

A Haplotype Block Model for Fine Mapping of Quantitative Trait Loci Regulating HIV-1 Pathogenesis

YUN ZHU, WEI HOU and RONGLING WU*

Department of Statistics, University of Florida, Gainesville, FL 32611, USA

(Received 12 December 2003; In final form 10 September 2004)

The dynamic change of human immunodeficiency virus type-1 (HIV-1) particles that cause AIDS displays considerable variation from patients to patients. It is likely that such variation in HIV-1 pathogenesis is correlated with the genetic architecture of hosts. Traditional genetic analysis of HIV-1 infection is based on various biochemical approaches, but it has been little successful because HIV-1 dynamics, as a complex trait, is under polygenic control and sensitive to environmental changes. Here, we present a novel model for integrating mathematical functions for HIV-1 dynamics that have been well constructed into a multivariate mixture model for genetic mapping. This integrative mapping model on the foundation of linkage disequilibrium (LD)-based haplotype block analysis provides unique power to precisely detect human quantitative trait loci (QTL) determining HIV-1 dynamics and facilitates positional cloning of target QTL. The model allows for a number of hypothesis tests for the effects of the dynamic QTL on the virion clearance rate, the infected cell life-span and the average viral generation time *in vivo*, all of which provide theoretical principles to guide the development of efficient gene therapy strategies.

Keywords: HIV-1; CD4 lymphocyte; Quantitative trait loci; AIDS

INTRODUCTION

During HIV-1 pathogenesis, an increased viral load is known to be closely linked with CD4 lymphocyte depletion and disease progression (Ho *et al.*, 1989; Patterson *et al.*, 1993), but little is clear about the genetic control of the kinetics of virus *in vivo*. The identification of specific genes underlying viral dynamics from the viral and host genome as well as genetic interactions between the two different genomes (Ameisen *et al.*, 2002) provides fundamental principles to guide the development of gene therapies for controlling and curbing AIDS. The development of an effective computational algorithm to search genes for viral kinetics therefore remains one of the most pressing challenges facing modern clinics.

The amount of infectious virus (viral load) reserved in hosts presents a strong dynamic feature and can be viewed as complex traits that are controlled by polygenes and sensitive to environmental stimuli (Lynch and Walsh, 1998). Wei *et al.* (1995) and Ho *et al.* (1995) observed pronounced variation in viral load trajectory among patients after the administration of an inhibitor of HIV-1 protease, implying the possible role of genes in regulating the balance of viral

production and clearance. Perelson *et al.* (1996) constructed a mathematical function to model the dynamic curves of HIV-1 for five different patients undergoing antiretroviral drug therapy. The estimated curve parameters from viral loads measured at a finite number of time points after antiviral treatment, such as the rate of loss of virus-producing cells and the rate for virion clearance, can well capture inter-patient variation and, therefore, genetic variation.

Genes controlling complex traits, or quantitative trait loci (QTL), can be detected using a genetic mapping strategy based on linkage analysis or linkage disequilibrium (LD) analysis (Lynch and Walsh, 1998). Whereas traditional linkage analysis uses recombination information only in pedigrees, LD analysis can use recombination information at the population level. Here, we propose a theoretical model for revealing how specific genetic determinants in humans affect HIV-1 infection pathogenesis by population-based LD analysis. We devise a general algorithm for implementing various mathematical models constructed for HIV-1 dynamics (Wei *et al.*, 1995; Ho *et al.*, 1995; Perelson *et al.*, 1996; Nowak and Bangham, 1996; Wu and Ding, 1999; Nowak and May, 2000) in the genetic mapping framework to detect and

*Corresponding author. Tel.: +1-352-392-3806. Fax: +1-352-392-8555. E-mail: rwu@stat.ufl.edu

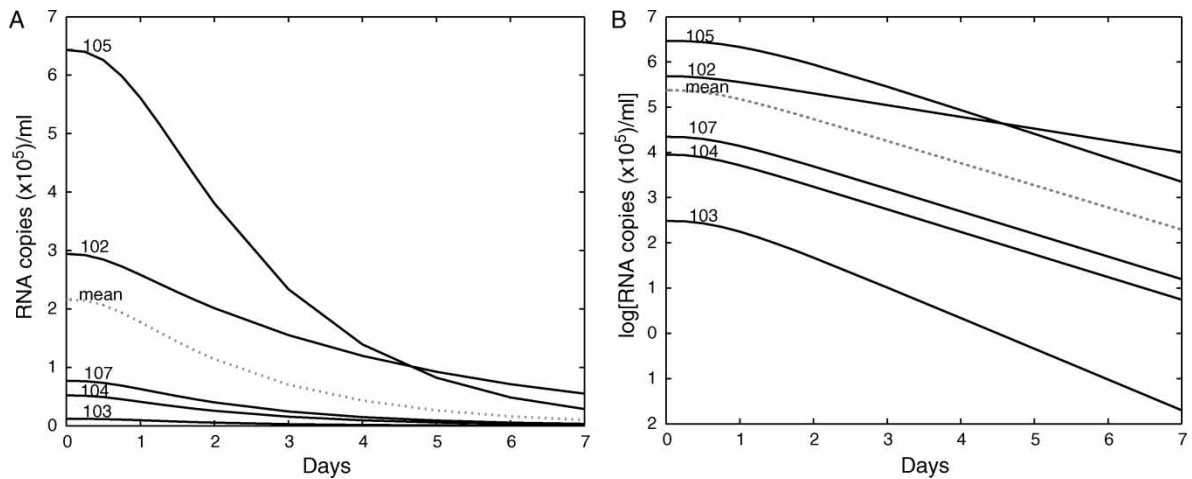


FIGURE 1 HIV-1 dynamic curves (plasma concentrations, copies per milliliter) for five patients 102, 103, 104, 105 and 107 measured at 16 time points. The dot curve represents the mean curve of the five patients (data from Perelson *et al.* 1996). The variation in viral load among the five patients after drug treatment decreases substantially with time for the untransformed curves (A), but it is roughly constant over time for the log-transformed curves (B). The constant variation is a prerequisite for the use of mathematically tractable AR(1) model to approximate the 16-dimensional residual (co)variance matrix (Σ) (Wu *et al.*, 2002; Ma *et al.*, 2002).

characterize the QTL responsible for HIV-1 disease progression. Our model allows for the tests of a number of hypotheses concerning the genetic control of clinically important variables in HIV/AIDS research.

THE MATHEMATICAL MODEL OF HIV-1 DYNAMICS

The viral data of Perelson *et al.* (1996) can be fit by a mathematical function derived to describe the time-dependent total concentration, $V(t)$, of plasma virions (including infectious, $V_I(t)$, and non-infectious, $V_{NI}(t)$) after antiviral treatment, assuming the steady state for a system. This function is expressed as

$$V(t) = V_0 e^{-ct} + \frac{cV_0}{c-\delta} \left\{ \frac{c}{c-\delta} [e^{-\delta t} - e^{-ct}] - \delta t e^{-ct} \right\}, \quad (1)$$

where V_0 is the concentration of viral particles in plasma at time 0 (the time of onset of the drug effect), δ is the rate of loss of virus-producing cells and c is the rate constant for virion clearance. We decompose $V(t)$ into two parts due to different virus compartments, i.e.

$$V_I(t) = V_0 e^{-ct}, \quad (2)$$

for infectious virions and

$$V_{NI}(t) = \frac{cV_0}{c-\delta} \left\{ \frac{c}{c-\delta} [e^{-\delta t} - e^{-ct}] - \delta t e^{-ct} \right\}, \quad (3)$$

for non-infectious virions. This model is different from the equation derived by Wei *et al.*, (1995) in which the loss of infected cells and virion clearance due to the use of inhibitors of HIV-1 protease cannot be distinguished.

Using the mathematical model for viral dynamics (Eq. (1)) and non-linear least squares fitting of the data, Perelson *et al.* (1996) obtained five distinct curves, redrawn in Figure 1, each corresponding to a different patient. These differences in curve shape can be explained by differences in the three estimated curve parameters (V_0, c, δ); for example, the estimated c values ranged from 2.06 to 3.81 day⁻¹, with a coefficient of variation (CV) of 21%, and the estimated δ values ranged from 0.26 to 0.68 day, with a CV of 27% (see Table I of Perelson *et al.*, 1996). These authors further estimated several clinically important variables which include the average life-spans ($1/c$) and average half-lives ($\ln 2/c$) of plasma virions, the average life-spans ($1/\delta$) and average half-lives ($\ln 2/\delta$) of

TABLE I Maximum likelihood estimates of the parameters describing the three dynamic curves, each corresponding to a QTL, and QTL allele frequency and marker-QTL linkage disequilibrium with 16 time points. The numbers in parentheses are the squared roots of the mean square errors of the estimates

Parameters	Genotype	True value	Heritability (H^2)	
			0.1	0.4
V_0	AA	643	650.8562 (33.0547)	646.9148 (11.2865)
	Aa	294	367.8426 (76.1271)	301.1568 (8.1070)
	aa	77	76.0635 (4.3980)	78.0755 (1.7433)
C	AA	2.06	2.5511 (0.5294)	2.6053 (0.5630)
	Aa	3.81	3.6322 (0.3327)	3.6124 (0.2975)
	aa	3.09	2.9395 (0.2684)	3.2112 (0.1930)
δ	AA	0.53	0.6475 (0.1300)	0.5377 (0.0121)
	Aa	0.26	0.3491 (0.0907)	0.2860 (0.0276)
	aa	0.5	0.6157 (0.1192)	0.5191 (0.0198)
ρ		0.6	0.5821 (0.0180)	0.5943 (0.0061)
$\sigma^{2(*)}$		5.06/0.84	4.8202 (0.2399)	0.8298 (0.0139)
p		0.6	0.6027 (0.0037)	0.6015 (0.0028)
q		0.6	0.6036 (0.0048)	0.5940 (0.0068)
D		0.08	0.0799 (0.0019)	0.0821 (0.0028)

* Different residual variances under different heritability levels.

productively infected cells, the average viral generation time (defined as the time from the release of a virion until it infects another cell and causes the release of a new generation of viral particles and calculated by $1/c + 1/\delta$), and the duration of the HIV-1 life cycle (defined as the time from the release of a virion until the release of its first progeny virus). As shown in Table II of Perelson *et al.* (1996), the five patients examined display remarkable variation in these clinical variables, with the coefficients of variation ranging from 8 to 36%. These discrepancies, in conjunction with other observations (Wei *et al.*, 1995; Ho *et al.*, 1995; Nowak and Bangham, 1996), suggest that QTL in humans may contribute substantially to HIV-1 dynamics *in vivo*.

STATISTICAL MAPPING MODELS

Functional Mapping

Unlike usual stationary traits, viral dynamics is a function-valued trait, whose genetic mapping presents a difficult statistical issue. More recently, we have proposed a novel statistical method for mapping QTL affecting function-valued traits by incorporating universal biological principles into the interval mapping framework (Wu *et al.*, 2002; Ma *et al.*, 2002; Wu *et al.*, 2004a, 2004b). Thus, different from the traditional treatment of directly estimating gene effects of a QTL at discrete time points, this method, called *functional mapping*, estimates the mathematical parameters for each QTL genotype that describe the shapes of dynamic curves. We extend this model to map QTL determining viral load trajectories.

Suppose there is a QTL with two alleles A and a for HIV dynamics. These two alleles form three QTL genotypes, AA , Aa and aa . If the three curves, each corresponding to one of these three genotypes, are different from each other, then we can say that this QTL determines HIV dynamics. Thus, by estimating the mathematical parameters of curves (e.g. V_0, c, δ), the effect of the QTL on dynamic changes can be estimated and tested.

Haplotype Block Mapping based on LD

Traditional LD analysis (or association studies) uses one marker (e.g. single nucleotide polymorphism or SNP) separately to predict one QTL in a non-random association with the marker. However, this method is limited for two reasons. First, separate use of individual SNPs is statistically less efficient and less powerful to detect a QTL than simultaneous use of multiple SNPs because the former fails to make use of association information among SNPs. Second, the number of individual SNPs, when they are used independently, should be large enough to identify the presence of QTL throughout the whole genome. For example, according to Kruglyak (1999) (the number of SNPs required for whole-genome association studies is $\geq 500,000$, which is expensive to genotype in practice.

Several recent empirical studies suggest that SNPs are not evenly distributed over the genome in terms of the extent of LD and that the structure of haplotype (a linear arrangement of non-alleles at linked loci) on a chromosome can be broken into a series of discrete haplotype blocks (Daly *et al.*, 2001; Patil *et al.*, 2001; Gabriel *et al.*, 2002; Dawson *et al.*, 2002; Phillips *et al.*, 2002). In each haplotype block, consecutive sites are in complete (or nearly complete) LD with each other and there is limited haplotype diversity due to little (coldspot) inter-site recombination. Adjacent blocks are separated by sites that show evidence of historical recombination (hotspot). It has generally been assumed that the presence of haplotype blocks provides evidence for fine-scale variation in recombination rates, with blocks corresponding to regions of reduced recombination, separated by recombination hotspots. Based on a study of the whole chromosome 21 (Patil *et al.*, 2001), 35,989 observed SNPs can be classified into different blocks with very low haplotype diversity and 80% of the variation in this chromosome can be described by only three SNPs per block.

Given the block-like pattern of LD distribution in the genome, it should be more efficient to locate allelic variants for a complex human disease trait based on haplotype blocks than individual SNPs to within a stretch of DNA that is amenable to positional cloning techniques. Because of the reported low haplotype diversity within blocks there is a possibility that very few haplotype-tagging SNPs (htSNPs) can be identified to detect common variants involved in human diseases. Suppose there are m SNPs genotyped from a chromosome, which are divided into n haplotype blocks. For block k ($k = 1, \dots, n$) comprised of n_k SNPs, we assume $n_{ht(k)}$ representative SNPs or htSNPs to uniquely identify the common haplotypes in this block. Our theoretical work has suggested that $n_{ht(k)}$ htSNPs accounting for 90% of the haplotype diversity do not significantly reduce the power to detect a QTL in the block. Therefore, our QTL mapping model here will be constructed on the basis of haplotype blocks composed of $n_{ht(k)}$ htSNPs for block k . If htSNPs in a block are detected to display strong association with a QTL, it is likely that this QTL is located within the same block based on the haplotype blocking theory.

Recently, a number of candidate genes, e.g. the MHC gene complex and the chemokine receptors genes, have been identified to regulate human immune function (Roger, 1998; Michael, 1999). Potential haplotype blocks in association with unknown linked loci predisposed to AIDS progression can be genotyped from these candidate regions. Denote the allele frequencies of $n_{ht(k)}$ htSNPs in block k by $p_{1(k)}, \dots, p_{n_{ht(k)}(k)}$, among which linkage disequilibria at different levels (digenic, trigenic, quadrigenic and so on) occur (Lou *et al.*, 2003). There is no LD between SNPs from different blocks. All these htSNPs in strong associations within haplotype blocks are

In the E step, the posterior probability for individual i to carry QTL genotype j is calculated using

$$\Pi_{j|i} = \frac{\omega_{j|i} f_j(\mathbf{y}_i)}{\sum_{j=0}^2 \omega_{j|i} f_j(\mathbf{y}_i)}, \quad (7)$$

by providing initiate values for the population genetic parameters contained in $\omega_{j|i}$ and quantitative genetic parameters contained in $f_j(\mathbf{y}_i)$.

In the M step, the calculated posterior probability is used to solve the haplotype frequencies expressed as

$$\hat{p}_{11} = \frac{1}{2n} \left[\sum_{i=1}^{n_2} (2\Pi_{i2} + \Pi_{i1}) + \sum_{i=1}^{n_1} (\Pi_{i2} + \psi\Pi_{i1}) \right], \quad (8)$$

$$\hat{p}_{10} = \frac{1}{2n} \left[\sum_{i=1}^{n_2} (2\Pi_{i0} + \Pi_{i1}) + \sum_{i=1}^{n_1} (\Pi_{i0} + (1 - \psi)\Pi_{i1}) \right], \quad (9)$$

$$\hat{p}_{01} = \frac{1}{2n} \left[\sum_{i=1}^{n_0} (2\Pi_{i2} + \Pi_{i1}) + \sum_{i=1}^{n_1} (\Pi_{i2} + (1 - \psi)\Pi_{i1}) \right], \quad (10)$$

$$\hat{p}_{00} = \frac{1}{2n} \left[\sum_{i=1}^{n_0} (2\Pi_{i0} + \Pi_{i1}) + \sum_{i=1}^{n_1} (\Pi_{i0} + \psi\Pi_{i1}) \right], \quad (11)$$

where $\psi = p_{11}p_{00}/(p_{11}p_{00} + p_{10}p_{01})$. Rather than estimating all the elements in \mathbf{h}_j and $\mathbf{\Sigma}$, our functional mapping proposed here estimates the curve parameters that model the time-specific mean vector specified by Eq. (1) for HIV dynamics and the parameters that model the (co)variance matrix $\mathbf{\Sigma}$ using a statistical approach, i.e. the first-order autoregressive [AR(1)] structure (Wu *et al.*, 2002; Ma *et al.*, 2002). The AR(1) model has two underlying assumptions, (i) the variance (σ^2) is constant over time, and (ii) the covariance decays exponentially in a proportion of ρ with time interval. Thus, the unknown vector ($\mathbf{\Omega}$) contains marker allele frequency (p), QTL allele frequency (q), marker-QTL disequilibrium (D), dynamic parameters (V_{oj}, c_j, δ_j) and (co)variance matrix-structuring parameters (ρ, σ^2).

The EM solution for quantitative genetic parameters can be difficult because the log-likelihood equations derived from the likelihood function incorporated with HIV dynamics and the AR(1) model have no unique solution. Without derivatives, the simplex method (Nelder and Mead, 1965) can be embedded within the EM algorithm to provide the estimates of (V_{oj}, c_j, δ_j) and (ρ, σ^2). Zhao *et al.* (2004) showed that the simplex method can provide faster computation for the curve parameters and (co)variance-structuring parameters for functional mapping than the EM algorithm. In this estimation, a complete loop is composed of Eqs. (7)–(11) and the simplex algorithm.

In order to make the variance stationarity assumption more realistic, a transformation approach of the effect

phenotypes \mathbf{y} can be used. As an example, we consider a log transformation in which we transform both \mathbf{y} and Eq. (1) to maintain the functional relationship between HIV loads and time (Wu *et al.*, 2004a). Carroll and Ruppert (1984) investigated a similar approach, though they allow the transformation used to be estimated by the data. In contrast, we suggest the data analyst consider a number of transformations until one is found that appropriately accounts for the particular features of the data being analyzed.

Haplotype-based Functional Mapping

Although the description of our model was based on a single SNP, the underpinning principle can be used for a haplotype-based analysis. As demonstrated for the human MHC by Ahmad *et al.* (2003), the patterns of LD may be haplotype-specific. Haplotype-specific LD patterns can be incorporated into our functional mapping model through deriving a multilocus LD mapping model. The idea for incorporating multilocus LD is described as follows.

Assume that there are n_{ht} htSNPs in a haplotype block. These htSNPs form $2^{n_{ht}}$ haplotypes each of which has a particular frequency in a population. However, because of the influences of various evolutionary forces and different recombination, some haplotypes may disappear (Jenisch *et al.*, 1999; Ahmad *et al.*, 2003; Mueller and Andreoli, 2004) so that there are actually m existing haplotypes in the population ($m < 2^{n_{ht}}$). We have constructed a multilocus LD analysis for QTL mapping in a natural population (Lou *et al.*, 2003). This analysis relies on the formulation of LD of different orders and can be directly used to map QTL for HIV dynamics based on haplotype blocks. The only difference between single SNP- and haplotype-based mapping strategies is in the form of conditional probabilities. There are three groups of conditional probabilities, expressed as $\omega_{j|2}$, $\omega_{j|1}$ and $\omega_{j|0}$, for the one SNP model, whereas there are such nine conditional probabilities, expressed as $\omega_{j|22}$, $\omega_{j|21}$, $\omega_{j|20}$, $\omega_{j|12}$, $\omega_{j|11}$, $\omega_{j|10}$, $\omega_{j|02}$, $\omega_{j|01}$ and $\omega_{j|00}$, for the two-SNP model. One can similarly derive the conditional probabilities for an arbitrary number of SNPs.

Although conditional probabilities are treated differently, modeling of the mean vector and residual (co)variance matrix is the same between the single SNP and multilocus LD mapping strategies. However, as shown below, the estimation precision of curve parameters and (co)variance-structuring parameters may be increased when multilocus LD mapping is used.

RESULTS

We perform simulation studies to investigate the statistical properties of our method. To reflect real HIV dynamic curves, we hypothesize curve parameters from the published literature. For example, Perelson *et al.* (1996) estimated curve parameters for five HIV-infected patients.

We used curve parameters estimated for these individuals, among which there are remarked differences in curve shapes (Fig. 1), to simulate viral load data at different time points for a given number of patients. The simulation conditions include:

- (i) There is a segregating QTL with allele *A* and *a* in the simulated population that determines HIV dynamics;
- (ii) The population is assumed to be at Hardy-Weinberg equilibrium, with allele frequencies of q for allele *A* and of $1 - q$ for allele *a*;
- (iii) The residual errors among different time points follow a multivariate normal distribution $MVN(\mathbf{0}, \Sigma)$, where Σ can be fit by the AR(1) model.

Under different levels of genetic control (defined by the heritability, H^2 , a proportion of genetic variance to total phenotypic variance), we simulated viral load data for 200 patients at 16 time points following the time schedule of Perelson *et al.* (1996) (every 2 h until the sixth hour, every 6 h until day 2, and every day until day 7) by taking the summations of genotypic values predicted by the theoretical curves (Eq. (1)) and residual errors obeying a multivariate normal distribution. This QTL is hidden in the simulated data set, which can be detected using known SNP markers in association with the QTL. The simulated data are analyzed by our model.

The results from our model suggest that the QTL responsible for HIV-1 dynamics can be detected using a SNP associated with the QTL, but the detection power is increased markedly from $H^2 = 0.1$ (47%) to 0.4 (100%). The curve parameters (V_0, c, δ) for each QTL genotype can be estimated with reasonably high accuracy, having the estimated values more consistent with the hypothesized values under a heritability of $H^2 = 0.4$ than 0.1 (Fig. 2). The QTL effects as reflected by the differences in

curve parameters can be estimated more precisely for a QTL displaying a greater proportion of the observed variation than that displaying a smaller proportion (Table I). The population genetic parameters of the QTL (including the allele frequencies of QTL variants and their LD with the marker) can be estimated with high precision using our closed form solution approach, not depending on the levels of heritability (Table I). According to our study, it seems that a sample of 200 can provide reasonable estimates of all parameters for a modest heritability of dynamic curve (e.g. $H^2 = 0.1$) although increased sample sizes and higher heritabilities can improve the estimation precision.

Our model provides an elegant means for testing the genetic control of clinically significant variables, such as the rate of loss of virus-producing cells (δ) and the rate for virion clearance (c). In the simulation example derived from the experiment of Perelson *et al.* (1996) we found that the QTL for overall HIV-1 dynamic curves might exert significant effects on the rate of virus-producing cells and the rate for virion clearance. The power to detect the effect of the QTL is much lower on rate c (18%) than rate δ (100%). The differences in these two rates among different QTL genotypes suggest that the QTL controlling overall HIV-1 dynamics affect the infected cells' and plasma virions' life-spans (and also average half-lives). A significance test is performed for the effect of the QTL on $1/c + 1/\delta$, suggesting that the QTL triggers an impact on the average viral generation time *in vivo*.

As described by Eq. (1), the total concentration of virions in plasma contains two components due to infectious (Eq. (2)) and non-infectious virions (Eq. (3)). The QTL detected from Eq. (1) governs the curves of the total viral load dynamics, which does not necessarily affect the dynamics of infectious and non-infectious virions, respectively. But the effects of the detected QTL

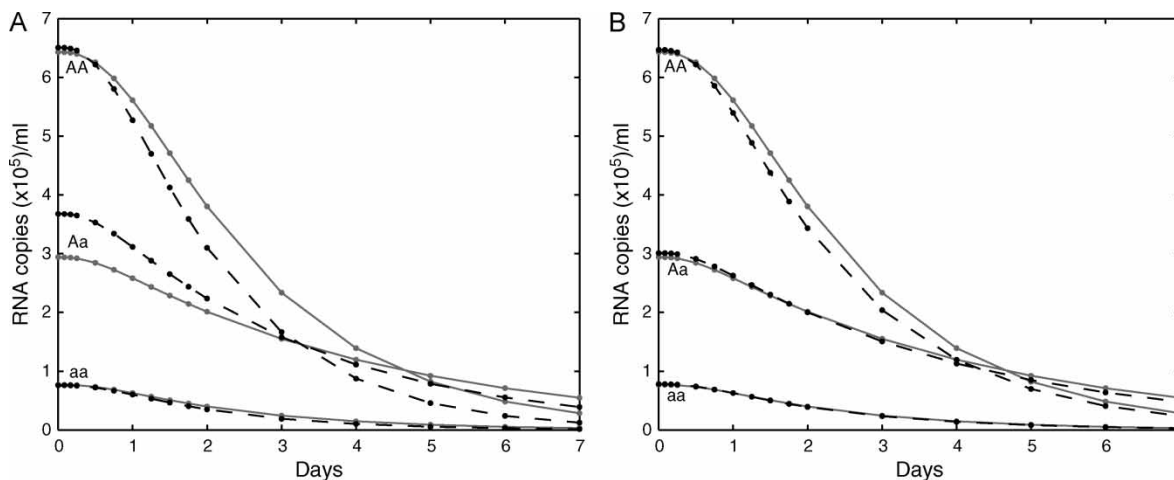


FIGURE 2 Estimated HIV dynamics curves (broken) for each of the three QTL genotypes, *AA*, *Aa* and *aa*, in comparison with the given curves (solid) used to simulate individual curves. The discrepancy between the estimated and given curves indicates the bias of our model to estimate the genetic control over viral load trajectories in human hosts. (A) Three curves under the heritability of $H^2 = 0.1$ can be distinguished in 47 of 100 simulation replicates. (B) Three curves under the heritability of $H^2 = 0.4$ can be distinguished in all simulation replicates. The log-likelihood ratio test statistics were estimated to infer the role of a QTL in shaping viral load trajectories. The dots denote the time points at which patients were measured in Perelson *et al.* (1996).

on these two compartments can be tested by formulating the null hypothesis that the three QTL genotypes have an identical infectious or non-infectious viral curve, i.e. $V_{I,AA}(t) = V_{I,Aa}(t) = V_{I,aa}(t)$ or $V_{NI,AA}(t) = V_{NI,Aa}(t) = V_{NI,aa}(t)$. Our test results suggest that the QTL detected from Eq. (1) also affect significantly the curves of infectious (Eq. (2)) and non-infectious viral dynamics (Eq. (3)).

We have also derived a more complex haplotype model in which multiple markers are used to infer an associated QTL. This extended haplotype-based QTL mapping can significantly improve our estimation precision (results not shown). An additional simulation was performed to investigate the advantages of this more complex model. As compared to the one-marker/one-QTL model, the simultaneous use of two markers can provide more precise estimates of all parameters, especially when individual markers and QTL display lower linkage disequilibria. There is a similar finding for the haplotype model composed of three or more SNPs.

DISCUSSION

We have derived a general genetic model for mapping QTL controlling HIV-1 dynamics in humans constructed from the biochemical and biophysical principles. The framework embraces two advanced mapping strategies—LD analysis based on the structure of haplotype blocks with power for fine scale mapping of QTL (Daly *et al.*, 2001; Patil *et al.*, 2001; Dawson *et al.*, 2002; Gabriel *et al.*, 2002; Phillips *et al.*, 2002) and functional mapping capable of detecting biologically meaningful QTL (Wu *et al.*, 2002; Ma *et al.*, 2002; Wu *et al.*, 2004a, 2004b). Our model allows for the incorporation of any biochemical and physiological properties related to HIV-1 infection pathogenesis. Its power is demonstrated by its ability to reliably estimate the QTL responsible for exponential curves of viral load and for the most important clinical variables. Its integration into haplotype block model framework facilitates the choice of the most informative SNPs (htSNPs) to genome-wide detect the human disease genes and positional cloning of target genes.

The derivation of our current model needs an assumption, i.e. there is only one QTL that affects HIV dynamics. This assumption is likely to be oversimplified given the possible polygenic control of this dynamics. It is essential to derive a multi-QTL model that integrates functional mapping and haplotype-based LD mapping in the gene identification framework for HIV load trajectories. Suppose there are two QTL that affect HIV dynamics. These two QTL form nine genotypes, designated by j_1j_2 ($j_1, j_2 = 2, 1, 0$), each corresponding to a curve. Thus, under the two-QTL model, we will need to estimate nine groups of curve parameters ($V_{ojj_2}, c_{jj_2}, \delta_{jj_2}$), but the residual (co)variance matrix can be modeled in the same way. The advantage of

the multi-QTL model is to allow for hypothesis tests of important genetic phenomena (Wu *et al.*, 2004b), such as QTL epistasis of different kinds and QTL linkage disequilibrium, in the control of HIV dynamics.

The model can be extended to include the pleiotropic effect of QTL on both the magnitude and shape of HIV-1 dynamic curves and the physical structure, body mass and behaviors of patients (Ameisen *et al.*, 2002). It can be further extended to model the effect of QTL interactions on HIV pathogenesis derived from the viral and human genomes. Perhaps the most appealing and powerful feature of our model is that its deployment in an HIV/AIDS research project can shed great light on the genetic control of HIV-1 infection pathogenesis and the development of efficient gene therapy for AIDS prevention and treatment. The software for the model presented in this paper is available upon request.

Acknowledgements

This work is supported by an Outstanding Young Investigator Award of the National Natural Science Foundation of China (30128017), a University of Florida Research Opportunity Fund (02050259) and a University of South Florida Biodefense grant (7222061-12) to R. W. The publication of this manuscript was approved as Journal Series No. R-10065 by the Florida Agricultural Experiment Station.

References

- Ahmad, T., *et al.* (2003) "Haplotype-specific linkage disequilibrium patterns define the genetic topography of the human MHC", *Hum. Mol. Genet.* **12**, 647–656.
- Ameisen, J.C., Lelievre, J.D. and Pleskoff, O. (2002) "HIV/host interactions: new lessons from the Red Queen's country", *AIDS* **16**, S25–S31.
- Carroll, R.J. and Ruppert, D. (1984) "Power-transformations when fitting theoretical models to data", *J. Am. Stat. Assoc.* **79**, 321–328.
- Daly, M.J., *et al.* (2001) "High-resolution haplotype structure in the human genome", *Nat. Genet.* **29**, 229–232.
- Dawson, E., *et al.* (2002) "A first-generation linkage disequilibrium map of human chromosome 22", *Nature* **418**, 544–548.
- Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) "Maximum likelihood from incomplete data via EM algorithm", *J. R. Stat. Soc. Ser. B* **39**, 1–38.
- Gabriel, S.B., *et al.* (2002) "The structure of haplotype blocks in the human genome", *Science* **296**, 2225–2229.
- Ho, D.D., Moudgil, T. and Alam, M. (1989) "Quantitation of human immunodeficiency virus type-1 in the blood of infected persons", *New Engl. J. Med.* **321**, 1621–1625.
- Ho, D.D., Neumann, A.U., Perelson, A.S., Chen, W., Leonard, J.M. and Markowitz, M. (1995) "Rapid Turnover of plasma virions and CD4 lymphocytes in HIV-1 infection", *Nature* **373**, 123–126.
- Jenisch, S., *et al.* (1999) "Corneodesmosin gene polymorphism demonstrates strong linkage disequilibrium with HLA and association with *Psoriasis vulgaris*", *Tissue Antigens* **54**, 439–449.
- Kruglyak, L. (1999) "Prospects for whole-genome linkage disequilibrium mapping of common disease genes", *Nat. Genet.* **22**, 139–144.
- Lou, X.-Y., *et al.* (2003) "A haplotype-based algorithm for multilocus linkage disequilibrium mapping of quantitative trait loci with epistasis", *Genetics* **163**, 1533–1548.
- Lynch, M. and Walsh, B. (1998) *Genetics and Analysis of Quantitative Traits* (Sinauer, Sunderland, MA).
- Ma, C.-X., Casella, G. and Wu, R.L. (2002) "Functional mapping of quantitative trait loci underlying the character process: a theoretical framework", *Genetics* **161**, 1751–1762.

- Michael, N.L. (1999) "Host genetic influences on HIV-1 pathogenesis", *Curr. Opin. Immun.* **11**, 466–474.
- Mueller, J.C. and Andreoli, C. (2004) "Plotting haplotype-specific linkage disequilibrium patterns by extended haplotype homozygosity", *Bioinformatics* **20**, 786–787.
- Nelder, J.A. and Mead, R. (1965) "A simplex method for function minimization", *Comput. J.* **7**, 308–313.
- Nowak, M.A. and Bangham, C.R.M. (1996) "Population dynamics of immune responses to persistent viruses", *Science* **272**, 74–79.
- Nowak, M.A. and May, R.M. (2000) *Virus Dynamics* (Oxford University, New York).
- Patil, N., *et al.* (2001) "Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21", *Science* **294**, 1719–1723.
- Patterson, B.K., *et al.* (1993) "Detection of HIV-1 DNA and messenger-RNA in individual cells by PCR-driven *in situ* Hybridization and flow-cytometry", *Science* **260**, 976–979.
- Perelson, A.S., Neumann, A.U., Markowitz, M., Leonard, J.M. and Ho, D.D. (1996) "HIV-1 dynamics *in vivo*: virion clearance rate, infected cell life-span, and viral generation time", *Science* **271**, 1582–1586.
- Phillips, M.S., *et al.* (2002) "Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots", *Nat. Genet* **33**, 382–387.
- Roger, M. (1998) "Influence of host genes on HIV-1 disease progression", *FASEB J.* **12**, 625–632.
- Wei, X.P., *et al.* (1995) "Viral dynamics in human immunodeficiency-virus type-1 infection", *Nature* **373**, 117–122.
- Wu, R.L. and Casella, G. (2005) *Statistical Genomics of Complex Traits: A Quantitative Trait Loci Perspective* (Springer, New York), in press.
- Wu, H. and Ding, A. (1999) "Population HIV-1 dynamics *in vivo*: applicable models and inferential tools for virological data from AIDS clinical trials", *Biometrics* **55**, 410–418.
- Wu, R.L., Ma, C.X., Chang, M., Littell, R.C., Wu, S.S., Yin, T., Huang, M., Wang, M. and Casella, G. (2002) "A logistic mixture model for characterizing genetic determinants causing differentiation in growth trajectories", *Genet. Res.* **79**, 235–245.
- Wu, R.L., Ma, C.X., Lin, M., Wang, Z.H. and Casella, G. (2004a) "Functional mapping of quantitative trait loci underlying growth trajectories using a transform-both-sides logistic model", *Biometrics*, accepted.
- Wu, R.L., Ma, C.-X., Lin, M. and Casella, G. (2004b) "A general framework for analyzing the genetic architecture of developmental characteristics", *Genetics* **166**, 1541–1551.
- Zhao, W., Wu, R.L., Ma, C.-X. and Casella, G. (2004) "A fast algorithm for functional mapping of complex traits", *Genetics* **167**, 2133–2137.