

A CONTINUOUS PHENOTYPE SPACE MODEL OF RNA VIRUS EVOLUTION WITHIN A HOST

ANDREI KOROBEGINIKOV

Centre de Recerca Matemàtica, Campus de Bellaterra
08193 Bellaterra, Barcelona, Spain
and

Departament de Matemàtiques, Universitat Autònoma de Barcelona
08193 Bellaterra, Barcelona, Spain

CONOR DEMPSEY

Department of Neuroscience, Columbia University
40 Haven Avenue
New York, NY 10032, USA

(Communicated by Yang Kuang)

ABSTRACT. Due to their very high replication and mutation rates, RNA viruses can serve as an excellent testing model for verifying hypothesis and addressing questions in evolutionary biology. In this paper, we suggest a simple deterministic mathematical model of the within-host viral dynamics, where a possibility for random mutations incorporates. This model assumes a continuous distribution of viral strains in a one-dimensional phenotype space where random mutations are modelled by Brownian motion (that is, by diffusion). Numerical simulations show that random mutations combined with competition for a resource result in evolution towards higher Darwinian fitness: a stable pulse traveling wave of evolution, moving towards higher levels of fitness, is formed in the phenotype space. The advantage of this model, compared with the previously constructed models, is that this model is mechanistic and is based on commonly accepted model of virus dynamics within a host, and thus it allows an incorporation of features of the real-life host-virus system such as immune response, antiviral therapy, etc.

1. Introduction. Viral evolution is probably the most significant single factor accountable for emergence of new and drug-resistant pathogens and preventing a development of effective drugs and vaccines. Furthermore, due to their extremely high replication rates (reaching and in some cases exceeding 10^5 per day) and high rates of mutation (of approximately 10^{-4} – 10^{-5} per nucleotide base per cycle of replication) combined with recombinogenic properties of reverse transcriptase [15, 18], RNA viruses can also serve as excellent models for addressing questions and verifying hypothesis in evolutionary biology. Despite its apparent significance, comparatively little mathematical work, in particular using deterministic modeling, has been done

2010 *Mathematics Subject Classification.* 92D30.

Key words and phrases. Evolution, mathematical modelling, HIV, random mutation, traveling wave.

A.K. is supported by the Ministry of Science and Innovation of Spain via Ramón y Cajal Fellowship RYC-2011-08061. This paper was prepared during C.D.'s summer student internship at MACSI; C.D. thanks MACSI for support and the opportunity to carry out this research.

so far in modeling viral evolution. The usual approach in this direction is to employ multi-strain models which explicitly assume the existence of a discrete [6, 16, 21] or continuous [7, 14, 19, 24] set of viral strains. These strains can be arranged into a discrete or continuous phenotype space (also known as variant space, strain space, or fitness space); among other advantages, the concept of phenotype space allows to define a distance between phenotypes [4, 5, 6]. The strains are either predetermined [2, 3, 8, 22], or emergence of new strains is assumed to be possible due to random mutations; the latter can be modeled by dispersion (diffusion) [7, 14, 24, 19] or its finite-difference equivalent [6, 21], or directly described by a stochastic process [16].

Tsimring *et al.* [24] suggested a model of viral evolution where random mutations are described by diffusion in a one-dimensional continuous fitness space. The model exhibits solutions in the form of a pulse-type traveling wave of evolution. An apparent deficiency of the Tsimring *et al.* model is that the model is phenomenological, and, as such, the model is not derived from biologically motivated hypotheses or assumptions, but is, instead, an equation with an expected dynamics. The model assumes a sort of “predation” of the more fit viral strains upon the less fit strains: the presence of more fit strains makes the proliferation of the less fit strains negative, while the presence of the less fit strains promotes the proliferation of the fitter strains. (Thus the model implies that there is no reproduction if the only one strain is present.)

To study co-evolution of a pathogen and an immune response (antigen drift) within a host Sasaki [20], Haraguchi and Sasaki [7], and Sasaki and Haraguchi [21] constructed a pathogen-antibodies model with a discrete or continuous one-dimensional strain space. In these studies, all strains were assumed equal and not interacting in any way; this made the concepts of the positive selection and the increasing fitness meaningless in this model framework. The model exhibits the evolution of antigen variants in the form of a pulse traveling wave in the strain space; on this basis, the authors came to a conclusion that the antigenic drift, by which the pathogen can continuously escape the immune defence, is driven by immune response. However, a more detailed analysis shows that the pulse wave observed in these studies is a result of superposition of two distinct traveling waves. The first of these waves is of the standard Kolmogorov-Fisher type [11] and describes the spread of the pathogen in the strain space by diffusion; this wave is followed by a delayed wave of specific immune response, which annihilates the strains which activate this response (and thus changes the shape of the first wave). For this model, neither existence of the first of these waves, nor its speed of propagation depends on the presence or absence of the second wave. Lin *et al.* [14] further develop the Sasaki model introducing the idea of a cross-immunity between the near strains into the model. The Lin *et al.* model shares the shortcomings of the Sasaki model (such as the same level of fitness of the strains and their interact only via cross immunity, which makes the idea of positive selection meaningless) and also exhibits a pulse traveling wave (which is also a superposition of two traveling waves, the first of each is of Kolmogorov-Fisher type).

There is also a significant literature focused on the dynamics of several strains with partial cross reaction (usually via cross-immunity), which circulate simultaneously in a population.

In this paper we suggest a mathematical model of the virus dynamics within a host, which is based on the biologically motivated Nowak-May and Wodarz *et al.*

models where a possibility for random mutations is incorporated. Simulations with this model demonstrate that random mutations combined with competition for a resource result in evolution towards higher Darwinian fitness.

2. Model. A basic model of virus dynamics within a host is comprises three interacting populations, namely susceptible cells, infected cells and free virus particles, and postulates that the free virus particles are able to infect the susceptible cells (which after an instance of infection move into the infected population), and that the infected cells produce the free virus particles. Such a model also assumes some form of reproduction or recruitment of the target cells. Seminal models of this type are due to Anderson and May [1], who formulated such a model to describe the spread of microparasites with a free-living infecting stage in a insects population, and Nowak and May [17], who suggested such a model to study HIV dynamics *in vivo*. In the Nowak-May model, the dynamics of the populations are governed by the following equations:

$$\begin{aligned}\frac{du(t)}{dt} &= b - \alpha u(t)x(t) - \sigma u(t), \\ \frac{dv(t)}{dt} &= \alpha u(t)x(t) - mv(t), \\ \frac{dx(t)}{dt} &= kv(t) - cx(t).\end{aligned}\tag{1}$$

Here, $u(t)$, $v(t)$ and $x(t)$ are populations of uninfected (susceptible) target cells, infected cells and free virus particles, respectively. The model postulates that there is a constant influx of the susceptible cells with rate b , that the rate of infection is proportional to the populations of susceptible cells and free virus particles, that the infected cells produce free virus particles at a rate $kv(t)$, and that average life spans of the susceptible cells, infected cells and free virus are $1/\sigma$, $1/m$ and $1/c$, respectively.

If the life span of free virus particles is considerably shorter than these of the susceptible and infected target cells (that is, if $c \gg \sigma, m$ hold), then the free virus population quickly converges to a quasi-equilibrium level proportional to the infected cell population [10, 25, 26]. This allows a reduction of the 3-dim Nowak and May model (1) to a 2-dimensional model. Indeed, if $dx(t)/dt = 0$ holds (at a quasi-equilibrium state), then $x(t) = kv(t)/c$. Hence the third equation in system (1) can be omitted, and the system takes the form

$$\begin{aligned}\frac{du(t)}{dt} &= b - \beta u(t)v(t) - \sigma u(t), \\ \frac{dv(t)}{dt} &= \beta u(t)v(t) - mv(t),\end{aligned}\tag{2}$$

where $\beta = \alpha k/c$. This model, which is due to Wodarz *et al.* [27], is probably the simplest model of virus dynamics, and this will be the basis for a model developed in this paper.

Properties of both these models are well-studied (see e.g. [12, 13, 25]). In particular, the generic properties of model (2) are entirely determined by the basic reproduction number of the infected cells $R_0 = b\beta/\sigma m$, which serves as a measure of Darwinian fitness of the virus for this model. Specifically, if $R_0 \leq 1$, then an infection-free equilibrium state $Q_0 = (u = b/\sigma, v = 0)$ is the only equilibrium state

of the model, and it is globally asymptotically stable (that is, for non-negative initial conditions the system eventually converge to the equilibrium state). If $R_0 > 1$, then, apart from equilibrium state Q_0 , the model has a positive (endemic) globally stable equilibrium state $Q^* = (\frac{b}{\sigma} \frac{1}{R_0}, \frac{\sigma}{\beta}(R_0 - 1))$, where both sub-populations coexist.

Let assume that several viral strain simultaneously coexist. Each strain is described by a set of parameters, and all possible values of these parameters form a phenotype space; we assume that this space is continuous. In model (2) framework, a strain is characterized by its basic reproduction number R_0 , which, as we mentioned, measures strains's Darwinian fitness. Therefore, for model (2) it suffices to consider a 1-dimensional phenotype space $M = \{s \in [0, \infty)\}$, where variable s is proportional to R_0 (and hence it can serve as a measure of fitness as well). Then $v(t, s)$ is the distribution of the infected population in the strain space, and the total infected population is $V(t) = \int_0^\infty v(t, s) ds$.

The basic reproduction number R_0 is proportional to the ratio of parameters β and m . For the sake of simplicity, let assume that m is the same for all strains, whereas β is a function of s . We assume in this paper that $\beta(s) = as$ (where $a > 0$); then the basic reproduction number is $R_0 = b\beta/\sigma m = bas/\sigma m$. New strains emerge as a result of random mutations, which in a continuous strain space can be modeled by dispersion. These assumptions lead to the following equations:

$$\begin{aligned} \frac{du(t)}{dt} &= b - u(t) \int_0^\infty \beta(s)v(t, s) ds - \sigma u(t), \\ \frac{\partial v(t, s)}{\partial t} &= \beta(s)u(t)v(t, s) - mv(t, s) + \mu \frac{\partial^2 v(t, s)}{\partial s^2}. \end{aligned} \quad (3)$$

The natural boundary condition for $v(t, s)$ at $s = +\infty$ is zero. The choice of a condition at $s = 0$ is not obvious; for convenience we use the no-flux condition $\frac{\partial v(t, 0)}{\partial s} = 0$. The system (3) should be complemented by non-negative initial conditions $u^0 = u(0)$ and $v^0(s) = v(0, s)$.

The variable u is measured in cells·mm⁻³; v is the density of cells in the phenotype space, and hence it is measured in cells·mm⁻³. The cell production rate b is measured in cells·mm⁻³·day⁻¹; m and σ are measured in day⁻¹, β and a are in mm³·cells⁻¹·day⁻¹, and μ (variance of the distribution of v over s) is in day⁻¹.

3. Results. Figures 1 to 4 show results of numerical simulations. For these simulations, we consider HIV as the case study, and the system parameters correspond to these for patient 2 in [9, 23]: $b = 20$ cells·mm⁻³·day⁻¹, $m = 0.8$ day⁻¹ and $\sigma = 0.02$ day⁻¹. For convenience of presentation, $a = 10^{-3}$ mm³·cells⁻¹·day⁻¹; that yields $\beta(s) = s/1000$ mm³·cells⁻¹·day⁻¹, and $R_0 = 1.25s$. In simulations, μ is equal to 10^{-8} , 10^{-7} and 10^{-6} day⁻¹.

To mimic a real life situation, in Figures 1 to 4 the initial virus load was assumed very low and the initial strains dispersion was narrow (see Fig. 3(a)):

$$v(0, s) = \begin{cases} 8(s - 0.9975) & \text{for } 0.9975 \leq s < 1.0, \\ 0.02 & \text{for } 1.0 \leq s \leq 1.005, \\ 8(1.0075 - s) & \text{for } 1.005 < s \leq 1.0075, \\ 0 & \text{otherwise.} \end{cases}$$

However, computations show that variations in magnitude or dispersion of the initial infective level have a negligible effect on the long-term dynamics, and that

for a relatively wide initial distribution its uppermost non-zero end only matters, whereas the rest of the initial non-zero interval relatively quickly disappears. It is noteworthy that the Tsimring *et al.* model exhibits the same effect [24].

Figure 1 shows the distribution of infected cells in the viral phenotype space in time for 10 years for $\mu = 10^{-6}$; formation of a pulse traveling wave moving towards increasing fitness (larger s) is clearly visible. (In this Figure, color corresponds to concentration; see the legend on the right side of Figure 1.) This Figure shows that random mutations in combination with competition for a “resource” (susceptible target cells in this case) are sufficient to initiate viral evolution. For $\mu = 5 \cdot 10^{-7} \text{ day}^{-1}$, the target cells (CD4^+ T helper cells in this case) level reached $200 \text{ cells}\cdot\text{mm}^{-3}$ in roughly 10 years, which roughly corresponds to clinical observations [17].

A noteworthy feature of the traveling wave in Fig. 1 is that the speed of evolution varies. (Figures 2 and 4, which show the dynamics of the uninfected and infected populations, also exhibit this phenomenon.) Figures 1 and 2 show that the speed of evolution depends on both the fitness of the viral strains and the abundance of the susceptible target cells: evolution goes faster as the fitness grows, until the susceptible population drops below a certain threshold; after this the shortage of susceptible cells slows evolution, and thereafter its speed remains approximately constant. In Figures 1, where $\mu = 10^{-6}$, this slow-down occurs after about 6.7 years after initial infection; for smaller μ evolution reaches this threshold later, but the change of the dynamics becomes even more evident (see curves for $\mu = 10^{-7}$ and 10^{-8} in Fig. 2).

For relatively low μ (e.g. for $\mu = 10^{-8}$ in Fig. 2), for $\mu = 10^{-8}$ the varying speed of evolution leads to the dynamics which has close qualitative resemblance with a typical progression of HIV infection. In Fig. 2, a relatively short initial transition period is followed by a longer period of slow evolution, which corresponds to a slow decreasing of the susceptible target cells (CD4^+ T cells) level. This period ends with a fast acceleration of evolution resulting in a fast drop of the susceptible CD4^+ T cells level. This picture resembles the asymptomatic stage of HIV infection, when CD4^+ T cells counts slowly decreasing, and the abrupt drop of the CD4^+ T cells level at the end of this stage leading to the development of AIDS.

Simulations also show an increase of the viral diversity; Figure 3(b) shows distributions of the infected population in the viral phenotype space after 10 years of evolution for $\mu = 10^{-6}$, 10^{-7} and 10^{-8} . It is noteworthy, that on a long run the viral diversity is nearly independent from its initial diversity.

4. Discussion and conclusion. Our objective was to construct a reasonably simple biologically motivated mathematical model of within-host RNA virus evolution. The suggested model is a straightforward extension of the Wodarz model of virus dynamics, where we assume the viral strains to be continuously distributed in a 1-dimensional phenotype space and incorporate a possibility of random mutation of the virus. Numerical simulations demonstrated that for this model random mutations combined with competition for a resource (the susceptible target cells in this case) result in evolution towards higher Darwinian fitness.

A noteworthy result is that for this model the speed of evolution is not constant, as it is for the Tsimring and the Sasaki models, but depends on (i) the fitness of strains and (ii) the abundance of susceptible target cells. This feature results in the dynamics, which qualitatively resembles the typical dynamics of HIV infection:

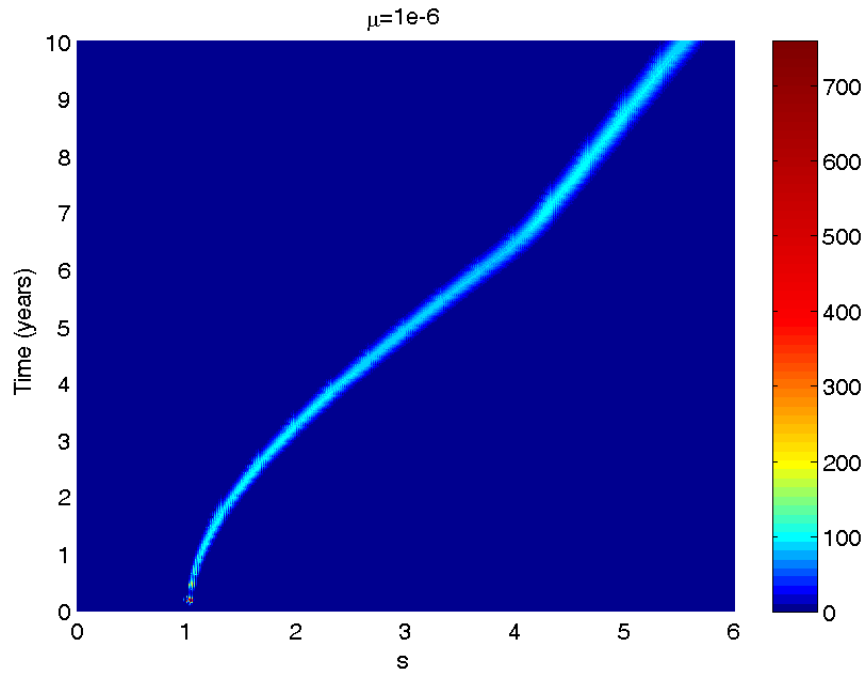


FIGURE 1. Distribution of the infected cells in the viral phenotype space in time for 10 years. Please note formation of a pulse-type traveling wave of evolution and the varying speed of the wave. Here, $b = 20 \text{ cells}\cdot\text{mm}^{-3}\cdot\text{day}^{-1}$, $m = 0.8 \text{ day}^{-1}$, $\sigma = 0.02 \text{ day}^{-1}$, $a = 10^{-3} \text{ mm}^3\cdot\text{cells}^{-1}\cdot\text{day}^{-1}$ and $\mu = 10^{-6}\text{day}^{-1}$. The colors corresponds to infected cells (and viral) concentration: dark blue is for zero concentration, while brighter colors are for non-zero concentrations (see a legend on the right-hand side).

following a short transition period, there is a prolonged period of relatively slow evolution (when the CD4^+ T cells level is slowly decreasing), which follows by a period of a fast acceleration of evolution (and hence by an abrupt drop of the CD4^+ T cells level). This period of slow evolution is similar to the asymptomatic stage of HIV infection, which ends with a rapid drop of the CD4^+ T cells level and a development of AIDS.

In this paper, for the sake of simplicity, immune response is not included in the model. A reason for this is that an objective was to demonstrate that immune response is not necessary to initiate viral evolution, and that competition for a limited resource is sufficient for natural selection. However, incorporating immune response into this model is a reasonably straightforward task. The obtained results indicate that immune response, when it is unable to eliminate virus, can accelerate evolution. Indeed, as we mentioned, the simulations demonstrate that a scarcity of the limiting resource slow evolution down, and hence a cutting the less fit strains of a distribution and thus providing more resources for newly emerged fitter strains should accelerate the speed of the traveling wave of evolution.

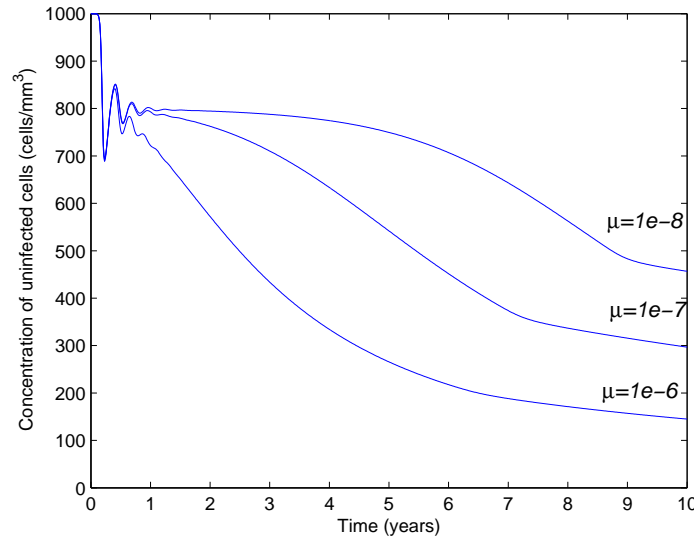


FIGURE 2. The dynamics of the uninfected $CD4^+$ T cells levels for 10 years for $\mu = 10^{-6}$, 10^{-7} and 10^{-8} .

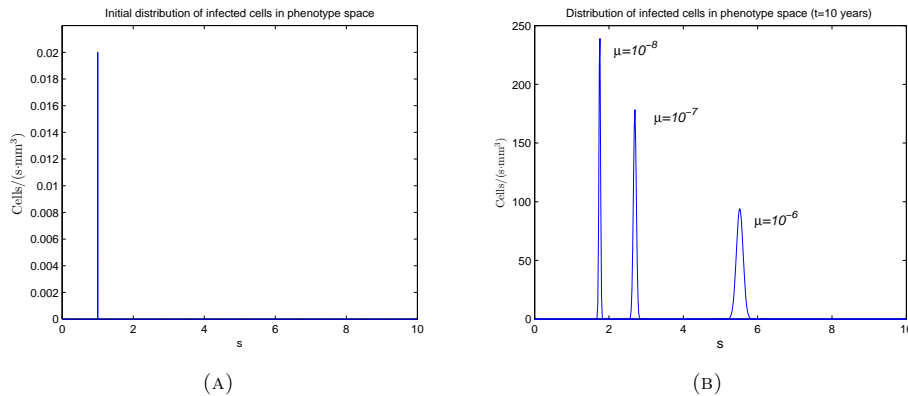


FIGURE 3. Distribution of the infected population in the phenotype space; here (a) is the initial distribution, and (b) are the distributions at $t = 10$ years for $\mu = 10^{-8}$, 10^{-7} and 10^{-6} . (Please note different vertical scales in (a) and (b).)

For simplicity we also assumed in this paper that incidence rate β (which for this model is a product of the rate of virus production by an infected cell, an average life span of a virus particle and the probability of infecting a cell by a virus particle) varies for different strains while the coefficient m is the same for all strains. We equally could assume the opposite, that is that $m(s)$ is a function of s while β is fixed. For instance, one could postulate $m(s) = \sigma + a/s$; this definition implies that the fitness is inversely proportional to the cytopathogenicity of the virus: that

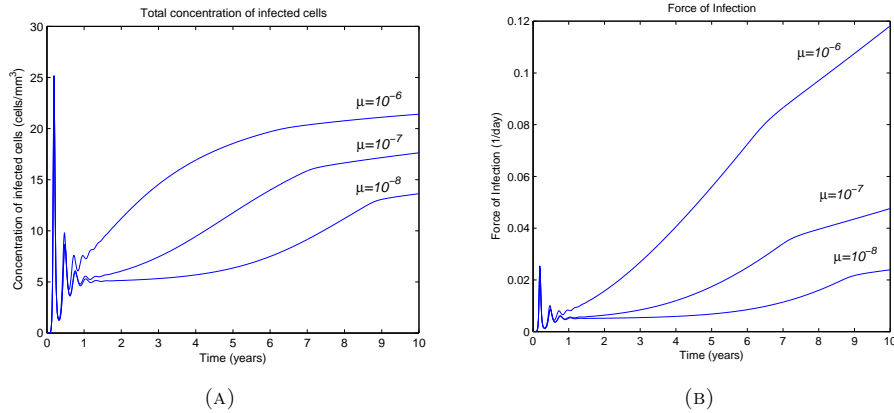


FIGURE 4. The dynamics of infected population (a) and the infective force (b) for 10 years for $\mu = 10^{-8}$, 10^{-7} and 10^{-6} .

is $\sigma < m$ holds for the wild strain, and $m(s)$ tends to σ as $s \rightarrow \infty$. However, for the Wodarz model (2), which serves as a basis for model (3), as well as for a more complex 3-dimensional Nowak-May model, the viral fitness is described by a single parameter R_0 , and, as a result, the latter assumption leads to an analogous outcome.

Details of fitness landscape is also left out of the scope of this paper. Instead we simply postulated a linear growth of fitness with variable s . While this assumption is hardly realistic, as on a large scale it implies an unlimited growth of the fitness, it is sensible as a “first approximation” in a limited region of the phenotype space, in particular taking into consideration that actual fitness landscapes for specific virus are an object of intensive research. Moreover, this assumption is sufficient to demonstrate that evolution goes toward increasing fitness and enables us to make a conclusion that for a “nonlinear” fitness landscape evolution goes to a local maximum of the fitness. However, more complicated landscapes can be incorporated into this model when specific details of these will be known.

REFERENCES

- [1] R. M. Anderson and R. M. May, [The population dynamics of microparasites and their invertebrate hosts](#), *Philos. Trans. R. Soc. Lond. Ser. B*, **291** (1981), 451–524.
- [2] V. Andreasen, [Dynamics of annual influenza A epidemics with immuno-selection](#), *J. Math. Biol.*, **46** (2003), 504–536.
- [3] V. Andreasen, S. Levin and J. Lin, [A model of influenza A drift evolution](#), *Z. Angew. Math. Mech.*, **76** (1996), 421–424.
- [4] M. F. Boni, J. R. Gog, V. Andreasen and M. W. Feldman, [Epidemic dynamics and antigenic evolution in a single season of influenza A](#), *Proc. R. Soc. B*, **273** (2006), 1307–1316.
- [5] V. Calvez, A. Korobeinikov and P. K. Maini, [Cluster formation for multi-strain infections with cross-immunity](#), *J. Theor. Biol.*, **233** (2005), 75–83.
- [6] J. R. Gog and B. T. Grenfell, [Dynamics and selection of many-strain pathogens](#), *Proc. Natl Acad. Sci. USA*, **99** (2002), 17209–17214.
- [7] Y. Haraguchi and A. Sasaki, [Evolutionary pattern of intra-host pathogen antigenic drift: Effect of crossreactivity in immune response](#), *Phil. Trans. R. Soc. B*, **352** (1997), 11–20.
- [8] T. Inoue, T. Kajiwara and T. Sasaki, [Global stability of models of humoral immunity against multiple viral strains](#), *Journal of Biological Dynamics*, **4** (2010), 282–295.

- [9] S. Iwami, T. Miura, S. Nakaoka and Y. Takeuchi, [Immune impairment in HIV infection: Existence of risky and immunodeficiency thresholds](#), *J. Theor. Biol.*, **260** (2009), 490–501.
- [10] Y. Iwasa, F. Michor and M. A. Nowak, [Virus evolution within patients increases pathogenicity](#), *J. Theor. Biol.*, **232** (2005), 17–26.
- [11] A. N. Kolmogorov, I. G. Petrovskii and N. S. Piskunov, A study of the equation of diffusion with increase in the quantity of matter, and its application to a biological problem. *Byul. Moskovskogo Gos. Univ.*, **1** (1937), 1–25. also in *Selected Works of A.N. Kolmogorov: Mathematics and Mechanics*, Kluwer, Dordrecht, (1991), 1–25.
- [12] A. Korobeinikov, [Global asymptotic properties of virus dynamics models with dose dependent parasite reproduction and virulence, and nonlinear incidence rate](#), *Math. Med. Biol.*, **26** (2009), 225–239.
- [13] A. Korobeinikov, [Stability of ecosystem: Global properties of a general prey-predator model](#), *Math. Med. Biol.*, **26** (2009), 309–321.
- [14] J. Lin, V. Andreasen, R. Casagrandi and S. A. Levin, [Traveling waves in a model of influenza A drift](#), *J. Theor. Biol.*, **222** (2003), 437–445.
- [15] L. M. Mansky and H. M. Temin, Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase, *Journal of Virology*, **69** (1995), 5087–5094.
- [16] M. A. Nowak, R. M. Anderson, A. R. McLean, T. F. W. Wolfs, J. Goudsmit and R. M. May, [Antigenic diversity threshold and the development of AIDS](#), *Science*, **254** (1991), 963–969.
- [17] M. A. Nowak and R. M. May, *Virus Dynamics*, Oxford University Press, 2000.
- [18] A. Rambaut, D. Posada, K. A. Crandall and E. C. Holmes, [The causes and consequences of HIV evolution](#), *Nature Reviews*, **5** (2004), 52–61. <http://tree.bio.ed.ac.uk/downloadPaper.php?id=242>.
- [19] J. Saldaña, S. F. Elena and R. V. Solé, [Coinfection and superinfection in RNA virus populations: A selection-mutation model](#), *Math. Biosci.*, **183** (2003), 135–160.
- [20] A. Sasaki, [Evolution of antigenic drift/switching: Continuously evading pathogens](#), *J. Theor. Biol.*, **168** (1994), 291–308.
- [21] A. Sasaki and Y. Haraguchi, Antigenic drift of viruses within a host: A finite site model with demographic stochasticity, *J. Mol. Evol.*, **51** (2000), 245–255.
- [22] M. O. Souza and J. P. Zubelli, [Global stability for a class of virus models with cytotoxic T lymphocyte immune response and antigenic variation](#), *Bull. Math. Biol.*, **73** (2011), 609–625.
- [23] M. A. Stafford et al., [Modeling plasma virus concentration during primary HIV infection](#), *J. Theor. Biol.*, **203** (2000), 285–301.
- [24] L. S. Tsimring, H. Levine and D. A. Kessler, [RNA virus evolution via a fitness-space model](#), *Phys. Rev. Lett.* **76** (1996), 4440–4443.
- [25] C. Vargas-De-León and A. Korobeinikov, [Global stability of a population dynamics model with inhibition and negative feedback](#), *Math. Med. Biol.*, **30** (2013), 65–72.
- [26] D. Wodarz, J. P. Christensen and A. R. Thomsen, [The importance of lytic and nonlytic immune responses in viral infections](#), *TRENDS in Immunology*, **23** (2002), 194–200.
- [27] D. Wodarz, P. Klenerman and M. A. Nowak, [Dynamics of cytotoxic T-lymphocyte exhaustion](#), *Proc. R. Soc. Lond. B* **265** (1998), 191–203.

Received August 01, 2013; Accepted October 16, 2013.

E-mail address: akorobeinikov@crm.cat

E-mail address: cd2656@columbia.edu