Imperial College London

Using semi-deterministic models to describe dynamics of multi-strain diseases

P. Minayev and N. Ferguson

(email: p.minayev@imperial.ac.uk)

Fully deterministic model

- The model consists of a set of classes representing parts of population immune to and infectious with virus strains.

- These classes can overlap, which means that the same parts of population can be immune to or can be infected and become infectious with multiple strains simultaneously. The processes of infection are described in uniform mixing approximation and immunity to a particular strain is assumed to be lifelong.

- Virus strains are described as sequences of antigens consisting of N_L loci, each of which can be occupied by one of N_A alleles. The total possible number of strains in the model is $N_S = N_A^{N_L}$

Dynamics of the part of population immune to a strain *i* is given by

$$\frac{dz_i}{dt} = (1 - z_i)\, l_i - \mu z_i, \qquad i = 1 \ldots N_S$$

$l_i = b y_i$ is the force of infection

b is the transmission coefficient and y_i is the proportion of infectious with strain *i*

μ is the natural birth/death rate (we suppose birth and death processes in the population to be at equilibrium)

Proportion of population completely susceptible to strain *i* is given by $(1 - z_i)$

To quantitatively characterize affinity of strains, we use common Hamming distance, i.e. we define the inter-strain distance as the number of loci occupied by different alleles. This value can take on discrete values from the set $[1,2,\ldots,N_L]$.

We also define a set of additional compartments, $w_i^{(k)}$,(k=*1*… (N_L-1)), which represent the proportion of immune to any strain *j* that shares alleles at the corresponding loci with strain *i* but has not more than *k* distinct alleles.

Dynamics of $w_i^{(k)}$ is described by the equations:

$$\frac{dw_i^{(k)}}{dt} = \left(1 - w_i^{(k)}\right) \sum_{j:d_{ij} \le k} l_j - m w_i^{(k)}, \qquad i = 1 \ldots N_S, \;\; k = 1 \ldots \left(N_L - 1\right)$$

where d_{ij} is the distance between strains *i* and *j*, and the summation is done over all strains, the distances between which are less than or equal to *k*.

By definition, we put $w_i^{(0)} = z_i$ and $w_i^{(N_L)} = 1$

Proportion of immune to strains sharing alleles with strain *i* but having exactly *k* distinct alleles is $\left(w_i^{(k)} - w_i^{(k-1)}\right)$

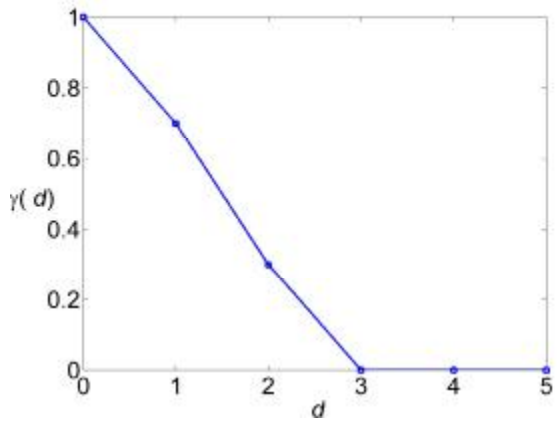
- The degree of cross-immunity against a new strain varies with the number of novel antigenic alleles in it and is described by a cross-immunity function $g(d)$ which decays with genetic distance *d*.

- Individuals who have been exposed to strain *j* and have immunity to strain *i* get infected with probability $(1-g(d))$, where *d* is the distance between strains *i* and *j.*

- The value of $g(d)$ equal to one corresponds to the full immune protection against a virus strain and the value of zero means complete susceptibility to it.

To keep the model simple, we use the following two-parametric form of $g(d)$:

$$\begin{cases} 1, & \text{if } d = 0 \\ b-(d-1)\cdot a, & \text{if } 1 \le d \le (b+a)/a < N_L \\ 0, & \text{otherwise} \end{cases}$$



- Mutation is characterized on population level by rate *m*.
- It is assumed that, as a result of mutation, hosts acquire infectiousness with a new strain and immediate mutations are possible between strains which differ from each other only by single allele, i.e. are separated by genetic distance equal to one.

Equations describing temporal evolution of the proportion of infectious with strains *i* have the form

$$\frac{dy_i}{dt} = l_i \sum_{j=0\ldots(N_L-1)} \left(w_i^{(j+1)} - w_i^{(j)}\right)\left(1-g\left(j+1\right)\right) - s\,y_i + \sum_{j=1\ldots N_S, j\neq i} m_{ij}\left(y_j - y_i\right), \quad i = 1\ldots N_S$$

s is the rate of loss of infectiousness of the host.
The last term describes mutation processes, $m_{ij} = m_{ji} = m$ if the distance between strains *i* and *j* is equal to one and $m_{ij} = m_{ji} = 0$ otherwise.

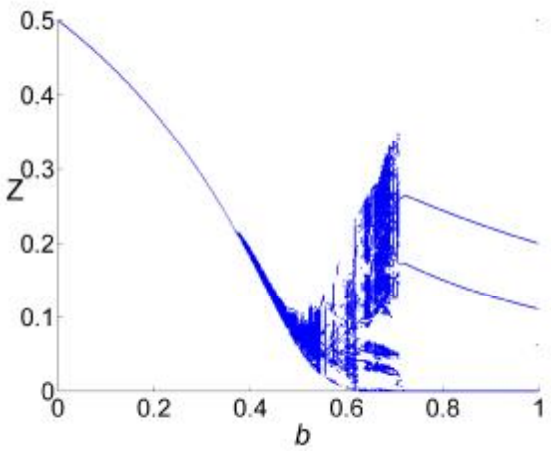
- The number of differential equations describing model dynamics is $C = \exp^{N_L \ln N_A}\left(N_L + 1\right)$
- The complexity of the model scales up exponentially with the number of loci in the virus genome and as a power function with the number of alleles.

Numerical analysis of fully deterministic model

The virus genotype consists of 5 loci and 3 alleles (the total possible number of different strains is 243). The mutation rate, *m*, for the virus is 10^{-4}/year. The infectivity of the virus is high, with the transmission coefficient, b, and recovery rate, s, equal to 100/year and 50/year respectively. The birth-death rate, m, is 0.014/year, which corresponds to life expectancy of approximately 70 years.

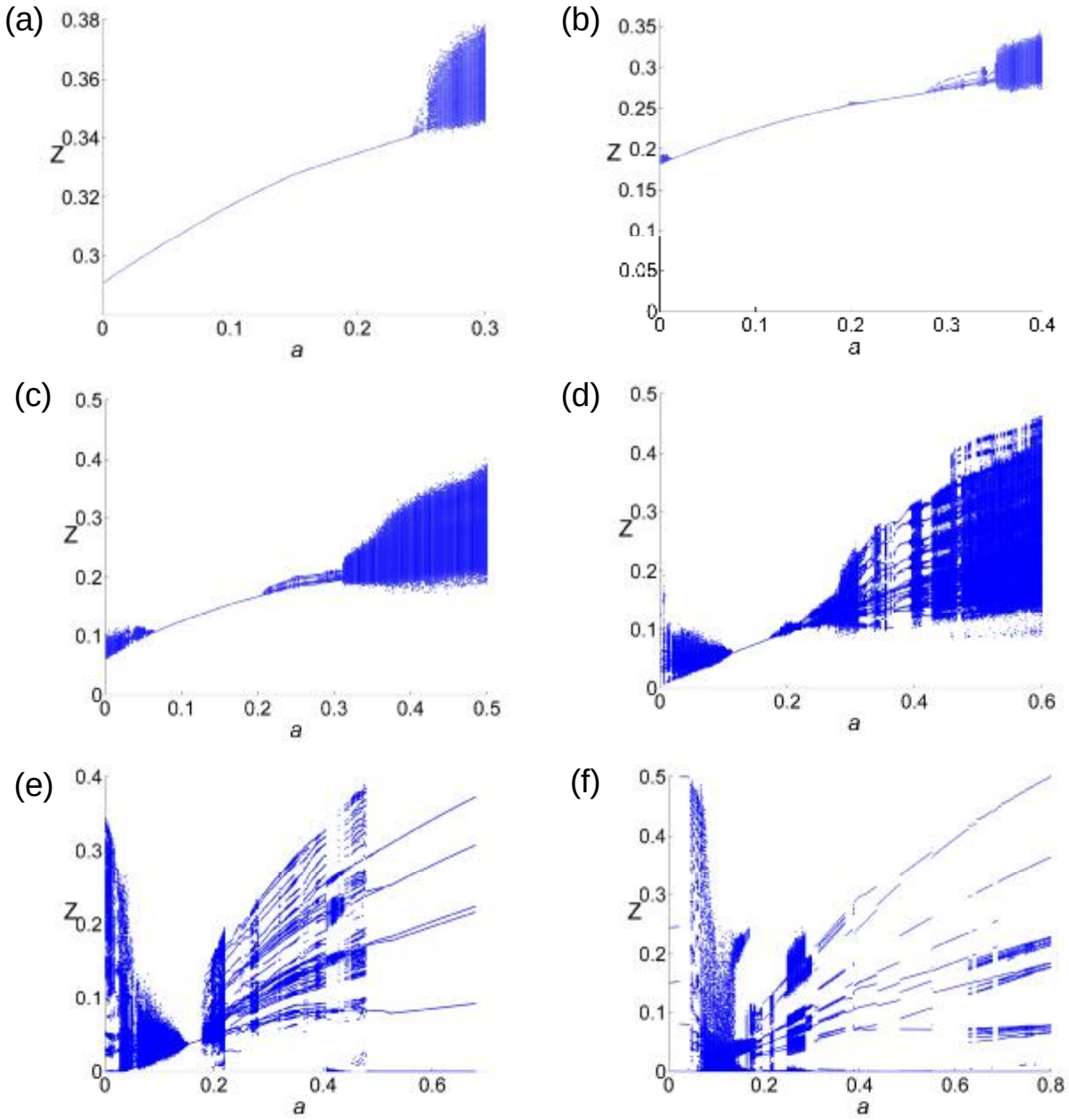
Simple case: the probability for a person exposed to one strain to be infected with another one is the same for all non-discordant strains regardless the genetic distance between them and is defined by single cross immunity parameter.

$g(d) = b = const, \quad 1 \le d < N_L; \quad g(N_L) = 0$



The model reveals three distinct types of dynamical behaviour:

- $0<b<b_L$, stable fixed point in the system phase space (fully symmetric equilibrium: all the strains coexist with the same prevalence that does not change in time)
- $b_L<b<b_U$, complex quasi-periodic and chaotic dynamics
- $b_U<b<1$, stable fixed point in the phase space (a structure of strain sets each having different prevalence is established; the strains from different sets are pairwise discordant)



Bifurcation diagrams which characterize the model dynamics depending on the cross immunity parameter *a*; the parameter *b* is fixed and is equal to
b=0.3 (a);
b=0.4 (b);
b=0.5 (c);
b=0.6 (d);
b=0.7 (e);
b=0.8 (f);

Semi-deterministic model

- Deterministic epidemic models governed by a set of differential equations have a major drawback: once emerged, a strain of infectious agent may never completely die out regenerating from arbitrarily small proportion of infected subpopulation. As a result, in the models with regular recruitment of susceptibles recurrent epidemic waves caused by the same strains and unrealistically high strain diversity and long life times can be observed.
- More biologically realistic models should take into account the probability of extinction of the disease once the proportion of infectious falls below a certain critical level.

We introduce the extinction threshold according to the following algorithm: if there is only one virus strain carrier in the population and no infection spread takes place during the mean infectious period, this strain is eliminated from the simulation, which means that the proportion of infectious with this strain is put to zero. Reemergence of extinct strains is possible due to mutation.

- Another factor which can play an important role in emergence of new strains and extinction of old ones is fluctuations in mutation processes.

Considering the probability of a single mutation as small, we describe mutation on population level as a stochastic Poisson process. We assume, as we did before, that after mutation hosts become infectious with mutated strains and immediate mutations can occur only between genetically closest strains.

In the modified model the infection and recovery processes as well as the processes of birth and death are described deterministically and the form of model equations for the parts of population immune to various virus strains remains unchanged.

$$\frac{dz_i}{dt} = (1 - z_i)\, l_i - \mu z_i, \qquad i = 1 \ldots N_S$$

$$\frac{dw_i^{(k)}}{dt} = \left(1 - w_i^{(k)}\right) \sum_{j:d_{ij} \le k} l_j - \mu w_i^{(k)}, \qquad i = 1 \ldots N_S, \;\; k = 1 \ldots (N_L - 1)$$

The equations for the infectious parts of population are superseded by the following:

$$\frac{dy_i}{dt} = l_i \sum_{j=0\ldots(N_L-1)} \left(w_i^{(j+1)} - w_i^{(j)}\right)\left(1-g\left(j+1\right)\right) - s\,y_i + SMT, \quad i = 1\ldots N_S$$

where *SMT* is a stochastic (Poisson) term describing mutation processes.

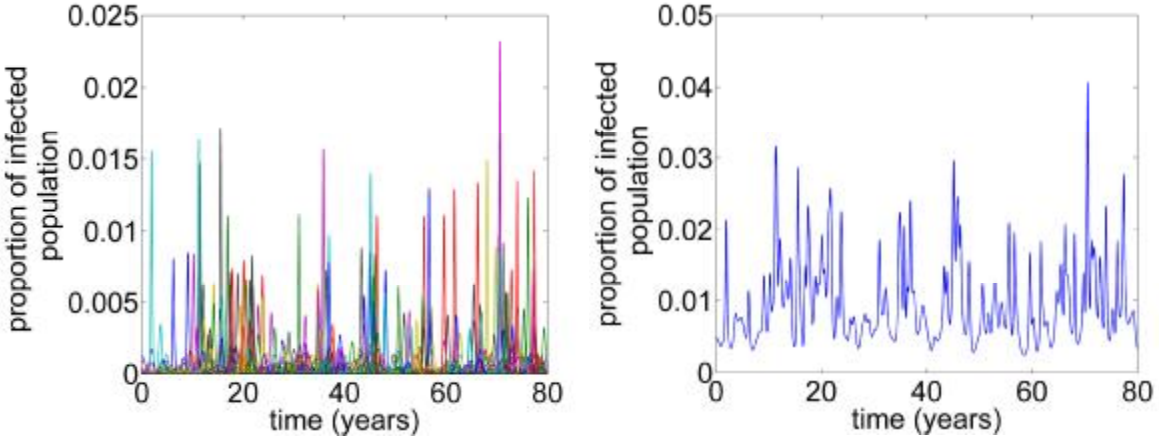
Numerical analysis of the semi-deterministic model dynamics

Simulations were done for two virus genotypes, one consisting of 10 loci and 2 alleles and allowing for 2^{10}=1024 strains in total and the other consisting of 7 loci and 3 alleles allowing for 3^7=2187 strains.

Cross immunity function is different from zero only within a rather narrow interval $0 \le d \le d_{max}$, where the distance characterizing cross immunity decay, d_{max}, has value of two to three meaning that two to three allele substitutions at different loci are enough to escape previously acquired immunity.

Immune resistance to the genetically closest strains is very high and characterized by the constant *b=0.95*.

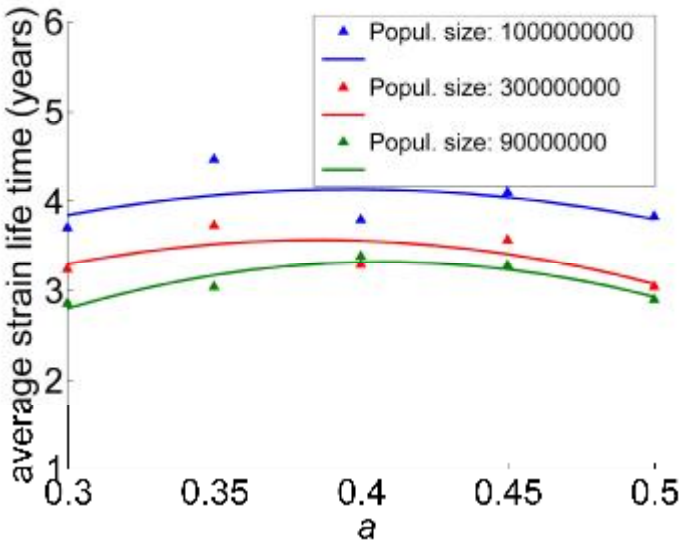
The dynamics of the system is essentially different from the one of the deterministic model.
In particular, regular periodic dynamics and equilibria are superseded by epidemic oscillations with chaotically varying amplitudes.
Temporal dynamics of the model possess an obvious trend: the average level of infection grows as cross immune response becomes more specific, which, in terms of this model, means increasing the cross immunity parameter *a*.



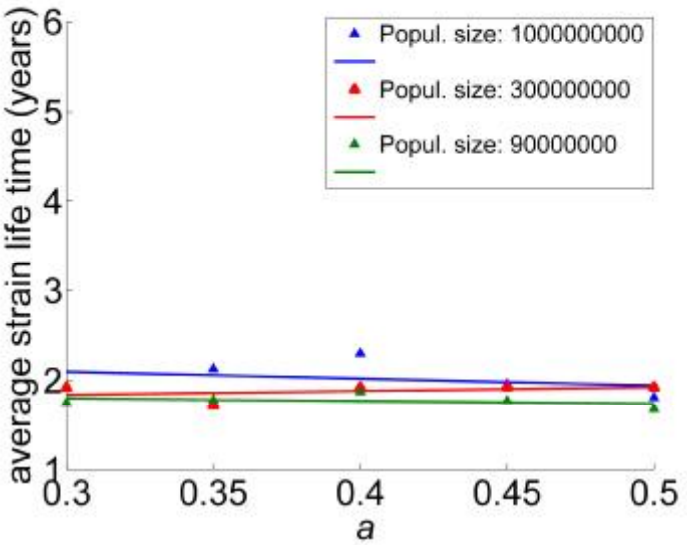
Example time series for the proportion of infected in one billion population (virus genotype consists of 10 loci and 2 alleles)

Mean strain life times in the case of the fully deterministic model can be extremely long as no factors, except for long term fitness deficit, prevent strains from extinction.
Allowing for extinction and treating mutation as a stochastic process eliminates this property of the model leading to mean strain life times of 1.5 years to several years.

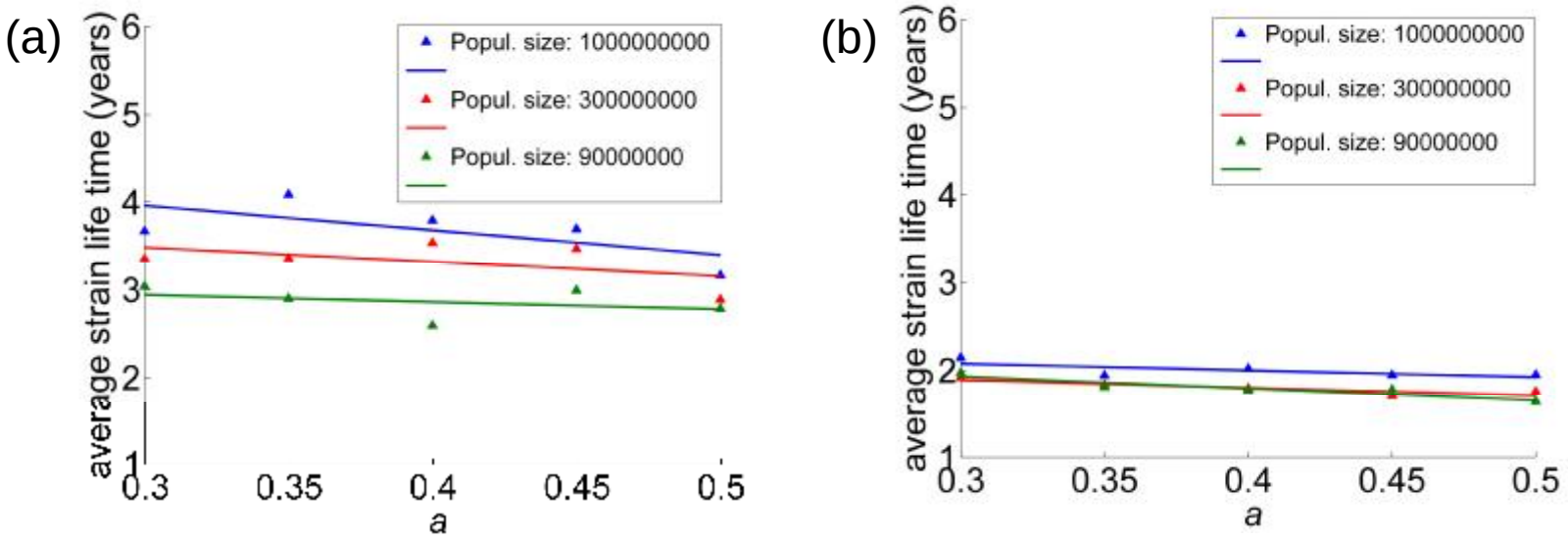
(a)



(b)

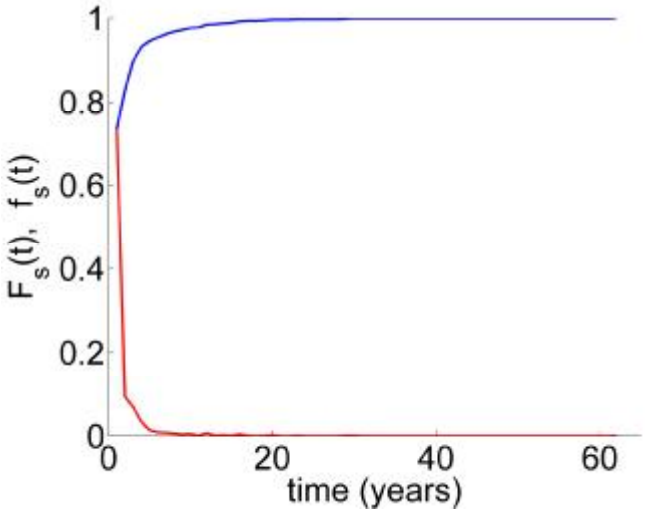


Average strain life time vs cross immunity parameter *a* (virus genotype consists of 10 loci and 2 alleles).
(a) transmission rate: b =*100*/year, recovery rate: s =*50*/year;
(b) transmission rate: b =*200*/year, recovery rate: s =*100*/year



Average strain life time vs cross immunity parameter *a* (virus genotype consists of 7 loci and 3 alleles).
(a) transmission rate: b =*100*/year, recovery rate: s =*50*/year;
(b) transmission rate: b =*200*/year, recovery rate: s =*100*/year

More detailed information about strain life times can be obtained from distribution and density functions. Cumulative distribution function $F_s(t)$ is defined as the percentage of strains having life times less than or equal to t and density function $f_s(t)$ is the percentage of strains whose life times fall within the interval $(t, t+dt)$, where dt is a constant, which we assume to be equal to one year.



Strain life time distribution, $F_s(t)$ (blue curve), and density, $f_s(t)$ (red curve), functions plotted for 90 million population (b =200/year and s =100/year).

For the case of b =100/year and s =50/year, about 84% of strains for 1 billion population and 88% of strains for 90 million population have their life times lower than 5 years. For the case of b =200/year and s =100/year this percentage is 95% and 96% for 1 billion and 90 million populations respectively.

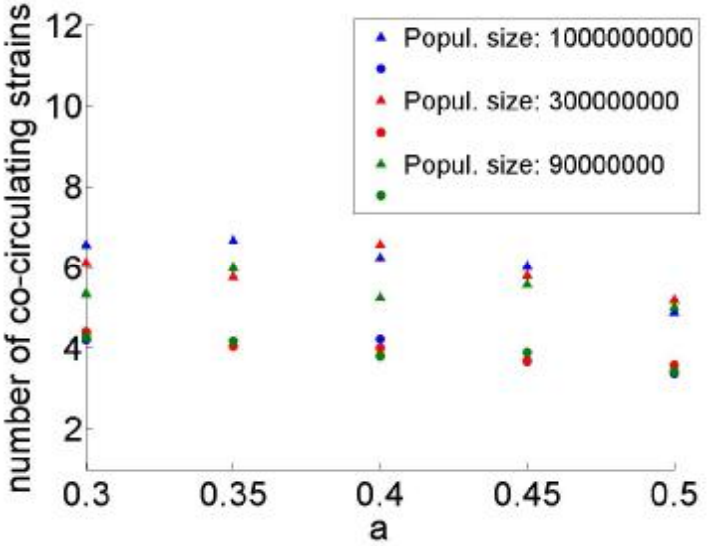
To characterize diversity of virus strains in accordance with their relative prevalences at a time *t,* the following weighted function is used: $n_w(t) = \sum_k y_k(t) / y_k^{(\max)}(t)$, where $y_k(t)$ and and $y^{(max)}(t)$ are respectively the proportion of infected with strain *k* and maximum proportion of infected at a time *t* and summation is done over all prevalent strains.

- Average weighted number of co-circulating strains is almost independent of the population size and the cross immunity parameter *a*, but is influenced by the values of infection and recovery rates and also by the genotype.

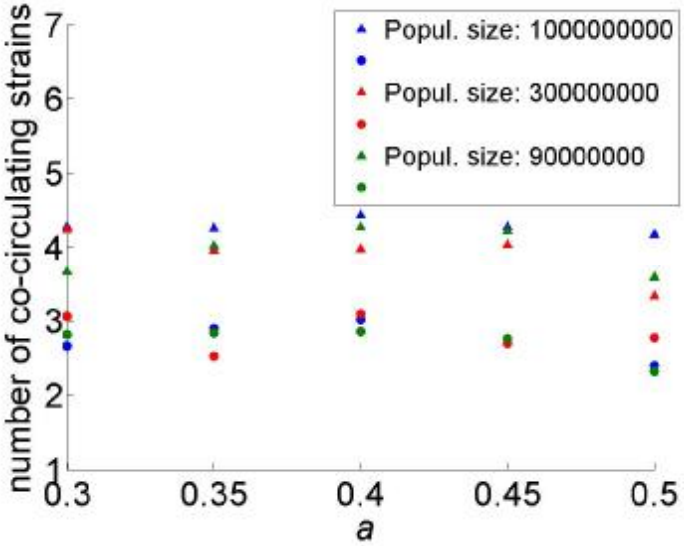
- Infectious agents of diseases characterized by higher transmissibility and recovery rates have less diversity and slightly less extinction rates compared to less transmissible and more prolonged infections.

For different transmission and recovery rates the difference in strain diversity is about 25% in the case of 10 loci 2 allele genotype and 30% in the case of 7 loci 3 allele genotype

(a)



(b)



Prevalence-weighted numbers of strains circulating yearly vs cross immunity parameter *a*. Color triangles correspond to the data for diseases with transmission and recovery rate equal to b = *100*/year and s = *50*/year, circles correspond to the data for diseases with b = *200*/year and s = *100*/year. Graph (a) is plotted for 10 locus 2 allele genotype and graph (b) is for 7 locus 3 allele genotype.

Conclusions

In this work we study a class of deterministic multi-strain epidemic models, which is an extension of the class first outlined by Gupta et al (Science, 1998. 280: p. 912-915) The extension includes description of mutation at deterministic level and cross immunity degree varying as a function of the number of novel antigenic alleles in new strains.

Dynamics of this model are much more complicated compared to the original model and are determined by the interplay of two parameters defining the function of cross immune response g *(d).*

- At low maximum values of cross immune response function g *(d)* (<0.5) dynamics of the model are characterized by fully symmetric equilibria and low amplitude chaos.
- At intermediate and high maximum values of g*(d)*(~>0.5) the dynamics are more diverse and include "windows" of regular and quasi-periodic regimes in chaotic areas as well as equilibria of a special type related to the self-organization of virus strains into pairwise discordant sets with different prevalences. Equilibrium solutions corresponding to the discordant strain structures are unstable with respect to variations of function g*(d).*

We make further extension of the deterministic model class taking into account processes of extinction and stochasticity of mutation. These two factors play crucial role in making biologically realistic estimates of mean strain life times and diversity in multi-strain systems.

- Population size has noticeable influence on strain life times for the diseases with moderate transmission rate (b =*100*/year) and duration of infection equal to a week approximately (s =*50*/year). The difference between strain life times in 1 billion and in 90 million populations is about 1 year. For highly infectious diseases (b =*200*/year) with the recovery period of about 3.7 days (s =*100*/year) the difference of strain life times in the populations of different sizes is only 3 months.
- Maximum life times of strains can be rather high (about a few decades). However the ratio of the number of long living strains to the number of all prevalent strains is negligible and the vast majority of strains have life times less than 5 years.
- Prevalence-weighted numbers of strains circulating yearly are approximately the same for all considered population sizes and are mainly defined by values of the transmission and recovery rates and the genotype (number of loci and alleles in modelled strains). Strains of diseases with higher transmission and recovery rates exhibit less diversity compared to less transmissible but more durational infections.