Multiscale selection in spatially structured populations

Hilje M. Doekes^{1,2}, Rutger Hermsen^{1,3,*}

1 Theoretical Biology, Department of Biology, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.

2 Laboratory of Genetics, Department of Plant Sciences, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands.

3 Centre for Complex Systems Studies, Utrecht University, Leuvenlaan 4, 3584 CE Utrecht, The Netherlands.

* r.hermsen@uu.nl

Abstract

The spatial structure of natural populations is key to many of their evolutionary processes. Formal theories analysing the interplay between natural selection and spatial structure have mostly focused on populations divided into distinct, non-overlapping groups. Most populations, however, are not structured in this way, but rather (self-)organise into dynamic patterns unfolding at various spatial scales. Here, we present a mathematical framework that quantifies how patterns and processes at different spatial scales contribute to natural selection in such populations. To that end, we define the Local Selection Differential (LSD): a measure of the selection acting on a trait within a given local environment. Based on the LSD, natural selection in a population can be decomposed into two parts: the contribution of local selection, acting *within* local environments, and the contribution of interlocal selection, acting *among* them. Varying the size of the local environments subsequently allows one to measure the contribution of each length scale. To illustrate the use of this new *multiscale* selection framework, we apply it to two simulation models of the evolution of traits known to be affected by spatial population structure: altruism and pathogen transmissibility. In both models, the spatial decomposition of selection reveals

that local and interlocal selection can have opposite signs, thus providing a mathematically rigorous underpinning to intuitive explanations of how processes at different spatial scales may compete. It furthermore identifies which length scales—and hence which patterns—are relevant for natural selection. The multiscale selection framework can thus be used to address complex questions on evolution in spatially structured populations.

1 Introduction

2 Spatial structure is the rule, rather than the exception, in biological populations. Examples span a
3 large range of scales. Many bacteria live in biofilms that are highly heterogeneous [1–3] and in which
4 interactions between bacteria are often limited to a range of a few microns [4–6]. At the same time, bushes
5 growing in semi-arid areas form intricate vegetation patterns that span tens to hundreds of meters [7,
6 8]. Spatial population structure may reflect heterogeneities in the abiotic environment, such as resource
7 availability, but can also arise from self-organisation through ecological interactions between individuals
8 [9].

9 Because spatial population structure determines with whom organisms interact and compete, it is
10 a key factor shaping evolution. A classical example of this is the evolution of altruism: behaviour that
11 negatively affects an individual’s own fitness, but increases the fitness of its interaction partners [10, 11].
12 It has long been recognised that a non-arbitrary interaction structure is necessary for altruism to evolve,
13 so that the behaviour of altruistic individuals preferentially benefits other altruists [10, 12–14]. A natural
14 way for such interaction structure to arise is through local interactions and local reproduction, which
15 leads to spatial assortment with altruistic individuals generally being close to other altruists [2, 15–19].

16 Most formal theoretical work on how spatial structure affects evolution has focused on populations
17 that are divided in distinct groups, *e.g.*, trait-groups in which selection is compartmentalised for periods
18 of time [20–22]. Selection is then considered to act at two levels, *within* and *between* groups, and the
19 selection pressures at these two levels can be quantified [20]. Selection pressures within and between
20 groups are not necessarily aligned. In the case of altruism, for example, selection within groups tends to
21 favour selfish behaviour (also called cheating or defecting), while selection between groups promotes
22 altruism [12, 23, 24]. Many biological populations, however, are not subdivided into distinct groups,

but are nevertheless structured in space. In such populations, selection can depend on spatial scale, because local selection pressures may differ from those observed at the level of the whole population [25, 26]. Returning to the example of altruism: locally cheaters might out-compete altruists even if altruism is favoured in the population as a whole due to emergent spatial patterning [15, 16, 19]. A formal treatment of selection pressures in such multiscale (rather than multilevel) populations, however, is currently lacking.

Here, we present a mathematically rigorous way to quantify natural selection at different spatial scales. A spatial decomposition of selection is derived which splits global selection into a local component, which describes the average selection *within* local environments, and an interlocal component, which describes the selection *among* these environments. To illustrate the use of this new framework, we apply it to two computational models of the evolution of traits known to be affected by spatial structure: altruism and pathogen transmissibility. We show how the spatial decomposition of selection captures the contribution to selection of processes and patterns at different scales. The multiscale selection framework furthermore allows us to identify which spatial scales are relevant to natural selection, rather than defining those scales *a priori*.

Results

A spatial decomposition of selection

Consider a spatially structured population of individuals that differ with respect to some phenotypic value, ϕ . This could be a quantitative trait value (e.g., an individual's investment in altruistic behaviour) or an indicator variable that is 1 for individuals that display a certain phenotype (e.g., altruism) or possess a certain gene, and 0 for those who do not. We are interested in the evolution of the mean value of ϕ over time. Over fifty years ago, George R. Price derived a highly general mathematical description of evolutionary change, showing that the change in mean value of ϕ over a given time interval due to selection is equal to the covariance between the phenotypic value and the relative fitness, w , of individuals [27]. This covariance, called the *selection differential* and denoted by S , provides a general measure of the strength and direction of natural selection.

The selection differential of Price's equation describes the effect of selection at the level of the whole

Fig 1. Illustration of the spatial decomposition of selection. (a) Spatially structured population of individuals that differ in some phenotypic characteristic. Local environments are defined as circular areas with a given radius. (b) Example of global and local selection pointing in different directions. The covariance between phenotype and fitness *within* all local environments is negative (*i.e.*, local selection is negative), as evident from the negative slopes of the red regression lines; nevertheless, the global covariance between phenotype and fitness is positive (*i.e.*, global selection is positive), as apparent from the positive slope of the black regression line. This is an example of Simpson's paradox. (c) The negative local selection is counteracted by a positive covariance between the mean phenotype and mean fitness of local environments (see the blue regression line). Local environments are weighted by their population density and mean fitness (size of points). This covariance represents the selection *among* environments, *i.e.*, the interlocal selection.

population. In spatially structured populations, however, this may fail to capture the whole story. For example, consider the hypothetical population in Fig 1a. Here, at the global scale the covariance between phenotype and fitness is positive (black regression line in Fig 1b), yet if the analysis is restricted to individuals within smaller-scale local environments (circles in Fig 1a) it is invariably negative (red lines in Fig 1b). Counter-intuitively, the effect of natural selection can thus be to reduce the mean of ϕ in every local environment, while driving it up globally; a spatial Simpson's paradox [23, 28].

To quantify selection at these smaller scales, we first need a mathematically rigorous definition of local environments. In this paper, we simply define local environments as circular areas (disks) with a given radius r (see Methods for a more general definition based on a *kernel function*). For any point in space, the local selection at scale r can now be measured as the covariance between phenotype and relative fitness of the individuals found within the local environment centred on this point; we call this

the Local Selection Differential (LSD; see Methods). Note that local environments may overlap and that they are not necessarily centred on individuals (Fig 1a).

Using this definition of local selection, for any scale r we can derive the following spatial decomposition of the selection differential:

$$S = \mathbf{mean}(\mathbf{LSD}) + \mathbf{Cov}(\mathbf{local\ mean}(\phi), \mathbf{local\ mean}(w)) \\ \equiv S_{\text{local}}(r) + S_{\text{interlocal}}(r), \quad (1)$$

which is the central result of this article (see Fig 1). To properly calculate the mean and the covariance in Eq 1, local environments have to be weighted according to the local density of individuals. A detailed derivation is provided in the Methods; here, we focus on the interpretation of the terms of Eq 1. As the average of the LSDs over all local environments, $S_{\text{local}}(r)$ describes the *local* component of selection, measuring the selection *within* local environments. The second term $S_{\text{interlocal}}(r)$, the (weighted) covariance between local *mean* phenotype and local *mean* fitness, can be interpreted as the *interlocal* component of selection and represents selection *among* environments.

The measure of local selection, $S_{\text{local}}(r)$, captures the effect of anything that happens within local environments of size r . In other words, it incorporates all mechanisms operating at length scales smaller than or equal to r . To identify how a *specific* length scale r contributes to selection, we should ask how the local selection component changes if we slightly expand the scale of the local environments from r to $r + dr$. That is, the contribution to selection of scale r is captured by the derivative of $S_{\text{local}}(r)$ with respect to r , which we denote by $s(r)$. Assuming that no two individuals can be at the exact same position in space, so that $S_{\text{local}}(0) = 0$, we can then write

$$S = \int_0^\infty s(r') dr', \quad S_{\text{local}}(r) = \int_0^r s(r') dr', \quad \text{and} \quad S_{\text{interlocal}}(r) = \int_r^\infty s(r') dr', \quad (2)$$

which decomposes S , $S_{\text{local}}(r)$ and $S_{\text{interlocal}}(r)$ into contributions of different scales.

From Eq 2 we can derive several general insight about the local and interlocal selection components. If local environments become larger and larger, then $S_{\text{local}}(r)$ approaches S . This makes sense: large “local” environments should capture the global population dynamics. At the same time $S_{\text{interlocal}}(r)$ approaches 0, which can be understood by noting that as environments become larger and larger, the local mean values

of phenotype and fitness approach the global mean values, eliminating variation between environments. If the environments are made ever smaller ($r \rightarrow 0$), $S_{\text{local}}(r)$ approaches 0 and $S_{\text{interlocal}}(r)$ approaches S . This also makes sense: very small local environments lack variation in phenotype and fitness, and hence local selection must vanish.

To illustrate the use of the multiscale selection framework (Eq 1–2), we apply it to two models of classical examples of multiscale selection: (i) the evolution of altruism, and (ii) the evolution of pathogen transmissibility.

Example I: Evolution of altruism aided by self-organising colonies

We model a population of individuals in two-dimensional space that reproduce, die, move around slowly and locally compete for resources (Fig 2a; Supplementary Text). Individuals are characterised by a continuous trait that represents their investment in altruism. Upon reproduction, this trait value is passed on from parent to offspring, at which time mutations are introduced with small probability. Altruistic behaviour directly reduces an individual’s reproduction rate, but benefits all individuals in the local social environment of the altruist. The effects of altruism and resource competition both depend on the distance between organisms, such that an individual that is close to others benefits more from their altruistic action, but also experiences stronger competition (see Supplementary Text for details). The distances beyond which the effects of altruism or resource competition become weak – the “interaction scales” – are denoted by σ_a and σ_{rc} .

If the interaction scale of altruism, σ_a , is sufficiently smaller than the scale of resource competition, σ_{rc} , the model population shows intriguing self-organisation (Fig 2b, S1 Movie): a Turing-like instability results in a hexagonal pattern of distinct colonies that display Darwinian dynamics of their own. In colonies with a high mean level of altruism, the density of individuals is high because they all benefit from the altruism of colony members. Over time, however, the level of altruism within a colony declines because mutants with lower levels of altruism are selected (“defectors” or “cheaters”). This decline eventually results in the demise of the colony, after which it is replaced by a newly-formed colony that originates from the binary fission of one of the surrounding colonies (see arrow in Fig 2b, S1 Movie). Crucially, colonies with a higher mean level of altruism are more likely to reproduce. These emergent colony dynamics are studied in depth in a companion paper [29].

Fig 2. Evolution of altruism. (a) Model illustration. Both resource competition and altruism are local processes. The range of resource competition, σ_{rc} , is larger than the range of altruism, σ_a . (b) Snapshot of part of the simulation plane (see S1 Movie for dynamics). The hexagonal lattice constant of the emerged colony pattern is $a = 8.4 \sigma_a$. The green arrow indicates a colony fission event. (c) Mean level of altruism over time, in a population that is well-mixed (grey) or spatially structured (blue). (d) Spatial decomposition of selection differential S at evolutionary equilibrium for varying length scales of the local environments. Space was scaled such that the range of altruism $\sigma_a = 1$. $S_{local}(r)$ and $S_{interlocal}(r)$ were calculated as averages over 10 000 instances of the simulation plane obtained between Time = 36 000 and 40 000 generations (shaded area in panel c). (e) Contribution to selection of different length scales. Red areas indicate a negative contribution to selection, green a positive contribution.

The simulations are initialised with individuals that do not invest in altruism. But, over the course of the simulation, the mean level of altruism increases until it reaches a stable value (Fig 2c). In contrast, if we destroy the self-organised pattern by mixing the population (*i.e.*, randomly assigning positions to individuals every time step), altruism does not evolve at all (Fig 2c). The emergent spatial patterns are hence crucial for the evolution of altruism, consistent with previous modelling work [15–17, 30, 31].

Once the mean level of altruism has stabilised, we should expect the global selection differential to equal zero on average because no directional selection remains. To average out fluctuations arising from the stochastic dynamics in the finite population, we take the mean of the selection differential over 4 000 generations (shaded area in Fig 2c), and indeed find that over this period, $S \approx 0$ (black line in Fig 2d).

119 However, the spatial decomposition of selection of Eq 1 reveals a completely different picture (Fig 2d). For
 120 r -values up to six times the altruism interaction range, the mean local selection, as measured by $S_{\text{local}}(r)$, is
 121 negative. This negative local selection is compensated by positive interlocal selection, $S_{\text{interlocal}}(r)$. These
 122 results capture and formalise the verbal explanation given above: within local environments, individuals
 123 with a lower level of altruism are selected because they benefit from the altruistic behaviour of others
 124 nearby while paying less costs; but local environments in which the mean level of altruism is low also
 125 tend to have low mean fitness, *i.e.*, selection *among* local environments favours higher levels of altruism.

126 Fig 2d illustrate how $S_{\text{local}}(r)$ and $S_{\text{interlocal}}(r)$ depend on the radius of the local environments, r . The
 127 effects observed for small and large r -values represent general properties of the spatial decomposition
 128 of selection (see Eq 2): for large r , $S_{\text{local}}(r)$ converges to the global selection differential (which is close
 129 to zero in this case) and $S_{\text{interlocal}}(r)$ converges to zero, while for small r , $S_{\text{local}}(r)$ declines because the
 130 variation in phenotype and fitness within the local environments is reduced.

131 Intuitively, one might expect that the scale associated with negative selection on altruism is tied to
 132 the size of single colonies. The shortest distance between colonies is given by the lattice constant of
 133 the emerging hexagonal colony pattern, $a = 8.4\sigma_a$ (see Fig 2b). Indeed, for local environments with
 134 radius $r < \frac{a}{2}$ (such that most disks contain individuals of one colony only), $S_{\text{local}}(r)$ is clearly negative
 135 (Fig 2d). However, when we consider the contribution of each length scale to selection, $s(r)$, we see that
 136 only very small length scales of $r < \sigma_a$ contribute negatively to selection, while length scales of $r \gtrsim \sigma_a$
 137 contribute positively (Fig 2e). This indicates that colonies are not homogeneous: even within a single
 138 colony we observe assortment of individuals with different investment in altruistic behaviour, and this
 139 assortment contributes positively to the selection of altruism. This is understandable: individuals that are
 140 very close together (distance $< \sigma_a$) experience a similar level of altruism and competition. In very small
 141 local environments ($r < \sigma_a$), cheaters hence must have an advantage over altruists because they pay less
 142 costs. But once the scale of local environments become similar to the range of altruistic interactions,
 143 individuals within the same local environment may experience different levels of altruism, and the effects
 144 of these heterogeneities start to contribute to local selection.

145 Looking further into which spatial scales are relevant for selection, we see that the scale for which the
 146 differences between S_{local} and $S_{\text{interlocal}}$ vanish is close to the lattice constant a (Fig 2d). This indicates
 147 that local environments that are large enough to contain individuals from more than one colony capture

the processes and patterns contributing significantly to global selection are; larger-scale patterns, such as the clear assortment at the colony level (see Fig. 2b), appear to have a negligible effect.

In conclusion, the multiscale selection framework allows us to mathematically show that local selection for altruism is negative, but that this is compensated by positive interlocal selection. It furthermore provides a way to quantify how specific length scales contribute to selection, thus revealing which patterns and processes are significant for natural selection. Specifically, it shows that the heterogeneity of colonies is relevant to selection, whereas the assortment at the colony level is not.

Example II: Evolution of pathogen transmissibility in an SI-model

As a second example, we consider the evolution of the transmission rate of an endemic pathogen in a spatially structured population of host individuals. This model is rooted in a long tradition of such epidemiological models; see *e.g.*, [32–36].

In the model, host individuals live on a 2D square simulation lattice (Fig 3). They can be either susceptible to infection, or infected. Susceptible individuals reproduce asexually, in which case the offspring is placed on a neighbouring lattice site. Each lattice site can hold at most one individual; susceptible individuals therefore locally compete for empty space. Infected individuals do not reproduce, and they die at a higher rate than susceptible individuals. The pathogen is transmitted locally at a rate that varies among pathogen variants. We consider the evolution of this pathogen transmissibility. For simplicity, each infected individual is considered to carry a single pathogen strain, and mutations instantaneously change the transmissibility of all pathogens within a single infected host (*i.e.*, newly arising pathogen variants rapidly sweep the within-host pathogen population). Details on the model implementation are provided in the Supplementary Text.

After initialisation, the simulated population quickly self-organises into spatial patterns: the infection chases patches of susceptible individuals in wave-like structures (Fig 3b, S2 Movie and S3 Movie). These patterns strongly influence the evolution of pathogen transmissibility (Fig 3c): If pattern formation is prevented by constantly mixing the population, pathogens with ever-increasing transmissibility are selected because pathogens with higher transmissibility spread faster among the available susceptible hosts. In spatially structured populations, however, the mean transmissibility eventually stabilises (blue lines in Fig 3c). This is explained by a feedback between evolution and the emergent spatial patterns:

176 pathogen strains shape their local environment, and this environment in turn affects the strain's local
 177 fitness [9, 25, 37]. Specifically, pathogens with very high transmissibility are affected by 'self-shading' [34,
 178 38]: they rapidly deplete the susceptible hosts in their vicinity and are then left with little opportunity to
 179 spread, locally resulting in low average pathogen fitness. In contrast, more prudent pathogens shape
 180 their environment in such a way that sufficient susceptible hosts remain available to allow the infection
 181 to continue spreading [35, 39].

Fig 3. Evolution of pathogen transmissibility. (a) Model illustration. (b) Snapshot of part of the simulation lattice for two different values of the reproduction rate of susceptible individuals, γ . Susceptible individuals are plotted in grey, infected individuals are coloured based on the transmissibility of the pathogen they carry. See S2 Movie and S3 Movie for dynamics. (c) Mean transmissibility of the pathogen over time in populations that are well-mixed (grey) or spatially structured, with different initial transmissibility values (blue) (d) Spatial decomposition of selection differential S at evolutionary equilibrium for varying length scales of the local environments. For both values of γ , S_{local} and $S_{\text{interlocal}}$ were calculated as averages over 10 000 instances of the simulation lattice obtained between Time = 9 500 and 10 000 generations. We define the critical scale of selection, r_C , as the length scale at which the contribution to selection switches from positive to negative (i.e., where $s(r) = dS_{\text{local}}/dr$ switches sign). (e) Critical scale of selection, r_C , plotted against size of the emerged patterns for different values of the susceptible reproduction rate γ . Pattern size was determined using the pairwise correlation function (see Supplementary Text).

182 We use the multiscale selection framework to quantify and formalise this self-shading. As before,
183 we allow the population to reach evolutionary equilibrium and then calculate $S_{\text{local}}(r)$ and $S_{\text{interlocal}}(r)$
184 over a range of r values (Fig 3d). While the global selection differential S is insignificant as expected
185 (because the global mean transmissibility no longer changes and mutations are unbiased), within local
186 environments selection still favours pathogens with high transmissibility: $S_{\text{local}}(r) > 0$ for small r (red line
187 in Fig 3d). This effect is counteracted by negative interlocal selection: $S_{\text{interlocal}}(r) < 0$ for small r (blue
188 line). The negative interlocal selection confirms that pathogens with higher transmissibility more often
189 reside in local environments in which susceptible host availability and hence pathogen spread is limited.

190 We can now further explore how the spatial patterns affect the evolutionary process. The size of the
191 spatial patterns that emerge in the population depends on several model parameters [35] including the
192 reproduction rate of the susceptible hosts: lower reproduction rates result in larger patterns (Fig 3b). To
193 demonstrate how these larger patterns are reflected in S_{local} and $S_{\text{interlocal}}$, we repeated our analysis with
194 a lower susceptible reproduction rate (right panel in Fig 3b, pink and light-blue lines in Fig 3d). Evidently,
195 the curves representing $S_{\text{local}}(r)$ and $S_{\text{interlocal}}(r)$ are stretched towards larger scales. This reflects that the
196 spatial scales relevant to selection depend on the size of the patterns in the population: after all, selection
197 for pathogen restraint can only be observed if the local environments in which selection is measured are
198 large enough to cover multiple patches.

199 To explore this relationship between multiscale selection and spatial pattern size, we define the critical
200 scale of selection r_C as the length scale at which the contribution of length scales to selection, $s(r)$,
201 switches sign (Fig 3d), such that scales smaller than r_C contribute positively to selection, and scales larger
202 than r_C contribute negatively. This critical scale of selection is an emergent property of the dynamics. By
203 repeating the analysis for a range of susceptible growth rates, we find that the critical scale of selection
204 is proportional to the size of the emergent patterns (Fig 3e). Hence, the S_{local} - and $S_{\text{interlocal}}$ -curves, and
205 specifically the critical scale of selection r_C , capture the length scale of the spatial structures that are
206 relevant for natural selection.

207 In conclusion, this second example illustrates that the multiscale selection framework can be used to
208 measure and quantify self-shading. It furthermore allows the identification of the scale(s) of population
209 structures relevant to natural selection.

210 Discussion

211 We have presented a new, multiscale selection framework that can be used to analyse evolution in spatially
212 structured populations. The framework is based on a spatial decomposition of selection (Eq 1–2) that
213 quantifies local and interlocal selection for any spatial scale. Two example models illustrated how this
214 framework can be used to measure the contribution to selection of processes and patterns at varying
215 scales, and thus to identify the spatial scales relevant to natural selection.

216 The spatial decomposition of selection demonstrates how natural selection in local environments can
217 substantially differ from the selection in the global population, even if an average is taken over all local
218 environments. Evolutionary studies based on local observations can hence provide an incomplete picture
219 of the evolutionary dynamics in the global population. The framework presented here provides a way
220 of determining whether a collection of local sampling areas is representative of the whole population
221 under selection. Namely, if the local sampling areas are large enough to capture all spatial structures and
222 processes relevant to selection, interlocal selection vanishes and the mean local selection differential (*i.e.*,
223 as measured within sampling areas) is equal to the global selection differential. Specifically, this means that
224 there should be no covariance between mean phenotype and mean fitness among local environments (*i.e.*,
225 $S_{\text{interlocal}} = 0$). A significant correlation between local mean phenotype and local mean fitness is a strong
226 indication that larger structures exist in the population that contribute to natural selection. However,
227 to reliably estimate S_{local} and $S_{\text{interlocal}}$, observations from many individuals are required, especially if
228 selection is weak. While such data is easily obtained in simulation studies, it remains to be seen if the
229 framework can also be successfully applied to observational or experimental data.

230 The examples of multiscale selection studied here—the evolution of altruism and of pathogen transmissibility—
231 were chosen because they are among the best-known examples of a feedback between spatial patterns
232 and eco-evolutionary dynamics, and because this feedback has also been confirmed experimentally. For
233 instance, increased population viscosity facilitates the evolution of altruistic public good production in
234 lab populations of *Pseudomonas aeruginosa* [40, 41], and several experiments have shown that increased
235 host population viscosity and/or localised pathogen spread select for lower virulence in an insect larval
236 virus [38] and bacterial viruses [42, 43]. The effect of spatial structure on natural selection is, however,
237 by no means limited to these two examples. It was first described in models of catalytic hypercycles of
238 self-replicating molecules, which give rise to self-organised rotating spirals that select for higher death

239 rates of the individuals constituting these spirals [44, 45]. Since then, many other examples have been
240 described [9, 25], including anticompetitor toxin production in bacteria [46–50]. Applying the multiscale
241 selection framework to such other examples could lead to new insights in these systems as well.

242 The spatial decomposition of selection in Eq 1 is analogous to the decomposition of selection into
243 within- and between-group components for distinct, non-overlapping groups derived by Price in 1972
244 [20]. It however provides a different perspective on the effects of (spatial) structure on selection. The
245 multilevel framework (*i.e.*, the distinct-groups approach) requires that distinct and non-overlapping
246 groups are defined. If the spatial structures are variable in time and space, such as the patches in the
247 model of pathogen transmissibility, this is infeasible and the multiscale approach is more appropriate. If
248 distinct groups can be easily recognised, such as in the altruism model, the group selection framework
249 can provide important insights into the selection pressures acting at different levels of organisation. This
250 is explored in depth in [29]. But even in this case the multiscale approach may provide additional insights,
251 for instance by showing that relevant structure exists within groups or in the spatial organisation of the
252 groups themselves.

253 While the multiscale decomposition of the selection differential is new and uniquely untangles local
254 and interlocal selection in spatially structured populations, it is not the only, nor the only correct way the
255 selection differential can be decomposed. Next to the distinct-groups multilevel approach discussed above,
256 the selection differential can also be decomposed into terms that capture the effect of an individual's own
257 character on its fitness, and the effect of its environment (the contextual analysis approach to multilevel
258 selection) [51], or in terms that capture the effect of an individuals behaviour on its own fitness and on the
259 fitness of related interaction partners (the inclusive fitness framework) [26, 52, 53]. Each decomposition of
260 the selection differential tells a potentially new story about the underlying mechanisms driving evolution.
261 The multiscale selection framework presented here is an addition to the toolbox available to address
262 complex evolutionary questions.

263 **Methods**

264 Below, we derive Eq 1. Full specifications of the simulation models, as well as computational details on
265 the calculation of the two terms of Eq 1, are provided in the Supplementary Text.

266 **Background: Definition of the selection differential S**

267 Consider a population in space that at some time t consists of n individuals. Let ϕ_i be the phenotype of
268 individual i , and W_i its fitness, defined as the number of offspring at some later time $t + \Delta t$, including the
269 individual itself if it survived over the time step Δt . Price's equation [27] now states that

$$\underbrace{\Delta \bar{\phi}}_{\text{change in mean value of } \phi} = \underbrace{\text{Cov}(\phi, w)}_{\text{change due to selection}} + \underbrace{\overline{w \Delta \phi}}_{\text{change due to transmission biases}}, \quad (3)$$

270 where $w_i = W_i / \bar{W}$ is the relative fitness of individual i over the time interval (we use the common
271 notation $\bar{c}$ to denote the population mean of characteristic c) and $\Delta \phi$ captures biases in the transmission
272 of phenotypic values from parents to offspring. Importantly, the first term of the Price's equation reveals
273 that the effect of selection on the mean phenotype is captured by the covariance between phenotype
274 and fitness; this term is also called the *selection differential*. The analysis presented here focuses on the
275 selection differential only; for a full version of the Price equation with overlapping generations, including
276 transmission and survival-bias effects, see *e.g.*, [54].

277 **Measuring selection in a local environment: the Local Selection Differential (LSD)**

278 For any point $\mathbf{m}$ in space, the local population density is defined as a conventional kernel density estimate:

$$D(\mathbf{m} | r) \equiv \sum_{i=1}^n K(\mathbf{m} - \mathbf{x}_i | r). \quad (4)$$

279 Here $\mathbf{x}_i$ is the position of individual i . Its contribution to the density at position $\mathbf{m}$ depends on its distance
280 to position $\mathbf{m}$ according to the kernel function $K(\mathbf{y} | r)$. The parameter r is the scale parameter (or band
281 width) or the kernel function. In effect, the kernel function defines the local environments: it determines
282 which organisms contribute to what extent to the environment around each point in space.

283 In this paper, we use disk-shaped kernel functions which include an individual in the local environment
284 only if its distance to the environment's midpoint is smaller than a given radius, r . Other reasonable
285 choices for the kernel function include bi-variate normal or exponential distributions, which weigh
286 individuals close to the focal point more heavily than those further away.

287 A proper kernel function is normalised (*i.e.*, the integral of $K(\mathbf{y} | r)$ over space is equal to one); it

288 follows that

$$\int_S D(\mathbf{m} | r) d\mathbf{m} = n, \quad (5)$$

289 where $\int_S$ represents the integral over the entire space.

290 For any characteristic c of individuals, such as phenotype or fitness, we define the *local mean* at point

291 $\mathbf{m}$ as

$$\{c | \mathbf{m}, r\}_1 \equiv \frac{\sum_i K(\mathbf{m} - \mathbf{x}_i | r) c_i}{D(\mathbf{m} | r)}, \quad (6)$$

292 which is the average of c over all individuals, weighted by the kernel function based on their position

293 relative to $\mathbf{m}$. We will often write $\{c\}_1$ as a shorthand for $\{c | \mathbf{m}, r\}_1$ to avoid clutter.

Analogous to the selection differential in Price's equation, we can now define the local selection around point $\mathbf{m}$ as the covariance between phenotype and relative fitness within the local environment, which we call the *Local Selection Differential (LSD)*:

$$\begin{aligned} S_l(\mathbf{m} | r) &\equiv \frac{\{\phi W | \mathbf{m}, r\}_1 - \{\phi | \mathbf{m}, r\}_1 \{W | \mathbf{m}, r\}_1}{\{W | \mathbf{m}, r\}_1} \\ &\equiv \text{Cov}_1(\phi, W / \{W\}_1 | \mathbf{m}, r). \end{aligned} \quad (7)$$

294 Note that the LSD is equal to the local covariance between ϕ and local relative fitness ($W / \{W\}_1$), *i.e.*, the

295 fitness of an individual relative to others in the local environment. Thus defined, it measures the effect of

296 selection on the change in the local mean of ϕ at position $\mathbf{m}$, $\{\phi | \mathbf{m}, r\}_1$.

297 **Decomposing the selection differential into local and interlocal selection**

For any function over space $g(\mathbf{m})$, we define the *spatial mean* as

$$\begin{aligned} \langle g(\mathbf{m}) | r \rangle_s &\equiv \frac{\int_S D(\mathbf{m} | r) g(\mathbf{m}) d\mathbf{m}}{\int_S D(\mathbf{m} | r) d\mathbf{m}} \\ &= \frac{1}{n} \int_S D(\mathbf{m} | r) g(\mathbf{m}) d\mathbf{m}. \end{aligned} \quad (8)$$

298 Note that this represents the average of $g(\mathbf{m})$ over the complete space, but that each position $\mathbf{m}$ is

299 weighted by the local density $D(\mathbf{m} | r)$. This weighting is equivalent to the weighting by group size in

300 Price's derivation of within- and between-group selection in a population consisting of distinct groups

[20, 21]. For readability, we will often write $\langle g(\mathbf{m}) \rangle_s$ for $\langle g(\mathbf{m}) | r \rangle_s$. Conveniently, given the definitions of Eq 6 and 8, the spatial mean of the local mean is simply the population mean, i.e., $\langle \{c\}_1 \rangle_s = \bar{c}$.

Using these definitions, can now derive the desired decomposition:

$$\begin{aligned} S &= \text{Cov}(\phi, w) = \overline{\phi w} - \overline{\phi} \overline{w} \\ &= \langle \{\phi w\}_1 \rangle_s - \langle \{\phi\}_1 \rangle_s \langle \{w\}_1 \rangle_s \\ &= \langle \{\phi w\}_1 - \{\phi\}_1 \{w\}_1 \rangle_s + \langle \{\phi\}_1 \{w\}_1 \rangle_s - \langle \{\phi\}_1 \rangle_s \langle \{w\}_1 \rangle_s \\ &= \langle \text{Cov}_1(\phi, w | \mathbf{m}, r) \rangle_s + \text{Cov}_s(\{\phi\}_1, \{w\}_1) \\ &\equiv S_{\text{local}}(r) + S_{\text{interlocal}}(r). \end{aligned} \tag{9}$$

Hence, S_{local} is the average of the LSD over all local environments (defined by their midpoints $\mathbf{m}$ and scale r), where each environment is weighted by its local density and its local mean fitness:

$$S_{\text{local}}(r) = \langle \{w\}_1 S_l(\mathbf{m} | r) \rangle_s. \tag{10}$$

It captures the total effect of selection *within* local environments. On the other hand, $S_{\text{interlocal}}(r)$ is the spatial covariance (defined in terms of spatial means) of the local mean phenotype and local mean fitness. This term captures the selection *among* environments.

Code availability

Simulation codes and analysis scripts are available from GitHub:
<https://github.com/rutgerhermsen/altruism.git> (altruism model) and
https://github.com/hiljedoekes/MultiscaleSelection_SI (SI model).

Acknowledgements

We thank Reinder Bosman for a preliminary analysis of the SI-model, and Laura van Schijndel for valuable discussions and comments on the manuscript. This work was supported by the Human Frontier Science Program, grant nr. RGY0072/2015 (<http://www.hfsp.org/funding/research-grants>).

316 Author Contributions

317 R.H. conceived the project with input from H.M.D. The mathematical derivations were performed
318 collaboratively by both authors. R.H. developed and analysed the altruism model with input from H.M.D.;
319 H.M.D developed and analysed the SI-model with input from R.H. H.M.D. wrote the first draft of the
320 manuscript, which was then substantially edited by both authors.

321 References

- 322 1. Stacy, A., McNally, L., Darch, S. E., Brown, S. P. & Whiteley, M. The biogeography of polymicrobial
323 infection. *Nature Reviews Microbiology* **14**, 93–105 (2016).
- 324 2. Nadell, C. D., Drescher, K. & Foster, K. R. Spatial structure, cooperation and competition in biofilms.
325 *Nature Reviews Microbiology* **14**, 589–600 (2016).
- 326 3. Yanni, D., Márquez-Zacarias, P., Yunker, P. J. & Ratcliff, W. C. Drivers of Spatial Structure in Social
327 Microbial Communities. *Current Biology* **29**, R545–R550 (2019).
- 328 4. Dal Co, A., van Vliet, S., Kiviet, D. J., Schlegel, S. & Ackermann, M. Short-range interactions govern
329 the dynamics and functions of microbial communities. *Nature Ecology & Evolution* **4**, 366–375 (2020).
- 330 5. Krishna Kumar, R. *et al.* Droplet printing reveals the importance of micron-scale structure for
331 bacterial ecology. *Nature Communications* **12**, 857 (2021).
- 332 6. Van Gestel, J. *et al.* Short-range quorum sensing controls horizontal gene transfer at micron scale in
333 bacterial communities. *Nature Communications* **12**, 2324 (2021).
- 334 7. Thiery, J. M., D’Herbes, J.-M. & Valentin, C. A Model Simulating the Genesis of Banded Vegetation
335 Patterns in Niger. *Journal of Ecology* **83**, 497–507 (1995).
- 336 8. Rietkerk, M., Dekker, S. C., Ruiter, P. C. d. & Koppel, J. v. d. Self-Organized Patchiness and Catastrophic
337 Shifts in Ecosystems. *Science* **305**, 1926–1929 (2004).
- 338 9. Lion, S. & Baalen, M. v. Self-structuring in spatial evolutionary ecology. *Ecology Letters* **11**, 277–295
339 (2008).
- 340 10. Hamilton, W. D. The genetical evolution of social behaviour. II. *Journal of Theoretical Biology* **7**,
341 17–52 (1964).

- 342 11. West, S. A., Griffin, A. S., Gardner, A. & Diggle, S. P. Social evolution theory for microorganisms.
343 *Nature Reviews Microbiology* **4**, 597–607 (2006).
- 344 12. Wilson, D. S. A theory of group selection. *Proceedings of the National Academy of Sciences* **72**, 143–146
345 (1975).
- 346 13. Lehmann, L. & Keller, L. The evolution of cooperation and altruism – a general framework and a
347 classification of models. *Journal of Evolutionary Biology* **19**, 1365–1376 (2006).
- 348 14. Fletcher, J. A. & Doebeli, M. A Simple and General Explanation for the Evolution of Altruism.
349 *Proceedings: Biological Sciences* **276**, 13–19 (2009).
- 350 15. Nowak, M. A. & May, R. M. Evolutionary games and spatial chaos. *Nature* **359**, 826–829 (1992).
- 351 16. Killingback, T., Doebeli, M. & Knowlton, N. Variable investment, the Continuous Prisoner’s Dilemma,
352 and the origin of cooperation. *Proceedings of the Royal Society B: Biological Sciences* **266**, 1723–1728
353 (1999).
- 354 17. Le Galliard, J.-F., Ferrière, R. & Dieckmann, U. The Adaptive Dynamics of Altruism in Spatially
355 Heterogeneous Populations. *Evolution* **57**, 1–17 (2003).
- 356 18. Nadell, C. D., Foster, K. R. & Xavier, J. B. Emergence of Spatial Structure in Cell Groups and the
357 Evolution of Cooperation. *PLoS Computational Biology* **6** (2010).
- 358 19. Colizzi, E. S. & Hogeweg, P. High cost enhances cooperation through the interplay between evolution
359 and self-organisation. *BMC Evolutionary Biology* **16**, 31 (2016).
- 360 20. Price, G. R. Extension of covariance selection mathematics. *Annals of Human Genetics* **35**, 485–490
361 (1972).
- 362 21. Wade, M. J. Soft selection, hard selection, kin selection, and group selection. *The American Naturalist*
363 **125**, 61–73 (1985).
- 364 22. Okasha, S. *Evolution and the Levels of Selection* (Oxford University Press, 2006).
- 365 23. Chuang, J. S., Rivoire, O. & Leibler, S. Simpson’s Paradox in a Synthetic Microbial System. *Science*
366 **323**, 272–275 (2009).
- 367 24. Cremer, J., Melbinger, A. & Frey, E. Growth dynamics and the evolution of cooperation in microbial
368 populations. *Scientific Reports* **2** (2012).

- 369 25. Johnson, C. R. & Boerlijst, M. C. Selection at the level of the community: the importance of spatial
370 structure. *Trends in Ecology & Evolution* **17**, 83–90 (2002).
- 371 26. Lion, S., Jansen, V. A. A. & Day, T. Evolution in structured populations: beyond the kin versus group
372 debate. *Trends in Ecology & Evolution* **26**, 193–201 (2011).
- 373 27. Price, G. R. Selection and covariance. *Nature* **227**, 520–521 (1970).
- 374 28. Simpson, E. H. The Interpretation of Interaction in Contingency Tables. *Journal of the Royal Statistical
375 Society. Series B (Methodological)* **13**, 238–241 (1951).
- 376 29. Hermsen, R. Emergent multilevel selection in a simple spatial model of the evolution of altruism.
377 preprint available on bioRxiv; <https://doi.org/10.1101/2021.11.30.469839> (2021).
- 378 30. Wakano, J. Y., Nowak, M. A. & Hauert, C. Spatial dynamics of ecological public goods. *Proceedings
379 of the National Academy of Sciences* **106**, 7910–7914 (2009).
- 380 31. Uppal, G. & Vural, D. C. Evolution of specialized microbial cooperation in dynamic fluids. *Journal of
381 Evolutionary Biology* **33**, 256–269 (2020).
- 382 32. Rand, D. A., Keeling, M. & Wilson, H. B. Invasion, stability and evolution to criticality in spatially
383 extended, artificial host–pathogen ecologies. *Proceedings of the Royal Society of London. Series B:
384 Biological Sciences* **259**, 55–63 (1995).
- 385 33. Haraguchi, Y. & Sasaki, A. The Evolution of Parasite Virulence and Transmission Rate in a Spatially
386 Structured Population. *Journal of Theoretical Biology* **203**, 85–96 (2000).
- 387 34. Boots, M. & Sasaki, A. ‘Small worlds’ and the evolution of virulence: infection occurs locally and at
388 a distance. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **266**, 1933–1938
389 (1999).
- 390 35. Goodnight, C. *et al.* Evolution in spatial predator–prey models and the “prudent predator”: The
391 inadequacy of steady-state organism fitness and the concept of individual and group selection.
392 *Complexity* **13**, 23–44 (2008).
- 393 36. Heilmann, S., Sneppen, K. & Krishna, S. Sustainability of Virulence in a Phage-Bacterial Ecosystem.
394 *Journal of Virology* **84**, 3016–3022 (2010).

37. Savill, N. J., Rohani, P. & Hogeweg, P. Self-Reinforcing Spatial Patterns Enslave Evolution in a Host-Parasitoid System. *Journal of Theoretical Biology* **188**, 11–20 (1997).
38. Boots, M. & Meador, M. Local Interactions Select for Lower Pathogen Infectivity. *Science* **315**, 1284–1286 (2007).
39. Van Baalen, M. & Sabelis, M. W. The Milker-Killer Dilemma in Spatially Structured Predator-Prey Interactions. *Oikos* **74**, 391–400 (1995).
40. Diggle, S., Griffin, A., Campbell, G. & West, S. Cooperation and conflict in quorum-sensing bacterial populations. *Nature* **450**, 411–414 (2007).
41. Kümmerli, R., Griffin, A. S., West, S. A., Buckling, A. & Harrison, F. Viscous medium promotes cooperation in the pathogenic bacterium *Pseudomonas aeruginosa*. *Proceedings of the Royal Society B: Biological Sciences* **276**, 3531–3538 (2009).
42. Kerr, B., Neuhauser, C., Bohannan, B. J. M. & Dean, A. M. Local migration promotes competitive restraint in a host–pathogen ‘tragedy of the commons’. *Nature* **442**, 75–78 (2006).
43. Berngruber, T. W., Lion, S. & Gandon, S. Spatial Structure, Transmission Modes and the Evolution of Viral Exploitation Strategies. *PLOS Pathogens* **11**, e1004810 (2015).
44. Boerlijst, M. C. & Hogeweg, P. Self-structuring and selection: spiral waves as a substrate for evolution. *Artificial life II Proceedings Volume X*, 255–276 (1991).
45. Boerlijst, M. C. & Hogeweg, P. Spiral wave structure in pre-biotic evolution: Hypercycles stable against parasites. *Physica D: Nonlinear Phenomena* **48**, 17–28 (1991).
46. Chao, L. & Levin, B. R. Structured habitats and the evolution of anticompetitor toxins in bacteria. *Proceedings of the National Academy of Sciences* **78**, 6324–6328 (1981).
47. Durrett, R. & Levin, S. Allelopathy in Spatially Distributed Populations. *Journal of Theoretical Biology* **185**, 165–171 (1997).
48. Kerr, B., Riley, M. A., Feldman, M. W. & Bohannan, B. J. M. Local dispersal promotes biodiversity in a real-life game of rock–paper–scissors. *Nature* **418**, 171–174 (2002).
49. Bucci, V., Nadell, C. D., Xavier, J. B., Day, A. E. T. & McPeck, E. M. A. The Evolution of Bacteriocin Production in Bacterial Biofilms. *The American Naturalist* **178**, E162–E173 (2011).

- 422 50. Doekes, H. M., Boer, R. J. d. & Hermesen, R. Toxin production spontaneously becomes regulated by
423 local cell density in evolving bacterial populations. *PLOS Computational Biology* **15**, e1007333 (2019).
- 424 51. Damuth, J. & Heisler, I. L. Alternative formulations of multilevel selection. *Biology and Philosophy* **3**,
425 407–430 (1988).
- 426 52. Marshall, J. A. R. Group selection and kin selection: Formally equivalent approaches. *Trends in*
427 *Ecology and Evolution* **26**, 325–332 (2011).
- 428 53. Gardner, A. Price’s equation made clear. *Philosophical Transactions of the Royal Society B: Biological*
429 *Sciences* **375**, 20190361 (2020).
- 430 54. Day, T. & Bonduriansky, R. A Unified Approach to the Evolutionary Consequences of Genetic and
431 Nongenetic Inheritance. *The American Naturalist* **178**, E18–E36 (2011).
- 432 55. Frigo, M. & Johnson, S. The Design and Implementation of FFTW3. *Proceedings of the IEEE* **93**,
433 216–231 (2005).

Supplementary Information

Supplementary Movies

S1 Movie. Dynamics of the altruism model. Simulation lattice points are coloured by the mean level of altruism of the individuals at that position, according to the colour scale shown in the movie and in Fig 2b. Empty lattice points are white. A high-quality version of this video is shared here: <https://doi.org/10.5281/zenodo.5608897>.

S2 Movie. Dynamics of the SI-model if reproduction of susceptibles is fast ($\gamma = 0.05$). Empty lattice points are coloured black, susceptible individuals white, and infected individuals are coloured according to the transmissibility of the pathogen they carry (see colour scale in Fig 3b).

S3 Movie. Dynamics of the SI-model if reproduction of susceptibles is slow ($\gamma = 0.02$). Same colour scheme as S2 Movie.

Supplementary Text

Individual-based model of the evolution of altruism

As a first example of multiscale selection, we consider the evolution of altruism in a spatially explicit individual-based model. A full analysis of this model is provided in a companion paper [29].

Model description

We model a population of individuals living in a 2D space with periodic boundary conditions. Each individual can reproduce asexually, die, and move in an undirected fashion. Individuals differ by one continuous phenotype, ϕ , that specifies their level of altruism. By itself, altruistic behaviour is costly: it directly reduces the individual's rate of reproduction. However, an altruistic individual does improve the living conditions of all individuals in its local neighbourhood (including itself), irrespective of their phenotype. Individuals locally compete for resources: the rate of reproduction of an individual decreases with the density of individuals in its local neighbourhood.

Both altruistic interactions and competition are modelled through a bi-variate normal (Gaussian) interaction kernel function $G(\mathbf{y} | \sigma)$, where σ is the standard deviation of the kernel. In terms of this function, we denote the total level of altruism experienced by individuals at position $\mathbf{y}$ by

$$A(\mathbf{y} | \sigma_a) \equiv \sum_{i=1}^n \phi_i G(\mathbf{x}_i - \mathbf{y} | \sigma_a). \quad (11)$$

Here, n is the population size. Although we do not intend to model a particular altruistic behaviour, it is convenient to envision $A(\mathbf{y} | \sigma_a)$ as the availability of a public good that is locally deposited by altruistic organisms. The scale σ_a can be interpreted as the interaction range or scale of altruism. Similarly, the resource competition experienced at position $\mathbf{y}$ is defined as

$$D_{rc}(\mathbf{y} | \sigma_{rc}) \equiv \sum_{i=1}^n G(\mathbf{x}_i - \mathbf{y} | \sigma_{rc}), \quad (12)$$

where σ_{rc} is the range or scale of resource competition. Note that $D_{rc}(\mathbf{y} | \sigma_{rc})$ is a kernel density estimate based on a Gaussian kernel with band width σ_{rc} ; in the derivation of the LSD the local population density was defined in the same way.

In the simulations, space is discretised using a regular square lattice with lattice constant δx , and time proceeds in discrete steps δt . The simulation construct the state of the system at time $t + \delta t$ through the following sequence of events:

1. Reproduction Each individual i that exists at time t reproduces with probability $g_i \delta t$, where the reproduction rate g_i is given by

$$g_i = g_0 \max \left[\left(1 - \underbrace{c \phi_i}_{\text{cost of altruism}} + \underbrace{\frac{b_{\max} A(\mathbf{x}_i | \sigma_a)}{b_{\max}/b_0 + A(\mathbf{x}_i | \sigma_a)}}_{\text{benefit from altruists nearby}} \right) \underbrace{\left(1 - \frac{D_{rc}(\mathbf{x}_i | \sigma_{rc})}{K} \right)}_{\text{competition}}, 0 \right]. \quad (13)$$

Here, g_0 is the basal reproduction rate, c the reproductive deficit per investment in altruism, $\mathbf{x}_i$ the position of the individual, and K is a factor scaling the local carrying capacity. The benefit of altruism is an increasing, saturating function of the local availability of “public good”: it is 0 if the local environment does not contain any altruists, and $b_{\max}$ if $A(\mathbf{x}_i | \sigma_a) \rightarrow \infty$. When an individual reproduces, a new organism is born and placed at the same lattice point as its parent.

477 **2. Heredity and mutation** The phenotype of the child is copied from its parent, but mutated with
 478 probability μ . The effect size of the mutation is drawn from an exponential probability distribution
 479 with mean m ; its sign is chosen positive or negative with equal probability. If a mutation would
 480 result in a negative phenotype, the phenotype is set to zero.

481 **3. Death** Individuals die with probability $d\delta t$. The death rate d is independent of phenotype.

482 **4. Motility** All individuals are displaced by adding a random number to both coordinates of their
 483 position. These random numbers are independently drawn from a discretised normal distribution
 484 with standard deviation $\sigma_d \equiv \sqrt{2k_D\delta t}$. This approximates movement by diffusion, with diffusion
 485 constant k_D .

486 Units and parameters

487 In the formulation above, we are free to choose the scaling of the phenotype ϕ ; this allows one to eliminate
 488 one parameter. We choose to set the cost to $c = 1$. We can eliminate another parameter by choosing
 489 suitable units of length; we choose to set $\sigma_a = 1$, so that all lengths become expressed in units of the
 490 interaction range of altruism. A last parameter can be eliminated by choosing convenient units of time;
 491 we set the death rate $d = 1$, so that time is measured in generations.

492 Other parameter values used in the simulations shown in the main text are listed in Table 1. While
 493 some parameters were chosen arbitrarily, the following rationale was followed: To resolve the altruistic
 494 interactions, the spacing of the lattice needs to be smaller than the range of altruism, $\sigma_a = 1$. We therefore
 495 set the lattice spacing to $\delta x = 0.1$. Similarly, to ensure that g_i and d can indeed be interpreted as rates,
 496 the time step should be chosen such that $g_i\delta t$ and $d\delta t$ are considerably smaller than 1. We therefore set
 497 $\delta t = 0.08$. The Turing-like instability causing the colony formation occurs only if the range of competition
 498 is significantly larger than the range of altruism (further explored in [29]); we chose $\sigma_{rc} = 4$. Since the
 499 evolution of altruism is favoured by positive genetic assortment (*i.e.*, close proximity of parents and
 500 offspring) [10], a low diffusion constant was chosen. Lastly, the parameter b_0 was chosen to ensure that
 501 the direct benefit reaped by an isolated individual from its own altruism is considerably smaller than
 502 the cost incurred. (Otherwise, the term altruism does not apply.) The benefit term in Eq 13 is bounded
 503 by $b_0 A(\mathbf{x}_i | \sigma_a) = b_0 \phi_i / (2\pi)$; hence, by choosing $b_0 = 1$, the benefit is guaranteed to be at least a factor 2π

Table 1. Parameters of altruism model

<i>Parameter</i>	<i>Description</i>	<i>Value</i>
σ_a	Scale of altruism	1 (by definition)
σ_{rc}	Scale of resource competition	4
g_0	Basal reproduction rate	5
c	Reproductive deficit of altruism	1 (by definition)
b_0	Basal reproductive benefit of altruism	1
$b_{\max}$	Maximal reproductive benefit of altruism	5
K	Scaling factor local carrying capacity	40
d	Death rate	1 (by definition)
k_D	Diffusion constant defining motility	0.04
μ	Mutation probability upon reproduction	10^{-3}
m	Mean effect size of mutations	5×10^{-3}
δx	Simulation lattice constant	0.1
δt	Simulation time steps	0.08
X	Linear size of the simulation lattice	1024 lattice points

smaller than the cost $c\phi_i = \phi_i$.

SI-model of the evolution of pathogen transmissibility

As a second example, we consider the evolution of pathogen transmissibility in a spatial SI-model. This model is based on several previous simulation models exploring pathogen evolution, see [32–35].

Model description

Consider a population of individuals living on a 2D square lattice. Each lattice point contains at most one individual. Individuals reproduce, die and move randomly. Individuals can be either susceptible to infection with a pathogen, or infected. Infected individuals no longer reproduce, and die at a higher rate than susceptible individuals. The infection spreads locally; the probability that an infected individual transmits the pathogen to a susceptible neighbour depends on the transmissibility of the pathogen that they carry, ϕ . We consider the evolution of this transmissibility.

Time in the model progresses in discrete time steps. Every time step, the following series of events

516 takes place:

517 **1. Reproduction** Since each lattice point can be occupied by at most one individual, individuals can
 518 reproduce only if one or more of their neighbouring lattice points is empty. Empty lattice points are
 519 repopulated by a susceptible individual with probability γn_S , where γ is the reproduction rate per
 520 susceptible individual and n_S is the number of susceptible individuals among the eight neighbours
 521 (including diagonals) surrounding the empty lattice point.

522 **2. Infection** For each susceptible individual i , let I_i be the set of infected individuals found on any of the
 523 eight lattice points surrounding the susceptible individual's position. Then individual i becomes
 524 infected with probability

$$P_i = \frac{\sum_{j \in I_i} \phi_j}{h + \sum_{j \in I_i} \phi_j}. \quad (14)$$

525 This infection probability is zero when the susceptible individual has no infected neighbours, and
 526 approaches 1 if many of the neighbours are infected with highly transmissible pathogens. If an
 527 infection takes place, one of the infected neighbours is chosen as the source of this new infection;
 528 the probability that a neighbour is chosen is proportional to the transmissibility of its pathogen, ϕ_j .
 529 The newly infected individual inherits the transmissibility value of the chosen infected neighbour;

530 **3. Mutation** Since we consider mutation of the pathogen rather than of the host individuals, mutations
 531 are not coupled to reproduction events of the hosts. Instead, each time step the transmissibility of
 532 the pathogen within each infected individual is mutated with a small probability μ . This resembles a
 533 within-host mutation of the pathogen that (instantly) sweeps the within-host pathogen population.
 534 If a mutation occurs, a new transmissibility ϕ'_i is chosen from a uniform distribution $[\phi_i - \lambda; \phi_i + \lambda]$.
 535 If $\phi'_i < 0$, it is set to zero.

536 **4. Death** Susceptible individuals die with probability δ_S . Infected individuals die with probability $\delta_I > \delta_S$.

537 **5. Motility** In a randomly assigned order, the contents of each lattice point are swapped with a randomly
 538 chosen neighbour with probability p_m . Individuals can hence change position either because they
 539 move themselves (with probability p_m) or because they are chosen in the swapping procedure
 540 of one of their neighbouring lattice points (empty or containing an individual). This procedure

Table 2. Parameters of SI model

<i>Parameter</i>	<i>Description</i>	<i>Value</i>
γ	Reproduction rate (per susceptible per time step)	0.05
δ_S	Death rate of susceptible individuals (per time step)	0.05
δ_I	Death rate of infected individuals (per time step)	0.2
h	Scaling factor of infection probability	20
p_m	Motility probability (per time step)	0.05
μ	Mutation rate (per time step)	0.005
λ	Maximal mutation step size	0.5
	Simulation lattice size	1024×1024 lattice points

approximates movement by random walk, while ensuring that the number of individuals does not change and that the rate of movement per individual does not depend on local population density.

Parameters

An overview of all model parameters and their values is provided in Table 2. Reproduction and death probabilities were set to values significantly smaller than 1 per time step, so that these probabilities can be interpreted to good approximation as rates of a Poisson process. To allow spatial patterns to arise, a small value was chosen for the motility probability p_m .

Scale of emergent spatial patterns

To measure the size of the emergent spatial patterns, we determine the pair correlation function $g(r)$ of a snapshot of the simulation. This function measures the mean density of individuals at distance r from a lattice point, given that this lattice point is occupied by some individual. It thus indicates whether the occupancy of a lattice point correlates with the occupancy of lattice points a distance r away. The function is normalised by the average density of the field. Hence, if $g(r) = 1$, the probability of finding other individuals at distance r from a given focal individual is equal to the probability of finding individuals in any random position; if $g(r) > 1$ the probability of finding other individuals at distance r from a focal individual is larger than one would expect from random sampling of the population.

For patch structures such as the emergent patterns studied here, $g(r)$ can be approximated by an

558 exponential function

$$g(r) = 1 + \exp(a - br). \quad (15)$$

559 Eq 15 was used to fit the observed pair correlation functions using least-squares fitting. The fitted
560 exponent b^{-1} was used as a measure of the size of the patches emerging in the field.

561 **Implementation of the calculations of $S_{\text{local}}(r)$ and $S_{\text{interlocal}}(r)$**

562 In the Methods section, we gave a general analytical derivation of the spatial decomposition of the
563 selection differential S into local and interlocal components S_{local} and $S_{\text{interlocal}}$. Below, we describe how
564 these quantities were calculated in practice.

565 When deriving the spatial decomposition of selection (Methods), we considered a population in
566 continuous space. In the simulations, however, space is discretised. This is easily dealt with by replacing
567 the integral over space in Eq 8 with a double summation over the coordinates of the 2D simulation space.

568 For both models, simulations were performed using periodic boundary conditions (a common choice
569 for these kinds of simulations). To calculate local population densities (Eq 4) one then must not only
570 take the individuals in the simulation lattice into account, but also their periodic images. The circular
571 convolution theorem provides an efficient way to do this using Discrete Fourier Transforms (DFTs). This
572 is a standard technique from signal processing; we shortly discuss it here for completeness.

573 For ease of notation we here present derivations for a 1D space; the 2D case follows analogously.

574 Assume that a sequence $o_0, o_1, \dots, o_{X-1}$ specifies the occupancy (number of individuals) at discrete
575 positions $j \in \{0, 1, \dots, X-1\}$ in a simulation field, and let $n = \sum_{j=0}^{X-1} o_j$ be the total number of individuals.
576 Because we assumed periodic boundary conditions, we should consider this finite field to represent a
577 finite stretch taken from an infinite field that is periodic with period X ; that is, the full field, including
578 periodic images, is a sequence $\tilde{o}_j$ with $j \in \mathbb{Z}$ defined by $\tilde{o}_j = o_{(j \bmod X)}$.

579 On the discretised space, the kernel function is represented by an infinite non-negative sequence
580 k_j with $j \in \mathbb{Z}$, whose sum converges to 1. Because the kernel function is a function of distance, it is
581 symmetric, i.e., $k_{-j} = k_j$. The kernel density d_j for positions $j \in \{0, 1, \dots, X-1\}$ (c.f. Eq 4) is now given by

$$d_j = \sum_{m=-\infty}^{\infty} \tilde{o}_m k_{(j-m)}. \quad (16)$$

Eq 16 can be rewritten as follows:

$$\begin{aligned}
 d_j &= \sum_{p=-\infty}^{\infty} \sum_{l=0}^{X-1} \tilde{o}_{l+pX} k_{j-l-pX} \\
 &= \sum_{l=0}^{X-1} \tilde{o}_l \sum_{p=-\infty}^{\infty} k_{j-l-pX} \\
 &= \sum_{l=0}^{X-1} \tilde{o}_l \hat{k}_{(j-l) \bmod X},
 \end{aligned} \tag{17}$$

where we introduced the *periodic summation* $\hat{k}_j$ for $j \in \{0, 1, \dots, X-1\}$ as

$$\hat{k}_j \equiv \sum_{p=-\infty}^{\infty} k_{j-pX}. \tag{18}$$

The sequences $\tilde{o}_j$ and $\hat{k}_j$ are both periodic with period X . Eq 17 is the *circular convolution* of these two sequences, which we denote by $\tilde{o} *_X \hat{k}$. By the *circular convolution theorem*, the DFT of $\tilde{o} *_X \hat{k}$ is equal to the element-wise product of the DFTs of $\tilde{o}$ and $\hat{k}$, i.e.,

$$\mathcal{F}(d_j) = \mathcal{F}(\tilde{o} *_X \hat{k}) = \mathcal{F}(\tilde{o}_j) \cdot \mathcal{F}(\hat{k}_j). \tag{19}$$

Using Eq 18 and 19, we can calculate the local density d_j by

1. calculating the periodic summation $\hat{k}_j$ of k_j ;
2. calculating the DFTs of $\hat{k}_j$ and $\tilde{o}_j$, and calculating their element-wise product;
3. calculating the inverse DFT of this element-wise product, which yields d_j .

Because algorithms for Fourier transformation are highly efficient [55], this procedure allows us to rapidly solve local densities for varying scales r . Fourier transformations were performed using the `fftw3`-library [55].

The local mean of a property c of individuals, $\{c|j, r\}_1$ (Eq 6), can be calculated in a very similar way. Let $\tilde{c}_j$ be the sequence of the sum of c -values of individuals at position j . In terms of this sequence, we can write

$$\{c|j, r\}_1 = \frac{\sum_{l=0}^{X-1} \tilde{c}_j \hat{k}_{(j-l) \bmod X}}{d_j} \tag{20}$$

596 In other words, the total local value of c , $\{c|j, r\}_{1,\text{tot}} \equiv d_j \{c|j, r\}_1$, is given by the circular convolution
 597 of the sequences $\tilde{c}$ and $\hat{k}$, and can hence be easily calculated using Fourier transformation as described
 598 above. To find the local mean, this total local value is divided by the local density, d_j . Note that, for some
 599 positions j , the local density might be very small or zero. To avoid numerical problems, division by local
 600 density is treated with caution in the calculations of $S_{\text{local}}(r)$ and $S_{\text{interlocal}}(r)$ (see below).

By definition, $S_{\text{local}}(r)$ is the spatial mean of the LSDs, weighted by the local mean fitness $\{w\}_1$. It is therefore possible to calculate $S_{\text{local}}(r)$ directly from the local selection differential, $S_1((x, y) | r)$:

$$\begin{aligned} S_{\text{local}}(r) &\equiv \langle \{w\}_1 S_1(j | r) \rangle_s \\ &= \frac{1}{n} \sum_j d_j \{w\}_1 S_1(j | r). \end{aligned} \quad (21)$$

However, noting that $\langle \{\phi w\}_1 \rangle_s = \overline{\phi w}$ we can also calculate $S_{\text{local}}(r)$ using Eq 10:

$$\begin{aligned} S_{\text{local}}(r) &= \langle \{\phi w\}_1 - \{\phi\}_1 \{w\}_1 \rangle_s \\ &= \overline{\phi w} - \langle \{\phi\}_1 \{w\}_1 \rangle_s \\ &= \overline{\phi w} - \frac{1}{n} \sum_j \frac{\{\phi\}_{1,\text{tot}} \{w\}_{1,\text{tot}}}{d_j}, \end{aligned} \quad (22)$$

601 which is less likely to cause numerical inaccuracies (if the local density $d_j = 0$ for some position j , the
 602 corresponding term in the summation is set to zero).

Using the definition of Eq 9, the value of $S_{\text{interlocal}}(r)$ can be calculated directly as the covariance between $\{\phi|j, r\}_1$ and $\{w|j, r\}_1$, in which each position j is weighted by the local density d_j . However, we can again derive an expression that is numerically more stable:

$$\begin{aligned} S_{\text{interlocal}}(r) &= \langle \{\phi\}_1 \{w\}_1 \rangle_s - \overline{\phi w} \\ &= \langle (\{\phi\}_1 - \overline{\phi}) (\{w\}_1 - \overline{w}) \rangle_s \\ &= \frac{1}{n} \sum_j \frac{(\{\phi\}_{1,\text{tot}} - d_j \overline{\phi})(\{w\}_{1,\text{tot}} - d_j \overline{w})}{d_j}, \end{aligned} \quad (23)$$

603 where again the summation term for index j is set to zero if the local density $d_j = 0$. Eq 22 and 23 were
 604 used in the calculations.