ELSEVIER

How to compute which genes control drug resistance dynamics

Yunqian Guo[1,3], Jiangtao Luo[2,3], Jianxin Wang[1], Yaqun Wang[2] and Rongling Wu[1,2]

[1] Center for Computational Biology, Beijing Forestry University, Beijing 100083, China
[2] Center for Statistical Genetics, Pennsylvania State University, Hershey, PA 17033, USA

Increasing evidence shows that genes have a pivotal role in affecting the dynamic pattern of viral loads in the body of a host. By reviewing the biochemical interactions between a virus and host cells as a dynamic system, we outline a computational approach for mapping the genetic control of virus dynamics. The approach integrates differential equations (DEs) to quantify the dynamic origin and behavior of a viral infection system. It enables geneticists to generate various testable hypotheses about the genetic control mechanisms for virus dynamics and infection. The experiment designed according to this approach will also enable researchers to gain insight into the role of genes in limiting virus abundance and the dynamics of viral drug resistance, facilitating the development of personalized medicines to eliminate viral infections.

Introduction

Antiretroviral drugs designed to treat infection by retroviruses, primarily HIV, act by inhibiting specific steps in the viral replication cycle [1–3]. Reverse-transcriptase inhibitors prevent the reverse transcription of viral genomic RNA into proviral DNA and thereby prevent the infection of new cells. Protease inhibitors influence the cleavage of viral polyproteins, resulting in the production of noninfectious virus particles. Both types of drug have proven effective in reducing the viral load of infected individuals, as demonstrated in Fig. 1. Before treatment is initiated, the viral load in the host is in a quasi-steady state (i.e. the viral load is constant over a short period of time) [1]. When a drug is administered, the viral load declines over several orders of magnitude after experiencing a transient shoulder phase. However, if monotherapy is used, resistant virus will rebound rapidly, in some cases within only a few weeks after the start of therapy. A phenomenon of drug resistance such as that shown in Fig. 1 presents a huge problem in the treatment of diseases, and has caused significant concern in the field of drug development [4,5].

Mathematical models have been widely used to study the decline of free virus in treated patients [6–10]. Meanwhile, there is increasing interest in modeling the dynamics of viral drug resistance using a system of differential equations (DEs) [1–3,11–14]. Mathematical models could be instrumental in shedding some light on the prediction of the emergence of drug-resistant virus and, ultimately, on the design of long-term therapy. Given the immense variation in the rate and pattern of rebound of resistant virus among different hosts [15–19], there is a pressing demand for the integration of genetic information into mathematical models for the precise prediction of the dynamics of decline in viral load during drug therapy and the rate of emergence of resistant virus. The increasing availability of single nucleotide polymorphism (SNP) data have made it possible 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 the trait [20,21].

In this opinion article, we describe a computational model for identifying specific host genes or quantitative trait loci (QTLs) that control the changes in viral load; we use these host QTLs to predict the dynamic pattern of viral drug resistance. This model integrates a group of DEs into functional mapping, a statistical framework originally derived to map dynamic QTLs for growth curves [22,23],

Corresponding author:. Wu, R. (rwu@hes.hmc.psu.edu)
[3] These authors contributed equally to this work.

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2011.02.004

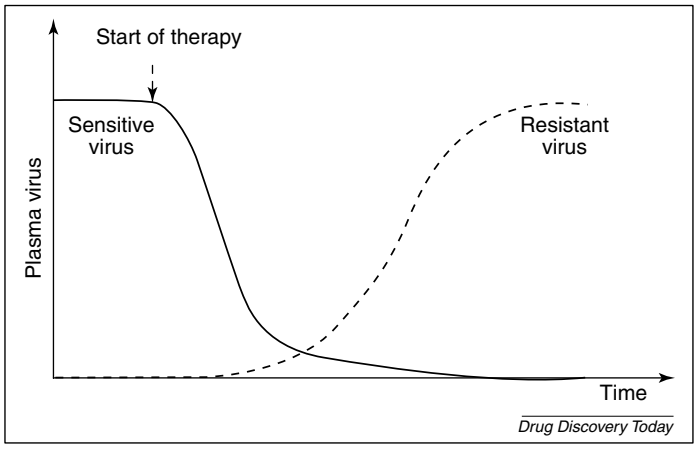

FIGURE 1

Plasma load profiles for sensitive (solid line) and resistance (dashed line) viruses in an infected patient from the start of monotherapy. Before treatment, the virus load in the patient's body is constant over a short period of time. After a drug is administered, the virus load declines drastically over several orders of magnitude after experiencing a transient shoulder phase. However, if monotherapy is used, resistant virus will rebound rapidly, leading to the failure of the original medicinal treatment. Redrawn from [1].

to study the genetic architecture of viral dynamics [24]. The model formulates a procedure for estimating parameters that define DEs based on the Markov chain properties of dynamic data. Here, we argue that DE-implemented models can be used to map QTLs and genetic interactions for the dynamics of viral drug resistance. A series of hypothesis tests about the genetic basis of viral dynamics and resistance can be formulated. Once implemented into genome-wide association studies aimed at revealing the inter-individual variability in response to viral infection [17–19], the model for mapping drug resistance dynamics will help to improve the understanding of the complex interactions between virus and its human host and, ultimately, draw the appropriate lessons for vaccine development to combat this particular virus.

Mapping a viral system via DEs

Luo *et al.* [24] pioneered the use of DEs to map QTLs that control the multifaceted behavior of a dynamic system by taking into account the emergent properties of that system. The innovative point of this mapping model lies not only in its precise quantification of biochemical interactions among different, but interconnected, components that comprise the system by exploiting mathematical properties of DE, but also in its identification of specific loci that govern these biochemical interactions and their impact on viral infection.

HIV infection: a dynamic system

HIV infection is the consequence of interactions between HIV and its human host, which can be viewed as a dynamic system. A basic model for describing virus dynamics includes three variables, uninfected cells, x; infected cells, y; and free virus particles, v. 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. With these assumptions, Bonhoeffer *et al.* [2] constructed a system of DE (Eqn (1)):

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

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. These equations can quantify how viral load in the host declines after an antiviral therapy is started and can further predict inter-individual variation in the decline trend if genetic information is incorporated.

QTLs for viral dynamics: can we map them?

Luo *et al.* [24] successfully advanced Bonhoeffer *et al.*'s dynamic modeling [2] of viral infection to the point that a comprehensive analysis of the role of genetic variation in viral control is now within reach. Briefly, Luo *et al.*'s [24] model is based on a natural human population at Hardy–Weinberg equilibrium (HWE), from which n patients are randomly sampled. All sampled patients are treated with antiviral drugs and then measured for three dynamic traits describing the system (Eqn (3)): uninfected cells, x, infected cells, y, and free virus particles, v, at a series of time points $(t_{i1}, \ldots, t_{iTi})$. Now suppose that there exist specific QTLs that control virus dynamics and resistance. These QTLs can be detected through molecular markers that are associated with them. Consider a marker (with two alleles, M and m) that is co-segregating with a QTL (with two alleles, A and a). Let p and $1-p$ denote the allele frequencies of M and m, and q and $1-q$ denote the allele frequencies of A and a. The linkage disequilibrium between the marker and QTL is denoted as D, which forms the basis of QTL mapping.

The four haplotypes for the marker and QTL, MA, Ma, mA and ma, have frequencies expressed as $p_{11} = pq + D$, $p_{10} = p(1-q) - D$, $p_{01} = (1-p)q - D$ and $p_{00} = (1-p)(1-q) + D$. The frequencies of marker–QTL diplotypes (i.e. the combination of maternally and paternally derived haplotypes) can be expressed as a product of the corresponding haplotype frequencies under the HWE assumption, from which joint marker–QTL genotype frequencies are derived. By observing marker genotypes, MM, Mm and mm, an unknown QTL genotype can be inferred from the conditional probability of the QTL genotype given a marker genotype [25].

At a given QTL, there are three genotypes, AA (coded as 2), Aa (coded as 1) and aa (coded as 0). For a given QTL genotype j, the parameters describing virus dynamics are denoted by $\Theta j = \{\lambda_j, d_j, \beta_j, a_j, k_j, u_j\}$ $(j = 2, 1, 0)$. By comparing these parameters among the three different QTL genotypes, one can determine whether and how this QTL affects the dynamic pattern of viral drug resistance.

Let $(\mathbf{x}_i, \mathbf{y}_i, \mathbf{v}_i) = \{x_i(t_\tau), y_i(t_\tau), v_i(t_\tau)\}_{\tau=1}^{Ti}$ denote the longitudinal viral data for patient i. The likelihood of phenotype and marker information $\mathbf{M}$ is formulated by the mixture transitional Markov model, expressed as Eqn (2):

$$L(x, y, v; M) = \prod_{i=1}^{n}\left[\sum_{j=0}^{2}\omega_{j|i}\, \boldsymbol{f}_j(\mathbf{x}_i, \mathbf{y}_i, \mathbf{v}_i; \boldsymbol{\Theta}_j, \boldsymbol{\Psi})\right] \qquad (2)$$

where $\omega_{j|i}$ is a mixture proportion that reflects the QTL genotype j of patient i, which can be inferred from its marker genotype [25],

and f_j ($\mathbf{x}_i$, $\mathbf{y}_i$, $\mathbf{v}_i$; Θ_j, Ψ) is a multivariate normal distribution with QTL genotype-specific mean vector specified by Θ_j and covariance matrix specified by Ψ. Given that a natural population is considered, $\omega_{j|i}$ is derived in terms of the frequencies of marker–QTL haplotypes [25].

Luo *et al.* [24] used the DE in Eqn (1) to model time-varying changes of viral loads for individual QTL genotypes by the DE parameters $\Theta_j = \{\lambda_j, d_j, \beta_j, a_j, k_j, u_j\}$. Various approaches have been available to model the longitudinal covariance structure by parameter Ψ, including parametric [22,26,27], nonparametric [28] and semiparametric [29] approaches. The standard expectation–maximization (EM) algorithm, as proposed by Dempster *et al.* [30] and Lander and Botstein [31], and Markov transitional properties of a dynamic system [32,33] were implemented to get the maximum likelihood estimates (MLEs) of all unknown parameters that describe the QTL segregation and effect on the system (Eqn (3)).

Testing QTL effects and genetic mechanisms for biochemical interactions

Luo *et al.* [24] formulated a hypothesis test for the existence of any specific QTL responsible for viral dynamics, described by a system of DEs (Eqn (3)). This hypothesis is based on the ratio of the likelihoods under the null hypothesis, H_0: $\Theta_j \equiv \Theta$ (there is no QTL), and the alternative hypothesis, H_1: at least one equality in the H_0 does not hold (there is a QTL). Empirical permutation tests are used to determine the critical threshold for claiming the existence of a significant QTL [34] because it is difficult to derive a theoretical formula for threshold determination.

Luo *et al.* [24] also showed how to test the co-segregation between a QTL and marker. After a significant QTL is claimed, one will need to test whether it can be detected by the marker considered using the hypotheses: H_0: $D = 0$ versus H_1: $D = 0$, whose log-likelihood ratio test statistic asymptotically follow the X^2 distribution with one degree of freedom. A significant D value implies that the putative QTL can be detected by this marker.

One of the most important merits of the DE-implemented mapping model is that it allows the test of genetic mechanisms for viral dynamics. Whether the QTL triggers a pleiotropic effect on three different types of cell can be tested. To do so, three null hypotheses for uninfected cells, infected cells and free virus particles are formulated as follows:

$$\begin{aligned}
&H_0 : (\lambda_j, d_j, \beta_j) \equiv (\lambda, d, \beta), \quad \text{for } j = 2, 1, 0,\\
&H_0 : (\beta_j, a_j) \equiv (\beta, a), \quad \text{for } j = 2, 1, 0,\\
&H_0 : (k_j, u_j) \equiv (k, u), \quad \text{for } j = 2, 1, 0.
\end{aligned}$$

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

New modeling method for mapping resistance genes

Luo *et al.*'s [24] dynamic model can be extended to study the genetic control of drug resistance dynamics. This needs a dynamic description of the biochemical and molecular mechanisms for drug resistance. It is well known that the emergence of drug-resistant virus in a therapy can be described by incorporating the difference between wild-type and mutant viruses into the equations, which is expressed as Eqn (3):

$$\begin{aligned}
\frac{dx}{dt} &= \lambda - dx - \beta_1 x v_1 - \beta_2 x v_2\\
\frac{dy_1}{dt} &= \beta_1(1-\mu) x v_1 + \beta_2 u x v_2 - a y_1\\
\frac{dy_2}{dt} &= \beta_1 u x v_1 + \beta_2(1-\mu) x v_2 - a y_2\\
\frac{dv_1}{dt} &= k_1 y_1 - u v_1\\
\frac{dv_2}{dt} &= k_2 y_2 - u v_2
\end{aligned} \tag{3}$$

Here, there are five variables: uninfected cells, x, cells infected by wild-type virus, y_1, cells infected by mutant virus, y_2, free wild-type virus, v_1, and free mutant virus, v_2. These five types of cell interact

BOX 1

Numerical simulation for drug resistance mapping.

A Monte Carlo simulation was performed to examine the statistical properties of the model for genetic mapping of the dynamics of viral drug resistance. By analyzing simulated data with the new model, we validate the practical efficacy and use of the model. A set of 100 random subjects was drawn from an HWE human population. All the subjects were genotyped for many markers, with the aim of detecting significant QTLs for viral resistance dynamics. Consider one-marker of two alleles, *M* and *m,* 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*. We assume a positive value of linkage disequilibrium ($D = 0.08$) between the marker and QTL so that their common alleles are in coupling phase.

Given that the QTL is assumed to influence viral resistance dynamics, as constructed by Eqn (3), the three QTL genotypes, *AA*, *Aa* and *aa*, will have different response systems for uninfected cells, x, infected cells, y_1 and y_2, and free virus particles, v_1 and v_2. Nine curve parameters $\{\lambda_j, d_j, \beta_{1j}, \beta_{2j}, m_j, a_j, k_{1j}, k_{2j}, \mu_j\}$ specify QTL genotype-specific systems, and these were chosen from their spaces of biological relevance [2]. For this simulation study, each subject was assumed to be measured for the five variables at 22 equally spaced time points. Using the genetic variance owing to the QTL for virus response at a middle measurement point, we calculated the residual variances for each of the five virus traits under different heritability levels, low (0.1) and high (0.4).

The new model was used to analyze the simulated data. Generally speaking, the QTL responsible for the dynamic system of viral resistance can be detected using our model. Owing to a closed form, derived haplotype frequencies and, therefore, the allele frequency of QTL and its association with a marker can be estimated. It seems that a modest sample size (100) is enough to estimate the curve parameters for virus resistance dynamics for each QTL genotype, even for a low heritability of viral loads (Table 1). As expected, the precision of all parameters can increase with increasing heritability level (Table 1). The dynamic behavior of the resistance system can be visualized by drawing the curves of QTL genotype-specific viral trajectories. Figure 2 illustrates the QTL genotype-specific curves of uninfected cells, x, cells infected by wild-type virus, y_1, cells infected by mutant virus, y_2, free wild-type virus, v_1 and free mutant virus, v_2, in a dynamic system from a random run of simulation. In general, the system can be reasonably estimated by the new model because the shapes of the estimated curves are broadly consistent with those of the true curves (Fig. 2).

with each other to determine the dynamic changes of drug-resistant virus in the body of a host.

The system (Eqn (3)) is defined by nine parameters {λ, d, β_1, β_2, m, a, k_1, k_2 and u} and the initial conditions for x, y_1, y_2, v_1 and v_2. Thus, the dynamic pattern of a resistance system can be determined and predicted by the change in these parameters and the initial conditions of x, y_1, y_2, v_1 and v_2.

By implementing a system of DEs (Eqn (3)) into likelihood (Eqn (2)), a new model can be derived, which is powerful enough to test whether a QTL triggers a pleiotropic effect on five different components of the dynamic system of viral drug resistance. In Box 1, we provide estimates of QTL genotype-specific DE parameters for a resistance dynamic system using the new model, validating the its statistical utility and usefulness. Below are several hypotheses that are particularly useful for testing the genetic mechanisms for drug resistance in terms of its biochemical underpinnings. Genetic pleiotropy for uninfected cells, cells infected by wild-type virus, cells infected by mutant virus, free wild-type virus and free mutant virus, respectively, can be tested by the following hypotheses:

$H_0 : (\lambda_j, d_j, \beta_{1j}, \beta_{2j}) \equiv (\lambda, d, \beta_1, \beta_2)$ uninfected cells,
$H_0 : (\mu_j, \beta_{1j}, \beta_{2j}, a_j) \equiv (\mu, \beta_1, \beta_2, a)$
cells infected by wild-type and mutant virus,
$H_0 : (k_{1j}, \mu_j) \equiv (k_1, \mu)$ free wild-type virus,
$H_0 : (k_{2j}, \mu_j) \equiv (k_2, \mu)$ free mutant virus.

where j (j = 2, 1, 0) denotes three different QTL genotypes. If all the null hypotheses are rejected, then this means that the QTL pleiotropically affect these five different aspects of viral dynamics. The pleiotropic effect of the QTL on any pair of five types of cell can also be tested accordingly. The critical threshold for these tests can be determined through simulation studies.

Several physiological important parameters define the dynamic system (Eqn (3)), including: (*i*) the average life-times, 1/d, 1/a, and 1/μ, of uninfected cells, infected cells, and free virus, respectively; (*ii*) the average number of virus particles or the burst size, k/a, yielded over the lifetime of a single infected cell; and (*iii*) basic reproductive ratios, that is, the average number of newly infected cells that arise from any one infected cell when most cells are uninfected, $R_1 = \beta_1\lambda k_1/(adu)$ for wild-type, and $R_2 = \beta_2\lambda k_2/(adu)$ for mutant virus.

All these enable one to test whether and how a QTL affects each of these physiological aspects of viral dynamics. Also, one can test whether a QTL triggers a pleiotropic effect on these aspects.

Although a one-marker/one-QTL was outlined for the sake of description, the idea can be readily extended to include multiple QTLs and their epistatic interactions [35]. Epistasis occurs when the phenotypic effect of a gene changes depending on the presence or absence of other genes in the genome. As one of the most relevant problems in understanding the nature and consequences of complex interactions within biological complex systems, epistasis has been extensively studied in the genetic analysis of viral infection [36,37]. A survey shows the pervasive existence of epistasis in RNA virus genomes [38]. Also, genetic epistasis between genes derived from viral and host genomes might have an important role in affecting viral dynamics [39]. Such genome–genome epistasis might have high-order interactions involving viral genes and host genes derived from a recipient and transmitter [40]. Dynamic modeling of these within-genome genetic interactions and genome–genome genetic interactions by DEs should be of great interest in unraveling the genetic secrets hidden in viral and viral resistance dynamics.

TABLE 1

The MLEs of parameters that define the dynamics of viral drug resistance for three different QTL genotypes and the association between the marker and QTL in a natural population, based on different assumptions of heritability of the simulated QTL[a]

Virus-host parameters							Genetic parameters		
	AA		*Aa*		*aa*			Given	MLE
	Given	MLE	Given	MLE	Given	MLE			
Heritability of the simulated QTL is H^2 = 0.10									
λ	33.000	35.1656(0.0168)	32.00	30.5237(0.0111)	36.00	33.9577(0.0149)	p	0.6	0.6021(0.0339)
d	0.1000	0.1186(0.0048)	0.300	0.3033(0.00031)	0.600	0.6634(0.0046)	q	0.6	0.5089(0.0447)
β_1	0.7000	0.6305(0.0003)	0.500	0.5094(0.0003)	0.900	0.8085(0.005)	D	0.08	0.0358(0.0352)
β_2	0.5000	0.5515(0.0005)	0.600	0.5016(0.0003)	0.340	0.3171(0.0005)	σ^2_x	0.2299	0.2347(0.0106)
μ	0.3000	0.3143(0.0039)	0.190	0.1746(0.0042)	0.260	0.27454(0.0066)	σ^2_{y1}	1.3566	1.3672(0.0591)
a	0.3500	0.3322(0.0019)	0.200	0.2587(0.0014)	0.300	0.3255(0.0019)	σ^2_{y2}	1.0944	1.1041(0.0450)
k_1	0.3700	0.4167(0.0013)	0.280	0.2619(0.0009)	0.200	0.22012(0.0015)	σ^2_{v1}	0.3382	0.3412(0.0138)
k_2	0.1980	0.1837(0.0017)	0.160	0.1782(0.0012)	0.138	0.1243(0.0018)	σ^2_{v2}	0.4332	0.4326(0.0180)
u	0.0550	0.0576(0.0015)	0.060	0.0658(0.0011)	0.020	0.0210(0.0019)			
Heritability of the simulated QTL is H^2 = 0.4									
λ	33.000	34.4361(0.0112)	32.00	32.9949(0.0096)	36.00	36.9951(0.0136)	p	0.6	0.6021(0.0339)
d	0.1000	0.1377(0.0034)	0.300	0.3207(0.00027)	0.600	0.6205(0.0039)	q	0.6	0.5529(0.0478)
β_1	0.7000	0.6843(0.0009)	0.500	0.5199(0.0003)	0.900	0.9297(0.0004)	D	0.08	0.0683(0.0311)
β_2	0.5000	0.5120(0.0004)	0.600	0.6199(0.0003)	0.340	0.3449(0.0005)	σ^2_x	0.1089	0.1141(0.0053)
μ	0.3000	0.3196(0.0020)	0.190	0.2023(0.0044)	0.260	0.2706(0.0071)	σ^2_{y1}	0.6426	0.6425(0.0281)
a	0.3500	0.3547(0.0011)	0.200	0.2118(0.0010)	0.300	0.3130(0.0016)	σ^2_{y2}	0.5184	0.5191(0.0212)
k_1	0.3700	0.3783(0.0008)	0.280	0.2901(0.0007)	0.200	0.2205(0.0011)	σ^2_{v1}	0.1602	0.1607(0.0065)
k_2	0.1980	0.2097(0.0010)	0.160	0.1798(0.0009)	0.138	0.1414(0.0015)	σ^2_{v2}	0.2052	0.2047(0.0085)
u	0.0550	0.0580(0.0010)	0.060	0.0700(0.0008)	0.02	0.0210(0.0015)			

[a] Numbers in parentheses are the standard deviations of the MLEs.

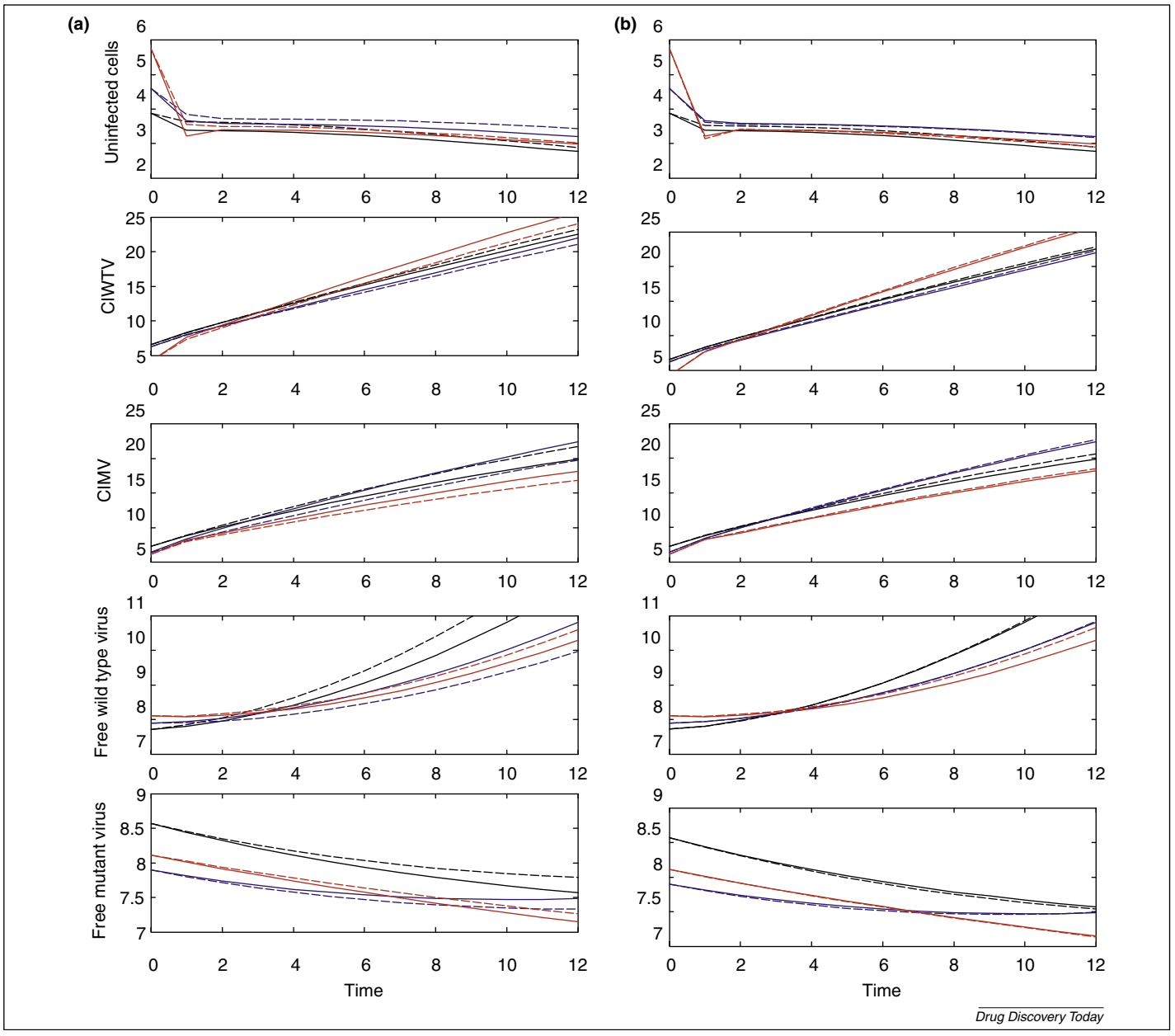

FIGURE 2

Estimated and true curves for a system of viral infection including uninfected cells, x, cells infected by wild-type virus (CIWTV), y_1, cells infected by mutant virus (CIMV), y_2, free wild-type virus, v_1, and free mutant virus, v_2, for three genotypes at a simulated QTL, *AA* (black), *Aa* (blue) and *aa* (red), under different heritability levels, 0.1 **(b)** and 0.4 **(a)**. The broad consistency between the estimated (solid) and true curves (broken) suggests that the model provides a good estimate of the dynamic system.

More recently, Li and Wu [41] proposed a family design including a set of random families, each composed of parents and their offspring, to estimate the linkage and linkage disequilibrium jointly between different loci. This design could overcome the problem of spurious associations owing to population structure and other evolutionary forces in traditional linkage disequilibrium mapping. Meanwhile, the family design enables the test and estimation of the effects on viral dynamics of genomic imprinting, a universal phenomenon by which certain genes are expressed or repressed depending on which parent they have been inherited from [42].

Concluding remarks

The resistance of a virus to an antiviral drug is one of the most important reasons for the failure of drug treatment to achieve complete viral suppression [4]. The identification of genes that control the dynamic pattern of viral drug resistance, such as the rate and pattern of rebound of resistant virus after drug therapy, will provide useful information for understanding the emergence of drug resistance and better predict treatment outcomes. Founded on Luo *et al.*'s model [24], we need to integrate a system of DEs for the dynamic change of viral drug resistance [2] into functional mapping developed by Wu *et al.* [22,23], which can shed light on the genetic mechanism of viral dynamics and resistance based on mathematical models.

The emergence of drug-resistant virus might be the result of the pre-existence of drug-resistant strains before the initiation of therapy or the generation of resistant virus during the course of treatment. It is important to identify which process is more likely

to be true (the drug-resistant virus either exists before the onset of therapy or is produced by residual virus replication during the course of antiviral treatment) because each requires different drug regimens to maximize the clinical benefits [12]. Rong *et al.* [14] developed a mathematical model to investigate analytically the mechanisms underlying the emergence of drug-resistant variants during antiviral treatment. This mathematical model should be incorporated into functional mapping, making it possible to test whether there is a specific QTL that determines each of these two processes and how they can be predicted using genetic information about the QTL detected.

The new computational model understands the genetic mechanisms underlying drug resistance from mathematical modeling. The model can be readily implemented in a practical experimental setting by sampling a group of patients randomly from a natural population. These patients are genotyped at an array of molecular markers and phenotyped for five variables, uninfected cells, x; cells infected by wild-type virus, y_1; cells infected by mutant virus, y_2; free wild-type virus, v_1; and free mutant virus, v_2 (Eqn (3)), at a series of time points. Not only can the new model be used to analyze the above experimental data to validate hypotheses about the genetic control of viral resistance dynamics, but designing and testing this model will also lead to testable experimental predictions. New successes in understanding pathogenesis will emerge as the model is used as a routine tool to decipher the genetic architecture of viral dynamics and resistance dynamics. The fundamental idea of the model will stimulate increased collaboration between modelers and geneticists in pathogenic diagnostics and therapeutics.

Acknowledgements

This work is partially supported by Joint Grant DMS/NIGMS-0540745, Changjiang Scholars Award, 'One-thousand Person Plan' Award, National Natural Science Foundation of China (Grant No. 30900854) and Beijing Forestry University Young Scientist Fund (Grant No. Blx2w8003).

References

1 Bonhoeffer, S. (1998) Models of viral kinetics and drug resistance in HIV-1 infection. *AIDS Patient Care STDS* 12, 769–774
2 Bonhoeffer, S. *et al.* (1999) Virus dynamics and drug therapy. *Proc. Natl. Acad. Sci. U. S. A.* 94, 6971–6976
3 Ribeiro, R.M. and Bonhoeffer, S. (2000) Production of resistant HIV mutants during antiretroviral therapy. *Proc. Natl. Acad. Sci. U. S. A.* 97, 7681–7686
4 Deeks, S.G. (2003) Treatment of antiretroviral-drug-resistant HIV-1 infection. *Lancet* 362, 2002–2011
5 Peet, N.P. (2010) Drug resistance: a growing problem. *Drug Discov. Today* 15, 583–586
6 Ho, D.D. *et al.* (1999) Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* 373, 123–126
7 Nowak, M.A. and May, R.M. (2000) *Virus Dynamics*. Oxford University Press
8 Perelson, A.S. (2002) Modeling viral and immune system dynamics. *Nat. Rev. Immunol.* 2, 28–36
9 Perelson, A.S. *et al.* (1997) Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature* 387, 188–191
10 Perelson, A.S. *et al.* (1996) HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* 271, 1582–1586
11 Bonhoeffer, S. *et al.* (1997) Human immunodeficiency virus drug therapy and virus load. *J. Virol.* 71, 6971–6976
12 Bonhoeffer, S. and Nowak, M.A. (1997) Pre-existence and emergence of drug resistance in HIV-1 infection. *Proc. R. Soc. B* 264, 631–637
13 Rong, L. *et al.* (2007) Modeling within-host HIV-1 dynamics and the evolution of drug resistance: trade-offs between viral enzyme function and drug susceptibility. *J. Theor. Biol.* 247, 804–818
14 Rong, L. *et al.* (2007) Emergence of HIV-1 drug resistance during antiretroviral treatment. *Bull. Math. Biol.* 69, 2027–2060
15 Shankarkumar, U. *et al.* (2009) HIV-1 evolution, drug resistance, and host genetics: the Indian scenario. *Virus Adaption Treat.* 1, 1–4
16 Fellay, J. *et al.* (2010) Host genetics and HIV-1: the final phase? *PLoS Pathog.* 6, e1001033
17 Fellay, J. *et al.* (2007) A whole-genome association study of major determinants for host control of HIV-1. *Science* 317, 944–947
18 Fellay, J. *et al.* (2009) Common genetic variation and the control of HIV-1 in humans. *PLoS Genet.* 5, e1000791
19 Limou, S. *et al.* (2009) Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS Genomewide Association Study 02). *J. Infect. Dis.* 199, 419–426
20 The International HapMap Consortium, (2003) The International HapMap Project. *Nature* 426, 789–794
21 Wu, R.L. and Lin, M. (2008) *Statistical and Computational Pharmacogenomics*. Chapman & Hall/CRC
22 Ma, C.-X. *et al.* (2002) Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. *Genetics* 161, 1751–1762
23 Wu, R.L. and Lin, M. (2006) Functional mapping – how to map and study the genetic architecture of dynamic complex traits. *Nat. Rev. Genet.* 7, 229–237
24 Luo, J.T. *et al.* (2010) A differential equation model for functional mapping of a virus cell dynamic system. *J. Math. Biol.* 65, 1–15
25 Wang, Z.H. and Wu, R.L. (2004) A statistical model for high-resolution mapping of quantitative trait loci determining HIV dynamics. *Stat. Med.* 23, 3033–3051
26 Zhao, W. *et al.* (2005) A nonstationary model for functional mapping of complex traits. *Bioinformatics* 21, 2469–2477
27 Zhao, W. *et al.* (2005) Structured antedependence models for functional mapping of multivariate longitudinal quantitative traits. *Stat. Appl. Mol. Genet. Biol.* 113610.2202/1544-6115
28 Yap, J. *et al.* (2009) Nonparametric modeling of covariance structure in functional mapping of quantitative trait loci. *Biometrics* 65, 1068–1077
29 Das, K. *et al.* A dynamic model for genome-wide association studies. *Hum. Genet.* doi:10.1007/s00439-011-0960-6
30 Dempster, A.P. *et al.* (1977) Maximum likelihood from incomplete data via EM algorithm. *J. R. Stat. Soc. B* 39, 1–38
31 Lander, E.S. and Botstein, D. (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121, 185–199
32 Bertsekas, D.P. (2003) *Nonlinear Programming* (2nd edn), Athena Scientific
33 Bremaud, P. (1999) *Markov Chains: Gibbs Fields, Monte Carlo Simulation and Queues*. Springer
34 Churchill, G.A. and Doerge, R.W. (1994) Empirical threshold values for quantitative trait mapping. *Genetics* 138, 963–971
35 Wu, S. *et al.* (2006) Multilocus linkage disequilibrium mapping of quantitative trait loci that affect HIV dynamics: a simulation approach. *Stat. Med.* 25, 3826–3849
36 Bonhoeffer, S. *et al.* (2004) Evidence for positive epistasis in HIV-1. *Science* 306, 1547–1550
37 da Silva, J. *et al.* (2010) Fitness epistasis and constraints on adaptation in a human immunodeficiency virus type 1 protein region. *Genetics* 185, 293–303
38 Elena, S.F. (2010) Simple genomes, complex interactions: epistasis in RNA virus. *Chaos* 20, 026106
39 Wang, Z.H. *et al.* (2005) A statistical model to analyse quantitative trait locus interactions for HIV dynamics from the virus and human genomes. *Stat. Med.* 25, 495–511
40 Li, Y. *et al.* (2009) A statistical model for genetic mapping of viral infection by integrating epidemiological behavior. *Stat. Appl. Genet. Mol. Biol.* 147510.2202/1544-6115
41 Li, Q. and Wu, R.L. (2009) A multilocus model for constructing a linkage disequilibrium map in human populations. *Stat. Appl. Genet. Mol. Biol.* 141910.2202/1544-6115
42 Wilkins, J.F. and Haig, D. (2003) What good is genomic imprinting: the function of parent-specific gene expression. *Nat. Rev. Genet.* 4, 359–368