

Dynamical Phase Transition in a Model for Evolution with Migration

Bartłomiej Waclaw, Rosalind J. Allen, and Martin R. Evans

SUPA, School of Physics and Astronomy, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, United Kingdom

(Received 3 August 2010; published 22 December 2010)

We study a simple quasispecies model for evolution in two different habitats, with different fitness landscapes, coupled through one-way migration. Our key finding is a dynamical phase transition at a critical value of the migration rate, at which the time to reach the steady state diverges. The genetic composition of the population is qualitatively different above and below the transition. Using results from localization theory, we show that the critical migration rate may be very small—demonstrating that evolutionary outcomes can be very sensitive to even a small amount of migration.

DOI: 10.1103/PhysRevLett.105.268101

PACS numbers: 87.23.Kg, 02.50.-r, 05.70.Fh, 87.10.-e

Biological dispersal—the movement of organisms between habitats—is a ubiquitous phenomenon with important and wide-ranging consequences. In the natural environment, organisms expand their ranges, colonize new habitats, and can undergo speciation if they become spatially isolated. Dispersal plays a key role in determining spatial and temporal patterns of genetic diversity in all organisms [1]. For sexual organisms, with low mutation rates, population subdivision into demes, connected by migration, can have important effects on genetic diversity [2,3], while in continuous space, transmission of unfit alleles can prevent the expansion of a species' range [4]. For asexual, rapidly evolving organisms such as bacteria and viruses, dispersal also facilitates the emergence of new diseases and resistance to known treatments. The “source-sink” paradigm [5,6], in which migration from a favorable habitat maintains organisms in an unfavorable one, has recently been used to explain the microbial genetics of urinary tract infections [7]. However, despite its importance, a general understanding of how migration affects mutation-selection balance in microbial systems is lacking. In particular, one would like to know how migration changes the proportions of different genotypes in evolving microbial populations.

In order to study the role of migration we introduce in this Letter a simple statistical physics model comprising two different environmental habitats coupled by oneway migration of asexual organisms from the primary to the secondary habitat as well as mutations, which give rise to clusters of closely related genotypes—“quasispecies.” We find that the model undergoes a dynamical phase transition: at a critical value of the migration rate, the time to reach the steady state diverges. For subcritical migration rates, the steady-state population in the secondary habitat is made up of a quasispecies “native” (best adapted) to this habitat, as well as other, nontrivial, quasispecies, which are not native to either habitat. Above the critical migration rate, the native quasispecies in the secondary habitat is wiped out by immigrants from the primary habitat. We use results from localization theory to gain further insight

into the transition and to show that even a small amount of migration can have an important effect on evolutionary dynamics.

In our model, organisms have M possible genotypes. N_i and n_i denote the abundance (number density) of organisms with genotype i in the primary and secondary habitat, respectively. The populations in the two habitats are thus described by the vectors $\vec{N} = \{N_i\}$ and $\vec{n} = \{n_i\}$. Organisms migrate from the primary to the secondary habitat with rate k . Within each habitat, mutations transform organisms from genotype i to j with rate γA_{ij} , where A_{ij} is a symmetric adjacency matrix, to be discussed later. Organisms of genotype i reproduce at a rate $\Phi_i - \sum_j N_j$ in the primary habitat and $\phi_i - \sum_j n_j$ in the secondary habitat. The vectors $\vec{\Phi} = \{\Phi_i\}$ and $\vec{\phi} = \{\phi_i\}$ thus describe the fitness landscapes (or the maximal growth rate for organisms with genotype i) in each habitat. The terms $-\sum_j N_j$ and $-\sum_j n_j$ in the growth rates account for population saturation due to finite resources, as in the logistic equation. This model is based on the para-mu-se (parallel mutation and selection) [8] version of quasispecies theory [9], widely discussed in the biological, chemical, and physical literature [10,11].

The time evolution of the system is governed by the following set of equations for $i = 1, \dots, M$:

$$\dot{N}_i = N_i \left(\Phi_i - \sum_j N_j \right) + \gamma \sum_j A_{ij} (N_j - N_i), \quad (1)$$

$$\dot{n}_i = n_i \left(\phi_i - \sum_j n_j \right) + \gamma \sum_j A_{ij} (n_j - n_i) + k N_i, \quad (2)$$

where we have assumed that the primary habitat is large, so that the loss of individuals due to migration has a negligible effect on its population [12]. For the calculations presented here, we suppose that the fitness values Φ_i , ϕ_i are independent random numbers drawn from a distribution $P(\varphi)$, common to both habitats. Thus genotypes which are well adapted in the primary habitat are likely to be maladapted in the secondary habitat. We also assume that the

mutation rate γ is below the error threshold [9,10], so that localized quasispecies are formed in both habitats in the absence of migration.

We first present an analytical solution for the steady state (which is independent of the initial condition for $k > 0$, as long as the mutation matrix A is irreducible). For the primary habitat, Eq. (1) can be linearized [8,9] by introducing new variables $\vec{X} = \vec{N} \exp(\int_0^t \sum_j N_j(t') dt')$. We have $d\vec{X}/dt = W\vec{X}$, where the matrix $W_{ij} = \delta_{ij}\Phi_i + \gamma\Delta_{ij}$, ($\Delta_{ij} = A_{ij} - \delta_{ij}\sum_k A_{ik}$ being the graph Laplacian). For large t , $\vec{X}(t) \cong e^{\Lambda_1 t} \vec{\Psi}_1$, where Λ_1 is the largest eigenvalue of W and $\vec{\Psi}_1$ is the corresponding eigenvector. Denoting $\vec{e} = (1, \dots, 1)$, the steady-state abundances \vec{N}^* are obtained by insertion into Eq. (1), yielding

$$\vec{N}^* = \frac{\Lambda_1}{(\vec{\Psi}_1^T \cdot \vec{e})} \vec{\Psi}_1. \quad (3)$$

We now determine the steady-state genotype abundances in the secondary habitat in a similar way by reducing Eq. (2) to a linear inhomogeneous equation $d\vec{x}/dt = V\vec{x} + k\vec{N}^* \exp(\int_0^t \sum_j n_j(t') dt')$, where $V_{ij} = \delta_{ij}\phi_i + \gamma\Delta_{ij}$ and $\vec{x} = \vec{n} \exp(\int_0^t \sum_j n_j(t') dt')$. We obtain for $t \rightarrow \infty$:

$$\vec{n}^* = k \sum_{\alpha=1}^M \frac{\vec{\psi}_\alpha^T \cdot \vec{N}^*}{n_{\text{tot}} - \lambda_\alpha} \vec{\psi}_\alpha, \quad (4)$$

where $\vec{\psi}_\alpha$ and λ_α are the eigenvectors and eigenvalues of V (ordered as $\lambda_1 > \lambda_2 > \dots$) and n_{tot} is the total steady-state population in the secondary habitat, which is determined self-consistently as the largest root of

$$n_{\text{tot}} = k \sum_{\alpha=1}^M \frac{(\vec{\psi}_\alpha^T \cdot \vec{N}^*)(\vec{\psi}_\alpha^T \cdot \vec{e})}{n_{\text{tot}} - \lambda_\alpha}. \quad (5)$$

The above equations hold for any fitness landscape and mutation matrix. Although most of our results are valid quite generally, in order to illustrate the most important features of the model, we now assume that the mutation graph is a one-dimensional closed chain, in which mutations are possible only between neighboring genotypes [i.e., $A_{ij} = 1$ if $i = (M + j \pm 1) \bmod M$, and zero otherwise]. We further suppose that the fitness can take only two values: 1 and 0 with probability p and $1 - p$, respectively. Since it has been suggested that viable genotypes form an interconnected network in genotype space [13], we shall consider the case $p \approx 1$, so that the fitness landscape is characterized by “islands” of fit genotypes separated by unfit ones.

Figure 1 shows how the steady-state abundances \vec{n}^* in the secondary habitat depend on the migration rate k . When k is very small, \vec{n}^* is peaked around the longest sequence of maximal fitness values: this peak corresponds to the “native” (or best-adapted) quasispecies for the secondary habitat. When k is very large (much larger than the mutation rate γ), the secondary habitat becomes

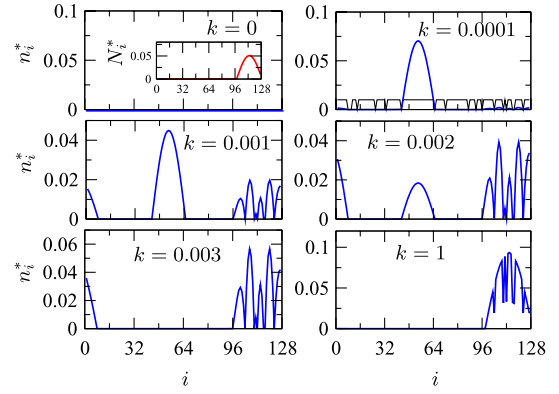


FIG. 1 (color online). Examples of steady-state genotype abundances in the secondary habitat n_i^* , for different values of the migration rate k where $p = 0.9$, $\gamma = 0.01$, and $M = 128$. These results were obtained by numerical self-consistent solution of Eqs. (3)–(5). The results agree with direct numerical solutions of Eqs. (1) and (2). The top left panel shows zero abundances ($n_i^* = 0$) in the absence of migration. The inset shows the abundances N_i^* in the primary habitat (red line). The top right panel shows the “native” steady state n_i^* (blue line), for very small migration rate, as well as the fitness landscape $\phi_i/100$ (black line). The other panels show n_i^* in the secondary habitat, for various values of the migration rate k .

dominated by immigrants from the primary habitat and \vec{n}^* tends to the primary-habitat steady state \vec{N}^* .

In contrast, for intermediate migration rates, the genetic composition in the secondary habitat is highly nontrivial. As k increases from zero, the quasispecies native to the secondary habitat is joined by additional quasispecies peaks. These are not native to either habitat but are instead determined by the overlap of eigenvectors in the primary and secondary habitats [as in Eq. (4)]. As the migration rate is increased slightly further from 0.002 to 0.003, these new peaks dominate completely and the native quasispecies of the secondary habitat disappears. This effect can be triggered by a very moderate change in the migration rate. The appearance of these new quasispecies peaks suggests that migration coupled to mutation can provide a mechanism for generation and maintenance of genetic diversity.

Figure 2 (left panel) shows that also the dynamics depends in a nontrivial way on the migration rate. The time to reach the steady state plotted as a function of k shows a striking maximum at $k_0 \approx 0.0027$, suggesting a critical slowing down and a likely dynamical phase transition. The approach to the steady state for $k \ll k_0$ is much slower than for $k \gg k_0$. Figure 3 illustrates the underlying reason for this. Here we plot snapshots of \vec{n} at various moments in time, for the same parameter set and fitness landscape, for migration rates below and above k_0 . For $k \ll k_0$, the immigrating population initially has the same composition as the primary habitat, but later the primary-habitat quasispecies peak is lost and the system undergoes a process of jumps between various local fitness maxima before finally settling in the global optimum. In contrast,

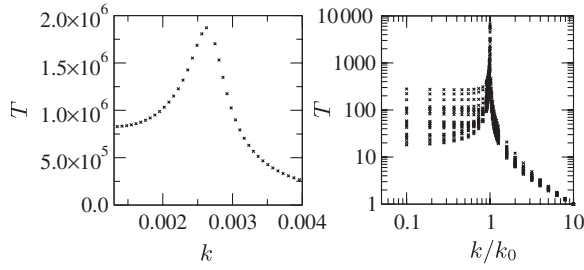


FIG. 2. Numerical results [from Eq. (2), using Eq. (3) for N_i^* and taking the secondary habitat to be initially empty] for the time T to reach the steady state, starting from $N_i = N_i^*$ and $n_i = 0$. Left panel: T as a function of migration rate k , for the same system as in Fig. 1. The steady state was assumed to have been reached when $\sum_i |n_i(t+1)/n_i(t) - 1|/M < 10^{-10}$. Right panel: $T(k/k_0)$, where k_0 is determined from Eq. (6), for $M = 64$, $\gamma = 0.01$, $p = 0.7$, normalized so that $T(10) = 1$. Results for 20 representative sets of $\vec{\Phi}$, $\vec{\phi}$ are presented on a log-log plot.

for $k \gg k_0$, the system rapidly relaxes to a steady state which overlaps strongly with that of the primary habitat.

Returning to our analytical expressions in the general case for \vec{n}^* , Eqs. (4) and (5), we can estimate the critical migration rate k_0 at which the dynamical phase transition takes place. Equation (4) expresses \vec{n}^* as a sum of eigenvectors $\vec{\psi}_\alpha$ for the secondary habitat, weighted by their overlap with \vec{N}^* . When $k \rightarrow 0$, $n_{\text{tot}} \rightarrow \lambda_1$, and $\vec{n}^* \rightarrow \vec{\psi}_1$. This is the native quasispecies solution for the secondary habitat. The phase transition occurs when this solution becomes dominated by the contributions from the other

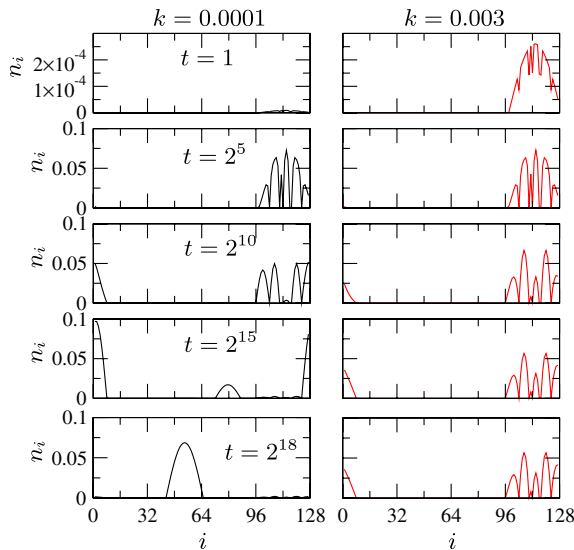


FIG. 3 (color online). Genetic composition in the secondary habitat $n_i(t)$, during the approach to steady state, for the same system as in Fig. 1, for two migration rates $k = 0.0001$ ($k \ll k_0$, black line, left) and $k = 0.003$ ($k \gg k_0$, red line, right, the same vertical scale), for $t = 1, \dots, 2^{18}$. The curves were obtained by solving Eqs. (1) and (2) numerically, taking the secondary habitat to be empty at $t = 0$.

terms [$\alpha = 2, \dots, M$ in Eqs. (4) and (5)], which arise from overlap with the primary-habitat solutions. This happens at a migration rate approximately given by

$$k_0 = \lambda_1 \left(\sum_{\alpha=2}^M \frac{(\vec{\psi}_\alpha^T \cdot \vec{N}^*)(\vec{\psi}_\alpha^T \cdot \vec{e})}{\lambda_1 - \lambda_\alpha} \right)^{-1}. \quad (6)$$

To show that this result indeed corresponds to the critical point, we will consider the matrix $J_{ij} = V_{ij} - \delta_{ij}n_{\text{tot}} - n_i^*$, which describes the dynamics of Eq. (2) linearized around the steady-state solution (4). The equation for the eigenvalues w of J_{ij} may be written as [14]

$$\sum_{\alpha=1}^M \frac{(\vec{\psi}_\alpha^T \cdot \vec{N}^*)(\vec{\psi}_\alpha^T \cdot \vec{e})}{n_{\text{tot}} - \lambda_\alpha} \left(\frac{1}{\lambda_\alpha - n_{\text{tot}} - w} - \frac{1}{n_{\text{tot}}} \right) = 0. \quad (7)$$

The relaxation time of the linearized system is given by $1/|w_1|$ where $w_1 < 0$ is the maximum eigenvalue of J_{ij} . In order to analyze the behavior of w_1 , we first determine the behavior of n_{tot} in the limit where $\delta \equiv (\vec{\psi}_1^T \cdot \vec{N}^*)(\vec{\psi}_1^T \cdot \vec{e})$ is small. Expanding Eq. (5) one obtains that $n_{\text{tot}} - \lambda_1$ is $O(\delta)$ for $k < k_0$, $O(\delta^{1/2})$ for $k = k_0$, and $O(1)$ for $k > k_0$. We see that the steady-state properties of the system (e.g., n_{tot}) behave very differently above and below the transition. Using the above behaviors of n_{tot} in (7) one finds $w_1 = O(1)$ for $k < k_0$ or $k > k_0$, whereas $w_1 = O(\delta^{1/2})$ for $k = k_0$. Thus in the limit $\vec{\psi}_1^T \vec{N}^* \rightarrow 0$, i.e., vanishing overlap between the quasispecies native to both habitats which is a generic situation for localized eigenvectors on random fitness landscapes, we have $w_1 \rightarrow 0$ at $k = k_0$ and the relaxation time diverges. Strictly, there is a true phase transition only in the limit $M \rightarrow \infty$, when $\vec{\psi}_1^T \vec{N}^* = 0$, but the diverging time scale is visible already for quite small systems, as shown in Fig. 2, right panel. This plot shows the time T to reach steady state as a function of k/k_0 , where k_0 is determined from (6). Each of ≈ 20 simulated fitness landscapes generates a slightly different curve $T(k/k_0)$, but all of the curves show a steep maximum at $k = k_0$.

For the one-dimensional model considered in our simulations, we can estimate the value of the critical migration rate k_0 . The eigenvector equation for $\vec{\Psi}_\alpha$ in the primary habitat maps onto a Schrödinger equation with random potential $U_j = -\Phi_j/\gamma$:

$$-(\Delta \vec{\Psi}_\alpha)_j + U_j(\vec{\Psi}_\alpha)_j = E_\alpha(\vec{\Psi}_\alpha)_j, \quad (8)$$

where $E_\alpha = -\Lambda_\alpha/\gamma$, and likewise for an eigenvector ψ_α in the secondary habitat. Equation (8) is essentially a 1D tight-binding electron model [15], in which $U_j = -1/\gamma$ with probability p and $U_j = 0$ with probability $1 - p$. Localization theory [16] tells us that for this problem the ground state eigenvector is localized, taking the form $\Psi_{1,j} \sim \sin(j\pi/w)$ on the longest run w of consecutive sites with $U_j = -1/\gamma$, and has eigenvalue $\Lambda_1 \approx 1 - \gamma\pi^2/w^2$. Eigenvectors corresponding to excited states are similarly localized on other, shorter potential wells. To estimate k_0 ,

we observe that the largest contribution to the sum in (6) comes from the eigenvector with the greatest overlap with \vec{N}^* , which we denote $\vec{\psi}_\beta$. Assuming that \vec{N}^* and $\vec{\psi}_\beta$ are localized on potential wells of length w and v , respectively, we can estimate that $(\vec{\psi}_\beta^T \cdot \vec{N}^*)(\vec{\psi}_\beta^T \cdot \vec{e}) \sim v/w$. The lengths w , v are the longest runs of $U_j = -1/\gamma$ in sequences of independent binary random numbers of length M and w , respectively; therefore, $w \simeq \ln(M(1-p))/\ln(1/p)$ and $v \simeq \ln(w(1-p))/\ln(1/p)$. For large M , v is much smaller than w , so $\lambda_1 - \lambda_\beta \simeq \gamma\pi^2/v^2$. Inserting this into Eq. (6), and setting $\epsilon = 1 - p$, we finally obtain

$$\langle k_0 \rangle \sim \gamma w/v^3 \simeq \frac{\gamma \epsilon^2 \ln(M\epsilon)}{(\ln \ln M \epsilon)^3}. \quad (9)$$

Remarkably, this rough estimate agrees up to a factor ≈ 2 with our simulation results. Here we have considered small ϵ , where multiple fit genomes lie close together in genotype space, and we see from (9) that $\langle k_0 \rangle$ is much smaller than γ for moderately large $M\epsilon$. This means that even low rates of migration can dramatically affect evolution in the secondary habitat [17].

In summary, we have shown that a dynamical phase transition occurs in a simple model for evolution with migration between two habitats. Bifurcations caused by migration have been observed in several models of sexual populations [3,4] but, to our knowledge, the present work is the first to consider the effects of migration on the evolutionary dynamics of asexual organisms from a quasispecies perspective. In our model, at the critical migration rate, the population in the secondary habitat becomes dominated by immigrants from the primary habitat. For subcritical migration rates, our quasispecies model also reveals that migration can provide a novel mechanism for creation and maintenance of genetic diversity.

To obtain analytical results and clear insights into the physics of the model, we have mainly considered a simple one-dimensional closed-chain representation of the genotype space and binary random fitness landscapes. As a step towards more complex and realistic representations of the genotype space and fitness landscape, we have also carried out numerical simulations for a continuous, uniform distribution of the fitness, as well as a hypercubic mutation graph. The latter corresponds to single-point mutations in a sequence of fixed length, with only two symbols allowed. Our key results (in particular the dynamical phase transition as a function of migration rate) remain valid in these cases, suggesting that our findings are likely to be of general significance, e.g., for sequences composed of 4 (nucleotide) or 20 (aminoacid) symbols. In particular, the phase transition will occur for any mutation matrix which allows for localized eigenstates, if the eigenvectors $\vec{\Psi}_1$, $\vec{\psi}_1$ in the two habitats do not overlap. It will be interesting to extend our work to empirical fitness landscapes generated from experimental data [18], and, inspired by existing

models for sexual organisms [3,4], and recent models in microbial ecology [19], to multiple connected habitats and spatially varying environments. Also, as advances in genomics make it possible to track microevolution of bacteria [20], we hope that the predictions made in this work will be verified experimentally, for example, in a system of coupled chemostats allowing well-controlled migration between different environments.

The authors are grateful to N. H. Barton, R. A. Blythe, M. Cates, A. Free, J.-M. Luck, and P. Warren for discussions. R. J. A. is supported by the Royal Society. This work was funded by the EPSRC under Grant No. EP/E030173.

-
- [1] A. K. Sakai *et al.*, *Annu. Rev. Ecol. Syst.* **32**, 305 (2001).
 - [2] R. Bürger, *J. Math. Biol.* **58**, 939 (2008); T. Nagylaki and Y. Lou, in *Tutorials in Mathematical Biosciences IV*, edited by A. Friedman (Springer, Berlin, 2008), Vol. 1922, p. 117, <http://www.springerlink.com/content/q34848r720459702/>.
 - [3] R. Lande, *Evolution (Lawrence, Kans.)* **33**, 234 (1979).
 - [4] M. Kirkpatrick and N. H. Barton, *Am. Nat.* **150**, 1 (1997).
 - [5] R. D. Holt, *Theor. Popul. Biol.* **28**, 181 (1985).
 - [6] H. R. Pulliam, *Am. Nat.* **132**, 652 (1988).
 - [7] S. Chattopadhyay *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 12 412 (2009).
 - [8] E. Baake and H. Wagner, *Genet. Res.* **78**, 93 (2001).
 - [9] M. Eigen and P. Schuster, *Naturwissenschaften* **64**, 541 (1977).
 - [10] C. O. Wilke, *BMC Evol. Biol.* **5**, 44 (2005); M. A. Nowak, *Trends Ecol. Evol.* **7**, 118 (1992).
 - [11] K. Jain and J. Krug, in *Structural Approaches to Sequence Evolution* (Springer, Berlin, 2007), p. 299, <http://www.springerlink.com/content/x286603435085643/>; L. Demetrius, *J. Chem. Phys.* **87**, 6939 (1987); I. Leuthäusser, *J. Stat. Phys.* **48**, 343 (1987); E. Baake, M. Baake, and H. Wagner, *Phys. Rev. Lett.* **78**, 559 (1997).
 - [12] The inclusion of an extra term $-kN_i$ in Eq. (1) to account for the loss of organisms from the primary habitat due to migration would be equivalent to adding a constant to the fitness landscape $\hat{\Phi}$ with no qualitative effect on the results.
 - [13] S. Gavrillets, *Am. Nat.* **154**, 1 (1999).
 - [14] See supplementary material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.105.268101> for a detailed proof.
 - [15] K. Ishii, *Prog. Theor. Phys. Suppl.* **53**, 77 (1973).
 - [16] I. M. Lifshitz, *Adv. Phys.* **13**, 483 (1964); *Sov. Phys. Usp.* **7**, 549 (1965).
 - [17] One can perform a similar analysis for $p \ll 1$, for which fit genomes are widely separated in genotype space, and check that $\langle k_0 \rangle \sim \lambda_1 = \sqrt{1 + 4\gamma^2} - 2\gamma$. The migration threshold is of order 1 in this limit, but the conclusion that migrants inundate better adapted genotypes at migration rate k_0 remains valid.
 - [18] D. M. Weinreich *et al.*, *Science* **312**, 111 (2006); F. J. Poelwijk *et al.*, *PLoS Comput. Biol.* **2**, e58 (2006).
 - [19] P. B. Warren, *Phys. Rev. E* **82**, 051922 (2010).
 - [20] S. R. Harris *et al.*, *Science* **327**, 469 (2010).