J Chem Biol (2010) 3:189–194
DOI 10.1007/s12154-010-0042-6

ORIGINAL ARTICLE

The two-mutant problem: clonal interference in evolutionary graph theory

Chris Paley · Sergei Taraskin · Stephen Elliott

Received: 27 October 2009 / Accepted: 21 May 2010 / Published online: 18 June 2010

Abstract In large asexual populations, clonal interference, whereby different beneficial mutations compete to fix in the population simultaneously, may be the norm. Results extrapolated from the spread of individual mutations in homogeneous backgrounds are found to be misleading in such situations: clonal interference severely inhibits the spread of beneficial mutations. In contrast with results gained in systems with just one mutation striving for fixation at any one time, the spatial structure of the population is found to be an important factor in determining the fixation probability when there are two beneficial mutations.

Keywords Clonal interference · Spatial population · Fixation probability · Beneficial mutations

Introduction

Recent work [1] has given us a newly thorough understanding of the molecular nature of mutations, applying whole-genome sequencing to detect de novo mutations [2] in the yeast *Saccharomyces cerevisiae*. They are, overwhelmingly, nucleotide substitutions, with transitions occurring approximately half as often as transversions. This and other work show that mutations that are deleterious [3–5] and advantageous [6, 7] are of high frequency, with deleterious mutations occurring of the order 10^{-8} per nucleotide, and advantageous mutations occurring of order 10^{-10} per cell per division. Therefore, it is not the case that a single mutation fixes on a homogeneous background, but, in fact, clonal interference [8] is of major importance. The current paper shows how clonal interference alters the probability of a given beneficial mutation establishing (fixing) in a population, which is a basic measure of the ability of that population to adapt.

Generally, the probability of fixation of a single mutation in a homogeneous population has been studied [9–11]. For a number of different model population structures (including fully connected systems, lattices and linear chains), it has been reported that the fixation probabilities of a mutation are the same [12], suggesting that the structure of a population is unimportant in evolutionary dynamics, a result that has been further bolstered by similar work on the evolution of cooperation [13] (also looking at the fixation of a single strategy in a homogeneous population). In real asexual populations, it is often the case that multiple advantageous mutations are present: clonal interference [8] is the term given to the process by which different asexual lineages compete with each other to establish. Here we demonstrate that, when multiple mutations are considered simultaneously, an extrapolation of the behaviour seen when mutations fix sequentially is inappropriate. Furthermore, we show that the spatial structure of the system becomes an important factor in determining the dynamics of the population.

In this paper, evolutionary graph theory [12] is used to study the simulated behaviour of interfering mutations in a population. This theory was invented to allow the easy study of the fixation behaviour of mutants in populations with differing spatial structures. The model is an extension of the Moran model [11], which becomes the special case of a population on a fully connected

C. Paley · S. Taraskin (✉) · S. Elliott
Department of Chemistry, University of Cambridge,
Lensfield Road, Cambridge, CB2 1EW, UK
e-mail: snt1000@cam.ac.uk

graph with self-loops. In fact, the same model was used by Williams and Bjerknes [14] to study the spread of cancer on a variety of different lattices. The Eden model [15], which is the Williams–Bjerknes model for the case of infinite fitness, is still actively used for studying the growth of rough surfaces [16, 17] and is related to the KPZ model [18, 19]. This simple model therefore belongs to a family of closely related models which have been found to be useful in a number of different areas of science.

In each of the above models, there are two types of particles (e.g. mutant/wild type, cancerous cell/healthy cell, interface/body, etc.). They do not consider the impact of clonal interference. In this paper, we therefore consider the simplest simulation case in which there are multiple mutants in the system: when a single beneficial mutation has partially spread through the system, a second mutation of the same advantage is introduced. We demonstrate that the presence of existing mutations is a significant fixation inhibitor for the second mutation.

Model

In evolutionary graph theory [12], a population consists of individuals which exist at the nodes of a network. Each individual in the population has a fitness and, on each turn, an individual is chosen to reproduce with probability proportional to its fitness. A clone of the chosen individual is produced, and this is placed onto a node connected to the parent by the network. In this paper, we consider the simplest case where the links between nodes are unweighted, and the offspring is placed onto any of the connected nodes with equal probability, replacing the existing individual, so conserving population size, N. Provided that the connectivity of each node is the same (i.e. each node is connected to Z other nodes: fully-connected systems and hypercubic lattices are examples of networks with this property), then a single mutant of fitness r placed on a background wild-type population (each such individual having a fitness of unity) will fix in the population (i.e. drive the wild-type population to extinction) with probability, ρ, where [12]:

$$\rho = \frac{1 - 1/r}{1 - 1/r^N} \,. \tag{1}$$

This is independent of the topology of the system, unlike the time taken for the mutation to fix [20]. If there is only one mutant in the system at any one time, it is possible to derive a simple formula for the rate of evolution, R, [12, 21] when u is the mutation rate of r-fitness mutations,

$$R = Nu\rho \,. \tag{2}$$

If the fixation probability is unchanged by the presence of other mutations in the system, then this equation will hold even in systems with multiple mutations spreading simultaneously. This is the case with neutral mutations [21, 22], where a lineage has a fitness independent of the number of neutral mutations (and the probability that any individual will be the ancestor of the entire population eventually is $1/N$). If the fixation probability is heavily influenced by the presence of other mutations in the system, then its calculation on a homogeneous background should be treated cautiously in systems where there may be a number of lineages of different fitnesses, and, in fact, Eq. 2 does not then hold.

Results

To investigate this, a single mutation of fitness r is introduced into a simulation system originally consisting of N wild-type individuals (each having a fitness of unity). The number of individuals having the mutation, m_{1t}, will potentially increase. If and when this first mutation has spread such that the mutant population consists of m_1 individuals, a second mutation having the same fitness effect (so an organism having both benefits will have a fitness of r^2 compared to the wild-type organism), is introduced into a randomly chosen individual. The invasion dynamics, when the second mutation occurs in a wild-type individual (Fig. 1), are very different to those when it occurs in an organism already carrying a beneficial mutation (Fig. 2). The major differences in dynamics can be seen at the interfaces between populations. When the mutation occurs in the wild-type population, the two mutant populations (black and red populations in Fig. 1) have the same fitness and the interface between them is dominated by drift. When an organism exists with both benefits (blue population in Fig. 2), the interface is characterized by the selective advantage of the double mutants (blue population) and their higher fitness, which leads to a higher fixation probability of double mutants. The double mutants are able to spread quickly through both wild-type and single mutant populations.

The probability that the second mutation fixes can be investigated numerically. Figure 3 shows two key results: (i) that the fixation probability deviates sub-

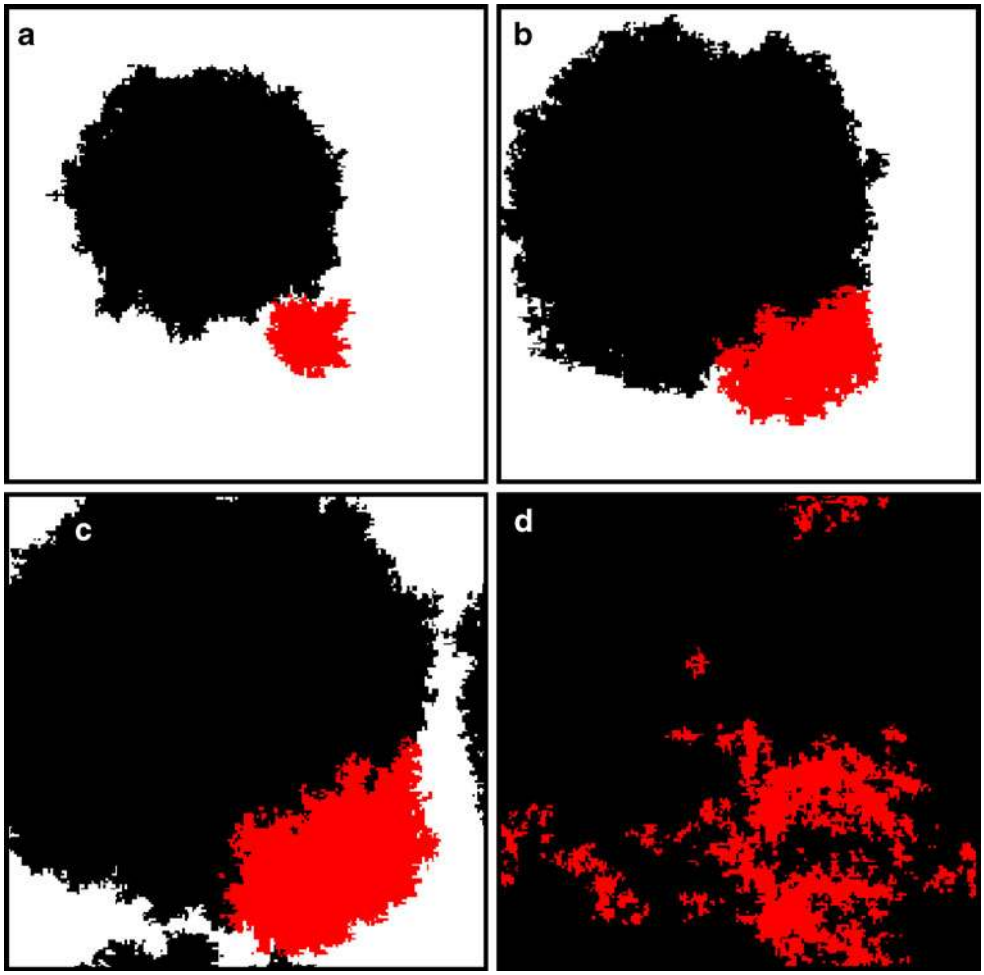

Fig. 1 Visualisation of the spread of a second invading mutant on a square lattice, initially starting outside the resident mutant population. The snapshots are taken at successive time points of a single simulation to demonstrate representative stages in the invasion. The *black cells* are the resident mutant population (m_{1t} individuals); the *red cells* are the second invasion (m_{2t} individuals). In (**a**), $m_{1t} = 10,000$, $m_{2t} = 8,63$. In (**b**), $m_{1t} = 18,000$, $m_{2t} = 2,476$. In (**c**), $m_{1t} = 26,000$, $m_{2t} = 4,756$. In (**d**), $m_{1t} = 35,000$, $m_{2t} = 5,000$. The system is a 200×200 square lattice ($N = 40,000$) with periodic boundary conditions; all the mutants have fitness $r = 2$. The wild types (*white*) have fitness $r = 1$

stantially from being independent of m_1 (the probability ρ predicted in Eq. 1 is found when $m_1 = 0$ or $m_1 = N$, i.e. $\rho(m_1 = 0) = \rho(m_1 = N) = (1 - 1/r)/(1 - 1/r^N)$); (ii) the spatial structure plays an important role in determining the level of clonal interference (i.e. the shape of the curve for $\rho(m_1)$), particularly as r increases. So, the presence of existing beneficial mutations in the system substantially inhibits the spread of other mutations, and the level of this inhibition is dependent on the structure (topology) of the system and thus Eq. 2 no longer holds.

The probability, $\rho^{(2)}$, that a (second-wave) mutation fixes in a population in which another mutation with the same beneficial advantage is spreading can be approximately evaluated. Indeed, the probability of the second-wave mutation fixing is the probability, p_{WTP}, that the second mutant occurs in the wild-type population (WTP), multiplied by the probability, $\rho^{(2)}_{\mathrm{WTP}}$, that it fixes, given that it occurred in a wild-type individual, added to the probability, p_{RMP}, that it occurred in an organism carrying the resident mutation, multiplied by the probability, $\rho^{(2)}_{\mathrm{RMP}}$, that it fixes, given that it occurred in a member of the resident mutant population (RMP). This can be written as:

$$\rho^{(2)} = p_{\mathrm{WTP}}\rho^{(2)}_{\mathrm{WTP}} + p_{\mathrm{RMP}}\rho^{(2)}_{\mathrm{RMP}} \,. \tag{3}$$

The probabilities $p_{\mathrm{WTP}} = (N - m_1)/N$ and $p_{\mathrm{RMP}} = m_1/N$ are known, and are the same for all networks.

Estimates of the conditional fixation probabilities, $\rho^{(2)}_{\mathrm{WTP}}$ and $\rho^{(2)}_{\mathrm{RMP}}$ can be obtained using several approximations. The first is that, once a substantial advantageous mutant population has built up, it is likely that the wild-type population will be driven to extinction. However, the likelihood of this depends on r. In the following analysis, it will be assumed that the wild-type population will be driven to extinction. This relies on r and m_1 being reasonably large (a small population of very fit mutants, or a large population of slightly advantageous mutants, will almost certainly fix).

For the linear chain, it is possible to derive an approximation for $\rho^{(2)}_{\mathrm{WTP}}$. If the second-wave mutation survives initial drift (probability $\simeq 1 - 1/r$ for $N \gg 1$,

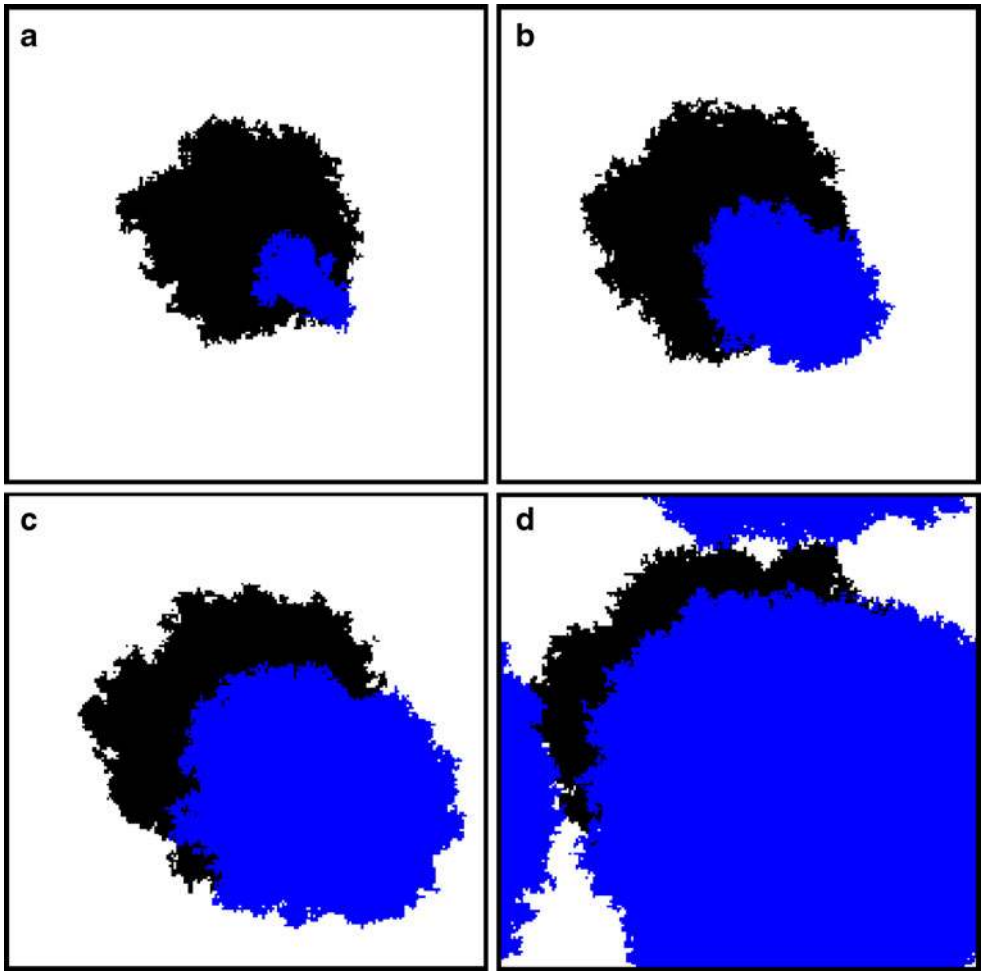

Fig. 2 Visualisation of the spread of a second invading mutant on a square lattice, with the second mutation occurring in an individual already possessing the first advantageous mutation. The snapshots are taken at successive time points of a single simulation to demonstrate representative stages in the invasion. The *black cells* are the resident mutant population (m_{1t} individuals); the *blue cells* are the second invasion (m_{2t} individuals). In (**a**), $m_{1t} = 5,717$, $m_{2t} = 1,000$. In (**b**), $m_{1t} = 5,998$, $m_{2t} = 4,000$. In (**c**), $m_{1t} = 5,467$, $m_{2t} = 10,000$. In (**d**), $m_{1t} = 3,854$, $m_{2t} = 28,000$. The system is a 200×200 square lattice ($N = 40,000$) with periodic boundary conditions. The mutants have fitness $r = 2$ (such that the blue individuals, which possess both mutations, have a fitness quadruple ($r^2 = 4$) that of the *white* wild types ($r = 1$))

which follows from Eq. 1), then it can be expected that it will travel at the same speed as the resident mutant population (as both have the same fitness and necessarily the same interface; see [20]). It can then be considered what m_{1t} and m_{2t} will be when there are no wild-type mutants left. The two mutant populations will each, on average, have taken half of the wild-type positions (initial population is of size $N - m_1$). When there are no wild types left, all mutations are comparatively neutral, and the probability that the second-wave mutation fixes is just the percentage of the nodes that second-wave mutants occupy. The probability that it fixes, given that it occurred in the wild-type population, can be written as:

$$\rho^{(2)}_{\mathrm{WTP}} = (1 - 1/r)\frac{N - m_1}{2N} \,. \tag{4}$$

If the second-wave mutation occurs in the resident mutant population, then it is likely that the second-wave mutation will never see a wild-type individual (given a substantial value of $m_1 \gg 1$). The interface of the resident mutants will pass through the wild types at the same rate that a growing second-wave mutant interface would pass through the resident mutant population, so that

$$\rho^{(2)}_{\mathrm{RMP}} = (1 - 1/r) \,. \tag{5}$$

If there is only a small resident mutant population, i.e. the inequality $m_1 \gg 1$ is violated, then this approximation will break down, underestimating the fixation probability, as the probability that the second-wave mutation occurs close to the wild-type boundary increases.

Using these approximate relations for conditional probabilities, we obtain the following estimate of the fixation probability for a second-wave mutation,

$$\rho^{(2)}_{1D} = \frac{m_1}{N}(1 - 1/r) + \frac{(N - m_1)^2}{2N^2}(1 - 1/r) \,. \tag{6}$$

Note that this equation has a minimum, whose value is given by:

$$\rho^{(2)}_{1D,\mathrm{min}} = \frac{1}{2}\left(1 - \frac{1}{r}\right) \,, \tag{7}$$

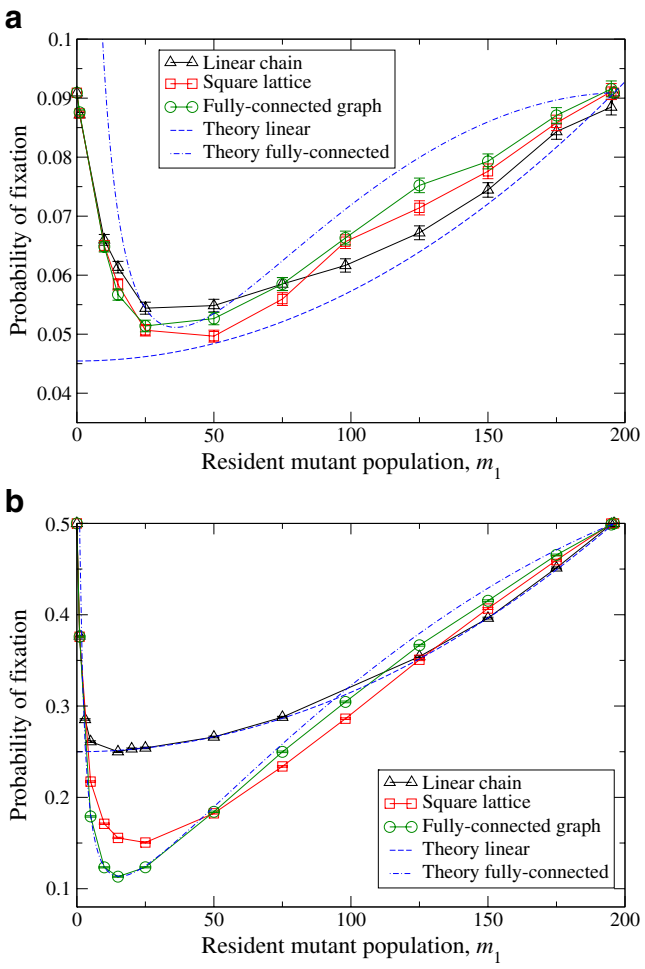

Fig. 3 Probability of a mutant fixing in a population when another mutant is already spreading. In (**a**), $r = 1.1$; in (**b**), $r = 2$. The population size is $N = 196$. Each data point was obtained from simulations on 500,000 initially wild-type populations. The *dashed-dotted line* is a plot of Eq. 9, and the *dashed line* is a plot of Eq. 6

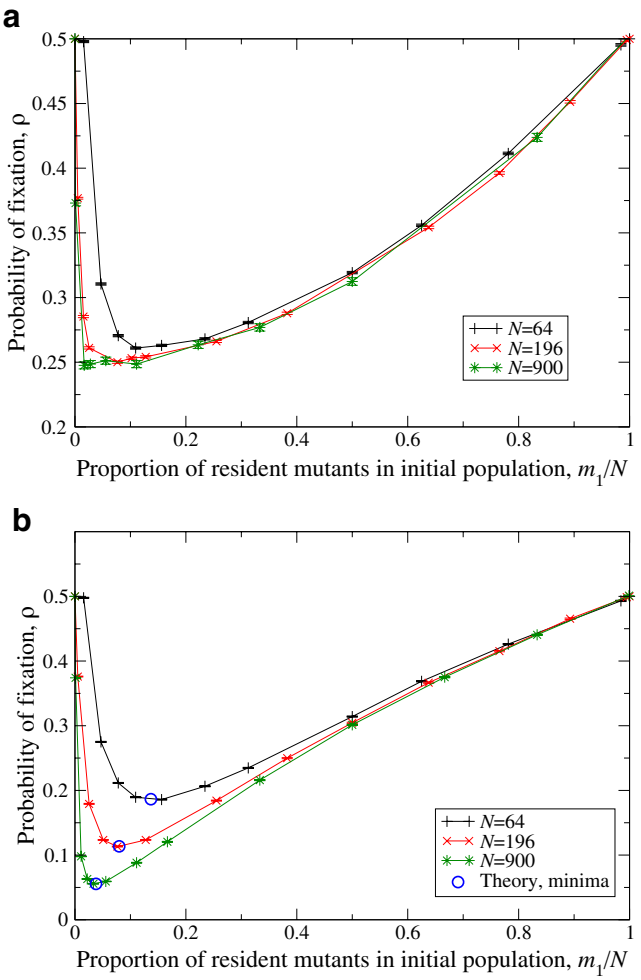

Fig. 4 Probability that a second mutant fixes, depending on the initial resident mutant population and the size of the population. In (**a**), the population network was a linear chain; in (**b**), it was fully connected. Mutant fitness is $r = 2$. Each data point was obtained from simulations on 500,000 initially wild-type populations, except for some of the $N = 900$ points for the linear chain, for which 50,000 runs were used due to computational constraints. The lines are guides for the eye. The *circles* in (**b**) are the theoretical predictions of the positions of the minima from Eq. 9

which occurs at $m_1 = 0$. The curves obtained by using Eq. 6 are plotted in Fig. 3 and they agree well with the results of simulations. As expected, the analytical approximations work best for substantial r, and break down when m_1 is very small (so the wild-type population sometimes fails to succumb, causing a break-down of the approximations used). It is noticeable that the minimum of the formula is independent of N (though it relies on the assumption that N is large). In fact, if Eq. 6 is re-scaled so that the independent variable is the proportion (m_1/N) of the population that has been conquered by the resident mutant, it is N-independent. Indeed, this behaviour can be seen for large N in Fig. 4.

For the fully-connected system, if the second-wave mutation falls outside the resident mutant population, it will have the same probability of fixing as any member of the resident mutant population. Therefore, $\rho^{(2)}_{\mathrm{WTP}} = 1/(m_1 + 1)$ (making the same approximation as for the linear chain that the wild-type population will succumb to the mutants).

If the second-wave mutation falls in the resident mutant population, the effects of competition with the wild-type population must be included in the fully-connected system (even if the wild types eventually succumb). Growth into the wild types can grant the second-wave mutation stability before it competes with just single mutants.

On the very first turn, once the second-wave mutant has been introduced, it is possible to consider the probability that the second-wave mutant population will increase versus the probability that it will decrease. The probability that it will increase is r^2/J, where $J = r^2 + (m_1 - 1)r + (N - m_1) \times 1$. The probability that the second-wave population will decrease is the sum of two terms: the probability that the resident mutants will increase by replacing the second-wave mutants, and the probability that the wild types will do so. This probability is $(m_1 - 1)r/[J(N-1)] + (N - m_1) \times 1/[J(N-1)]$. Provided that $N \gg 1$, the ratio of the probability that the population will decrease to the probability that it will increase can be written as $[N + (r-1)m_1]/Nr^2$. This ratio can be put into the original formula for the fixation probability in evolutionary graph theory (see Eq. 1), to give:

$$\rho^{(2)}_{\mathrm{RMP}} = 1 - \frac{N + (r-1)m_1}{Nr^2} \,. \tag{8}$$

This must be an overestimate: as time continues, the fraction of the population that is wild type will decrease, and so the probability that the second-wave mutation population grows will decrease. It is worst at intermediate values of m_1 (when the growth of the resident mutant population at the expense of the wild-type population would be expected to be fastest, see [20]). Putting these approximations together, the probability of the second-wave mutation fixing in a fully-connected system can be written as:

$$\rho^{(2)}_{\mathrm{FCG}} = \frac{N - m_1}{N} \frac{1}{m_1 + 1} + \frac{m_1}{N} \left(1 - \frac{N + (r-1)m_1}{Nr^2} \right) . \tag{9}$$

Again, the approximation is good for high r, and breaks down for small values of m_1. The minimum of $\rho^{(2)}_{\mathrm{FCG}}$ is now N-dependent, with the minimum decreasing with increasing population size. The positions of the minima for different values of N given by Eq. 9, and found numerically, agree well with the numerical results from simulations (see the circles in Fig. 4). A square-lattice system also exhibits a deeper minimum for larger populations (not shown).

It should also be noted that, even when there is only one beneficial mutation in the system at any instant, the presence of clonal lineages with differing deleterious mutational loads will create diverse populations as backgrounds for the production and spread of beneficial mutations. Therefore, even in populations with very low beneficial mutation rates, the background diversity of the population should be taken into account when calculating fixation probabilities.

Conclusions

To conclude, we have demonstrated that, in biological populations with two mutations present simultaneously, the simple expressions for the mutation fixation probability and the rate of evolution given by Eq. 1 and Eq. 2, respectively, break down, along with the conclusion that the fixation probability is largely independent of spatial structure. The presence of an existing beneficial mutant population acts as an inhibitor to the spread of other advantageous genes. As there are usually multiple mutations in real biological systems at any one time, as suggested by the experimental results described above, the rate of adaptation in asexual populations may be expected to be significantly slower than would be predicted theoretically by considering only a single mutation.

Acknowledgements CJP would like to thank the Engineering and Physical Sciences Research Council (UK) for financial support. The Cambridge University Condor programme provided computational resources.

References

1. Lynch M et al (2008) Proc Natl Acad Sci 105:9272
2. Kondrashov AS (2008) Proc Natl Acad Sci 105:9133
3. Denver DR, Morris K, Lynch M, Thomas WK (2004) Nature 430:679
4. Haag-Liautard C, et al (2007) Nature 445:82
5. Nachman MW, Crowell SL (2000) Genetics 156:297
6. Imhof M, Schlötterer C (2001) PNAS 98:1113
7. Kessen R, Bataillon T (2006) Nat Genet 38:484
8. Gerrish P, Lenski R (1998) Genetica 102/103:127
9. Haldane J (1927) Proc Camb Philos Soc 23:838
10. Fisher R (1930) The genetical theory of natural selection. Clarendon Press, Oxford
11. Moran P (1958) Proc Camb Phil Soc 54:60
12. Lieberman E, Hauert C, Nowak M (2005) Nature 433:312
13. Taylor P, Day T, Wild G (2007) Nature 447:469
14. Williams T, Bjerknes R (1972) Nature 236:19
15. Eden M (1961) In: Neyman J (ed) Proceedings of the fourth Berkeley symposium on mathematics, statistics and probability. University of California Press, California
16. Meakin P (1993) Phys Rep 235:189
17. Irurzun I, Horowitz C, Albano E (2005) Phys Rev E 72: 036116
18. Kardar M, Parisi G, Zhang Y (1986) Phys Rev Lett 56:889
19. Cieplak M, Maritan A, Banavar J (1996) Phys Rev Lett 76:3754
20. Paley C, Taraskin S, Elliott S (2007) Phys Rev Lett 98:098103
21. Nowak M (2006) Evolutionary dynamics: exploring the equations of life. Belknap Press, Cambridge, MA
22. Kimura M (1968) Nature 217:624