SL1. Motion suppresses selection on the cycle

Let $M$ be the transition matrix for an evolutionary process on $S^N$, where $S = \{ A, B \}$. For any state, $s$, let $|s|$ denote the number of mutants (type $A$) in $s$. We assume that for each $s \in S^N$, $M_{ss'} = 0$ whenever $|s'| - |s| \neq 0, \pm 1$ (which holds for the processes we consider in the main text). Thus, the Markov chain on $S^N$, which we denote by $\{X_n\}_{n \geq 0}$, has the property that at most one mutant is added to or subtracted from the population at each step. For any non-absorbing state, $s \in S^N$, we define the forward bias by

$$p_{s;+} := \sum_{s' \in S^N} \frac{M_{ss'}}{|s'| = |s| + 1}.$$

$$p_{s;-} := \sum_{s' \in S^N} \frac{M_{ss'}}{|s'| = |s| - 1}.$$

$$\gamma_s := \frac{p_{s;+}}{p_{s;-}}.$$

Let $\rho_s$ denote the probability of ending up in the all-mutant absorbing state when starting from state $s$. Recall that motion on the graph is modeled as a stochastic shuffle, $\mu \in \Delta(\mathcal{S}_N)$, at each time step, where $\mathcal{S}_N$ denotes the symmetric group on $N$ letters. (In general, $\mu$ can change at each update step.) If $\mu$ and $\nu$ are stochastic shuffles, then the sequence of $\mu$ followed by $\nu$, $\nu \circ \mu$, is again a stochastic shuffle, with

$$\nu \circ \mu (\pi) = \sum_{\sigma \in \Delta(\mathcal{S}_N)} \mu(\sigma) \nu (\pi) .$$

In words, the probability of realizing $\pi$ as an effective two-shuffle outcome is the probability of all paths (sequences) of shuffles of length two that lead to $\pi$. This property extends to sequences of shuffles of any length. Thus, even motion that is iterated many times between each time step can always be captured by a stochastic shuffle. With this definition in place, the main technical lemma we need is the following:

**Lemma 1.** Suppose that $P$ and $Q$ are Markov chains on $\{0, 1, \ldots, m\}$ with the property that $\gamma_s^P \leq \gamma_s \leq \gamma_s^Q$ for each $s \in S^N$. Then, for every non-absorbing initial state, $s \in S^N$, we have $\rho_s^P \leq \rho_s \leq \rho_s^Q$.

**Proof.** By the recurrence relation for fixation probabilities,

$$\sum_{s' \neq s} M_{ss'} \rho_s = \sum_{s' \neq s} M_{ss'} \rho_s'.$$

Therefore, whenever $s$ is not an absorbing state and $s' \neq s$, replacing $M_{ss'}$ by $\frac{1}{1-M_{ss'}}$ does not change these fixation probabilities; we may assume then that the probability of staying put in a given state is 0 in any Markov chain for which we care about fixation probabilities. In other words, we have

$$p_{s;+} = \frac{\gamma_s}{1 + \gamma_s};$$

$$p_{s;-} = \frac{1}{1 + \gamma_s}.$$
Consider the function \( f_P : \{0, 1, \ldots, N\} \rightarrow \mathbb{R} \) defined by
\[
f_P(m) = \begin{cases} 
0 & m = 0, \\
1 + \frac{1}{\gamma_1} + \frac{1}{\gamma_1 \gamma_2} + \cdots + \frac{1}{\gamma_1 \cdots \gamma_{m-1}} & m > 0.
\end{cases}
\]
(SI.5)

Since \( f_P(m+1) = f_P(m) + \frac{1}{\gamma_1 \cdots \gamma_m} \) for each \( m < N - 1 \), we have
\[
\mathbb{E} [f_P(X_{n+1}) \mid X_n = m] = \frac{\gamma_m^P f_P(m + 1) + \frac{1}{\gamma_m} f_P(m)}{1 + \gamma_m} = f_P(m).
\]
(SI.6)

Therefore, by Doob’s optional stopping theorem, we see that \( \rho_m^P = f_P(m) / f_P(N) \). Now, since
\[
\mathbb{E} [f_P(|X_{n+1}|) \mid X_n = s] = \frac{\gamma_s f_P(|s| + 1) + \frac{1}{\gamma_s} f_P(|s| - 1)}{1 + \gamma_s} 
\geq \frac{\gamma_s^P f_P(|s| + 1) + \frac{1}{\gamma_s} f_P(|s| - 1)}{1 + \gamma_s} = f_P(|s|),
\]
(SI.7)
it follows once again from Doob’s optional stopping theorem that \( \rho_s \geq f_P(|s|) / f_P(N) = \rho_1^Q \). An analogous argument with the inequalities reversed gives \( \rho_s \leq \rho_1^Q \), which completes the proof. □

**SI.1.1. Mixed BD and DB updating.** Consider a mixed update rule in which, at each time step, there is a DB update with probability \( \delta \) and a BD update with probability \( 1 - \delta \). For each state, \( s \), we have \( p^\delta_{s;+} = (1 - \delta) p^\BD_{s;+} + \delta p^\BD_{s;-} \) and \( p^\delta_{s;-} = (1 - \delta) p^\BD_{s;-} + \delta p^\BD_{s;+} \), so the forward bias is
\[
\gamma^\delta_s = \frac{(1 - \delta) p^\BD_{s;+} + \delta p^\BD_{s;-}}{(1 - \delta) p^\BD_{s;-} + \delta p^\BD_{s;+}}.
\]
(SI.8)

If \( 1 < |s| < N - 1 \) and \( s' \) is chosen so that \( |s'| = |s| \) but all mutants in \( s' \) are in a single cluster, then \( x_A(s') = x_B(s') = 0 \) and \( y_A(s') = y_B(s') = 1 \); a simple calculation then gives \( \gamma^\delta_{s'} = r \). If \( r \geq 1 \), then
\[
\gamma^\delta_s \leq \frac{(1 - \delta) \frac{x_A + y_A}{m + N - m} + \delta \left( \frac{x_B + \frac{2r}{1+r} y_B}{N - 2x_A - 2y_A} \right)}{(1 - \delta) \frac{x_A + y_A}{m + N - m} + \delta \left( \frac{x_A + \frac{2r}{1+r} y_A}{N - 2x_A - 2y_A} \right)} 
\leq \frac{(1 - \delta) \frac{x_A + y_A}{m + N - m} + \delta \left( \frac{2r}{1+r} \frac{x_B + y_B}{N - 2x_A - 2y_A} \right)}{(1 - \delta) \frac{x_A + y_A}{m + N - m} + \delta \left( \frac{2r}{1+r} \frac{x_A + y_A}{N - 2x_A - 2y_A} \right)} 
= r
\]
(SI.9)
since \( x_A + y_A = x_B + y_B \) on the cycle. Thus, \( \gamma^\delta_s \leq \gamma^\delta_{s'} \) when \( r \geq 1 \). Similarly, we see that \( \gamma^\delta_s \geq 1 \) when \( r \geq 1 \). When \( r \leq 1 \), we get the opposite inequalities, namely \( r = \gamma^\delta_{s'} \leq \gamma^\delta_s \leq 1 \). Therefore, for any \( r > 0 \),
\[
\min \{ \gamma^\delta_{s'}, 1 \} \leq \gamma^\delta_s \leq \max \{ \gamma^\delta_{s'}, 1 \}.
\]
(SI.10)
The biases for the two remaining cases, \( |s| = 1 \) and \( |s| = N - 1 \), are
\[
\gamma^\delta_s = \begin{cases} 
(1 - \delta) \frac{r}{m + N - m} + \delta \left( \frac{\frac{2r}{1+r}}{N - 2x_A - 2y_A} \right) & |s| = 1, \\
(1 - \delta) \frac{r}{m + N - m} + \delta \left( \frac{\frac{2r}{1+r}}{N - 2x_A - 2y_A} \right) & |s| = N - 1.
\end{cases}
\]
(SI.11)

While these biases need not be equal to \( r \), Eq. (SI.10) still holds since \( s = s' \) in these two cases.

It follows at once from Eq. (SI.10) and Lemma 1 that motion on the cycle suppresses selection under mixed BD and DB updating since any such (non-trivial) motion can disrupt clusters. In particular, this suppression holds for \( \delta = 0 \) (pure BD updating) and \( \delta = 1 \) (pure DB updating). Note, however, that for
δ = 0, i.e. pure BD updating, motion does not alter fixation probabilities on the cycle (but one could still formally consider this boundary case to be a “weak” suppressor of selection for consistency).

SI.2. Simulations on one- and two-dimensional lattices

SI.2.1. BD updating. We simulate BD updating on the cycle $10^6$ times for $N = 25$, thirty values of $r$ evenly distributed on $[0.5,1.5]$, and different values of $d$. At first, we constrain ourselves to one shuffle per update, $J = 1$. As predicted, the fixation probability is identical to the well-mixed result, but the times to fixation and absorption can be different. Generally speaking, when the motion is independent of the state and position of the individuals, the time to absorption or fixation is longer than in the well-mixed case. However, if $\mu$ depends on the state rather than being drawn from the uniform distribution, we can provide an algorithm for the shuffles that give the fastest dynamics, even faster than the well-mixed case. The fastest realizations for BD updating are those that minimize $p^0$ (the probability that the mutant count is neither increased nor decreased) at every update. It is easily seen that the desired $\mu(s)$ maximizes $c$ for all $s$, that is, creates as many clusters as possible. This can easily be accomplished by an algorithm that ensures no two mutants are adjacent when $m \leq N/2$, and no two resident types are adjacent when $m > N/2$. We plot $p^0$ versus $m$ for this strategy and the well-mixed case in Fig. SI.3. Since the biases are the same in the two cases, and because $p^0$ for the well-mixed case dominates $p^0$ for the algorithm, we actually expect the algorithm on the cluster to be faster than dynamics in the unstructured, well-mixed case (see Fig. SI.4).

SI.2.2. DB updating. Similarly, we simulate DB updating for the same parameters and iterations described in the previous section. The behavior of the conditional fixation time is shown in Fig. 2(b). Quantitative differences arise due to the fact that the forward bias is different under DB and BD updating when motion is present, as described in the main text. This difference in bias implies that the fixation probability for an invading mutant is different under DB updating when compared to the well-mixed and BD cases. The fixation probability for an invading mutant is plotted in Fig. 2(a). In the presence of random swapping, the fixation probability with motion is dominated by the fixation probability without motion.

In considering the lattice, all of the properties and parameters described in this and the previous section hold. When individuals are chosen to swap, the distance $d$ is measured in units of lattice steps, or in the $L^1$ (taxicab) norm. The probability that the mutant will fix is depicted in Fig. 3 in the main text.

SI.2.3. Multiple shuffles per update. When $J$ is large, it is more natural to examine the role of various quantities as the percentage of swapped sites, $J/N$, is varied; if $d \ll N$, it can take a very large number of swaps to accomplish the same relative mixing as a small number of swaps when $d = O(N)$. We fix values of $d$ and average over ten values of $N \in [25,50]$ in Fig. SI.5, which examines the dependence of the fixation probability and conditional fixation time under DB updating on the amount of random swapping (as well as the distance between swapped individuals). As discussed in the main text, the nature of the swapping implies that there is a critical value of $J/N$, beyond which there is no further effect in the average on fixation probabilities and times, which corresponds to the percolation limit. Depending on the value of $d$, we estimate that between 10 and 20 percent of individuals in our population must be swapped before this completely-randomized condition is met, leading to the fastest dynamics and lowest fixation probabilities.

SI.3. Weighted graphs

We now illustrate different types of motion on weighted graphs. For simplicity, we assume that the population is updated according to a BD rule; similar behavior can be observed DB updating. Suppose that $\Gamma$ is a weighted graph with $N$ vertices. To each pair of vertices, $i$ and $j$, $\Gamma$ associates a weight, $\Gamma_{ij} \geq 0$. BD updating on $\Gamma$ (without motion) is defined as follows [1]: First, an individual is chosen for reproduction with probability proportional to (relative) fitness. If $i$ is chosen to reproduce, then the offspring of $i$ replaces $j$ (who dies) with probability proportional to $\Gamma_{ij}$. Without a loss of generality, we may assume that $\sum_{j=1}^{N} \Gamma_{ij} = 1$ for each $i$, meaning $\Gamma_{ij}$ is the probability of offspring dispersal to location $j$ from location $i$.

Let $M$ be the transition matrix on $S^N$ induced by this process, where $S = \{A,B\}$. If $\mu \in \Delta(\mathcal{S}_N)$ is the shuffle used before each BD update, then the transition matrix with motion induced by $\mu$ is given by

$$M^\mu_{ss'} = \sum_{\pi \in \mathcal{S}_N} \mu(\pi) M_{\pi(s)s'}.$$  

(SI.12)
Let $\rho_s$ denote the probability of ending in the all-$A$ absorbing state when starting in state $s$. For each $s \in S^N$,

$$\rho_s = \sum_{s' \in S^N} M^h_{ss'} \rho_{s'},$$

(SI.13)

with the boundary conditions $\rho_{(A,A,...,A)} = 1$ and $\rho_{(B,B,...,B)} = 0$. For small $N$, one can solve for these fixation probabilities directly. Doing so gives Figs. 4 and 5 in the main text (where, by “fixation probability,” we mean the average fixation probability over all starting positions of the mutant; see Eq. (2.1)).

SI.4. Dynamic graphs

We simulate the Moran process on a dynamic graph induced by metric balls of radius $R$ on $N$ players, with the particular values $N = 49$ and $R = 1.5/7$, $3/7$, $4.5/7$. The values were picked based on conforming to a “thermodynamic” scaling $R = R^* / \sqrt{N}$ and approximating the values of $R$ such that the limit of a very dense (not quite complete) network was realized ($R = 4.5/7$), as well as a much sparser network ($R = 1.5/7$). This value was chosen so that the chance of a player being completely isolated (sharing no edges) was small. This probability cannot be eliminated on a finite dynamical graph, so in our simulations we have stipulated that any realization that passes through a transient state with an isolated player is discarded to avoid the possible bias induced by “sparing” death to an individual who cannot be replaced by any neighbors. Our results are from 25,000 realizations (without isolated players) for 25 values of $r \in [0.8, 1.5]$.

The ambient space that was used was the flat unit torus, that is, $[0, 1]^2$ with edges identified. In principle, any automorphism of the torus could be used to rearrange the players, and the study of the effect of different classes of automorphisms would be very interesting. As an initial study, however, we chose the map

$$\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_t \\ y_t \end{pmatrix} + \varepsilon \begin{pmatrix} \sin(2\pi x_t) \\ \sin(2\pi y_t) \end{pmatrix}.$$  

(SI.14)

This map has chaotic mixing properties and is a variant of V.I. Arnold’s CAT (continuous automorphism of the torus) map with a small perturbation term ($\varepsilon \ll 1$) to avoid the possibility of a stable fixed point for some initial conditions. The perturbation has no significant effect on the dynamics, which can be checked by omitting it and simulating the dynamics, taking great care to avoid the possibility of a fixed point “trapping” all the individuals. The map always has a positive Lyapunov exponent equal to the golden ratio, $\phi$, and is therefore chaotic for all initial conditions. There are no complicated regions of phase space such as elliptic islands that would be expected in genuine continuous fluid flows, so the map is, in some sense, very simple among the set of chaotic toral automorphisms. Its ergodic properties ensure that if the simulation runs forever (beyond the expected time to fixation of either species), each individual meets each other individual infinitely often. Such a map ensures that we are examining a very dynamic structure, with minimal possible correlation between edges in the adjacency matrices between two consecutive time steps.

The results are shown in Fig. 6 in the main text. As the connectivity of the graph decreases from that of a complete graph (through decreasing $R$), we observe suppression of selection due to the map.

References

Figure SI.1. Heterogeneity induced by motion. Without motion, the fixation probability of a mutant at vertex 2 is the same as that of a mutant at vertex 5 since the cycle is homogeneous. Once motion is introduced, these fixation probabilities need not coincide. As an example, we consider deterministic motion in which individual 2 acquires the type of individual 1, 3 acquires the type of 2, and 1 acquires the type of 3. In terms of the symmetric group, $S_6$, this motion is represented by the 3-cycle $(123)$. For DB updating when the mutant has fitness $r = 2$ relative to the resident, the fixation probability is $\approx 0.3351$ in (a) and $\approx 0.3703$ in (b), both rounded to four digits beyond the decimal point. Therefore, motion can introduce heterogeneity into a homogeneous graph.
Figure SI.2. Forward bias, $\gamma$, for DB updating is decreased by motion. The example is that of five mutants arranged in different configurations on any cycle with $N \geq 15$. Thick black and red lines show the neutral limit, corresponding to $\gamma = 1$, and the well-mixed limit, corresponding to $\gamma = r$, respectively. The blue lines give the maximum and minimum possible $\gamma$, realized when $x_A = 0; y_A = 1$ (a single cluster of mutants) and $x_A = 5; y_A = 0$ (five mutants isolated amongst large clusters of resident type), respectively. The shaded, light blue region demarcates some other possible values of $\gamma$ that can be realized when the five mutants are dispersed in other configurations.
Figure SI.3. Motion increases the frequency of birth and death events that change the abundance of types. The probability to neither create nor absorb a mutant under BD updating, $p^0$, is shown versus $m$ for $N = 20$. The black curve demonstrates the well-mixed case, whereas the red curve shows the fastest possible dynamics, given by an algorithm that maximizes $c$ for every value of $m$. 
Figure SI.4. An algorithm that uses motion to maximize the number of clusters, $c$, of mutants between every update step can accelerate the evolutionary process to the point that the conditional time to fixation is even less than that of a well-mixed (unstructured) population.
Figure S1.5. The effects of multiple shuffles on evolutionary dynamics on a cycle. Fixation probability (top) and conditional fixation time (bottom) are depicted for the death-birth process with a single mutant of relative fitness $r = 1.5$ in a cycle with motion. During each update step, an individual is selected at random and swaps with another individual who is at most $d$ loci away. This swapping procedure is repeated $J$ times before a reproduction event occurs. The curves shown have $d = 1$ (black circles), $d = 3$ (purple triangles), and $d = 8$ (blue squares). The horizontal axis denotes $J/N$, the frequency of shuffles. The value on the vertical axis is obtained by fixing $J/N$ and averaging over ten values of $N$ (to avoid the fact that a certain combination of $d$ and $J$ will have a stronger or weaker impact on evolutionary dynamics depending on the actual size, $N$, of the cycle). After a certain amount of shuffling, no further effect is seen, and the types are randomly distributed with uniform probability across the cycle, corresponding to the percolation limit.