Journal of Theoretical Biology 417 (2017) 61–67

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Bacterial resistance to antibodies: a model evolutionary study

Lawrence S. Schulman

Physics Department, Clarkson University, Potsdam, NY 13699-5820, USA

ARTICLE INFO

Keywords:
Evolution
Horizontal gene transfer
Tangled nature
Ecology

ABSTRACT

The tangled nature model of evolution (reviewed in the main text) is adapted for use in the study of antibody resistance acquired by horizontal gene transfer. Exchanges of DNA and the acquisition of resistant gene sequences are considered. For the parameters used, resistant strains rapidly proliferate and dominate, although initial intense antibiotic treatment can occasionally prevent this. Variation in genome distribution appears to be long tailed. If this is reflected in nature, the occurrence of resistant bacterial strains can be expected, as well as considerable variation in patient outcomes.

1. Introduction

Bacteria are becoming immune to antibiotics. This is in the newspapers,[1] although it is hardly news, except for the details. There are mathematical models of ways to administer antibiotics (e.g., Obolski and Hadany, 2012; Sotto and Lavigne, 2012; Nguyen et al., 2014; Ternent et al., 2015) but it is clear that if humanity is to have control over infectious disease a better understanding of bacterial evolution is needed. The most troubling aspect has been horizontal gene transfer (e.g., Francino, 2012; Barlow et al., 2012 ; Berglund and Frank, 2012) in which bacteria not only mutate but pass on entire genes (or more) to other bacteria. In particular they can pass on the resistance to particular antibiotics. (The literature is enormous, and the citations listed are only a small fraction of those that address these problems.)

In general an antibiotic operates by disabling some vital process or essential component of a bacterium. Resistance corresponds to finding an alternative process or component. There are several mechanisms by which such alternative routes can be found. The simplest (conceptually) is by mutation, including the presence of a mutant form already in the population and selected by application of the antibiotic. Other methods involve swapping or adding. One way this can happen is that a piece of DNA from another bacterium may be acquired and provide the immunity. Alternatively, the piece of DNA need not come from a living bacterium, but is rather obtained from whatever medium the bacterium finds itself in. The latter processes (i.e., other than mutation) are known as *horizontal gene transfer*, abbreviated HGT.

In the present article we model the evolutionary processes behind a bacterium's acquisition of resistance, both mutation and horizontal gene transfer. The model—which we adapt to our present purposes—is known as the "tangled nature model" (TNM) and has been used in studies of evolution and ecology (Christensen et al., 2002; di Collobiano et al., 2003; Jensen, 2004; Anderson and Jensen, 2005; Laird and Jensen, 2006; Lawson et al., 2006; Lawson and Jensen, 2006; Jensen and Arcaute, 2010; Becker and Sibani, 2014; Sevim and Rikvold, 2005; Canko et al., 2015; Wosniack et al., 2017). In Section 2 we give an introduction to the TNM; following that (Section 3) we describe our adaptations, allowing it to model the development of antibody resistance. For this purpose we have introduced a number of innovations, the need for some of which has also come up in previous studies (e.g., Andersen and Sibani, 2016). In particular we introduce the following: a) correlation of genotype and phenotype, expressed here as the correlation of the preference matrix (to be called J) and the hamming distance between genomes; b) enrichment of an environment for the support of particular genomes; c) reduction in fertility of a replacement genome. All these reflect biological realities. It will be seen that in general our model does not yield one of the positive features of the TNM, namely the emergence of punctuated equilibrium. This is irrelevant in the present study. Our goal is not to model global evolution, but rather a (literally) pathological situation. And finally Section 3 also describes our specific steps in the modeling of horizontal gene transfer. As will be seen, we do not model changes in genome size. Once our procedures are described we are able to present, in Section 4, the results of our investigations. The last Section 5 is a discussion of strengths and weaknesses of our modeling.

2. The tangled nature model (TNM)

The evolution is an evolution of genomes. For each time step,

E-mail address: schulman@clarkson.edu.

[1] See. e.g., http://www.nytimes.com/2016/05/27/health/infection-raises-specter-of-superbugs-resistant-to-all-antibiotics.html?_r=0 or http://www.nytimes.com/2016/09/22/health/united-nations-drug-resistant-superbugs-antibiotics.html.

http://dx.doi.org/10.1016/j.jtbi.2017.01.022
Received 30 May 2016; Received in revised form 2 January 2017; Accepted 16 January 2017
Available online 17 January 2017

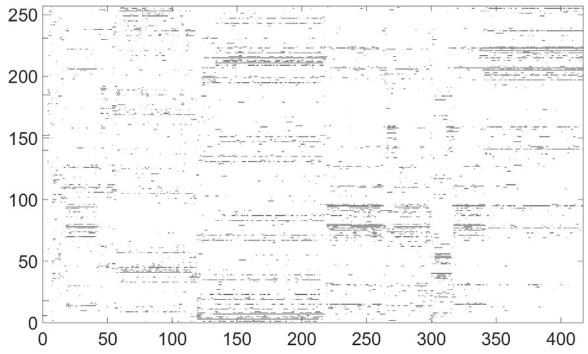

Fig. 1. The figure shows which genomes have at least one individual alive at the indicated time. Time, the horizontal axis, is measured in generations, while the vertical axis is the number, k_α, associated with genome-α. The parameters for this run are L=8, $p_{\text{kill}} = 0.097$, $p_{\text{mutate}} = 0.552/L$, $\mu = 0.22$, c=0.0166. One generation, defined as, $\langle N\rangle/p_{\text{kill}}$, is 479 time steps.

individuals possessing a particular genome can reproduce, mutate or die. Those genomes that are more successful have large populations (many individuals with that genome), those that are less successful have fewer or zero individuals possessing that genome. The determinant of success or failure, and the principal factor deciding whether or not reproduction takes place, is the preference matrix, J, which measures the mutual compatibility of an existing population. It is fixed for a given simulation. Here are the details.

A *genome* of length L is a string of L zeros and ones. For any given genome (α) and at time-t, there is a certain number of individuals alive who posses that genome, call this number $n_\alpha(t)$. The *state* of the system is described by the set of integers, $\{n_\alpha(t)\}$. The genome α is assigned the integer $k_\alpha \equiv 1 + \sum_j \alpha_j 2^{L-j}$, so that $1 \le k_\alpha \le 2^L$ (where α_j is the j^{th} component of α). The total number alive at time-t is $N(t) \equiv \sum_\alpha n_\alpha(t)$.

The procedure for going forward one time step varies slightly in the literature. We adopt the following rules for asexual evolution.[2]

First, a living *individual* is randomly chosen (so if $n_\beta > n_\gamma$, it is more likely that an individual with genome β is chosen). Suppose the genome of the chosen individual is α. The first test to be applied is whether or not to kill this individual. This is done with probability p_{kill}. If the individual is killed then $n_\alpha(t+1) = n_\alpha(t) - 1$, and we move to the next time step. If not, the individual is given the opportunity to reproduce. This is where additional features of the model enter.

Governing the model is a matrix J that can be interpreted in various ways. It measures the compatibility of pairs of genomes, their preferences. We will take $-1 \le J_{\alpha\beta} \le 1$ where $J_{\alpha\beta} > 0$ implies that α is more likely to reproduce if β is well populated. (There is a slight abuse of terminology here: to say that α does something means that an individual with genome α does it.) One might suppose that if α eats β then $J_{\alpha\beta} > 0$, while $J_{\beta\alpha} < 0$. This might or might not be true. It could happen that α's predation has in fact a positive effect on β in (say) culling herds and effectively improving the fitness of those that remain. Thus J does not necessarily reflect the trophic food chain.

The matrix J is produced in the following way.[3] Two sets of random numbers are produced, f_α and g_α, uniformly distributed on the interval $[-1, +1]$. (Note that, e.g., f_α can also be written as f_{k_α} with k_α related to α as described above. Both notations are used.) Next, only a (randomly selected) fraction θ of the f_α are kept; the remainder are set equal to zero. The matrix $J_{\alpha\beta}$ is defined as

$$J_{\alpha\beta} \equiv f_\alpha g_{\alpha\oplus\beta}, \tag{1}$$

where "$\oplus$" is the exclusive or, XOR, and α and β are length-L strings of zeros and ones. Once determined, J is fixed for the entire simulation.

For the decision of whether or not to reproduce two more parameters are needed, μ and c; μ is a kind of chemical potential and c is effectively a normalization for J. Using them, as well as the state of the system ($\{n_\gamma(t)\}$), one calculates

$$H_\alpha \equiv \frac{1}{cN}\sum_\beta J_{\alpha\beta} n_\beta - \mu N. \tag{2}$$

Now finally we can state the reproduction decision: Given that an individual with genome α has not been killed, it reproduces with probability

$$p_{\text{reproduce}} = \frac{1}{1+\exp(-H_\alpha)} \tag{3}$$

At this point another possibility enters. The new individual is allowed to be slightly different. There is a chance that an offspring can differ from its progenitor. With probability p_{mutate} each 0 or 1 can switch. (It is reasonable to take $p_{\text{mutate}} = \text{const}/L$ so that the overall probability of mutation is roughly the same as the size of the genome changes.)

Those are the rules for asexual reproduction in the TNM. With the right parameter values one obtains punctuated equilibrium as *emergent* with no need for significant external events, like the asteroid that supposedly killed the dinosaurs. Many other studies have been based on this model, including the introduction of additional features, such as spatial dependence. Typical results of using this model are shown in Fig. 1, although in some of the cited articles much longer runs as well as larger L values are used.

2.1. Elephants do not become giraffes: the correlation of phenotype and genotype

This model will ultimately be adapted to study horizontal gene transfer, but first we consider a modification that is also suitable for the original TNM model. Adaptations of this sort have been considered elsewhere (Andersen and Sibani, 2016), but we use a different method.

The problem addressed is that the rule for mutation given above takes no account of phenotype. The latter is implicit in the preference (J) matrix, but mutations (i.e., exchanges $0 \leftrightarrow 1$) bear no a priori relation to J. What is wanted is a correlation between the phenotype and possible mutations, whose likelihood is determined by the hamming distance between genomes.[4]

To create this correlation we employed a random process. The idea is to pick two pairs. The pair with the greater hamming distance is given the smaller (algebraically) J-value. Specifically, define $D(k_\alpha, k_\beta)$ to be the hamming distance between sequences α and β. Now pick 4 random integers between 1 and 2^L (say $n_1, \ldots, n_4$) and examine $D(n_1, n_2)$, $D(n_3, n_4)$, $J(n_1, n_2)$ and $J(n_3, n_4)$. If $D(n_1, n_2) < D(n_3, n_4)$, and $J(n_1, n_2) < J(n_3, n_4)$ then there would be improved correlation if the J values were switched, that is $J(n_1, n_2) \leftrightarrow J(n_3, n_4)$. In words, satisfying the inequality in D implies that n_1 and n_2 are more nearly alike than n_3 and n_4. Therefore $J(n_1, n_2)$ should be larger than $J(n_3, n_4)$. By doing this random process for a reasonable number of times a correlation can be established. It was found that for $M \times 2^L$ attempts, with $M \sim 1000$, one could obtain good results. For large L this procedure could be done on the fly as new matrix elements of J are encountered, although for the small values of L that we used this was not a problem. This process may change the number of zero elements in J, but insignificantly.

3. Adapting the model to study antibody resistance

All the genomes represent bacteria and those with a certain signature (subsequence) are the ones killed by the antibiotic. Not all

[2] Regarding many details we differ from the formulations in other articles. Notation can differ (our α is often called a), rules for constructing J can differ, as can the rules for mutation. However, the qualitative results are unchanged

[3] Our method is closest to that in Hall et al. (2002).

[4] The hamming distance between two sequences of zeros and ones is the minimum number of changes to bring one sequence to be equal to the other.