithm (Li and Wu 2009) facilitates the estimation of many parameters at the same time. Also, a multi-locus model allows the test of the role of genetic interference in recombination events between adjacent intervals. Although linkage disequilibrium mapping has proven to be powerful for the high-resolution of QTLs, it often gives spurious results due to population structure and other evolutionary forces. A new genetic design that samples a set of random families, each composed of parents and their offspring, can overcome this limitation of linkage disequilibrium mapping (Li and Wu 2009; Wu et al. 2002). This design allows the simultaneous estimation of the linkage and linkage disequilibrium between different genes, thus making it possible to construct a genome-wide linkage disequilibrium map for gene discovery.

Our model focuses on the identification of genes for a dynamic system of viral changes in a host's body before the administration of an anti-viral drug. When the patients are treated with a drug, the equilibrium state of the system will be violated, from which a new equilibrium will be generated. Bonhoeffer et al. (1999) described a series of DE that specify the dynamic change of the system after drug treatment. The current model can be readily extended to model the genetic control of viral declines in a response to the anti-viral drug and half-lives of infected cells in the body. Perhaps, the most promising aspect of the new model is that it, when incorporated with the dynamics of virus' drug resistance, can provide scientific guidance for drug delivery and development by characterizing genes for drug resistance. The emergence of drug-resistant virus presents a main problem with antiviral therapy. A system of DE that captures the essential dynamics of resistance is given in the literature (Ribeiro and Bonhoeffer 2000; Wodarz and Nowak 2000). With the idea presented in this article, they can be readily incorporated into the functional mapping model, in a hope to achieve the maximum prevention of virus resistance to drugs by determining an optimal administration dose and time for individual patients based on their genetic makeups.

Acknowledgments We thank two anonymous referees for their constructive comments on the manuscript which led to a better presentation of it. This work is partially supported by NSF Grants 0619080 and 0620286 and NSF/NIH Joint grant DMS/NIGMS-0540745.

References

- Bertsekas DP (2003) *Nonlinear programming*, 2nd edn. Athena Scientific, Belmont
- Bonhoeffer S, Coffin JM, Nowak MA (1997) Human immunodeficiency virus drug therapy and virus load. *J Virol* 71:6971–6976
- Bonhoeffer S, May RM, Shaw GM, Nowak MA (1999) Virus dynamics and drug therapy. *Proc Natl Acad Sci USA* 94:6971–6976
- Bremaud P (1999) *Markov chains: Gibbs fields, Monte Carlo simulation and queues*. Springer, New York
- Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. *Genetics* 138:963–971
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via EM algorithm. *J R Stat Soc B* 39:1–38
- Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M (1999) Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* 373:123–126
- Lander ES, Bostein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185–199
- Little RJA, Rubin DB (2002) *Statistical analysis with missing data*, 2nd edn. Wiley, New York
- Li Q, Wu RL (2009) A multilocus model for constructing a linkage disequilibrium map in human populations. *Stat Appl Mol Genet Biol* 8:Article 18
- Ma CX, Casella G, Wu RL (2002) Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. *Genetics* 161:1751–1762
- Nowak MA, May RM (2000) *Virus dynamics*. Oxford University, New York
- Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD (1996) HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* 271:1582–1586
- Perelson AS, Essunger P, Cao Y, Vesanen M, Hurlley A, Saksela K, Markowitz M, Ho DD (1997) Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature* 387:188–191
- Ribeiro RM, Bonhoeffer S (2000) Production of resistant HIV mutants during antiretroviral therapy. *Proc Natl Acad Sci USA* 97:7681–7686
- Sedaghat AR, Dinoso JB, Shen L, Wilke CO, Siliciano RF (2008) Decay dynamics of HIV-1 depend on the inhibited stages of the viral life cycle. *Proc Natl Acad Sci* 105:4832–4837
- The International HapMap Consortium (2003) The International HapMap Project. *Nature* 426:789–794
- Wang ZH, Wu RL (2004) A statistical model for high-resolution mapping of quantitative trait loci determining HIV dynamics. *Stat Med* 23:3033–3051
- Wang ZH, Hou W, Wu RL (2005) A statistical model to analyze quantitative trait locus interactions for HIV dynamics from the virus and human genomes. *Stat Med* 25:495–511
- Wei X, Ghosh SK, Taylor ME, Johnson VA, Emini EA, Deutsch P, Lifson JD, Bonhoeffer S, Nowak MA, Hahn BH, Saag MS, Shaw GM (2004) Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* 23:3033–3051
- Wodarz D, Nowak MA (2000) HIV therapy: managing resistance. *Proc Natl Acad Sci USA* 97:8193–8195
- Wu RL, Lin M (2006) Functional mapping—how to study the genetic architecture of dynamic complex traits. *Nat Rev Genet* 7:229–237
- Wu RL, Ma C-X, Casella G (2002) Joint linkage and linkage disequilibrium mapping of quantitative trait loci in natural populations. *Genetics* 160:779–792
- Wu RL, Ma CX, Lin M, Casella G (2004a) A general framework for analyzing the genetic architecture of developmental characteristics. *Genetics* 166:1541–1551
- Wu RL, Ma CX, Lin M, Wang ZH, Casella G (2004b) Functional mapping of growth quantitative trait loci using a transform-both-sides logistic model. *Biometrics* 60:729–738
- Wu RL, Wang ZH, Zhao W, Cheverud JM (2004c) A mechanistic model for genetic machinery of ontogenetic growth. *Genetics* 168:2383–2394
- Wu S, Yang J, Wu RL (2006) Multilocus linkage disequilibrium mapping of quantitative trait loci that affect HIV dynamics: a simulation approach. *Stat Med* 25:3826–3849
- Zhao W, Hou W, Littell RC, Wu RL (2005) Structured antedependence models for functional mapping of multivariate longitudinal quantitative traits. *Stat Appl Mol Genet Biol* 4:Article 33