

Critical Mutation Rate in a Population with Horizontal Gene Transfer

Elizabeth Aston¹, Alastair Channon¹, Roman V. Belavkin²,
Danna Gifford³, Rok Krašovec³ and Christopher G. Knight³

¹ School of Computing and Mathematics, Keele University, ST5 5BG, UK

² Middlesex University, London, NW4 4BT, UK

³ Faculty of Life Sciences, The University of Manchester, M13 9PT, UK

{e.j.aston, a.d.channon}@keele.ac.uk, r.belavkin@mdx.ac.uk, {danna.gifford, rok.krasovec, chris.knight}@manchester.ac.uk

Abstract

Horizontal gene transfer (HGT) enables segments of DNA to be transferred between individuals in a population in addition to from parent to child. It is a prominent process in bacterial reproduction. Existing *in silico* models have succeeded in predicting when HGT will occur in evolving bacterial populations, and have utilised the concept of HGT in evolutionary algorithms. Here we present a genetic algorithm designed to model the process of bacterial evolution in a fitness landscape in which individuals with greater mutational robustness can outcompete those with higher fitness when a critical mutation rate (CMR) is exceeded. We show that the CMR has an exponential dependence on population size and can be lowered by HGT in both clonal and non-clonal populations. A population reproducing clonally has a higher CMR than one in which individuals undergo crossover. Allowing HGT only from donors with a non-zero fitness prevents HGT from lowering the CMR. In all cases the change in CMR with population size is greater for populations with 100 individuals or less. This represents a significant stage in bacterial evolution; smaller populations will exist when a population is founded or near to extinction. This will also be the case if a subset of the population is considered as a population in its own right, for example, the sub population of resistant bacteria that emerges due to the introduction of antibiotic resistance genes. Understanding the effect of mutation at such a critical stage is key to predicting the likely fate of a population.

Introduction

It is estimated that up to 32% of the bacterial genome has been acquired by horizontal gene transfer (HGT) (Koonin et al., 2001), during which segments of genetic material are transferred between individuals as opposed to inherited from a parent (vertical gene transfer) (Mozhayskiy and Tagkopoulos, 2012). It was first identified in 1928 when it was noted that virulence could be passed horizontally from virulent to non-virulent strains of *Streptococcus pneumoniae* via transformation (Griffith, 1928; Ravenhall et al., 2015). HGT has also been shown to occur via transduction and conjugation (Tatum and Lederberg, 1947; Zinder and Lederberg, 1952; Jones and Sneath, 1970), increasing genetic diversity in bacteria above the level that would be accessible via mutation

alone (Nielsen et al., 2014). The majority of HGT events lead only to minimal changes in recipient fitness, an exception being antibiotic resistance; the presence of antibiotics leads to strong positive selection for resistance genes (Nielsen et al., 2014). Advances in whole genome sequencing have enabled the amount of genetic material acquired via HGT to be quantified; Ochman et al. (2000) review a distribution of estimates of between 0 and 12.8% of the genome designated as horizontally acquired in a variety of species, while Cooper (2007) gives an estimate of 1×10^{-4} genes transferred horizontally per cell per generation in *Escherichia coli*.

In a similar manner to mutation, HGT occurs at low frequencies and depends on conferred fitness gains to persist (Kimura and Ohta, 1969; Johnsen et al., 2011). However, the rate HGT happens can be orders of magnitude higher than per gene point mutations (Puigbò et al., 2014; van Dijk and Hogeweg, 2015). For example, mutation rates in wild type *E. coli* are estimated to be between 2×10^{-8} and 2×10^{-10} per generation, per base pair (Drake et al., 1998; Mozhayskiy and Tagkopoulos, 2011) and 3×10^{-7} per gene (Mozhayskiy and Tagkopoulos, 2011, 2012). Experimentally observed HGT rates between bacteria in natural environments vary between 10^{-7} and 10^{-11} per generation, per cell (Baur et al., 1996; Mozhayskiy and Tagkopoulos, 2012) and in some cases the rate can go as high as 10^{-3} to 10^{-1} (McDaniel et al., 2010; Mozhayskiy and Tagkopoulos, 2012). Bacteria can increase their mutation rate in response to stress (Krašovec et al., 2014); the fact that mismatch-repair-deficient cells also have significantly increased recombination and HGT rates suggests there is a strong correlation between HGT, stress, and adaptation (Raz and Tannenbaum, 2010).

Evolving Prokaryotic Populations: Modelling Horizontal Gene Transfer

The gene content of microbes can change in a relatively short amount of time (van Dijk and Hogeweg, 2015). *In silico* evolution enables each parameter to be tested independently or in combination as a means of understanding the un-

derlying mechanisms of evolution dependent on parameters such as population size, generation time, HGT frequency, and change in relative fitness of recipient cells (Hindr e et al., 2012). For example, modelling has facilitated prediction of HGT detectability in bacterial populations growing both *in vitro* and *in vivo* (Nielsen et al., 2014).

Microbial GA was one of the first genetic algorithms (GAs) to be inspired by bacterial evolution (Harvey, 1996, 2011; Tomko et al., 2013). During the reproduction step, if two individuals picked at random to be parents are called A and B, whilst the offspring is called C, then C will replace the weaker one of the parents. If C is the product of sexual recombination between A and B, however, then ~50% of C's genetic material is from A, and ~50% from B. This is indistinguishable from B remaining in the population, but with ~50% of its original genetic material replaced by material copied and passed over from A; it can be considered as an excessive case of HGT from A (the fitter) to B (the weaker). The Pseudo-Bacterial GA (PBGA) (Nawa et al., 1997) and the Bacterial Evolutionary Algorithm (Nawa and Furuhashi, 1998) both use a genetic operator called the 'Bacterial Operator' (Tomko et al., 2013). This operator mimics gene transduction, the process by which bacteria can horizontally transmit parts of their genome via a viral vector. In the Unconstrained Genetic Algorithm (Tomko et al., 2013), HGT is done by the process of individuals swapping genes as opposed to one gene overwriting another, ensuring the frequency of genes remains constant. Within the reproduction step, the genotypes of each offspring are broken up and put into separate locus categories known as "bags". From these pieces, a new population is constructed by randomly choosing genes from each bag to emulate the process of HGT. While *in silico* modelling of horizontal information transfer has provided insight into the importance of HGT as an evolutionary process, the role of HGT in conjunction with mutation-selection balance (Kimura and Maruyama, 1966; Bull et al., 2005), remains unclear.

Critical Mutation Rate: Horizontal Gene Transfer and Mutation-Selection Balance

There is evolutionary pressure to evolve sequences that are both fit and robust to mutation (Jones and Sneath, 1970; Lenski et al., 2006; Sanju an et al., 2007). At high mutation rates, individuals with greater mutational robustness can outcompete those with higher fitness, a concept known as "survival-of-the-flattest" (Wilke, 2005); in terms of fitness landscapes, narrow, high fitness peaks may be lost, while the population remains on broader, lower peaks. At the optimal mutation rate there will be survival-of-the-fittest, while the CMR is the lowest mutation rate at which fitter individuals are outcompeted by those with greater mutational robustness. The higher the CMR, the greater the robustness of the population as a whole. The CMR has been shown to have an exponential dependence on population size in both hap-

loid (Channon et al., 2011) and diploid *in silico* populations (Aston et al., 2013). This exponential relationship is maintained when the parameters of a genetic algorithm modelled on the biological process of meiosis are set to ranges appropriate for eukaryotic organisms (Aston et al., 2016). The exponential relationship was identified by curve-fitting using R with a least squares method; it was found that the data for population size and CMR could be approximated by an exponential function, where $y = A - B * e^{-((m/C)^D)}$ (with m being population size). Furthermore, increasing the number of genes and the length of each gene in each individual's genome to the appropriate range for the model plant *Arabidopsis thaliana* lead to CMRs within one order of magnitude of existing measurements of biological mutation rate; it has been demonstrated that, in a system in which an individual's fitness is dependent on the minimum fitness of its n constituent genes, it is possible to input biologically realistic parameter values for a specific eukaryotic organism into the simulation model and produce a CMR within the range of current biological estimates of mutation rate for that organism (Aston et al., 2016). While the relationship between population size and CMR has been established in eukaryotes, its applicability to prokaryotes undergoing HGT remains an open question.

Hypothesis 1

The exponential relationship between CMR and population size means that much lower mutation rates are tolerated by small populations. While bacteria are typically associated with large population sizes, they will exist in much smaller populations when the population is founded, near to extinction, or if a subset of the population is considered as a population in its own right. For example, the introduction of antibiotic resistance genes into a population via HGT will lead to a sub population of resistant bacteria. We hypothesise that the exponential relationship previously observed for eukaryotes will also be observable for populations subject to both vertical and horizontal gene transfer. Ochoa and Harvey (1998) observed that recombination can push the population in a genetic algorithm over the error threshold when the mutation rate is high, inducing error catastrophe; it is expected that increasing the amount of donor sequence transferred to a recipient individual will also lead to a reduction in the CMR.

Hypothesis 2

Experimentally observed HGT rates between bacteria in natural environments vary between 10^{-7} and 10^{-11} per generation, per cell (Baur et al., 1996; Mozhayskiy and Tagkopoulos, 2012); in some cases the rate can go as high as 10^{-3} to 10^{-1} (McDaniel et al., 2010; Mozhayskiy and Tagkopoulos, 2012). The rate of HGT will determine how much exchange of genetic material occurs within one generation; it may be possible for one bacterial cell to receive genetic ma-

terial from more than one donor. In terms of a genetic algorithm with a constant sequence length, it is possible that the genetic material from one donor cell may overwrite the genetic material from another donor cell. Transfer of large segments of sequence may increase the chance of an overlap in donor sequences overwriting the recipient's sequence. We therefore hypothesise that increasing the rate of HGT by increasing the number of donor individuals that transfer part of their sequence via HGT to one recipient per reproduction will have a smaller effect on the CMR than increasing the length of the sequence transferred; it is expected that the increased possibility of donor sequences overwriting previous donor sequences will counteract the effect of increasing sequence length. It should be noted that over one complete generation an individual may be selected as a donor more than once and may therefore transfer its genetic material to multiple recipients (see Simulation Model).

Hypothesis 3

Bacteria reproduce clonally rather than exchange their genetic material via recombination during reproduction (Tatum and Lederberg, 1947). Ochoa and Harvey (1998)'s observation that recombination can force the population over the error threshold at high mutation rates suggests that removing the recombination step and allowing the population to reproduce clonally will have the opposite effect. We therefore hypothesise that removing recombination will lead to an increase in the CMR. The introduction of clonality is not expected to affect the influence of population size on the CMR therefore we hypothesise that the exponential relationship will be maintained.

Hypothesis 4

Donor individuals are either the result of crossover between the two fittest individuals (of the three chosen during the reproduction step) or clonal reproduction of the fittest individual (see Simulation Model). Consistent with the model used in previous studies (Channon et al., 2011; Aston et al., 2013, 2016), the simulation model does not allow individual death; individuals can have a fitness of zero but will still be included in the reproduction step. This means it is possible for an individual with zero fitness to be a HGT donor, with the potential to reduce the fitness of a recipient on one of the peaks. To determine the effect of this on the CMR, the donor individual must come from the subset of the population on the peaks. We hypothesise that only allowing donors to have non-zero fitness will increase the CMR; this is expected to be most significant for longer sequences. The exponential relationship between population size and CMR is expected to be maintained.

Simulation Model

The model used a two-peak fitness landscape (Figure 1), with the height of peak 0 constant at 15 and the radius 2,

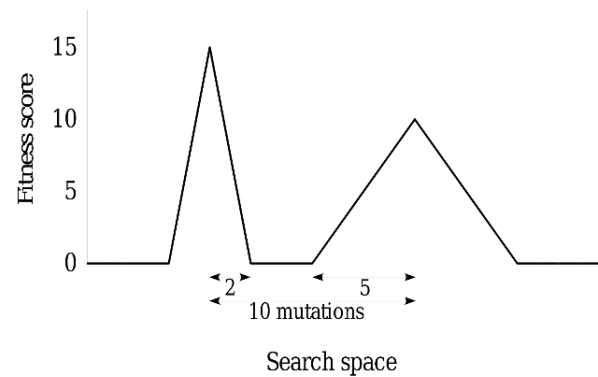


Figure 1: **Two-peak fitness landscape**, with one narrow peak of high fitness (Peak 0), and one broader peak of lower fitness (Peak 1). The fitness score is relative, and the width and distance between the peaks are given in terms of Hamming distance. Diagram adapted from Wilke (2005).

the height of peak 1 constant at 10 and the radius 5, and the Hamming distance between the peaks set at 10 as per previous studies (Channon et al., 2011; Aston et al., 2013, 2016). It should be noted that the model uses relative fitness scores therefore the absolute value of the height of the peaks is not important. Each individual was represented as a single (haploid) sequence of alphabet size 4 (representing the bases of DNA, i.e., A, C, G, and T) and length L ; the sequence space is therefore 4^L . L was kept constant at 1000 bases (b). For simplicity, peak 0 was set to be all 0s and peak 1 was randomly generated to be Hamming distance 10 away. Hamming distance is calculated by simply comparing one sequence with another, base by base. If the bases do not match, the Hamming distance is incremented. Fitness values were determined based on the Hamming distance of each individual from the top of each peak. The fitness value of an individual relative to peak 0 was compared with its fitness value relative to peak 1 and the highest of these selected to give the individual an overall fitness value. The simulation was initialised with half of the population on each peak and run for a range of population sizes. The CMR was recorded as the mutation rate at which 95% of 2000 runs lost peak 0 within 10,000 generations (as per (Channon et al., 2011; Aston et al., 2013, 2016)). The CMR was calculated for each 100 runs to produce 20 CMR values from which the mean was calculated. Selection was based on fitness-dependent tournament selection between selected parents. Three individuals were selected at random from the population. When reproduction was non-clonal, two of the three selected individuals were chosen as parents in a single point crossover which replaced the third individual with the resulting child. In the case of clonal reproduction, the child was overwritten entirely by the fittest parent. The individual to be replaced was determined based on the fitnesses of the three selected individuals: there was an equally small chance of

either of the two fittest of the three being replaced (25%), and a larger chance of replacing the least fit. This ratio ensures that there is potential for any individual to be chosen for replacement, allowing loss of the fittest peak. The use of tournament selection ensures that selection is independent of the precise shape of the landscape; it does not matter exactly how much fitter one individual is compared to another, tournament selection depends on the relative fitness. The child was then allowed to receive a portion of a sequence from a donor individual (HGT). The amount transferred was kept constant during each experiment and set to a range of amounts from 1 b quadrupling up to 256 b (0.1 - 25.6% of the sequence). The section of the donor sequence to be transferred was selected at random. The child sequence (the recipient) was then overwritten with the section of donor sequence. This ensured sequences remained fixed in length consistent with existing work (Channon et al., 2011; Aston et al., 2013, 2016). Fixed sequence lengths also ensure that one sequence is directly comparable to another in terms of the simulation and fitness landscape defined in the paper. The donor was either chosen randomly from the population, or was selected from a subset of the population consisting of individuals with non-zero fitness. The child was then subject to one round of point mutation at a given per-base mutation rate. These steps were repeated until each individual in the population had been chosen exactly once to undergo reproduction (or there were less than three remaining to select); this represents one generation. The rate of HGT was kept constant (unless stated otherwise), with each individual in the population subject to tournament selection with the possibility of being a HGT recipient once per generation. To test the effect of varying the rate of HGT, the amount of the sequence to be transferred was either kept constant at either 1 b (0.1%) or 16 b (1.6%) with the number of donor individuals per recipient set to equal 4, 16, or 32 (to allow for the possibility of donor sequence to overwrite up to just over half of the recipient sequence (32 transfers of 16 b)). Over one complete generation an individual may be selected as a donor more than once and may therefore transfer its genetic material to multiple recipients.

Results

Horizontal Gene Transfer Reduces CMR in a Population with Crossover

Figure 2 shows the relationship between population size and CMR for a haploid *in silico* population both in the presence and absence of HGT. The amount of genetic material transferred from donor to recipient was kept constant for the duration of each run and set to equal from 1 b quadrupling up to 256 b. Individuals were 1000 b long therefore this represents between 0.1% and 25.6% of the sequence transferred to a recipient. It is apparent from Figure 2 that enabling HGT does not affect the CMR if the size of the sequence transferred is small (16 b or less). However, in agreement with Hypoth-

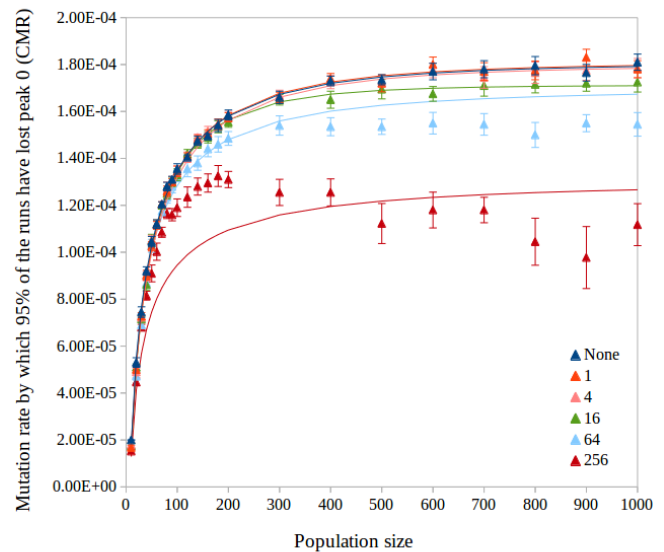


Figure 2: **CMR plotted against population size for varying size of donor sequence transferred via HGT.** Donor individuals transferred between 1 b and 256 b of genetic material (given in the legend) per reproduction. The exponential lines were obtained by curve-fitting using R with a least squares method. Error bars represent +/- 95% confidence intervals.

esis 1, transfer of longer segments of sequence lowers the CMR. The exponential relationship is maintained, although the relationship between CMR and population size is less consistent for populations with approximately 300 individuals and above when the size of the sequence transferred reaches 25.6% of the donor sequence. Two factor ANOVA indicates that the null hypothesis, that there is no significant effect of population size and the amount of genetic material transferred on the CMR, can be rejected ($p=1.54 \times 10^{-49}$, $F=165.9$, $df = 5$).

Increase in the Rate of Horizontal Gene Transfer Has Minimal Effect on the Relationship Between Population Size and CMR

Figure 3 shows the relationship between population size and CMR when the number of HGT donors per recipient individual is increased. Each recipient received 0.1% or 1.6% of the sequence from 4, 16, or 32 donors selected at random from the population. Each individual in the population had the potential to be a recipient once per generation. The relationship between the CMR and population size is maintained as the number of donors increases. Hypothesis 2 stated that increasing the rate of HGT by increasing the number of donor individuals that transfer part of their sequence via HGT to one recipient per reproduction will have a smaller effect on the CMR than increasing the length of the sequence transferred; Figure 3 confirms this is the case. Two factor ANOVA indicates that the null hypothesis, that

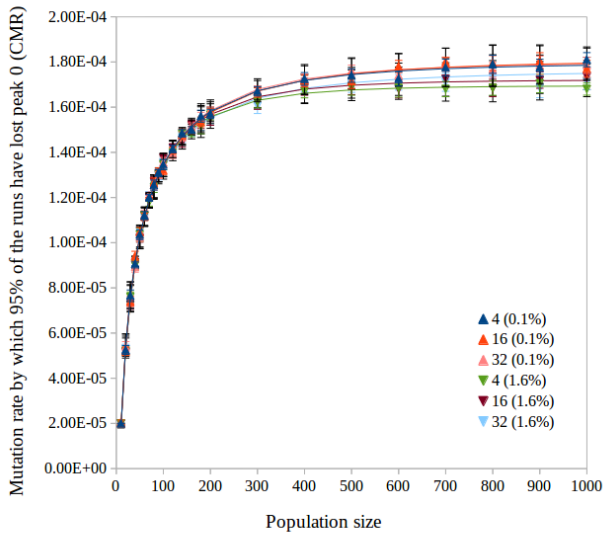


Figure 3: **CMR plotted against population size for varying rate of HGT when 0.1% or 1.6% of the donor sequence is transferred.** Recipient individuals received 1 b or 16 b (0.1 - 1.6%) of genetic material from 4, 16, and 32 donors per reproduction (given in the legend). The exponential lines were obtained by curve-fitting using R with a least squares method. Error bars represent +/- 95% confidence intervals.

there is no significant effect of the number of HGT donors per recipient on the CMR for each population size, can fail to be rejected when the amount of the sequence transferred is 0.1% ($p=0.548$, $F=0.606$, $df=2$). However, when the amount of the sequence transferred is 1.6%, the null hypothesis can be rejected ($p=0.003$, $F=6.62$, $df=2$); varying the population size and the number of HGT events has an effect on the CMR for the longer length of sequence transferred that is minimal but statistically significant.

Clonal Reproduction Increases the CMR

Figure 4 shows the relationship between population size and CMR for a population reproducing clonally both in the presence and absence of HGT compared to the CMR for a population reproducing with crossover (represented by the lower results in Figure 4 and presented separately in Figure 2). Individuals were subject to tournament selection before undergoing HGT, with the fittest of three individuals most likely to replace the least fit to simulate clonal reproduction. The amount of genetic material transferred from donor to recipient was kept constant for the duration of each run and set to equal from 1 b quadrupling up to 256 b, with individuals 1000 b in length. The relationship between population size and CMR remained exponential when clonal reproduction was introduced. However, the overall magnitude of the CMR was higher than that for the population with crossover. This is in agreement with Hypothesis 3 which expected the introduction of recombination to lead to an increase in the

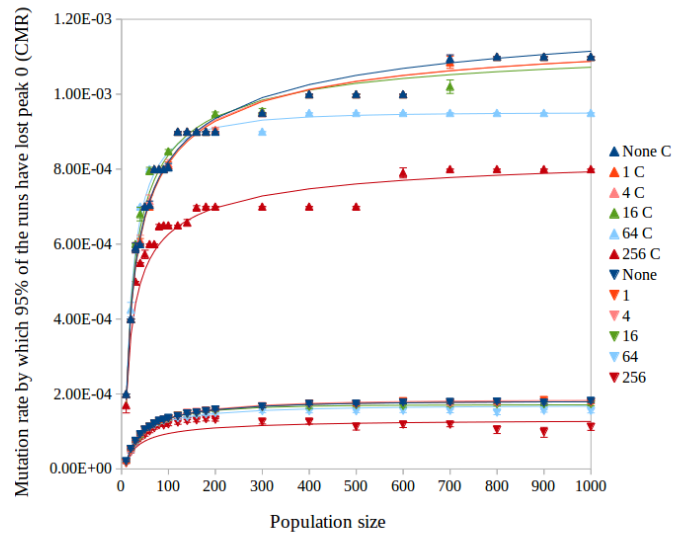


Figure 4: **Comparison of CMR plotted against population size when reproducing clonally (upper curves) or with crossover (lower curves).** Donor individuals transferred between 1 b and 256 b of genetic material (given in the legend, appended with C for clonal) per reproduction. The exponential lines were obtained by curve-fitting using R with a least squares method. Error bars represent +/- 95% confidence intervals. Note the lower curves are taken from Figure 2.

CMR. Two factor ANOVA indicates that the null hypothesis, that there is no significant effect of population size and the amount of genetic material transferred on the CMR in a clonal population, can be rejected ($p=2.82 \times 10^{-40}$, $F=105.3$, $df=5$). A comparison of the clonal and non-clonal results for 16 b (the mid point of the range of lengths of sequence transferred) indicates that the null hypothesis, that there is no significant effect of population size and clonality on the CMR, can be rejected ($p=5.31 \times 10^{-15}$, $F=350.3$, $df=1$).

Horizontal Gene Transfer from Donors with Non-zero Fitness Increases CMR for Longer Sequences in a Clonal Population

Figure 5 shows the relationship between population size and CMR when individuals in a clonal population (C) or one undergoing crossover select a donor for HGT either randomly or from a subset of the population with non-zero fitness (+). In each case, the exponential relationship between CMR and population size is maintained. In agreement with Hypothesis 4, allowing HGT to proceed in a clonal population only with donors with non-zero fitness lead to an increase in the CMR when the size of the sequence transferred was greater than 16 b long; two factor ANOVA indicates that the null hypothesis, that there is no significant effect of the restriction on donors on the CMR for each population size in a clonal population, can be rejected ($p=3.02 \times 10^{-08}$, $F=69.3$, $df=1$).

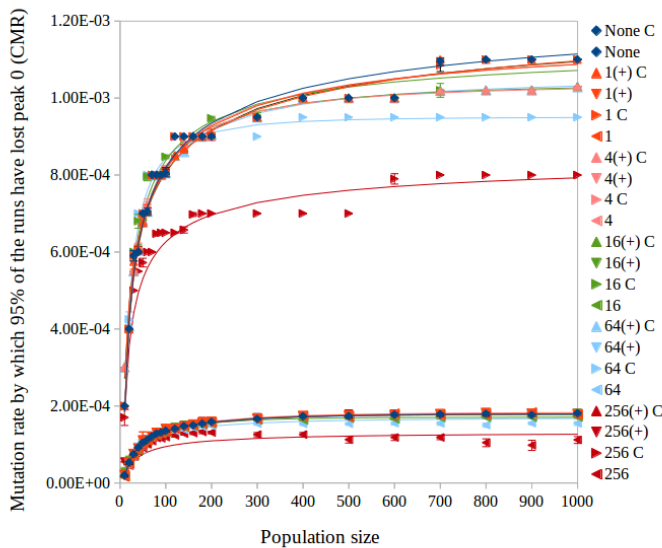


Figure 5: **Comparison of CMR plotted against population size for a population reproducing clonally (C) (upper curves) or with crossover (lower curves) with HGT donors selected either randomly or by fitness (+).** Donor individuals transferred between 1 b and 256 b of genetic material (given in the legend) per reproduction. The exponential lines were obtained by curve-fitting using R with a least squares method. Error bars represent \pm 95% confidence intervals. Note the overlap in results for 16 b and below, and the lack of overlap for 64 b and 256 b (both clonal and non-clonal). Also note the overlap in data for 64(+)*C* and 4(+)*C*, and that all of the data for (+) is close together.

However, in a population with crossover, the null hypothesis can fail to be rejected ($p=0.789$, $F=0.073$, $df=1$). As per the case where donors were selected at random, introducing clonality to a population in which donors have non-zero fitness lead to an increase in the CMR; the null hypothesis, that there is no significant effect of population size and clonality on the CMR, can be rejected ($p=2.99 \times 10^{-15}$, $F=370.2$, $df=1$).

Discussion

Enabling a population of *in silico* individuals to transfer between 0.1% and 25.6% of their sequence to a recipient has an effect on the CMR when longer sequences are transferred (Figure 2); transfer of greater than 1.6% of the donor sequence to the recipient per reproduction lead to a decrease in CMR. The exponential relationship between population size and CMR is maintained with the addition of HGT; the change in CMR with population size is greater for populations with 100 individuals or less. However, when the size of the sequence transferred reaches 25.6% of the donor individual, the relationship between CMR and population size is less consistent for populations with approximately 300 individuals and above. Transfer of genetic information horizon-

tally effectively introduces an additional recombination step. Recombination can push the population in a genetic algorithm over the error threshold when the mutation rate is high (such as when it is close to the CMR). HGT of longer sequences from the donor to the recipient will lead to a greater proportion of the recipient consisting of donor sequence; the increase in recombination pushes the population over the CMR at a lower magnitude.

Figure 3 shows that increasing the number of HGT events and therefore increasing the number of donors per recipient, while keeping the amount of the sequence transferred constant at 1 b, has no significant effect on the CMR. Increasing the length of the sequence transferred to 16 b (the sequence length above which there was a reduction in the CMR in Figures 3 and 4) and then increasing the number of transfers to each recipient leads to a minimal (yet statistically significant) reduction in the CMR. This is due to the fact that there is no guarantee that, for example, a recipient receiving two 16 b sequences from two donors will have 32 b of new genetic material within its sequence; the position of the sequence to be transferred is selected at random from along the length of the sequence therefore there may be an overlap. For short sequences of only 1 b, while an overlap is less likely, when an overlap does occur one sequence will completely replace the other. The effect of increasing the length of the sequence transferred via HGT observed in Figure 2 was not repeated in Figure 3. It should be noted that while for each reproduction step there is one recipient and potentially multiple donors, over one complete generation an individual may be selected as a donor more than once and may therefore transfer its genetic material to multiple recipients. An increase in the rate of transfer may be of relevance to a biological environment where an increase in HGT leads to an abundance of DNA to be taken up by recipient bacteria; in terms of the simulation model, such an increase in the amount of DNA transferred will have minimal effect on the CMR.

Keeping the number of HGT donors per recipient individual constant at 1, the relationship between population size and CMR remained exponential when clonal reproduction was introduced (Figure 4). The CMR increased in comparison to the CMR for reproduction with crossover in the presence of HGT. Recombination can reduce the diversity in a population and enhance fitness when the mutation rate is low (Ochoa and Harvey, 1998). However, at higher mutation rates (such as close to the CMR), it can lead to the population being pushed over the CMR; the addition of HGT, which involves information transfer between individuals in a similar manner to recombination, was seen to lower the CMR in Figure 2. Removal of the recombination step per individual per reproduction by introducing clonality had the opposite effect; individuals on peak 0 are less able to share genetic information with individuals in parts of the fitness landscape representing lower fitness when there is only HGT.

While transfer of genetic information between individu-

als in a population can be beneficial, it has the potential to lead to the spread of genes that are detrimental to fitness. In terms of the two-peak fitness landscape (Figure 1), an individual on peak 0 transferring its sequence to a recipient with zero fitness has the potential to lead to an increase in the fitness of the recipient. Conversely, an individual with zero fitness transferring its sequence to an individual on peak 0 may cause the recipient to lose peak 0 and therefore reduce the CMR. While the effect on the fitness of the recipient is dependent on the portion of the sequence transferred from the donor, a donor on the top of peak 0 is more likely to increase the fitness of a HGT recipient than a donor that is far away from peak 0 in the landscape. Restricting HGT to a subset of donors with non-zero fitness means that recipient individuals only receive genetic material from individuals on peak 0 or peak 1; they will not undergo HGT with individuals that have lost both of the peaks. Without this restriction, individuals from anywhere in the fitness landscape may be picked as a donor as the simulation model enables individuals with zero fitness to be selected in the reproduction step; it does not allow individual death. HGT only from donors on the peaks prevents the reduction in the CMR observed for transfer of long sequences in both clonal and non-clonal populations (Figure 5). However, the restriction of donors to those with only non-zero fitness only had a statistically significant effect on the CMR for the clonal population.

Enabling genetic information to be transferred both vertically and horizontally affects the magnitude of the CMR at which survival-of-the-flattest occurs. While bacteria are typically associated with large population sizes, they will exist in much smaller populations when the population is founded, near to extinction, or if a subset of the population is considered as a population in its own right. Understanding the effect of HGT on the CMR and the potential for survival-of-the-flattest is crucial to understanding the likely fate of bacterial sub populations or founding populations. For example, a sub population of resistant bacteria that will emerge due to introduction of antibiotic resistance genes into a population via HGT may be unable to optimally reproduce above a CMR increasing exponentially with population size. Further work will be required to enable development of our *in silico* simulation model to predict the outcome of a real biological population of evolving bacterial cells. This will involve use of prokaryotic-length genomes made up of a biologically realistic number of genes, of which a subset can be transferred horizontally, to output CMRs that can then be compared with existing mutation rates observed for bacterial species; it is expected that biological mutation rates will have evolved to be close to optimal, likely to be close to but not exceeding the CMR. This will enable us to further refine both the simulation model and its parameter values to close the gap between artificial simulation and prokaryotic evolution.

Acknowledgements

This work was supported by the Biotechnology and Biological Sciences Research Council [grant numbers BB/M020975/1, BB/M021106/1, BB/M021157/1].

References

- Aston, E., Channon, A., Day, C., and Knight, C. G. (2013). Critical mutation rate has an exponential dependence on population size in haploid and diploid populations. *PLoS ONE*, 8(12):e83438.
- Aston, E., Channon, A. D., Belavkin, R. V., Krasovec, R., and Knight, C. G. (2016). Critical mutation rate has an exponential dependence on population size for eukaryotic-length genomes. MIT Press.
- Baur, B., Hanselmann, K., Schlimme, W., and Jenni, B. (1996). Genetic transformation in freshwater: *Escherichia coli* is able to develop natural competence. *Applied and Environmental Microbiology*, 62(10):3673–3678.
- Bull, J. J., Meyers, L. A., and Lachmann, M. (2005). Quasispecies made simple. *PLoS Computational Biology*, 1(6):e61.
- Channon, A., Aston, E., Day, C., Belavkin, R. V., and Knight, C. G. (2011). Critical mutation rate has an exponential dependence on population size. In *Advances in Artificial Life, ECAL 2011: Proceedings of the Eleventh European Conference on the Synthesis and Simulation of Living Systems*.
- Cooper, T. F. (2007). Recombination speeds adaptation by reducing competition between beneficial mutations in populations of *Escherichia coli*. *PLoS Biology*, 5(9):e225.
- Drake, J. W., Charlesworth, B., Charlesworth, D., and Crow, J. F. (1998). Rates of spontaneous mutation. *Genetics*, 148:1667–1686.
- Griffith, F. (1928). The significance of pneumococcal types. *The Journal of Hygiene*, 27:113–159.
- Harvey, I. (1996). The microbial genetic algorithm.
- Harvey, I. (2011). The microbial genetic algorithm. In *In Kampis, G., Karsai, E., and Szathmari, E., editors, ECAL 2009, Part II. LNCS 5778*.
- Hindré, T., Knibbe, C., Beslon, G., and Schneider, D. (2012). New insights into bacterial adaptation through *in vivo* and *in silico* experimental evolution. *Nature Reviews Microbiology*, 10:352–365.

- Johnsen, P. J., Townsend, J. P., Bøhn, T., Simonsen, G. S., Sundsfjord, A., and Nielsen, K. M. (2011). Retrospective evidence for a biological cost of vancomycin resistance in the absence of glycopeptide selective pressures. *Journal of Antimicrobial Chemotherapy*, 66:608–610.
- Jones, D. and Sneath, P. H. (1970). Genetic transfer and bacterial taxonomy. *Bacteriological Reviews*, 34:40–81.
- Kimura, M. and Maruyama, T. (1966). The mutational load with epistatic gene interactions in fitness. *Genetics*, 54:1337–1351.
- Kimura, M. and Ohta, T. (1969). Average number of generations until extinction of an individual mutant gene in a finite population. *Genetics*, 61:763–771.
- Koonin, E. V., Makarova, K. S., and Aravind, L. (2001). Horizontal gene transfer in prokaryotes: quantification and classification. *Annual Review of Microbiology*, 55:709–742.
- Krašovec, R., Belavkin, R. V., Aston, J. A. D., Channon, A. D., Aston, E., Rash, B. M., Kadirvel, M., Forbes, S., and Knight, C. G. (2014). Mutation rate plasticity in rifampicin resistance depends on escherichia coli cell-cell interactions. *Nature Communications*, 5:3742.
- Lenski, R., Barrick, J. E., and Ofria, C. (2006). Balancing robustness and evolvability. *PLoS Biology*, 4:2190–2192.
- McDaniel, L., Young, E., Delaney, J., Ruhnau, F., Ritchie, K., and Paul, J. (2010). High frequency of horizontal gene transfer in the oceans. *Nature*, 330(6000):1.
- Mozhayskiy, V. and Tagkopoulos, I. (2011). In silico evolution of multi-scale microbial systems in the presence of mobile genetic elements and horizontal gene transfer. *Lecture Notes in Computer Science*, 6674:262.
- Mozhayskiy, V. and Tagkopoulos, I. (2012). Horizontal gene transfer dynamics and distribution of fitness effects during microbial in silico evolution. *BMC Bioinformatics*, 13(Suppl 10):S13.
- Nawa, N. and Furuhashi, T. (1998). Bacterial evolutionary algorithm for fuzzy system design. In *IEEE Conference on Systems, Man, and Cybernetics*, pages 2424–2429.
- Nawa, N., Hashiyama, T., Furuhashi, T., and Uchikawa, Y. (1997). A study on fuzzy rules discovery using pseudo-bacterial genetic algorithm with adaptive operator. *Proceedings of 1997 IEEE International Conference on Evolutionary Computation (ICEC 97)*, pages 589–593.
- Nielsen, K. M., Bøhn, T., and Townsend, J. P. (2014). Detecting rare gene transfer events in bacterial populations. *Frontiers in Microbiology*, 4(415).
- Ochman, H., Lawrence, J. G., and Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature*, 405:299–304.
- Ochoa, G. and Harvey, I. (1998). Recombination and error thresholds in finite populations. In *Foundations of Genetic Algorithms (FOGA-5)*, pages 245–264.
- Puigbò, P., Lobkovsky, A. E., Kristensen, D. M., Wolf, Y. I., and Koonin, E. V. (2014). Genomes in turmoil: quantification of genome dynamics in prokaryote supergenomes. *BMC Biology*, 12:66.
- Ravenhall, M., Škunca, N., Lassalle, F., and Dessimoz, C. (2015). Inferring horizontal gene transfer. *PLoS Computational Biology*, 11(5):e1004095.
- Raz, Y. and Tannenbaum, E. (2010). The influence of horizontal gene transfer on the mean fitness of unicellular populations in static environments. *Genetics*, 185:327–337.
- Sanjuán, R., Cuevas, J. M., Furió, V., Holmes, E. C., and Moya, A. (2007). Selection for robustness in mutagenized RNA viruses. *PLoS Genetics*, 3(6):e93.
- Tatum, E. L. and Lederberg, J. (1947). Gene recombination in the bacterium escherichia coli. *Journal of Bacteriology*, 53:673–684.
- Tomko, N., Harvey, I., and Philippides, A. (2013). Unconstrain the population: The benefits of horizontal gene transfer in genetic algorithms. *SmartData*, pages 117–127.
- van Dijk, B. and Hogeweg, P. (2015). In silico gene-level evolution explains microbial population diversity through differential gene mobility. *Genome Biology and Evolution*, 8(1):176–188.
- Wilke, C. O. (2005). Quasispecies theory in the context of population genetics. *BMC Evolutionary Biology*, 5:44.
- Zinder, N. D. and Lederberg, J. (1952). Genetic exchange in salmonella. *Journal of Bacteriology*, 64:679–699.