Evolutionary dynamics of escape from biomedical intervention

Yoh Iwasa[1], Franziska Michor[2] and Martin A. Nowak[2*]

[1]*Department of Biology, Kyushu University, Fukuoka 812-8581, Japan*
[2]*Program for Evolutionary Dynamics, Harvard University, Cambridge, MA 02138, USA*

Viruses, bacteria, eukaryotic parasites, cancer cells, agricultural pests and other inconvenient animates have an unfortunate tendency to escape from selection pressures that are meant to control them. Chemotherapy, anti-viral drugs or antibiotics fail because their targets do not hold still, but evolve resistance. A major problem in developing vaccines is that microbes evolve and escape from immune responses. The fundamental question is the following: if a genetically diverse population of replicating organisms is challenged with a selection pressure that has the potential to eradicate it, what is the probability that this population will produce escape mutants? Here, we use multi-type branching processes to describe the accumulation of mutants in independent lineages. We calculate escape dynamics for arbitrary mutation networks and fitness landscapes. Our theory shows how to estimate the probability of success or failure of biomedical intervention, such as drug treatment and vaccination, against rapidly evolving organisms.

Keywords: resistance; drug treatment; vaccination; HIV; cancer; multi-type branching process

1. INTRODUCTION

A successful treatment of an infectious disease is one that reduces its basic reproductive ratio to less than one (Anderson & May 1992). On a population level, this means that each infected host produces, on average, less than one newly infected host. A vaccination strategy that achieves this goal can ultimately remove the infectious agent from the population. Within an infected host, a successful intervention has to ensure that each microbe produces, on average, less than one offspring; for example, each virally infected cell has to produce less than one newly infected cell (Nowak & May 2000). In this case, the host can be cured from infection. Vaccines and anti-microbial drugs work towards these goals. Similarly, anti-cancer therapy aims to reduce the basic reproductive ratio of cancer cells to less than one, thereby eliminating cancer from the patient.

Success is difficult, however, because microbes and cancer cells reproduce fast, take advantage of new mutations and develop resistance (Levin *et al.* 1999; Levin 2001; Singh Sidhu *et al.* 2002). Bacterial resistance against antibiotics is a growing challenge to public health (Bonhoeffer *et al.* 1997; Lenski 1998; Levin *et al.* 2000; Chang & Roth 2001; Levy 2001). The human immunodeficiency virus (HIV) can rapidly evolve to escape from anti-viral drugs or immune responses in individual patients (Nowak *et al.* 1991; Condra *et al.* 1995; Martinez-Picado *et al.* 2000; Richman 2001). Emergence of resistance is a main reason for failure of cancer therapy (Ozols 1989; Sawyers 2001). Successful vaccines of the past were directed against organisms with little antigenic variation, whereas many of the current efforts in vaccine development target highly diverse organisms (Levin *et al.* 1999; Frank 2002). The struggle to develop an HIV or malaria vaccine, for example, is determined by the problem of inducing sufficiently strong and broad immunity in a vaccinated individual to sustain challenges by genetically diverse microbe populations without producing vaccine escape mutants (Desrosiers 1998; Letvin 1998; Dittmer *et al.* 1999; Amara *et al.* 2001; Gaschen *et al.* 2002; Ho & Huang 2002; Richie & Saul 2002). The influenza vaccine has to be modified every year to keep up with antigenic variation (Plotkin *et al.* 2002). Agricultural pests are opposed by breeding durable resistance into plants that can only be overcome by several mutations of the pathogen genome (McDonald & Linde 2002).

The fundamental problem of evolution of escape is the following. Consider a genetically diverse microbe population of size N, which is subjected to some biomedical intervention such as drug treatment or vaccine-induced immunity. In the presence of intervention, sensitive mutants have a basic reproductive ratio of less than one, but resistant mutants have a basic reproductive ratio in excess of one. Successful intervention means that the microbe population is driven to extinction before a substantial number of escape mutants has accumulated that could maintain the infection. Failure is evolution of escape. We will calculate the chances for success or failure depending on the population size of microbes (or cancer cells), the efficacy of the intervention, the mutation rate, the number of mutations required for escape and the selective conditions prior to intervention. Our theory will specify the contribution to the risk of escape that comes from the pre-existence of resistant mutants before intervention versus emergence of resistant mutants during intervention (Bonhoeffer & Nowak 1997). We will quantify the relative importance of escape pathways that require a certain number of mutational events, and allow for the possibility that intermediate mutants have different residual reproductive rates during intervention and can be subject to different selection pressures before intervention.

2. A MULTI-TYPE BRANCHING PROCESS

Suppose that n point mutations in some crucial positions in the genome are relevant for escape. The genomes

*Author for correspondence (martin_nowak@harvard.edu).

Proc. R. Soc. Lond. B (2003) **270**, 2573–2578
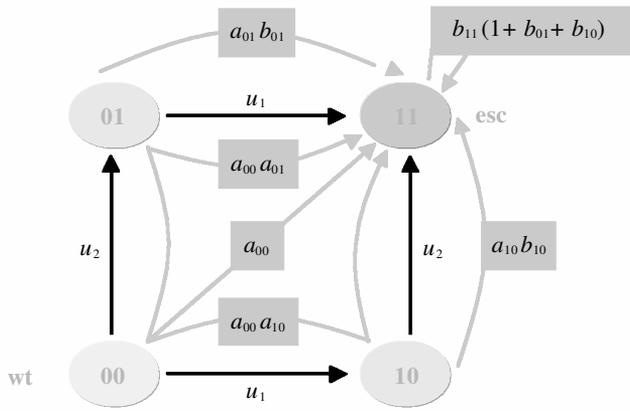DOI 10.1098/rspb.2003.2539

2573

© 2003 The Royal Society

Figure 1. Two mutations to escape. The mutation rates in the two positions are u_1 and u_2. Prior to intervention, the fitness values of the four mutants are $w_{00} = 1$ and $0 \leqslant w_{01}, w_{10}, w_{11} < 1$. Intervention leads to basic reproductive ratios $0 \leqslant R_{00}, R_{01}, R_{10} < 1$ and $R_{11} > 1$. Thus 00 is wild-type, 00, 01 and 10 are sensitive to intervention, while 11 is resistant. Let $a_i = R_i/(1 - R_i)$ and $b_i = 1/(1 - w_i)$. Prior to intervention, the mutants have the following frequencies: $x_{00} \approx 1$, $x_{01} \approx u_2 b_{01}$, $x_{10} \approx u_1 b_{10}$, $x_{11} \approx u_1 u_2 b_{11}(1 + b_{01} + b_{10})$. During intervention, escape mutants are generated from sensitive mutants at rates $\xi_{00} = u_1 u_2 a_{00}(1 + a_{01} + a_{10})$, $\xi_{01} = u_1 a_{01}$ and $\xi_{10} = u_2 a_{10}$. For the probability of success, we obtain $P = \exp[-Nz(\xi_{00} + x_{01}\xi_{01} + x_{10}\xi_{10} + x_{11})]$, where N is the total population size and z is the probability of escape starting with a single 11 type. If $R_{11} \gg 1$, then $z \approx 1$. We can write $P = \exp(-NzCu_1 u_2)$ with a risk coefficient $C = a_{00}(1 + a_{10} + a_{01}) + a_{01}b_{01} + a_{10}b_{10} + b_{11}(1 + b_{10} + b_{01})$. The selective scenario prior to intervention (w-values) and the residual reproduction during intervention (R-values) provide symmetric contributions for the odds of escape. The shades of grey indicate the level of sensitivity/escape, from sensitive (dark) to escape (pale).

can be described as binary strings of length n. All possible mutants are enumerated by $i = 0, \ldots, m$, where $m = 2^n - 1$. Treatment reduces the basic reproductive ratios, R_i, of mutants i. Sensitive mutants have $R_i < 1$, whereas escape mutants have $R_i > 1$. In each generation, a mutant i individual produces a random number of offspring following a Poisson distribution, with mean R_i. We want to calculate the probability, P, that treatment is successful against a population of microbes or cancer cells of size N. For example, N denotes the total amount of viruses challenging a vaccinated individual, or the total population size of bacteria present in a patient at the time of initiating antibiotic therapy. The distribution of mutants is determined by the mutation–selection balance prior to intervention, with fitness values given by $w_0, \ldots, w_m$; without loss of generality, we set $0 \leqslant w_i \leqslant 1$.

The general calculation is based on multi-type branching processes (Seneta 1970; Athreya & Ney 1972) and shown in Appendix A. Let us consider a specific example (figure 1). Suppose that two point mutations confer resistance. The mutation rates in these two positions are given by u_1 and u_2. In the presence of treatment, the basic reproductive ratios of the four types 00, 01, 10 and 11, are given by R_{00}, R_{01}, R_{10} and R_{11}. We have $R_{11} > 1$, while all other values are less than one. Prior to intervention, the fitness values are given by w_{00}, w_{01}, w_{10} and w_{11}. The wild-type has fitness $w_{00} = 1$, while all other fitness values

are less than one. For the probability of successful intervention we obtain

$$P = \exp(-NCu_1 u_2 z).$$

The risk coefficient, C, is given by

$$C = a_{00}(1 + a_{01} + a_{10}) + a_{01}b_{01} + a_{10}b_{10} + b_{11}(1 + b_{01} + b_{10}),$$

where $a_i = R_i/(1 - R_i)$ and $b_i = 1/(1 - w_i)$. The parameter z denotes the probability that a single 11 mutant leads to escape. For a discrete time branching process, where the number of offspring per individual follows a Poisson distribution with mean R_{11}, the escape probability z can be calculated as the root of the transcendental equation $\log(1 - z) = -R_{11}z$. For example, if $R_{11} = 2$ we obtain $z = 0.797$. A discrete time branching process with a geometric distribution or a continuous time branching process, where individuals either produce one offspring or die, leads to $z = 1 - 1/R_{11}$. In any case, if $R_{11} \gg 1$ then $z \approx 1$.

Let us now consider the situation where n mutations are required for escape, but all sequences except the escape mutant, $11 \ldots 1$, have the same basic reproductive ratio, $R < 1$, during treatment, and all mutants except the wild-type, $00 \ldots 0$, have the same relative fitness, $w < 1$, in the absence of treatment (figure 2a). Let $a = R/(1 - R)$ and $b = 1/(1 - w)$. The probability of successful intervention is given by

$$P = \exp(-NC_n u^n z).$$

For the risk coefficient, we obtain

$$C_n = \sum_{i=0}^{n} \binom{n}{i} f_{n-i}(a) f_i(b).$$

The function f is recursively defined as

$$f_0(x) = 1,$$

$$f_i(x) = x \sum_{j=0}^{i-1} \binom{i}{j} f_j(x).$$

For perfect intervention, $R = 0$, and maximum prior selection, $w = 0$, we find risk coefficients $C_1 = 1$, $C_2 = 3$, $C_3 = 13$, $C_4 = 75$, $\ldots$, $C_n = f_n(1)$. These numbers count all possible jumps from the $00..0$ to the $11..1$ sequence, including single and multiple mutations. Interestingly, multiple simultaneous mutations cannot be ignored in the calculation, because intermediate mutants have frequencies of the order of the mutation rate (Seneta 1970). If the n mutations must occur in a particular order (figure 2b), then the risk coefficient is given by

$$C_n = [a^2(1 + a)^{n-1} - b^2(1 + b)^{n-1}]/[a - b].$$

For perfect intervention, $R = 0$, and maximum prior selection, $w = 0$, we have $C_n = 2^{n-1}$, counting all possible jumps.

The model can be applied to any mutation–selection network describing any fitness or mutation landscape, also including multiple escape mutants and neutral networks (figure 2c). In general, the probability of successful intervention is of the form $P = \exp(-NC_n u^n z)$. Let us define the critical population size $N^* = 1/(C_n u^n z)$. If $N = N^*$, then the probability of success is $1/e$. If $N \gg N^*$, success is nearly impossible. If $N \ll N^*$, success is almost certain. Let us now discuss two examples that show how the theory can be applied to specific biological situations.
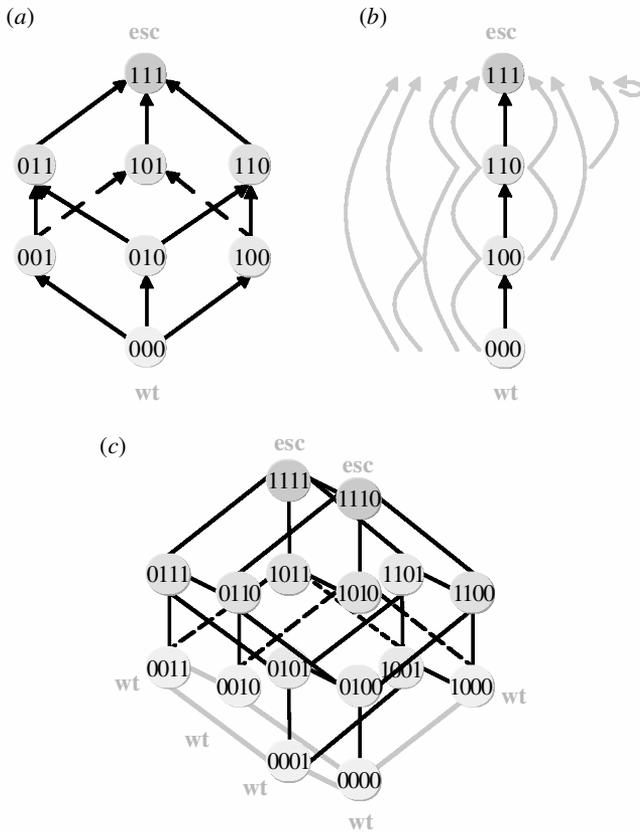
(a)

(b)

(c)

Figure 2. Mutation–selection networks and trajectories of escape. Assume that individual mutations occur at the same rate, u. (a) Three mutations are required for escape. They can occur in any order. The calculation has to consider 13 trajectories from 000 to 111, three trajectories from 001 to 111 and so on. Suppose that all variants (except 111) have the same basic reproductive ratio, $R < 1$, during treatment and all variants (except 000) have the same fitness, $w < 1$, prior to treatment. Let $a = R/(1 - R)$ and $b = 1/(1 - w)$. The risk coefficient is $C = a + 6a^2 + 6a^3 + 3b(a + 2a^2) + 3a(b + 2b^2) + b + 6b^2 + 6b^3$. (b) The three escape mutations must occur in a particular order. The eight relevant trajectories are shown. The risk coefficient is $C = a_{000}[1 + a_{100}(1 + a_{110}) + a_{110}] + b_{100}a_{100}(1 + a_{110}) + b_{110}(1 + b_{100})a_{110} + b_{111}[1 + b_{110} (1 + b_{100}) + b_{100}]$. (c) An escape situation that includes a neutral network of wild-type virus and two escape mutants. There are three shortest routes to escape: $1000 \rightarrow 1110$, $0010 \rightarrow 1110$ and $0011 \rightarrow 1111$. Each route has three trajectories. The risk coefficient is $C = \frac{3}{5}[a(1 + a) + 2ab + b(1 + b)]$, assuming the same R- and w-values for the corresponding mutants. The shades of grey indicate the level of sensitivity/escape, from sensitive (dark) to escape (pale). Abbreviations: esc, escape mutant; wt, wild-type.

3. EXAMPLE 1: HIV THERAPY

In HIV therapy, single point mutations confer resistance to non-nucleoside reverse transcriptase inhibitors, while resistance to protease inhibitors or nucleoside/nucleotide reverse transcriptase inhibitors usually evolves gradually by accumulation of multiple mutations (Larder & Kemp 1989; Nowak & May 2000; Condra et al. 1995; Martinez-Picado et al. 2000; Casado et al. 2001; Richman 2001). Consider a particular combination therapy for which $n = 3$ point mutations are required for resistance. The wild-type is denoted by 000, whereas the escape mutant is given by 111. In the best possible scenario, wild-type

and all intermediate mutants cannot replicate during therapy, $R = 0$, and all mutants have minimum fitness prior to therapy, $w = 0$. In this case, the critical virus population size is given by $N^* = 9 \times 10^{12}$. This assumes that the mutations have to occur in a particular order, otherwise $N^* = 3 \times 10^{12}$. Patients with a virus load below N^* will not evolve resistance. If, by contrast, the mutants have only a 10% selective disadvantage compared with wild-type prior to therapy, $w = 0.9$, then $N^* = 3 \times 10^{10}$. If the intermediate mutants, 001 and 011, have a 1% selective disadvantage ($R = 0.99$) and the 011 mutant has a basic reproductive ratio of $R = 0.9$, then $N^* = 4 \times 10^8$. For comparison, a perfect intervention ($R = 0$, $w = 0$) that requires only $n = 2$ point mutations has a critical population size of $N^* = 6 \times 10^8$. Figure 3 gives further examples and also shows how the risk of escape is distributed between the evolution of escape mutants following intervention and the pre-existence of escape mutants. The success of anti-HIV therapy crucially depends on: (i) the minimum number of point mutations required for escape; (ii) the virus load in the patient; (iii) the residual replication of wild-type and intermediate mutants during therapy; and (iv) the selective pressure against relevant mutations prior to treatment. Our calculation quantifies the individual contributions of these four factors and thus allows us to estimate the probabilities of success of particular interventions.

In previous work (Nowak et al. 1997; Bonhoeffer & Nowak 1997; Ribeiro et al. 1998; Nowak & May 2000), we calculated the expected frequency of resistant mutants prior to treatment. We showed that the probability that resistant mutants are being generated *de novo* during effective therapy is less than the probability that such mutants are already present before therapy. This result holds for the limit of 'effective therapy', which is defined as reducing the basic reproductive ratios of all sensitive mutants to well below one ($R_i \ll 1$). In the present paper, we show how pre-existing resistance and *de novo* resistance contribute to escape dynamics for any choice of the fitness landscape, R_i.

4. EXAMPLE 2: ANTI-CANCER THERAPY

Cancer cells can evolve resistance to chemotherapy by inactivation of tumour suppressor genes (Ozols 1989; Lowe et al. 1994; McCurrach et al. 1997; Sawyers 2001). Suppose that, at the beginning of therapy, both alleles of a certain gene are wild-type. Inactivation occurs by a combination of point mutations, at rate u, and loss of heterozygosity, at rate p (figure 4). Assume that cells with at least one wild-type allele are sensitive to chemotherapy and have a reproductive ratio of $R < 1$ during treatment. Cells that have no wild-type allele can escape from chemotherapy. The maximum cell number of a cancer that can be contained by chemotherapy is $N^* = 1/[a(1 + 2a) \times u(u + 2p)]$, where $a = R/(1 - R)$. For example, if $R = 0.9$, $u = 10^{-7}$ and $p = 10^{-6}$, we have $N^* = 3 \times 10^{10}$. If the number of cancer cells in a patient is below this threshold, then the therapy will be successful. Many cancers, however, have mutations in genes that cause chromosomal instability resulting in increased rates of loss of heterozygosity (Lengauer et al. 1998; Nowak et al. 2002). If, for example, $p = 10^{-2}$ (Lengauer et al. 1998), then the critical number of cancer cells still compatible with success is reduced to $N^* = 3 \times 10^6$. Thus the response to chemo-

| | 000 | 001 | 011 | 111 | N^* | pre-existence |
	wt	mut	mut	esc		
R	0	0	0	>1	9×10^{12}	1.00
w	1	0	0	0		
R	0.1	0.1	0.1	>1	7×10^{12}	0.90
w	1	0.1	0.1	0.1		
R	0.1	0.5	0.9	>1	1×10^{11}	0.11
w	1	0.9	0.5	0.1		
R	0.9	0.9	0.9	>1	4×10^{10}	0.005
w	1	0.1	0.1	0.1		
R	0.1	0.1	0.1	>1	3×10^{10}	0.99
w	1	0.9	0.9	0.9		
R	0.9	0.9	0.9	>1	9×10^{9}	0.30
w	1	0.9	0.9	0.9		
R	0.1	0.1	0.9	>1	4×10^{8}	0.11
w	1	0.99	0.99	0.1		
R	0	0	>1		6×10^{8}	1.00
w	1	0	0			

Figure 3. Anti-viral therapy in HIV infection often fails because the virus evolves resistance in individual patients. Usually, several point mutations are required for escape from combination therapy including reverse transcriptase and protease inhibitors. In this figure, we show various treatment scenarios in which two or three point mutations lead to escape. There is a maximum total virus load, N^*, compatible with success. If the virus load in a patient, N, exceeds N^*, then evolution of resistance is certain. A perfect intervention requiring three point mutations for escape leads to $N^* = 9 \times 10^{12}$, whereas a perfect intervention requiring two point mutations leads to $N^* = 6 \times 10^8$. The figure also shows the contribution to the probability of escape that comes from pre-existing escape mutants, $b_3[1 + b_2(1 + b_1) + b_1]/C$. Here, C is the same as in figure 2b. For perfect treatment, $R = 0$, all escape is due to pre-existing mutants. For less than perfect treatment, $R > 0$, there is considerable probability that resistance evolves after initiation of therapy. In these examples, we assume that the mutations must occur in a particular order; if this is not the case, then the estimates for the maximum population size are slightly lower. Mutation rate $u = 0.00003$ per base. The shades of grey indicate the level of sensitivity/escape, from sensitive (dark) to escape (pale).

therapy crucially depends on whether or not a particular cancer has already evolved some form of genetic instability.

5. LIMITATIONS

Our theory can be applied to the spread of replicating organisms both in single individuals and in populations. The calculations hold for arbitrarily complex fitness landscapes, including epistatic interaction among mutants. The model can be extended to include spatial compartments with different extinction probabilities. There can be latently infected cells or latent cancer cells. The fitness values of individual mutants during intervention can be time-dependent.

There are some limitations to our approach. The theory is based on multi-type branching processes that describe the accumulation of mutations in independent lineages. Therefore, we cannot describe recombination, horizontal gene transfer or frequency-dependent fitness. All of these phenomena can be important in certain situations of escape dynamics. Our study provides an analytical theory for the evolutionary dynamics of escape. It is a point of departure and comparison for more specific and complex models that deal with particular situations (Lipsitch 1999,

2001; Lipsitch *et al*. 2000). We also note that the relationship between resistance and treatment failure can be complicated. Sometimes treatment fails in the absence of resistance mutations, while at other times treatment remains succesful despite the presence of resistant mutants. Such questions require an analysis of the usually nonlinear and frequency-dependent infection dynamics of particular situations (Levin *et al*. 1999; Nowak & May 2000; Lipsitch 2001).

We also note that HIV, like other retroviruses, is diploid and has the ability to recombine. Our theory cannot deal with recombination. It is unclear, however, to what extent recombination contributes to the escape of HIV from selection pressures exerted by anti-viral therapy or vaccination, because recombining two different mutations requires the superinfection of the same cell with two different virions. Superinfection might be unlikely in patients with low virus load.

6. CONCLUSIONS

We have calculated evolutionary escape dynamics for a population of organisms challenged by a biomedical intervention that has the potential to eradicate it. Initially, the
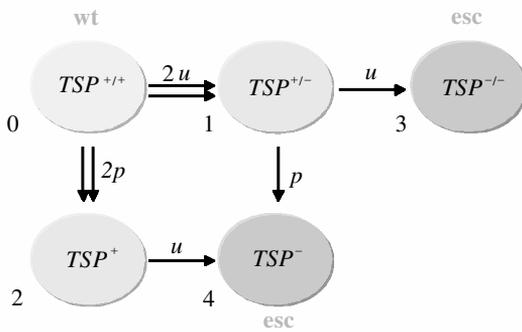
Figure 4. Cancer cells can evolve resistance to chemotherapy by mutations of specific genes such as *p53*, *RAS*, etc. Genes can be mutated by subtle nucleotide changes that occur at rate *u* per gene per cell division, or by gross chromosomal changes or recombination events that lead to loss of heterozygosity, occurring at rate *p* per gene per cell division. Let us consider the example of a tumour suppressor gene. Suppose that all cancer cells have two functioning alleles of a tumour suppressor gene prior to chemotherapy. Cells and their reproductive rates, $R_0, ..., R_4$, are enumerated as given in the figure. Let $R = R_0 = R_1 = R_2$ and $a = R/(1 - R)$. We have $N^* = 1/[a(1 + 2a)u(u + 2p)]$. Suppose $u = 10^{-7}$ and $R = 0.9$. For $p = 10^{-6}$ we have $N^* = 3 \times 10^{10}$. For $p = 10^{-2}$ we have $N^* = 3 \times 10^6$. Thus, cancers with chromosomal instability, defined by an increased rate *p*, have a much lower critical population size for escape from chemotherapy. The same prediction holds for cancers with increased point mutation rates. The shades of grey indicate the level of sensitivity/escape, from sensitive (dark) to escape (pale).

population can be genetically heterogeneous: partial or full escape mutants can be present at the time of starting the intervention. This initial distribution can be the consequence of a mutation–selection process prior to intervention. We show how to calculate the probability of escape for any initial distribution and any fitness landscape during intervention. The efficacy of an intervention can be characterized in terms of a critical population size, N^*, that is compatible with success. For example, if the mutation rate is the same for all steps, then the critical population size is proportional to $1/(C_n u^n)$, where C_n is a risk coefficient that can be calculated combinatorially and u^n denotes the *n*th power of the mutation rate. Here, *n* is the number of mutational steps required for escape. The risk coefficient, C_n, essentially sums up all possible forward trajectories from wild-type to escape mutant. There are two sets of parameters that determine C_n: the basic reproductive ratios, R_i, of mutants during intervention and their fitness values, w_i, prior to intervention. Thus, we can quantify how escape depends on the pre-existence versus emergence of resistant mutants. The primary objective of biomedical intervention against a variable pathogen population is to maximize the minimum distance between wild-type and escape mutant.

APPENDIX A

Suppose that the microbial or cancer cell genomes are described by sequences of *n* bits. The wild-type is 00 ... 0 and the escape mutant is 11..1. The escape mutant has index *m*. For all sensitive mutants, $i = 0, ..., m - 1$, we have $R_i < 1$. For the escape mutant, we have $R_m > 1$. In

the time interval of one generation, an individual *i* produces a random number of offspring following a Poisson distribution with mean R_i. Most offspring are of the same type, *i*, but some are mutants, *j*. The mutation rate from *i* to *j* is given by u_{ij}. If *u* is the mutation rate per bit, then $u_{ij} = u^{h_{ij}}(1 - u)^{n - h_{ij}}$. The Hamming distance, h_{ij}, denotes the number of bits that differ between *i* and *j*. We have $u \ll 1$.

Let ξ_i denote the probability of escape starting from one individual of type *i*. The corresponding probability of extinction is $1 - \xi_i$. For extinction, all offspring lineages generated from this one individual have to become extinct. Thus, we have

$$1 - \xi_i = \left[\sum_{k = 0}^{\infty} (1 - \xi_i)^k \frac{[(1 + u_{ii})R_i]^k}{k!} e^{-(1 + u_{ii})R_i} \right]$$
$$\times \prod_{j \neq i} \left[\sum_{k = 0}^{\infty} (1 - \xi_j)^k \frac{(u_{ij}R_i)^k}{k!} e^{-u_{ij}R_i} \right]$$
$$= \exp\left(-\xi_i R_i - \sum_{j = 0}^{m} \xi_j u_{ij} R_i \right).$$

Here, the diagonal elements of the matrix are modified to $u_{ii} = -\sum_{k \neq i} u_{ik}$. The probability ξ_i is a positive solution of the equation

$$\log(1 - \xi_i) = -\xi_i R_i - \sum_{j = 0}^{m} \xi_j u_{ij} R_i.$$

For the escape mutant, *m*, we can neglect the sum and obtain $\xi_m = z$ as a positive root of $\log(1 - z) = -zR_m$. For sensitive mutants, we cannot neglect the sum, but linearize $\log(1 - \xi_i) \approx -\xi_i$, because $\xi_i \ll 1$. Then we have

$$\xi_i = \left(zu_{im} + \sum_{j = 0}^{m - 1} \xi_j u_{ij} \right) R_i/(1 - R_i).$$

Focusing only on the terms with the smallest power of *u*, we obtain

$$\xi_i = zu^{h_{im}} \sum_{p:i \to m} v(p).$$

The sum is calculated over all paths, $p:i = k_1 \to k_2 \to ... \to k_g = m$, which progress monotonically from *i* to *m* (the Hamming distance between k_i and *m* decreases). The value of path *p* is $v(p) = a_{k_1} a_{k_2} \cdots a_{k_{g-1}}$, where $a_i = R_i/(1 - R_i)$. Because the mutation rate is small, we have to consider only those paths that proceed towards the escape mutant(s) with the minimum number of mutational steps. Suppose that we want to calculate escape dynamics from the sensitive mutant 000 to the escape mutant 111. The following are examples of paths with minimum length: (i) a sequence of three one-step mutations, $000 \to 001 \to 101 \to 111$; (ii) a double mutation followed by a one-step mutation, $000 \to 110 \to 111$; and (iii) a triple-mutation, $000 \to 111$. The following two examples do not have minimum length: (iv) $000 \to 011 \to 101 \to 111$ or (v) $000 \to 010 \to 011 \to 001 \to 111$.

Let us calculate the initial distribution of mutants at the time of intervention assuming a mutation selection equilibrium from a quasi-species equation (Eigen & Schuster 1977). Denote by x_0 the frequency of the wild-type, $0 = 00 ... 0$. For all other mutants we have $dx_i/dt = -(1 - w_i)x_i + \sum_{j = 0}^m u_{ji}x_j + u_{0i}x_0$. For $u \ll 1$, we obtain at the equilibrium

$$x_i = u^{h_{0i}} \sum_{q:0 \to i} \tilde{v}(q).$$

The sum is calculated over all paths, $q:0 = k_1 \to k_2 \to \dots \to k_g = i$, which progress monotonically from 0 to i. The value of path q is $\tilde{v}(q) = b_{k_2} b_{k_3} \dots b_{k_g}$, where $b_i = 1/(1 - w_i)$.

The risk of escape per microbe is $c = \Sigma_{i=0}^{m} \xi_i x_i$. For a population of N microbes, the probability of escape is $1 - \exp(-Nc)$. Using the expression of X_i and ξ_i, we can derive the formulas in the text. If all sensitive mutants have the same reproductive rate, R, during intervention and $a = R/(1 - R)$, then $\xi_i = z u^{h_{im}} f_i(a)$. Here, $f_i(a)$ is the sum of the value of all paths between two sequences that differ in i sites, given that the value of a path with k transitions is a^k. Similarly, if all strains except the wild-type have the same fitness, w, before intervention and $b = 1/(1 - w)$, then $x_i = u^{h_{0i}} f_i(b)$.

REFERENCES

Amara, R. R. (and 21 others) 2001 Control of a mucosal challenge and prevention of AIDS by a multiprotein DNA/MVA vaccine. *Science* **292**, 69–74.

Anderson, R. M. & May, R. M. 1992 *Infectious diseases of humans*. Oxford University Press.

Athreya, K. B. & Ney, P. E. 1972 *Branching processes*. New York: Springer.

Bonhoeffer, S. & Nowak, M. A. 1997 Pre-existence and emergence of drug resistance in HIV-1 infection. *Proc. R. Soc. Lond.* B **264**, 631–637. (DOI 10.1098/rspb.1997.0089.)

Bonhoeffer, S., Lipsitch, M. & Levin, B. R. 1997 Evaluating treatment protocols to prevent antibiotic resistance. *Proc. Natl Acad. Sci. USA* **94**, 12 106–12 111.

Casado, J. L., Dronda, F., Hertogs, K., Sabido, R., Antela, A., Marti-Belda, P., Dehertogh, P. & Moreno, S. 2001 Efficacy, tolerance, and pharmacokinetics of the combination of stavudine, nevirapine, nelfinavir, and saquinavir as salvage regimen after ritonavir or indinavir failure. *AIDS Res. Hum. Retroviruses* **17**, 93–98.

Chang, G. & Roth, C. B. 2001 Structure of MsbA from *E. coli*: a homolog of the multidrug resistance ATP binding cassette (ABC) transporters. *Science* **293**, 1793–1800.

Condra, J. H. (and 15 others) 1995 *In-vivo* emergence of HIV-1 variants resistant to multiple protease inhibitors. *Nature* **374**, 569–571.

Desrosiers, R. C. 1998 Prospects for live attenuated HIV. *Nature Med.* **4**, 982.

Dittmer, U., Brooks, D. M. & Hasenkrug, K. J. 1999 Requirement for multiple lymphocyte subsets in protection by a live attenuated vaccine against retroviral infection. *Nature Med.* **5**, 189–193.

Eigen, M. & Schuster, P. 1977 The hypercycle. A principle of natural self-organisation. Part A. Emergence of the hypercycle. *Naturwissenschaften* **64**, 541–565.

Frank, S. A. 2002 *Immunology and evolution of infectious disease*. Princeton University Press.

Gaschen, B. (and 10 others) 2002 Diversity considerations in HIV-1 vaccine selection. *Science* **296**, 2354–2360.

Ho, D. & Huang, Y. 2002 The HIV-1 vaccine race. *Cell* **110**, 135–138.

Larder, B. A. & Kemp, S. D. 1989 Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). *Science* **246**, 1155–1158.

Lengauer, C., Kinzler, K. W. & Vogelstein, B. 1998 Genetic instabilities in human cancers. *Nature* **396**, 643–649.

Lenski, R. E. 1998 Bacterial evolution and the cost of antibiotic resistance. *Int. Microbiol.* **1**, 265–270.

Letvin, N. L. 1998 Progress in the development of an HIV-1 vaccine. *Science* **280**, 1875–1880.

Levin, B. R. 2001 Minimizing potential resistance: a population dynamics view. *Clin. Inf. Dis.* **33**, 161–169.

Levin, B. R., Lipsitch, M. & Bonhoeffer, S. 1999 Population biology, evolution, and infectious disease: convergence and synthesis. *Science* **283**, 806–809.

Levin, B. R., Perrot, V. & Walker, N. 2000 Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. *Genetics* **154**, 985–997.

Levy, S. B. 2001 Antibiotic resistance: consequences of inaction. *Clin. Inf. Dis.* **33**, 124–129.

Lipsitch, M. 1999 Bacterial vaccines and serotype replacement: lessons from *Haemophilus influenzae* and prospects for *Streptococcus pneumoniae*. *Emerging Inf. Dis.* **5**, 336–345.

Lipsitch, M. 2001 The ups and downs of antimicrobial resistance. *Trends Microbiol.* **9**, 438–444.

Lipsitch, M., Bergstrom, C. T. & Levin, B. R. 2000 The epidemiology of antibiotic resistance in hospitals: paradoxes and prescriptions. *Proc. Natl Acad. Sci. USA* **97**, 1938–1943.

Lowe, S. W., Bodis, S., McClatchey, A., Remington, L., Ruley, H. E., Fisher, D. E., Housman, D. E. & Jacks, T. 1994 Status and the efficacy of cancer therapy *in vivo*. *Science* **266**, 807–810.

McCurrach, M. E., Connor, T. M., Knudson, C. M., Korsmeyer, S. J. & Lowe, S. W. 1997 *bax*-Deficiency promotes drug resistance and oncogenic transformation by attenuating p53-dependent apoptosis. *Proc. Natl Acad. Sci. USA* **94**, 2345–2349.

McDonald, B. A. & Linde, C. 2002 Pathogen population genetics, evolutionary potential, and durable resistance. *A. Rev. Phytopathol.* **40**, 349–379.

Martinez-Picado, J. (and 15 others) 2000 Antiretroviral resistance during successful therapy of HIV type 1 infection. *Proc. Natl Acad. Sci. USA* **97**, 10 948–10 953.

Nowak, M. A. & May, R. M. 2000 *Virus dynamics*. Oxford University Press.

Nowak, M. A., Anderson, R. M., McLean, A. R., Wolfs, T., Goudsmit, J. & May, R. M. 1991 Antigenic diversity thresholds and the development of AIDS. *Science* **254**, 963–969.

Nowak, M. A., Bonhoeffer, S., Shaw, G. M. & May, R. M. 1997 Anti-viral drug treatment: dynamics of resistance in free virus and infected cell populations. *J. Theor. Biol.* **184**, 203–217.

Nowak, M. A., Komarova, N. L., Sengupta, A., Jallepalli, P. V., Shih, I. M., Vogelstein, B. & Lengauer, C. 2002 The role of chromosomal instability in tumor initiation. *Proc. Natl Acad. Sci. USA* **99**, 16 226–16 231.

Ozols, R. F. 1989 *Drug resistance in cancer therapy*. Dordrecht, The Netherlands: Kluwer.

Plotkin, J. B., Dushoff, J. & Levin, S. A. 2002 Hemagglutinin sequence clusters and the antigenic evolution of influenza A virus. *Proc. Natl Acad. Sci. USA* **99**, 6263–6268.

Ribeiro, R. M., Bonhoeffer, S. & Nowak, M. A. 1998 The frequency of resistant mutant virus before antiviral therapy. *AIDS* **12**, 461–465.

Richie, T. L. & Saul, A. 2002 Progress and challenges for malaria vaccines. *Nature* **415**, 694–701.

Richman, D. D. 2001 HIV chemotherapy. *Nature* **410**, 995–1001.

Sawyers, C. L. 2001 Research on resistance to cancer drug gleevec. *Science* **294**, 1834.

Seneta, E. 1970 Population growth and multi-type Galton–Watson process. *Nature* **225**, 776.

Singh Sidhu, A. B., Verdier-Pinard, D. & Fidock, D. A. 2002 Chloroquine resistance in *Plasmodium falciparum* malaria parasites conferred by *pfcrt* mutations. *Science* **298**, 210–213.