Preprint núm. 1220
April 2016

Mathematical modeling of the development of specialization in viral evolution

A. Korobeinikov, A. Nurtay

MATHEMATICAL MODELING OF THE DEVELOPMENT OF SPECIALIZATION IN VIRAL EVOLUTION

ABSTRACT. The aim of this notice is to verify the hypothesis that the development of specializations, through adaptation towards more efficient exploiting of a particular niche at the expenses of a reduced ability to exploit other niches, is a major route for the appearance of new biological species. We consider viral evolution as the case study and construct a deterministic mathematical model of viral evolution. To allow an opportunity of specialization, several types of resources (target cells in this case) are included into the model. In this model, viral subtypes are assumed to be distributed in a n-dimensional continuous phenotype space, where random mutations are modeled by diffusion, and niches are defined as local maxima of the fitness landscape. For virus, the niches correspond to different types of target cells, and the fitness can be interpreted as the efficacy of exploiting a particular type of target cells. Numerical results are obtained and presented.

It was long suggested that the development of specialization by a biological species, that is, increasing the efficacy of exploiting a particular niche at the expenses of reduced ability of exploiting other niches, can result in the appearance of new specific qualities and ultimately to the emergence of a new species [1, 2, 3, 4]. Moreover, specialization appears to be a major route of new species formation. To explore and illustrate this hypothesis, we construct a mathematical model of evolution where a possibility of specialization is incorporated.

In this notice, we consider the development of specialization by virus as a case study. Viruses is a convenient case study because viral evolution is comparatively fast, whereas the evolution of hosts (even if these are bacteria) is usually considerably slower. This difference in time scales allows to consider these two processes separately neglecting co-evolution of the host. Moreover, development of specialization by a virus appears to be common: specialization of HIV [5, 6] and strong connections of bacteriophages to a specific bacterial species [7, 8, 9, 10, 11] are examples of this. Finally, we can use a recently suggested model of viral evolution in [16], which can be easily extended to study the development of specialization.

A model which we formulate in this notice is a development of a model of viral evolution suggested by A. Korobeinikov and C. Dumpsey [16]. This model, in turn, is based on the Nowak-May model of HIV dynamics [18], which utilizes the concept of the Anderson-May model of a parasite with a free infectious stage [12] and was specifically developed to describe the dynamics of HIV within a host.

2010 *Mathematics Subject Classification.* 92D15.

Key words and phrases. evolutionary biology; viral evolution; phenotype space; peak traveling wave; Nowak-May model; Darwinian fitness; fitness landscape.

The model comprises three interacting populations represented by three variables, namely the healthy (uninfected) target cells (which in this model are CD4+, or T helper cells) of concentration $x(t)$, infected target cells of concentration $y(t)$, and free virus particles of concentration $v(t)$. The model postulates that the uninfected target cells enter the system at a constant rate λ, and are dying at a per capita rate δ. (That is, an average life span of uninfected target cells is $1/\delta$.) The uninfected target cells can be infected by free virus particles at a rate proportional to concentrations of both, target cells and free virus, αxv, where α is a proportionality coefficient. Immediately after an infection, the target cell moves to the class of infected cells. The infected cells die at a per capita rate γ. Free virus particles are produced by the infected cells at rate $ky(t)$ and are removed at a per capita rate ζ. The model equations are

$$\begin{aligned} \frac{\mathrm{d}x(t)}{\mathrm{d}t} &= \lambda - \alpha\, x(t)v(t) - \delta\, x(t), \\ \frac{\mathrm{d}y(t)}{\mathrm{d}t} &= \alpha\, x(t)v(t) - \gamma\, y(t), \\ \frac{\mathrm{d}v(t)}{\mathrm{d}t} &= k\, y(t) - \zeta\, v(t). \end{aligned} \tag{1}$$

An average life span of a free virus particle is usually considerably shorter than that of the infected or healthy target cells, and hence variable $v(t)$ in system (1) is fast compared with $x(t)$ and $y(t)$. This consideration allows to reduce the order of this model [20]. Since variable $v(t)$ is fast, then it can be expected that its value quickly reaches a quasi-equilibrium level, where $dv/dt \approx 0$ holds. (The globally asymptotic stability, or at least local asymptotic stability of the slow manifold corresponding to the quasi-equilibrium states is necessary for the convergence of a fast variable to a quasi-equilibrium state. However, for system (1) this condition holds [14, 15].) At a quasi-equilibrium state, $v(t) \approx \frac{k}{\zeta} y(t)$ holds, and system (1) can be immediately reduced to two equations, namely

$$\begin{aligned} \frac{\mathrm{d}x(t)}{\mathrm{d}t} &= \lambda - \beta\, x(t)y(t) - \delta\, x(t), \\ \frac{\mathrm{d}y(t)}{\mathrm{d}t} &= \beta\, x(t)y(t) - \gamma\, y(t), \end{aligned} \tag{2}$$

where $\beta = \alpha k/\zeta$. Model (2), which sometimes is referred to as the Wodarz model, was used as a basis for the model of viral evolution introduced in [16].

It may be worthy of noting that system (2) is equivalent to a nutrient-bacteria model (or a chemostat model) and a nutrient-plankton model. For these models, $y(t)$ stays for concentration of bacteria or plankton and $x(t)$ is for concentration of the nutrient [17, 19]. (Therefore, model of viral evolution in [16], which is based upon this model, can be suitable to study bacterial evolution as well.)

In order to model biological evolution, one has to assume, firstly, the existence of multiple variants (or phenotypes) of a studied species. Arranging these variants

CRM Preprint Series number 1220

into a phenotype (or variant) space allows to define a distance between the variants [13]. A natural methods to define the distance is to assume that the distance between variants i and j is inversely proportional to the probability of mutation of variant i into variant j. A concept of distance defined in this or another way is necessary to model random mutations. A specific approach to the random mutation modelling depends on the type of model that one wishes to obtain: thus, in a stochastic or probabilistic model one can employ a random process, such as a random walk, with a given distribution; in a deterministic model one can use an integral operator or, simpler, diffusion. Moreover, the above-mentioned definition of the distance between variants reflects the fact that mutations occur in a genotype (rather then in the phenotype) space, and potentially allows to construct a projection of the genotype space onto the phenotype space. (Such a projection must preserve the Darvinian fitness of the variants as well as the probability of mutation of a given variant to a variant with a certain fitness. Therefore, the projection, the fitness landscape and a method of modelling random mutations are strongly interconnected. In this paper a fitness is arbitrary defined, and hence this issue is left out of the scope.)

Model (2) does not include the viral population explicitly, and hence it was assumed in [16] that the infected cell population is distributed according to viral variant that they produce. (A distribution of viral variants is proportional to the corresponding distribution of infected cells.) In the model (2) framework, a viral phenotype is characterized by two parameters, namely β and γ. They are independent, and it can be assumed that each of these is continuously distributed (in contrast to a discrete distribution of genotypes). Therefore, the phenotype space can be assumed continuous and two-dimensional. In a continuous phenotype space random mutations can be modelled by diffusion (dispersion), which corresponds to random walk, or Brownian motion for a stochastic or probabilistic model.

It was assumed in [16] that viral mutations occur at the moment of viral production within an infected cell. That is, a cell infected by type i can with some probability produce a virion of type j. This set of assumption leads to the following partial integro-differential equations:

$$\begin{aligned} \frac{\mathrm{d}x(t)}{\mathrm{d}t} &= \lambda - x(t)\int_0^\infty \beta(s)\, y(s,t)\, ds - \delta\, x(t), \\ \frac{\partial y(s,t)}{\partial t} &= \beta(s)\, x(s)\, y(s,t) - \sigma(s)\, y(s,t) + \mu\Delta y(s,t). \end{aligned} \tag{3}$$

The model should be complemented with a fitness landscape, that is functions $\beta(s)$ and $\sigma(s)$, which define the Darwinian fitness, must be specified.

It is noteworthy that this model is based on a set of clearly identified assumptions and hypotheses (the "first principles") and, therefore, the results and parameters can be easily and immediately interpreted. In other words, the model is

mechanistic (as far as the Nowak-May model is mechanistic), and for this reason constitutes a sound base for further development.

A natural way to model the development of specialization is introducing several resources, with a possibility for specializing in exploiting one of these, into this model. In model (3) framework this can be done by introducing several types of target cells (or several nutrients, in the case of a nutrient-bacteria model). Let us assume that the virus can infect k differen types of cells, of concentrations $x_i(t)$ (where $i = 1, \ldots, k$), respectively. These assumptions lead to the following system of partial integro-differential equations:

$$\begin{aligned} \frac{\mathrm{d}x_i(t)}{\mathrm{d}t} &= \lambda_i - x_i(t) \int_{\bar{s}\in S} \alpha_i(\bar{s})\, y(\bar{s},t)\, d\bar{s} - \delta_i\, m_i(t), \qquad i = 1, \ldots, k; \\ \frac{\partial y(\bar{s},t)}{\partial t} &= \sum_{i=1}^{n} \alpha_i(\bar{s})\, y(\bar{s},t)\, x_i(t) - \sigma(\bar{s})\, y(\bar{s},t) + \mu\, \Delta y(\bar{s},t), \end{aligned} \tag{4}$$

where $\bar{s} = (s_1, s_2, ..., s_k, s_{k+1})$, and $\Delta y(\bar{s},t)$ is diffusion in the $k+1$ dimensional phenotype space. To close the system, we have to define functions α_i $(i = 1, \ldots, k)$ and σ (and hence define a fitness landscape). System (4) has to be complemented by initial and boundary conditions.

In this model, a viral type is described by $k+1$ parameters, namely α_i $(i = 1, \ldots, k)$ and σ, and hence the phenotype space S is $k+1$ dimensional. For simplicity we assume that the phenotype space is the positive quadrant of the $k+1$ dimensional real space, that is $S = \{\bar{s} = (s_1, s_2, ..., s_i, ..., s_k, s_{k+1}) \in \mathbb{R}^{k+1}\}$. We consider non-negative region of k-dimensional space as the feasible region. At $s_i = \infty$ for all $i = 1, 2, ..., k$ the natural boundary conditions for $y(s_i, t)$ are zero. At the other boundaries (that is, at the coordinate planes) the no-flux boundary conditions $\left.\frac{\partial y(\bar{s},t)}{\partial s_i}\right|_{s_i+0} = 0$ can be applied. Initial conditions are $y(\bar{s}, 0)$ and $x_i(0)$, where $y(\bar{s}, 0)$ and $x_i(0)$ are non-negative.

For this model, the fitness of a phenotype $\bar{s}^* = (s_1^*, s_2^*, ..., s_k^*, s_{k+1}^*)$ at moment t can be defined as

$$F(\bar{s}^*, t) = \frac{1}{\sigma(\bar{s}^*)} \sum_{i=1}^{k} \alpha_i(\bar{s}^*) m_i(t).$$

Please compare this with the basic reproduction number of the viral type, which for this model is

$$R_0(\bar{s}) = \frac{1}{\sigma(\bar{s}^*)} \sum_{i=1}^{k} \frac{\lambda_i \alpha_i(\bar{s})}{\delta_i}. \tag{5}$$

To allow the development of specialization, we have to assume a possibility for virus to increase the efficacy of exploiting some of the types of the target cells at the expense of decreasing efficacy of exploiting the others. This can be done by defining parameters $\alpha_i(\bar{s})$.

CRM Preprint Series number 1220

To illustrate that developing of specialization leads to segregation of subpopulations and to the appearance of new species, we run simulations with this model. For simplicity, we limit the model to two types of target cells, and assume that the mortality rate of the infected cells is the same for all phenotypes. These assumptions allow to use a two-dimensional phenotype space $S = \{(r, s) \in \mathbb{R}^2_+\}$.

The interaction between the variables is described by following equations:

$$\begin{aligned}
\frac{\mathrm{d}x_1(t)}{\mathrm{d}t} &= \lambda_1 - x_1(t) \int_0^\infty \int_0^\infty \alpha(r,s)\, y(r,s,t)\, dr\, ds - \delta_1\, x_1(t), \\
\frac{\mathrm{d}x_2(t)}{\mathrm{d}t} &= \lambda_2 - x_2(t) \int_0^\infty \int_0^\infty \beta(r,s)\, y(r,s,t)\, dr\, ds - \delta_2\, x_2(t), \\
\frac{\partial y(r,s,t)}{\partial t} &= (\alpha(r,s)\, x_1(t) + \beta(r,s)\, x_2(t))\, y(r,s,t) \\
&\qquad - \sigma\, y(r,s,t) + \mu\, \Delta y(r,s,t),
\end{aligned} \tag{6}$$

where $\Delta y(r,s,t) \equiv \partial^2 y(r,s,t)/\partial r^2 + \partial^2 y(r,s,t)/\partial s^2$.

For comprehensibility, we take $\lambda_1 = \lambda_2 = \lambda$ and $\delta_1 = \delta_2 = \delta$, and consider symmetric niches defining efficacies $\alpha(r,s)$ and $\beta(r,s)$ symmetric with respect to line $r = s$; that is, $\alpha(r,s) = \beta(s,r)$. Hence, the fitness landscape (5)

$$R_0(r,s) = \frac{1}{\sigma}\left(\frac{\lambda_1}{\delta_1}\alpha(r,s) + \frac{\lambda_2}{\delta_2}\beta(r,s)\right) = \frac{\lambda}{\delta\sigma}\left(\alpha(r,s) + \beta(r,s)\right) \tag{7}$$

is symmetric with respect to line $r = s$ as well. The niches correspond to (local) maxima of the fitness landscape.

In simulations, we take an average duration of an infected cells generation (gnr) as a time unit. Hence, obviously, death rate of the infected cells $\sigma = 1\, gnr^{-1}$. The influx rates of target cells $\lambda = \lambda_1 = \lambda_2 = 2\, cells/gnr$, and the death rates are $\delta = \delta_1 = \delta_2 = 0.2\, gnr^{-1}$. The coefficient of diffusion (which is assumed to be proportional to the mutation rate) is $\mu = 0.2\, var^2/gen$, where var is a unit of measurement in the phenotype space. Also, $x_1(0) = x_2(0) = 10\, cells$, and the initial value $y(r,s,0)$ is equal to zero everywhere but at a point:

$$v(r,s,0) = \begin{cases} 10^3 & \text{for } (r,s) = (1/2, 1/2), \\ 0 & \text{otherwise.} \end{cases}$$

The no-flux boundary conditions $\partial y(r,s,t)/\partial r = \partial y(r,s,t)/\partial s = 0$ are applied at all boundaries.

To allow the development of specialization, in the simulations we use

$$\begin{aligned}
\alpha(r,s) &= 10 \cdot \exp\left(-10((r-0.25)^2 + (s-0.75)^2)\right), \\
\beta(r,s) &= 10 \cdot \exp\left(-10((s-0.25)^2 + (r-0.75)^2)\right).
\end{aligned}$$

CRM Preprint Series number 1220

The corresponding fitness landscape that we use is shown in Figure 1a. Figures $1b$ to $1f$ show the distribution of the infected population $y(r, s, t)$ by viral phenotype for generations number 1 ($1b$), 5000 ($1c$), 9000 ($1d$), 10000 ($1e$), and 30000 ($1f$) for this landscape. Segregation of the population into two sub-populations which then further evolve towards deeper specialization in consuming a specific resource is clearly seen in Figures $1d - 1f$. These results show the development of specialization is a route for the emergence of sub-species.

We illustrate these results using a symmetric fitness landscape and parameters which are chosen for illustrative purposes only and are unrelated to a real-life situation. However, our simulations demonstrate that a particular choice of the landscape and the parameters does not affect the principal qualitative outcome and conclusions. In particular, the diffusion coefficient μ (which is proportional to mutation rate) does not affect the qualitative outcome in a deterministic case, but influences the speed of the evolutionary processes.

Acknowledgements. A. Nurtay is funded by la Caixa Foundation. A. Korobeinikov is supported by the Ministerio de Economía y Competitividad of Spain (MINECO), via grant MTM2015-71509-C2-1-R and Ramón y Cajal Fellowship RYC-2011-08061, by CONACYT (Mexico) via grant N219614, and by AGAUR (Generalitat de Catalunya) via grant 2014SGR-1307.

References

[1] D.J. Futuyma and G. Moreno, The Evolution of Ecological Specialization. *Annual Review of Ecology and Systematics* **19** (1988), 207–233.

[2] R. Levins, Evolution in Changing Environments. *NJ: Princeton Univ. Press* (1968).

[3] R.H. MacArthur, Geographical Ecology. *Harper & Row* (1972).

[4] R.H. MacArthur and R. Levins, The Limiting Similarity, Convergence, and Divergence of Coexisting Species. *The American Naturalist, University of Chicago Press for American Society of Naturalists* **101**, no. 921 (Sep. - Oct., 1967), 377–385.

[5] R.M. Zorzenon dos Santos, S. Coutinho, On the dynamics of the evolution of the HIV infection. *Physical Review Letters* **87** (2001), 168102.

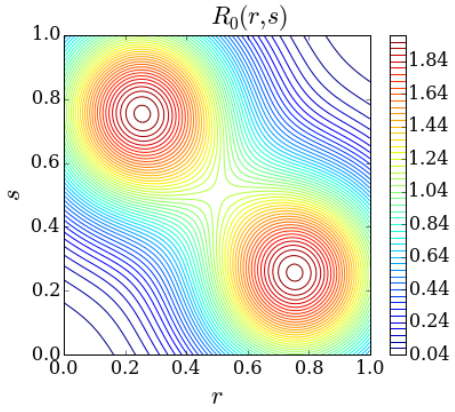
[6] G. Huanga, Y. Takeuchib, A. Korobeinikov, HIV evolution and progression of the infection to AIDS. *Journal of Theoretical Biology* **307** (August 2012), 149–159.

[7] K.H. Hoffmann, B. Rodriguez-Brito, M. Breitbart, D. Bangor, F. Angly, B. Felts, J. Nulton, F. Rohwer, P. Salamon, Power law rank-abundance models for marine phage communities. *FEMS Microbiol Lett.* **273** (2007), 224–228.

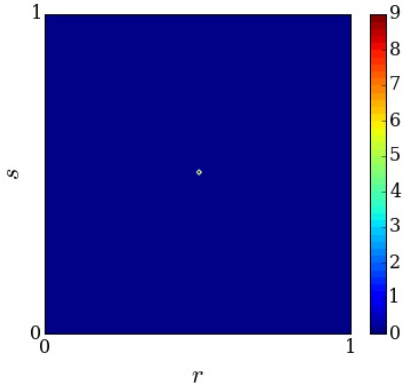
[8] E. Beretta and Y. Kuang, Modeling and analysis of a marine bacteriophage infection. *Mathematical Biosciences***149**, (April 1998), no. 157–76.

[9] S.M. Faruque, M.M. Rahman, Asadulghani, K.M.N. Islam and J.J. Mekalanos, Lysogenic conversion of environmental *Vibrio mimicus* strains by CTXÎŠ. *Infect. Immun.* **67** (November 1999), no. 11, 5723–5729.

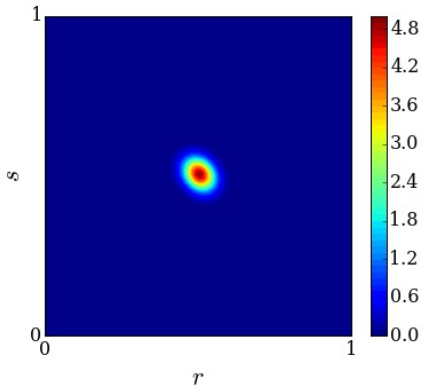
[10] K. Holmfeldt, M. Middelboe, O. Nybroe and L. Riemann, Large variabilities in host strain susceptibility and phage host range govern interactions between lytic marine phages and their Flavobacterium hosts. *Appl. Environ. Microbiol*, **73** (November 2007), no. 21, 6730–6739.



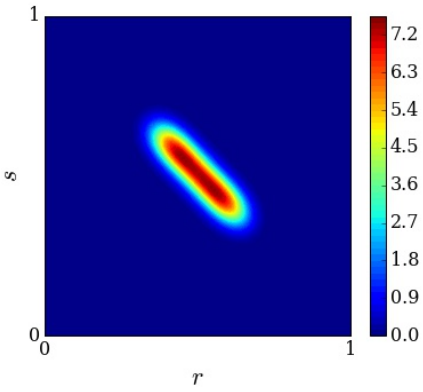
(A) $R_0(r, s)$



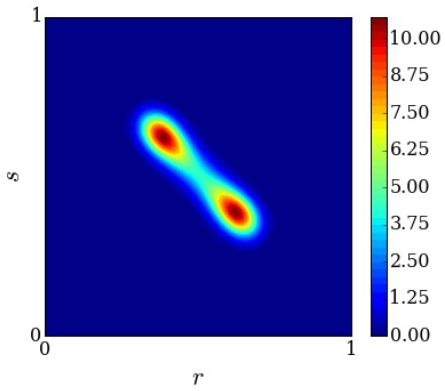
(B) t = 1



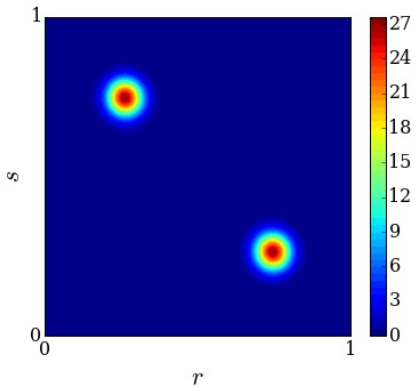
(C) t = 5000



(D) t = 9000



(E) t = 10000



(F) t = 30000

FIGURE 1. The fitness landscape used in simulations (Fig. 1a), and densities of distribution of viral variants $y(r, s, t)$ in phenotype space after given numbers of parasite generations (time t) (Fig. 1$b-$ 1f) Segregation of the viral population into two sub-populations according to target cells type is seen in Fig. $1d - 1f$. (Note, that maximum values of $y(r, s, t)$ in each plot are different.)

CRM Preprint Series number 1220

[11] M. Mehrotra, G. Wang, and W.M. Johnson, Multiplex PCR for detection of genes for *Staphylococcus aureus* enterotoxins, exfoliative toxins, toxic shock syndrome toxin 1, and methicillin resistance. *J. Clin. Microbiol*, **38** (March 2000), no. 3, 1032–1035.
[12] 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, [10.1098/rstb.1981.0005].
[13] V. Calvez, A. Korobeinikov and P.K. Maini, Cluster formation for multi-strain infections with cross-immunity. *J. Theor. Biol.*, **233** (2005), 75–83, (MR2122456)[10.1016/j.jtbi.2004.09.016].
[14] A. Korobeinikov, Global properties of basic virus dynamics models. *Bulletin of Mathematical Biology*, **66** (2004), no. 4, 879–883.
[15] 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, [10.1093/imammb/dqp009].
[16] A. Korobeinikov and C. Dempsey, A continuous phenotype space model of RNA virus evolution within a host. *Math. Biosci. Eng.* **11** (2014), no. 4, 919-927.
[17] A. Korobeinikov and S.V. Petrovskii, Toward a general theory of ecosystem stability: plankton–nutrient interaction as a paradigm, in "*Aspects of Mathematical Modelling*", Roger J. Hosking and Ezio Venturino (Editors), Birkhäuser, Basel, 2008, 27–40.
[18] (MR2009143) M.A. Nowak and R.M. May, *Virus Dynamics*, Oxford University Press, 2000.
[19] H.L. Smith and P. Waltman, The Theory of the Chemostat: Dynamics of Microbial Competition Cambridge University Press, 1995.
[20] 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, [10.1016/S1471-4906(02)02189-0].

Andrei Korobeinikov and Anel Nurtay
Centre de Recerca Matemática
Campus de Bellaterra, Edifici C
08193 Bellaterra, Barcelona, Spain
and
Departament de Matemàtiques,
Universitat Autònoma de Barcelona,
Campus de Bellaterra, Edifici C
08193 Barcelona, Spain
E-mail address: akorobeinikov@crm.cat
anurtay@crm.cat

CRM Preprint Series number 1220

CENTRE DE RECERCA MATEMÀTICA
Facultat de Ciències, UAB
08193 Bellaterra
Barcelona, Spain
www.crm.cat