Ecological scaffolding and the evolution of individuality: the transition from cells to multicellular life

— Supplementary material —

Andrew J. Black, Pierrick Bourrat & Paul B. Rainey

May 29, 2019

1 Single type model

To construct the within-patch model we begin with a stochastic individual-based approach [2]. This means we specify the dynamics of the individual cells and hence the population is modelled as a continuous-time Markov chain from which a further approximating model can then be derived.

Here we consider the dynamics of just a single patch, so all quantities are assumed local to that patch. Cells are classified according to their growth rate. The number of the ith type in the patch is denoted G_i. Each type then reproduces at rate β_i by consuming some resource, E. Each patch is assumed to initially contain some limited amount of this resource $N \gg 1$.

At every reproduction event there is some probability, p, that a cell produces a mutant with a different growth rate $\beta_j = \beta_i + \xi$ where ξ is a random variable encoding the possible mutations. The rate of reproduction is assumed to follow simple mass-action dynamics which scales with the initial amount of resource in the patch. Thus as the resource is consumed, the rate of growth slows, with the initial resource scaling the maximum number of cells the patch supports.

These reproduction dynamics can be summarised by the following two events and rates,

$$\begin{aligned} G_i + E &\to G_i + G_i \quad \text{at rate} \quad \frac{(1-p)\beta_i G_i E}{N}, \\ G_i + E &\to G_i + G_j \quad \text{at rate} \quad \frac{p\beta_i G_i E}{N}, \end{aligned} \tag{1}$$

where $i = 1, \ldots, m(t)$, and $m(t)$ is the number of distinct cells types in the patch at time t. We also assume that cells of each type die at a constant rate, independent of its type,

$$G_i \to \phi \quad \text{at rate} \quad \gamma G_i. \tag{2}$$

The parameter γ only scales time, so this is set equal to 1. Thus the phenotype of a cell of type i is specified only by its growth rate, β_i.

The distribution of ξ can be chosen in many ways. Herein we adopt the simplest model where ξ is a Bernoulli random variable with probability of success ν, where a successful trial increases the growth rate by a fixed amount μ and otherwise decreases it by the same. Such a scheme can accommodate symmetric ($\nu = 0.5$) or asymmetric cases ($\nu < 0.5$), where most mutations would be deleterious. The initial conditions for the process are $m = 1$, $G_1(0) = 1$ and $E(0) = N$.

As defined above, the within-patch dynamics are modelled as a continuous-time Markov chain with an unbounded state space [6]. This cannot be solved analytically, but can be be simulated using Gillespie's algorithm [5]. This produces exact stochastic realisations of the process, but this can be very slow depending on the initial amount of resource in the patch. Also, these dynamics can

be quite complex, with extinction at the beginning being a possibility as well as large fluctuations in the time for the exponential growth phase to begin.

As we need to solve this model for many patches over many generations, we require a quicker approach, at the cost of some simplification. Thus we derive an approximation to this individual based model. As p is assumed to be small, but N large, there is a separation in time-scales that can be exploited. Thus we can approximate the dynamics with a piecewise deterministic process where the times of the creation of new mutant types are modelled stochastically, but the growth dynamics between the introduction of new types are modelled deterministically [4]. Note that we cannot just approximate the above model by a simple deterministic one. This is because the initial mutations cause the number of a particular type to go from $0 \to 1$ and this discrete behaviour cannot be captured using just ordinary differential equations (ODEs).

1.1 Piecewise deterministic approximation

First note that the rates of the events are density dependent (in the mathematical sense defined by [8], not the ecological sense). Next, we define the proportions of the ith cell type, $x_i(t) = G_i(t)/N$ for $i = 1, \ldots, m(t)$ and the resource $y(t) = E(t)/N$, within the patch. We can then approximate the full model above with a piecewise deterministic process where mutations that create new types occur as an inhomogeneous Poisson process with rate [1]

$$\lambda(t) = Np\, y(t) \sum_{i=1}^{m} \beta_i x_i(t), \tag{3}$$

and between mutations, the dynamics evolve as a set of ODEs,

$$\begin{aligned} \frac{dy}{dt} &= -y \sum_{i=1}^{m} \beta_i x_i \\ \frac{dx_i}{dt} &= \beta_i y x_i - \gamma x_i \quad i = 1, \ldots, m(t). \end{aligned} \tag{4}$$

Note that we have neglected a term of $1 - p$ in the Eqs (4) as this is close to 1. As each patch is colonised by a single cell, the initial conditions are $m = 1$, $x_1(0) = 1/N = \epsilon$ and $y(0) = 1$. When a new mutant type is created, the state of the process (and hence the number of of ODEs to be solved) is updated as,

$$\begin{aligned} m &\leftarrow m + 1, \\ y(t) &\leftarrow y(t) - \epsilon, \\ x_m(t) &= \epsilon. \end{aligned} \tag{5}$$

This approximation can be made in a number of ways resulting in slightly different formulations of the above equations. Here we have chosen the simplest, so we do not include the effect of later mutations between already existing types in the ODEs (4). We also limit the the number of mutant types to 2 (with growth rates $\pm\mu$ compared to the founding cell's growth rate) and ignore back mutations, which will be limited anyway by the small numbers of mutants produced overall due to the bottleneck and patch ecology.

To solve for the dynamics of a patch we need to determine the mutation times, i.e., we need to sample from the inhomogeneous Poisson process with rate given by Eq. (3). There are a number of possible ways of doing this, but we use a simple cumulative distribution function (CDF) inversion method because, as shown below, this allows the use of basic ODE solvers for the whole process. The expectation of the Poisson random variable with rate (3) is

$$E[N_t] = \Lambda(t) = \int_0^t \lambda(s) ds.$$

It can be shown that the inter event time, X_j, conditioned on the first j times $t_1, \ldots, t_j$ has CDF [7, 9],

$$F_{t_i}(x) = 1 - \exp(-\Lambda(t_i + x) + \Lambda(t_i)). \tag{6}$$

To use a CDF-inverse method to generate the times from (6) we then need to solve the equation [3]

$$\int_t^{t+x} \lambda(y)dy = -\ln(u)$$

where $u \sim \mathrm{U}(0,1)$ is a uniform random variable, t is the current time and x the time until the next event. Using Eqs (3) and (4) we see that

$$\lambda(t) = -Np\frac{dy}{dt}. \tag{7}$$

hence

$$-Np\int_t^{t+x} \frac{dy}{dt}dy = -Np(y(t+x) - y(t)) = \ln(u). \tag{8}$$

Hence Eq. (7) can be solved simultaneously with the equations (4) for the proportions. The procedure is to first generate u then starting with the initial condition we integrate the combined set of equations (4) and (7) until equation (8) is true, or the dispersal time, T, is reached. This is easily accomplished using an event function passed to the ODE solver. In this paper we use MATLAB's built in function `ODE45`. Once the time has been determined the state is updated according to Eq. (5).

1.2 Dynamics

Figure 1 shows four realisations of the within patch model with different founding cell growth rates. The final panel shows the trajectory resulting from the optimal growth rate for the dispersal time of $T = 30$, but note that the peak does not coincide with $t = 30$, but occurs slightly before. This figure also illustrates that slower growth rates lead to smaller *peak* populations in an absolute sense.

Another way to interpret the evolutionary dynamics of this model is by thinking in terms of a fitness landscape. This is possible because the accumulation of mutants in the patch is small, and so the number of cells at the time of dispersal is strongly determined by the growth rate of the founding cell. A fitness landscape is derived by plotting the patch size (number of G cells at the time of dispersal) as a function of the growth rate for different (fixed) dispersal times with $p = 0$, i.e., no mutation. Figure 2 shows this for $T = 10, 15$ and 30. As patch fitness is proportional to patch size, larger patches at the time of dispersal will be more likely to be selected. For a given T the system will evolve up the fitness curve; for example if $\beta = 1.8$ initially and $T = 30$ (blue curve), then this predicts that the average β in the system will decrease, as this leads to larger patches. Conversely, if $T = 10$ then the growth rate increases. The dashed line shows the imposed maximum growth rate, and hence if $T = 10$, this is where evolution will stop. If the maximum growth rate was unbounded then the growth rate would continue to rise, reaching an equilibrium around 2.5. By choosing a slightly slower dispersal time, $T = 15$, the equilibrium growth rate is reached at $\beta < 2$.

This fitness landscape view also emphasizes that whether we observe fitness decoupling or not depends on the dispersal time scale *relative* to the initial cell growth rate. It also shows that shorter times allow for fitter patches in an absolute sense as larger cell growth rates lead to bigger overall sizes. This is simply due to the particular form of the birth / death dynamics assumed in the model.

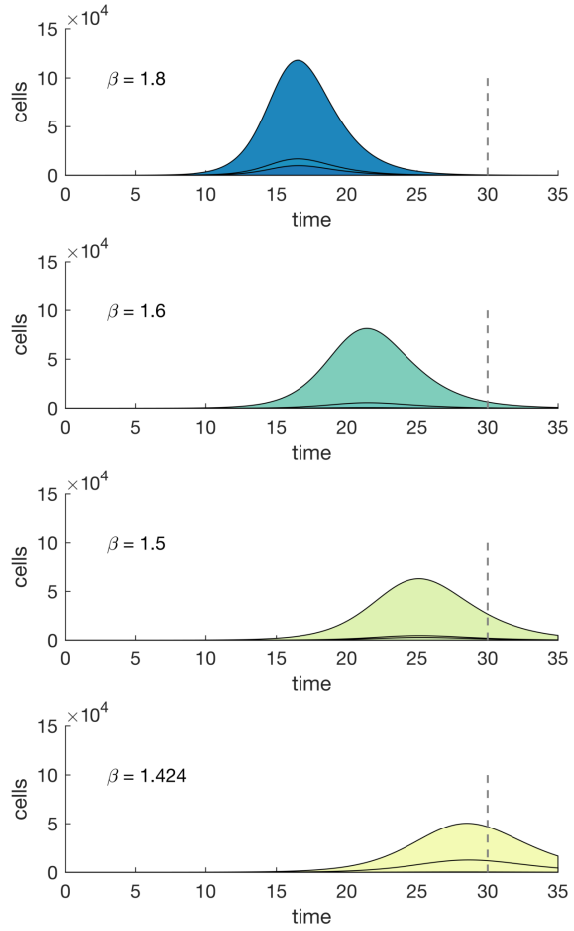

Figure 1: Realisations of within-patch dynamics with mutations for different initial cell growth rates. $\beta = 1.42$ is the optimal growth rate for $T = 30$. The curves are colour filled according to their growth rates as in the main text (darker blues represent faster rates and lighter greens slower rates). Other parameters: $N = 10^6$, $p = 10^{-2}$, $\nu = 0.5$ and $\mu = 0.02$.

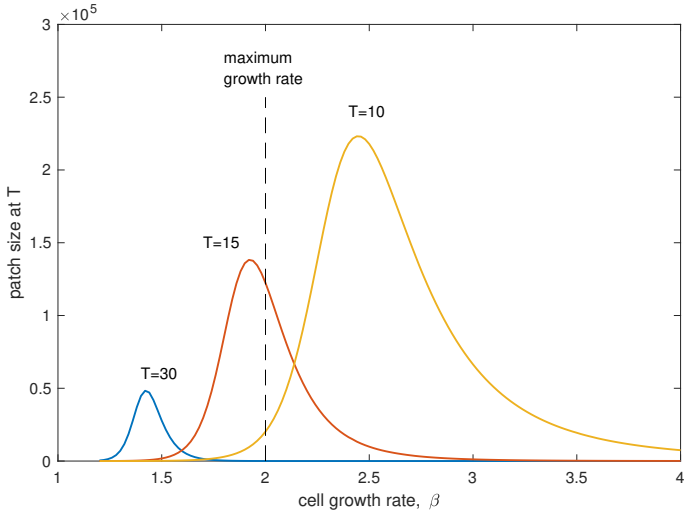

Figure 2: Fitness landscape for the single type model with no mutations. For the indicated dispersal times ($T = 10$, yellow; $T = 15$, red; and $T = 30$, blue) the patch size (population at the dispersal time) is plotted against the initial cell growth rate. The maximum possible growth rate that is enforced in our models is shown by the dashed lined.

2 Model with sterile types

We now extend the model from the previous section to also include a type of sterile cell, denoted S. When a G cells now reproduces, there is a probability, q, that instead of reproducing to create another G, an S is created instead. This is a simple form of stochastic developmental program. Each S cell depletes resources at a rate d and dies at rate γ; they are not able to reproduce themselves or be dispersed. Mutations that create new G types occur similarly to the previous model. If an S cell is not produced at a reproduction event (which happens with probability $1-q$) then with probability p a mutation happens to create a new G type. The phenotype of the G cells is now two-dimensional, quantified by the growth rate, β and the probability of S production, q. The individual cell dynamics can be summarised by the events and rates given in Table 1.

Event	Transition	Rate
Reproduction	$(E, G_i) \to (E-1, G_i+1)$	$(1-p)(1-q_i)\beta_i G_i E$
Mutation	$(E, G_i) \to (E-1, G_j+1)$	$p(1-q_i)\beta_i G_i E$
Creation of S	$(E, S) \to (E-1, S+1)$	$q_i\beta_i G_i E$
Depletion of E	$E \to E-1$	dSE
Death of G type	$G_i \to G_i - 1$	γG_i
Death of S	$S \to S-1$	γS

Table 1: Events, transitions and rates that define the model with sterile types. Only components of the state that change in a given transition are shown, all others remain fixed.

As the phenotype space is two-dimensional, the process of simulation for this model is much more involved than for the first. To make the model tractable we again make a piecewise approximation so that the creation of new types in the system is stochastic, but the growth dynamics are deterministic. As in the previous model, we also limit the number of mutant types to the first two created. The S cells only interact with G cells via competition for resources within a patch and influence dispersal when their numbers have grown to an appreciable proportion. Thus their growth is modelled as deterministic rather than stochastic.

When new mutant types are created, the mutant phenotype, (β_j, q_j), is related to the original, (β_i, q_i), by

$$\begin{aligned} \beta_j &= \beta_i + \beta_r \sin(2\pi u), \\ q_j &= q_i + q_r \cos(2\pi u), \end{aligned} \tag{9}$$

where $u \sim U(0,1)$ and β_r and q_r specify the magnitude of mutation. Essentially this creates a mutant with parameters in a random ellipse about the old values. When calculating the mutant type we also enforce that $0 \geq \beta_j > 2$ and $0 \geq q_j > 1$, where the upper limit on β is assumed due to physical constraints on the growth process. If a mutant is created with a phenotype outside these bounds then the new parameters are truncated to fall within the bounds above.

The approximate model is constructed as follows. As before, we define proportions of the constituents within a patch as $x_i(t) = G_i(t)/N$, $y(t) = E(t)/N$ and $z(t) = S(t)/N$. Mutations to create new types occur as an inhomogeneous Poisson process with rate

$$\lambda(t) = Npy(t) \sum_{i=1}^{m} (1-q_i)\beta_i x_i(t). \tag{10}$$

Between mutations, the dynamics evolve as,

$$\begin{aligned} \frac{dy}{dt} &= -y \sum_{i=1}^{m} \beta_i x_i - dyz, \\ \frac{dx_i}{dt} &= \beta_i(1-q_i) y x_i - \gamma x_i, \quad i = 1, \ldots, m(t), \\ \frac{dz}{dt} &= y \sum_{i=1}^{m} q_i \beta_i x_i - \gamma z. \end{aligned} \tag{11}$$

As $p \ll 1$, we have dropped the $1-p$ term that should appear in the second equation.

To facilitate solving for the mutation times, we add another equations for the time derivative of $\lambda(t)$ in terms of the other state variables,

$$\frac{d\lambda}{dt} = Npy \sum_{i=1}^{m} \beta_i (q_i - 1) x_i = \sum_{i=1}^{m} \left(\frac{dx_i}{dt} + \gamma x_i \right). \tag{12}$$

The mutation times are then generated in the same way as described in Section 1.1.

2.1 Dispersal process

For this model we assume that only the G cells can be dispersed. Hence, the bottleneck enforced by the dispersal mechanism means a patch is always seeded from a single G type cell.

We assume we have a system composed of $k = 1, \ldots, K$ patches. Then let $x_i^{(k)}$ be the proportion of type G_i and $y^{(k)}$ be the proportion of type S in patch k. Each patch is founded by a single G cell with phenotype $(\beta, q)^{(k)}$. These then reproduce, mutate, and create S cells until the dispersal time, T, at which point we sample from the resulting populations to create the next generation. This occurs in two steps:

1. Randomly select a patch k, in proportion to its weight, w_k, which is a function of the proportion of its constituents, $x_i^{(k)}$ and $y^{(k)}$.
2. From the patch selected in part 1, randomly select a G cell from the patch.

We consider two ways to assign a weight to patches for the first step. Most simply, if we take $w_k = \sum_{i=1}^{m} x_k$, then the dispersal process is as in the first model, i.e., the probability of choosing a patch is proportional only to the number of G cells, so patches with more G cells at the time of dispersal are more likely to be sampled. The other function we consider is

$$w_k = (1 + 200 y^{(k)}) \sum_{i=1}^{m} x_i^{(k)}. \tag{13}$$

This can be interpreted as the S cells aiding the dispersal from the patch, for example by attracting the dispersal agent. The constant multiplying the $y^{(k)}$ term represents the strength of this effect.

2.2 Dynamics

To understand the effect of the production of S cells on the overall dynamics, trajectories are shown in Figure 3 without mutation ($p = 0$) for two different values of q. We see that $q > 0$ leads to the production of S cells, but that the the number of G cells peaks at a later time and at a lower overall number. This is because production of S cells decreases the effective growth rate of G cells as well as consuming the resource used for growth (without reproducing themselves) so there is less for G cells to use for reproduction. If dispersal time is long compared with peak time, increasing q causes an increase in patch fitness by both increasing the number of G and S at the time of dispersal.

As with the previous model, because the number of mutant G cells in a patch is kept small by the bottleneck, we can adopt a fitness landscape approach for understanding the evolutionary dynamics. This is analogous to that shown in Figure 2, but now two-dimensional. Figure 4A and B show contour plots for the number of S and G in a patch at the time of dispersal (assumed slow, $T = 30$) as a function of the initial growth rate, β, and q. Figure 4C also shows the patch fitness calculated from Eq. (13). This shows that, as expected, the number of G cells is maximised by a growth rate of 1.45 and $q = 0$. Thus if patches are selected based only on the number of G cells, then this is the evolutionary outcome as seen in the simulations presented in the main text. In

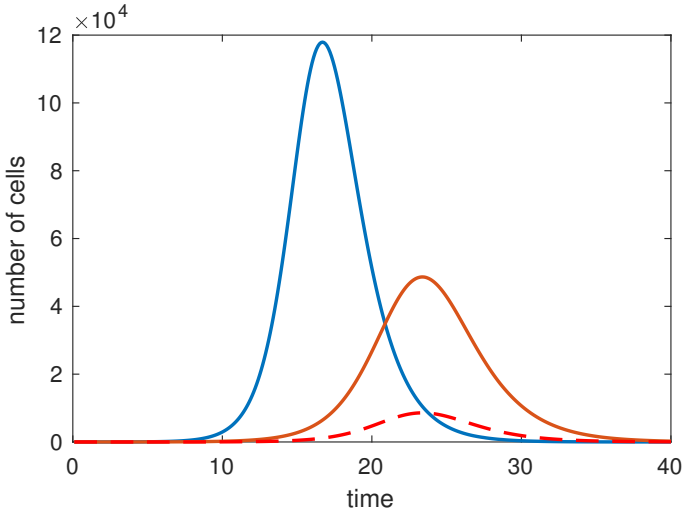

Figure 3: Trajectories from the model with sterile cells and $p = 0$. The blue line is the number of G cells with $\beta = 1.8$ and $q = 0$, hence no S cells are produced. The red lines show the number of G cells (solid) and S cells (dashed) for parameters $\beta = 1.8$ and $q = 0.15$. Other parameters: $N = 10^6$, $d = 2$ and $\gamma = 1$.

contrast to the behaviour shown in (A), the number of S cells is always maximised by increasing both q and β. If the patch fitness is a function of these two numbers then this is maximised at intermediate values (indicated by $*$). Also note that the landscape is flat in the direction of q around the maximum, which is why fluctuations about the steady state seen in the simulations are large.

Figure 5 shows the same quantities as Figure 4, but for fast dispersal ($T = 10$). The behaviour of the number of G and S cells follows a similar pattern to before, but at larger values of β. Figure 5C shows that if we impose a maximum growth rate on cells then patch fitness is maximised by maximising β and minimising q ($*$), but if no such maximum is imposed then a non-zero q and larger β maximises patch fitness ($*$). This is analogous to the behaviour shown in Figure 2, but in a two-dimensional setting.

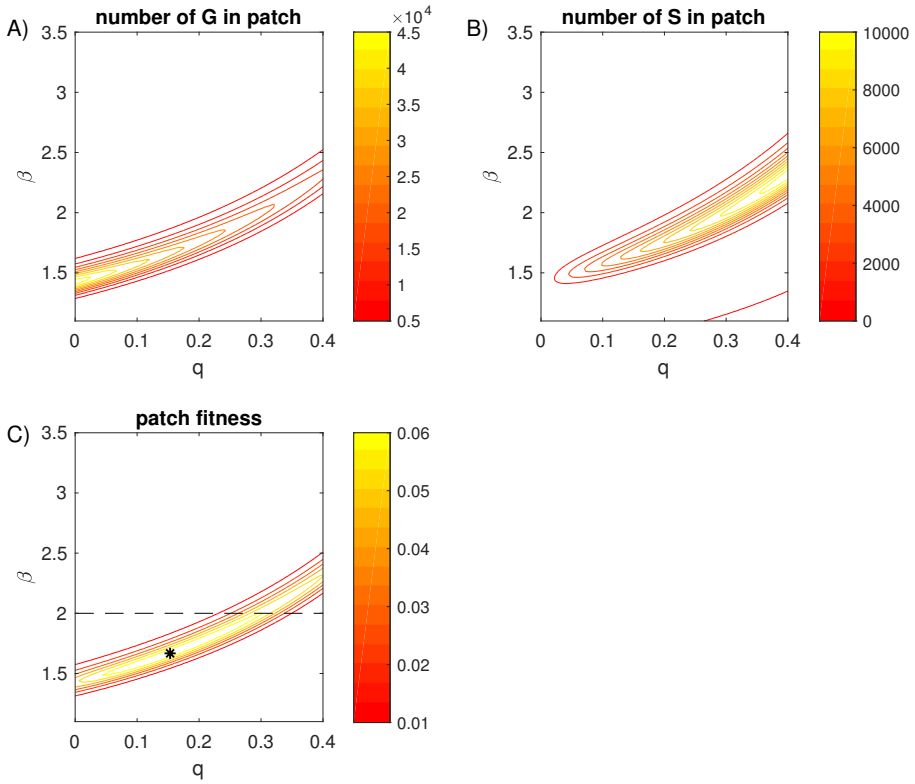

Figure 4: Fitness landscapes assuming slow dispersal ($T = 30$). (A) shows the number of G cells in a patch at the time of dispersal as a function of the initial growth rate, β, and q. (B) shows how the number of S cells at the time of dispersal changes as a function of β and q. (C) shows the patch fitness, calculated from Eq. (13). For slow dispersal we see this is maximised at $q = 0.15$, $\beta = 1.66$ $(*)$. Other parameters: $N = 10^6$, $d = 2$ and $\gamma = 1$.

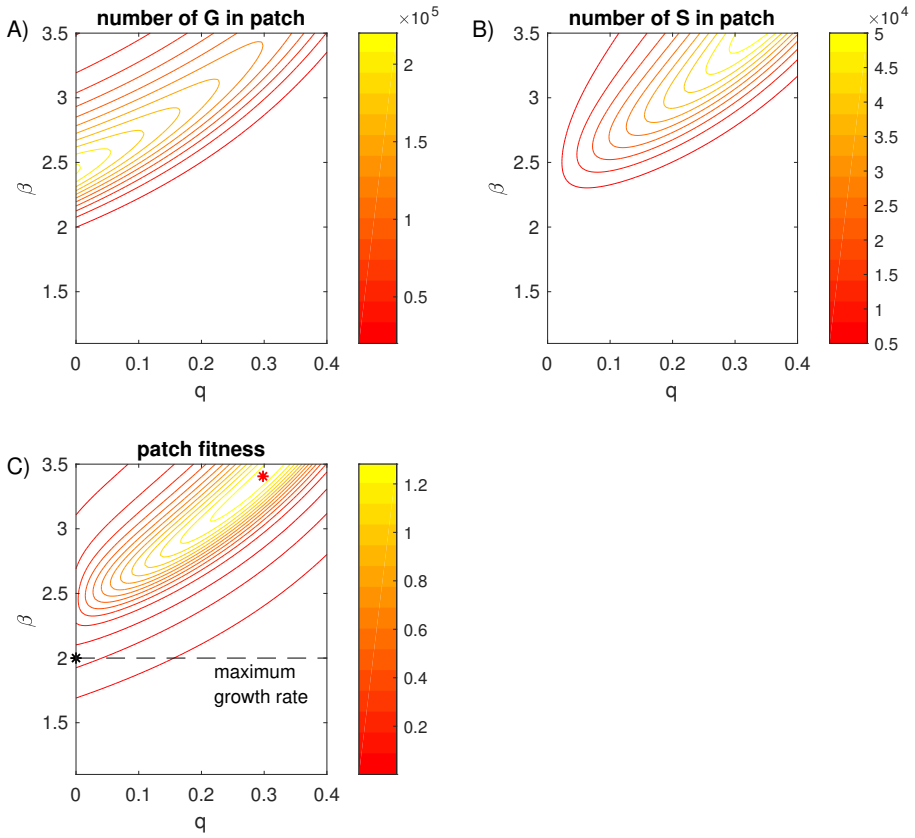

Figure 5: Fitness landscapes assuming fast dispersal ($T = 10$). This shows the same quantities as the Figure 4. If β is limited to a maximum of 2 (shown by the dashed line) then patch fitness is maximised at $\beta = 2$ and $q = 0$ ($*$) as seen in the simulations of the evolutionary model shown in the main text. If there was no limit to the growth rate then we would see evolution to a state with non-zero $q = 0.3$ and $\beta = 3.4$ ($*$).

References

[1] K. Ball, T. G. Kurtz, L. Popovic, and G. Rempala. Asymptotic analysis of multiscale approximations to reaction networks. *Ann. Appl. Probab.*, 16:1925–1961, 2006.

[2] A. J. Black and A. J. McKane. Stochastic formulation of ecological models and their applications. *Trends Ecol. Evo.*, 27:337–345, 2012.

[3] E. Çinlar. *Introduction to Stochastic Processes.* Prentice-Hall, New Jersey, 1975.

[4] M. H. A. Davis. Piecewise-deterministic Markov processes: a general class of nondiffusion stochastic models. *J. R. Statist. Soc. B*, 46:353–388, 1984.

[5] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22:403–434, 1976.

[6] S. Karlin and H. M. Taylor. *A First Course in Stochastic Processes.* Academic Press, New York, NY, 1975.

[7] D. P. Kroese, T. Taimre, and Z. I. Botev. *Handbook of Monte Carlo Methods.* Wiley, 2011.

[8] T. G. Kurtz. Solutions of ordinary differential equations as limits of pure jump markov processes. *J. Appl. Prob.*, 7:49–58, 1970.

[9] R. Pasupathy. Generating nonhomogeneous Poisson processes. *Wiley Encyclopedia of Operations Research and Management Science*, 2011.