

Results and discussion

Steven Court

September 2, 2016

Abstract

My work over the last few months has been to simulate and attempt to reproduce Bartek's experimental set-up in order to understand the most important features. This fell into 2 parts: the first was to investigate the evolutionary dynamics of the system, the second to investigate the travelling waves observed in the experiments. Here I summarise the main things that we have learned and reference files from the directory "Working notes" so as to keep this as brief as possible and not duplicate everything. For notes on the code and data analysis scripts, see accompanying document NotesOnCode.

1 Evolutionary dynamics

1.1 Introduction

Bartek's original PRL work that inspired these experiments was quite abstract: a pure statistical physics model that looked at times for resistance to emerge in flat vs exponential antibiotic gradients, with or without fitness valleys, with the model being designed such that resistance always emerged. The experimental set-up he has developed may not be best for investigating times to resistance emergence, but is designed for determining the likelihood of resistance emerging and with earlier wells basically "frozen in time", one could in principle construct the evolutionary pathway.

The initial aim of our simulations is to simulate Bartek's experimental set-up exactly. They are individual-based and use experimentally determined MICs and fitnesses (Marcusson2009 []) for the specific choice of ciprofloxacin and E. coli. With this we can test, in a less abstract manner than other works, the dynamics of antibiotic resistance in a heterogeneous environment. Is it possible to test any of Bartek's PRL predictions with this? Does the Marcusson2009 data have any fitness valley for instance?

In my mind, some important questions are:

- (1) Is the evolutionary trajectory for a given environment reproducible, or are there many equally probable routes?
- (2) How strongly do these routes depend on the environment? (This is important but is ignored entirely by the evolutionary-steering paper, Weinreich and others).
- (3) What are the probabilities of resistance emerging, i.e. probability of reaching some genotype or the final well.
- (4) How often do we see sub-global-optimal solutions being found (i.e. in the simulations with a maximum ciprofloxacin concentration of 20,000 ng/ml, the globally most resistant and fit genotype, 11100, is the 2nd most probable to emerge ??).
- (5) (Future work) How does this all change when we implement a set-up more akin to a tissue (24 wells) already occupied by pathogen (at carrying capacity) and we apply an antibiotic at one end and allow it to diffuse through the wells. I expect very different dynamics from the original case in which mutants colonise empty space: it would only kill everything once the MIC was reached, but in the time it takes for the antibiotic to build-up to this level it would be creating the space and selective pressure for evolution to occur with mutants likely to arise at various points in experiment and spread in both directions.

Other works have used the strong selection weak mutation (SSWM) model to look at "steering evolution" so as to minimise the chance of evolving to the most resistant state, and for calculating likelihood of evolutionary trajectories (e.g. Weinreich finds only few pathways possible with this very unsatisfactory model []). Experimentally, Kishnoy has used a morbidostat [] to constantly challenge a population and sees that evolution appears to be quite predictable (but if you're in such tight control of the environment, maybe this isn't so surprising). Bartek's work is a far more realistic and interesting set-up to investigate the reproducibility of evolution.

1.2 Summary of results

After an initial period of calibration and testing (see below) I modified Bartek’s code to include the mutations and strains studied in Marcussen2009 [1] along with their fitnesses (relative to wild type in the absence of antibiotics) and MIC values. I also included a ciprofloxacin-dependent mutation rate that roughly matched the increase that is observed experimentally (see SimNotes-3.pdf, Section 1.1).

The first observation is that for a given ciprofloxacin gradient, there can be a large number of possible evolutionary trajectories and end-points, but that a small subset of these trajectories are predominantly followed. For instance, studying the evolutionary trajectories of all cells present in the final well of 200 experiments at xxx ng/ml shows that there are more than 100 realised trajectories, but that only xx of them account for the trajectories followed by xx% of all cells (see SimNotes-7 figure 2 and XX). Altering this profile (different max value or shape) can produce significant changes in the outcome and lead to different genotypes being more likely to emerge. For instance, including a “plateau” of some number of final wells at the highest concentration encourages further selection for the fittest strain (XX).

One large discrepancy however that existed between the simulations and experiments was that in terms of the likelihood of evolving resistance, the simulations were far more likely to see resistance emerging than the experiments. We convinced ourselves that this was not due to bad parameter choices or mutation rates, but must reflect something in the experiments that wasn’t being accounted for – most likely the effect of filamentation [2]. To investigate this the code was modified to include an MIC-dependent cutoff for motility: if the antibiotic concentration in a well is larger than some fraction of the cell’s MIC, the cell is immotile and although it may continue to grow (reproduce) as an immotile filament, it cannot produce mutant offspring. The effect is to greatly reduce the chance of reaching the final well, as well as reducing the diversity of evolutionary trajectories that are viable. If we assume this mechanism is biologically realistic, looking at the data for exponential profiles with maximum concentrations of 300 and 1000 ng/ml and comparing this with Bartek’s experiments suggests that the motility cutoff is somewhere between 0.1 and 0.2 of a strain’s MIC. (See SimNotes-10, Section 3 and SimNotes-11).

1.3 Experimental information still lacking

There is still a lot that is unknown about the experimental system. Some of these may be unimportant, while others may need to be known if we are to simulate the system as accurately as possible. It may be possible to estimate some of these through our monte carlo optimisation procedure.

- We do not know how ciprofloxacin affects motility, and if this is different for different strains.
- Do not know the rate of occurrence and full significance of filamentation in the experiments.
- We do not know exactly how ciprofloxacin affects mutation rate, and again if this is the same for all strains.
- Exactly how much more likely are the marR and acrR mutations in comparison to the others?

1.4 Informing experiments

When the experiments and DNA sequencing are up and running, these simulations could be used to inform them:

- For various ciprofloxacin profiles, we have seen that the simulations can indicate what is the earliest well worth sampling from. Shallow gradients for example will accumulate less mutations in earlier wells.
- One could also design various alternative concentration profiles for a given maximum concentration to see which facilitate or hinder evolution.

2 Travelling waves

2.1 Introduction

Bartek noticed that in some of his experiments there were very well-defined travelling waves. Although a huge amount of theoretical work exists on the properties of these only a surprisingly small number of studies have tried to produce these waves experimentally. Those that we are aware of [] are not particularly interesting with the exception of the recent study [] in which transition from a “pulled” to “pushed” wave occurs with the inclusion of the Allee effect. However it does not look at the wave profiles which is something that we wished to do.

2.2 Summary

We found that using the correct growth curve $g(n)$ (obtained experimentally from a single-well experiment with non-mixed LB media) it is possible to fairly accurately predict the shape of the travelling wave seen in Bartek’s experiment. We can do this by using the full individual-based, stochastic simulation with correct well and channel dimensions and diffusion coefficient of *E. coli* but, perhaps more interestingly, we can achieve the same result by simply solving the set of ODEs corresponding to the 1-dimensional discrete Fisher equation (which ignores well and channel dimensions, in-well diffusion and does not model bacteria individually). See SimNotes-8 section 3 and Friday_slides.pdf

We also proposed a simple mapping that says, ignoring diffusion (which seems to be fine to do in the case of non-mixed LB), we can predict the shape of the travelling wave simply by knowing the growth curve $N(t)$ of a single well – we just need to rescale time by $(-v)$ where v is the speed of the wave. Thus it appears that horizontally stretching or compressing the growth curve in this way, and flipping it about the y-axis can provide an approximation to the shape of the travelling wave. See SimNotes-8 section 4 and Friday_slides.pdf

Finite and discrete system: We used the simulations to test if we might be able to experimentally validate the theoretically predicted change in speed due to finite size effects (carrying capacity) and found that in principle this might be possible. The wavespeed varied by more than 70% as we increased the carrying capacity of the wells by 4 orders of magnitude (SimNotes-3, Fig. 4). We also found that the speed of the wave with the discrete-system prediction does not exactly match that found in the simulations, probably due to the form of $g(n)$ in the non-mixed case falling very rapidly in comparison to logistic growth, meaning that the validity of the linearisation in which we only need $g(0)$ breaks down. Indeed, using the discrete system prediction the wave speed no longer depends on \sqrt{D} which is the case in the continuous model (SimNotes-8, section 5).

To test the growth curve–wave profile mapping, Bartek is currently repeating the experiments with different growth media, as well as mixtures of media so as to alter the shape of the growth curve via some diauxic shift. He has discovered some odd phenomenon such as (i) significant variability across experiments, (ii) pellicle formation in some but not all experiments (possibly due to oxygen), (iii) the growth curve mapping seems to fail entirely for MOPS medium and (iv) odd but reproducible kinks in his growth curves; the single-well growth curves taken from the full experiment do not exactly match that observed in the isolated single-well experiments.

3 Monte Carlo parameter estimation using tau-leaping algorithm

In order to determine unknown model parameters we thought it may be possible to do some sort of Monte Carlo parameter estimation: we would give the simulation an initial set of parameters and compare the outcome with some experimental observations, defining a similarity distance between them. Altering the parameters and re-running the simulation a new distance measure would be calculated and the new parameter set would be accepted if the distance was decreased and be rejected (or accepted with some probability p) if the distance had increased. To do this, we need the simulations to be far more efficient than the individual-based model and so I re-wrote the program, approximating the system by using a simple tau-leaping algorithm.

As a proof of principle, I used as the parameter set an array of diffusion constants, each corresponding to individual channels, and I generated an “experimental” well population data set using the deterministic set of ODEs governing the discrete Fisher system with all diffusion parameters equal. As a similarity index I calculate the square of the differences between the populations of each well at a given time point, and sum this for many time points throughout the simulation.

Allowing the procedure to run raised a number of issues. The main issue is that for this similarity distance measurement, and the choice of parameters (23 diffusion constants), parameter sets that are very different from the true (known) values can still produce small distances and the procedure may get stuck at one of these many points. Another issue is that there is a large range in possible distances calculated using the same parameter set purely due to the stochastic nature of the simulation. The ranges from different parameter sets can overlap substantially with that of the true set meaning we should ideally be performing averages over many runs for each parameter set. SimNotes-10 section 2 shows this. Figure 2 shows how large the stochastic spread in distances can be for a given parameter set and figure 3 demonstrates exactly how this can arise.

It may be possible to correct for these problems but I did not spend time doing so since it would be a lot of work and ultimately we are not interested in fitting diffusion parameters. I don’t think this idea is unfeasible however, just that we have looked at a complicated test case with too many parameters. I think if we know exactly what features we are trying to understand this could be useful. For instance, if we have good experimental data on the likelihood of finding specific mutated genes in given wells, and have some idea of what model parameters we know least well (perhaps how motility, or the mutation rate changes with ciprofloxacin) we could define a better similarity measurement and use a much smaller parameter set.

4 Other findings / calibration

We spent some time calibrating the well and channel dimensions as well as the diffusion constant so that the wave speeds matched those found in the experiments and the diffusion matched Bartek's methyl blue experiments. A well width and depth of 3.25 mm and 11.0 mm, and a channel width and depth of 1.5 mm and 3.75 mm along with the literature value of the diffusion constant, $D = 1.4\text{--}1.5 \text{ mm}^2/\text{h}$ was found to match experiments.

We looked at the distribution of cells in each well and found that statistically cells do reside nearer top of wells since they take some time to diffuse the full depth, but not to the extent observed in Bartek's experiments, which now appears to be due to the formation of pelicles (possible due to presence of oxygen).

We attempted to reproduce the growth curves of the non-mixed LB single-well experiment but using the well-mixed data by explicitly modelling nutrients. This did not however significantly alter the growth curves (see SimNotes-9.pdf) and further experimentation revealed that Bartek's growth curves are very mysterious!

We looked at the distribution of distances travelled by cells and found that actually the majority of cells are expected to be found outside of their well of birth. After 3 days around 40% have travelled no wells, 35% of cells have moved 1 well away from their well of birth, and a further 8% have moved 2 wells (see SimNotes-10).

Before I implemented the realistic strain data I saw that increasing the ciprofloxacin gradient did not necessarily result in a drop in the likelihood of reaching the final well as I had expected. This was due to the ciprofloxacin increasing the mutation rate. That is, it may not always be obvious or easy to separate the two possible features that (1) a stronger selection gradient may be increasing the rate of evolution in some particular system, as stated in Bartek's and Bob Austin's papers, or (2) the fact that the causative force of this selection gradient is actually explicitly increasing the rate of mutation (and thus evolution).