1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29

**Tipping the mutation-selection balance:
Limited migration increases the frequency of deleterious mutants**

Jacob D Cooper ^{*,†}, Claudia Neuhauser[‡], Antony M Dean[§], Benjamin Kerr ^{*,†}

^{*} *Department of Biology, University of Washington, Seattle, WA*

[†] *BEACON Center for the Study of Evolution in Action, University of Washington, Seattle, WA*

[‡] *Biomedical Informatics and Computational Biology, University of Minnesota, Rochester, MN*

[§] *College of Ecology and Evolution, Sun Yat-sen University, Guangzhou, China*

30 **Running title:**
31 Tipping the Mutation-Selection Balance
32

33 **Corresponding author:**
34 Jacob D Cooper
35 University of Washington
36 Department of Biology
37 Box 351800
38 Seattle, WA 98195-1800

39 phone: (206) 221-7026

40 email: yankel@uw.edu
41

42 **Keywords:**
43 population genetics
44 fitness landscape
45 adaptive valley crossing
46 spatial structure
47 moment closure
48
49
50
51
52
53
54

55 **Abstract:** Typical mutation-selection models assume well-mixed populations, but dispersal and
56 migration within many natural populations is spatially limited. Such limitations can lead to enhanced
57 variation among locations as different types become clustered in different places. Such clustering
58 weakens competition between unlike types relative to competition between like types; thus, the rate by
59 which a fitter type displaces an inferior competitor can be affected by the spatial scale of movement. In
60 this paper, we use a birth-death model to show that limited migration can affect asexual populations by
61 creating competitive refugia. We use a moment closure approach to show that as population structure
62 is introduced by limiting migration, the equilibrium frequency of deleterious mutants increases. We
63 support and extend the model through stochastic simulation, and we use a spatially explicit cellular
64 automaton approach to corroborate the results. We discuss the implications of these results for standing
65 variation in structured populations and adaptive valley crossing in Wright’s “shifting balance” process.

78 Most mutations affecting fitness appear to be deleterious (see review by Eyre-Walker and Keightley,
79 2007). A deleterious mutation is expected to persist in a population at a level influenced by the rate at
80 which it is generated and the strength of selection against it. This mutation-selection balance was first
81 developed mathematically by Haldane and Fisher in the 1920's in models that assumed well-mixed
82 populations (Fisher, 1930; Haldane, 1927). However, many natural populations are not well mixed:
83 individuals may not disperse, and even if they do, dispersal or migration is often restricted to nearby
84 locations (Evans et al., 2009; Howells et al., 2013; Martin and Canham, 2010). Such limited movement
85 may influence the proportion of deleterious mutants at equilibrium in several ways. In mating diploid
86 populations, the Wahlund effect (in which population-level heterozygosity is depressed when
87 subpopulations differ in allele frequency) combines with dominance relationships among genotypes to
88 influence the frequency of deleterious mutant alleles (Roze and Rousset, 2004; Whitlock, 2002). In
89 haploid asexual models, limiting migration increases between-deme variation and decreases within-
90 deme variation, but the extent to which this shift in variation affects evolution is unclear.

91 Limitations to migration are not predicted to affect the equilibrium frequency of deleterious
92 mutants in asexual populations when fitness is independent of local composition and density. For
93 instance, Whitlock (2002) finds no effect of migration under a "hard selection" scheme (in which
94 absolute fitness is determined solely by genotype, and thus demes of different compositions may differ
95 in productivity). However, in "soft selection" regimes (in which relative fitness within a deme depends
96 on genotype, but each deme's productivity is the same regardless of composition), demes enriched for
97 mutants are as productive as demes enriched for wild types. Such mutant-rich demes may serve as
98 competitive refugia. Thus, in soft selection schemes, limiting migration can increase the frequencies of
99 deleterious mutants (Roze and Rousset, 2004; Whitlock, 2002).

100 As mutation, selection and migration occur in a subdivided population, both first-order
101 moments (i.e., the mean) and higher-order moments (i.e., variance, skew, kurtosis, etc.) of allele
102 frequencies across demes can change. Previous models have estimated higher-order moments (or
103 related quantities like F_{ST}) in terms of first-order moments under an assumption of weak selection. In
104 this paper, we take a different approach. We build an ecological model of a subdivided population, in
105 which higher-order moments are dynamic variables. No assumptions about the strength of selection or
106 mutation are required. Using this model, we find that limited migration increases the fraction of
107 mutants at mutation-selection balance. However, our moment-closure approach (in which we express
108 third-order moments in terms of lower-order moments) is exact only under total migration. Thus, our
109 analytical results are accurate when there is minimal subdivision. Similar moment closure approaches
110 have been used to model ecological neutrality, competition, and stability (Bolker and Pacala, 1997;
111 Haegeman and Loreau, 2011; Neuhauser, 2002; Vanpeteghem and Haegeman, 2010). We use
112 computer simulations to confirm that the fraction of mutants at equilibrium increases under limited
113 migration (where the mathematical analysis is approximate). The simulations also show spatial
114 segregation of types, suggesting that mutant-rich areas act as competitive refugia.

115

116 **MUTATION-SELECTION BALANCE IN A SUBDIVIDED POPULATION**

117 In our model, a population inhabits a metapopulation of patches. Space is implicit in this model; all
118 patches are equally “far” from any given patch. Migration between patches occurs at birth with a
119 specified probability. When the probability is one, every offspring migrates to a random patch, and the
120 population is essentially well mixed. When the probability is lowered slightly from one, there is a
121 small chance an offspring will stay in its natal patch, and thus a modicum of spatial structure is
122 introduced.

123

124 **Terminology and Life Cycle:** Consider two genotypes W and M , for wild type and mutant,
 125 respectively, inhabiting a metapopulation with an infinite number of patches. The population size of
 126 each patch is finite. In all that follows, genotype indices i and j will be used where $i, j \in \{W, M\}$ and
 127 $i \neq j$. The per capita birth rate of genotype i is given by $F_i(n_i, n_j) = f_i - \beta_i(n_i + \alpha_{ij}n_j)$, where n_i and
 128 n_j are the numbers of genotype i and j in the patch, f_i is the intrinsic growth rate of genotype i , β_i
 129 measures the effect of intra-genotypic competition, and α_{ij} is an inter-genotypic conversion factor (i.e.,
 130 one individual of genotype j counts as α_{ij} individuals of genotype i). Genotype i dies with rate δ_i .
 131 Mutation from genotype i to j occurs during the birth process with probability $\mu_{i \rightarrow j}$. Migration also
 132 occurs at birth, when genotype i migrates to a random patch with probability m_i . The population
 133 evolves stochastically in continuous time.

134

135 **Moment Dynamics:** Let $N_i(t)$ be the expected number of genotype i per patch at time t . For
 136 typographical convenience, we drop the explicit reference to time dependence in our notation for the
 137 terms and equations that follow (e.g., $N_i(t)$ is written N_i). In Appendix 1 we show that

$$138 \quad \frac{dN_i}{dt} = (1 - \mu_{i \rightarrow j}) N_i F_i(N_{i|i}, N_{j|i}) + \mu_{j \rightarrow i} N_j F_j(N_{j|j}, N_{i|j}) - \delta_i N_i, \quad (1)$$

139 where $N_{i|j}$ is the expected number of individuals of genotype i in the patch of a randomly chosen
 140 individual of genotype j , with $i, j \in \{W, M\}$.

141 It can be shown that $N_{i|i} = N_i + \sigma_i^2 / N_i$, where σ_i^2 is the variance in the number of genotype i .
 142 When individuals of the given genotype are uniformly distributed (i.e., variance is zero), this reduces
 143 to the mean N_i . Similarly, $N_{i|j} = N_i + C / N_j$, where C is the covariance between the numbers of
 144 genotypes i and j . When the two genotypes are independently distributed (i.e., covariance is zero) this

145 term reduces to the mean N_i . Covariance may be positive, indicating association between types, or
 146 negative, indicating segregation of types.

147 Thus the dynamics of the first order moments N_i and N_j rely on second order moments σ_i^2 , σ_j^2 ,
 148 and C . The equations governing the dynamics of these second order moments involve third order
 149 moments, the differential equations for the third order moments involve fourth order moments, and so
 150 on. Our task is similar to Hercules' battle with the Hydra (in spirit, not magnitude!). With each Hydra
 151 head Hercules sliced off, new heads popped up in its place. For each moment dynamical equation we
 152 describe, the description of new, higher-order moment equations becomes necessary. We must find a
 153 way to stem the endless flow of higher-order moments. Hercules seared the necks of the Hydra to
 154 prevent the regrowth of the heads; we close our system of differential equations by a second-order
 155 moment closure technique. We approximate third-order moments in terms of lower-order moments
 156 (see Appendix 1 for details), thus sealing the endless flow. Our moment closure approximation is exact
 157 when migration is absolute (i.e., $m_W = m_M = 1$), and we are not limited by assumptions of near
 158 neutrality (Neuhauser, 2002). With this approximation, the dynamics for the second order moments are
 159 given by:

160

$$\begin{aligned} \frac{d\sigma_i^2}{dt} = \frac{dN_i}{dt} + 2\delta_i(N_i - \sigma_i^2) + 2(1 - m_i)(1 - \mu_{i \rightarrow j})\{f_i\sigma_i^2 - \beta_i(N_i + 2N_i\sigma_i^2) - \beta_i\alpha_{ij}(N_iC + N_j\sigma_i^2)\} \\ + 2(1 - m_j)\mu_{j \rightarrow i}\{f_jC - \beta_j2N_jC - \beta_j\alpha_{ji}(N_iC + N_j\sigma_i^2)\} \end{aligned} \quad (2)$$

161

162

$$\begin{aligned} \frac{dC}{dt} = -(\delta_i + \delta_j)C + (1 - m_i)(1 - \mu_{i \rightarrow j})\{f_iC - \beta_i2N_iC - \beta_i\alpha_{ij}(N_jC + N_i\sigma_j^2)\} \\ + (1 - m_j)\mu_{j \rightarrow i}\{f_j\sigma_j^2 - \beta_j(N_j + 2N_j\sigma_j^2) - \beta_j\alpha_{ji}(N_jC + N_i\sigma_j^2)\} \\ + (1 - m_j)(1 - \mu_{j \rightarrow i})\{f_jC - \beta_j2N_jC - \beta_j\alpha_{ji}(N_iC + N_j\sigma_i^2)\} \\ + (1 - m_i)\mu_{i \rightarrow j}\{f_i\sigma_i^2 - \beta_i(N_i + 2N_i\sigma_i^2) - \beta_i\alpha_{ij}(N_iC + N_j\sigma_i^2)\} \end{aligned} \quad (3)$$

163

164

165 **Mutation-Selection Balance:** Our dynamical system contains many parameters. To simplify matters,
 166 we assume $m_W = m_M = m$, $f_W = f_M = f$, $\beta_W = \beta_M = \beta$, $\alpha_{WM} = \alpha_{MW} = 1$, $\mu_{W \rightarrow M} = \mu$, and
 167 $\mu_{M \rightarrow W} = 0$. Thus, we assume our genotypes are identical in all parameters except their death rates,
 168 which define a W to M mutation as deleterious (i.e. $\delta_M > \delta_W > 0$), and their mutation rates.
 169 Consequently, we only consider viability selection in this analysis, though we simulate other
 170 possibilities below. We have also assumed that intra-genotypic competition is identical to inter-
 171 genotypic competition (the α parameters are set to unity), and that back mutation does not occur. This
 172 might be realistic if the mutation from wild type to the mutant involves a deletion, but even if this
 173 mutation is a base substitution, the density of mutants is often so low that back mutation does not
 174 greatly affect our results (see simulations below).

175 In Appendix 2, we derive the mutant fraction of the population at equilibrium under full
 176 migration ($m = 1$). Because the fraction of mutants cannot be greater than one, there are parameter
 177 constraints on the analysis to ensure mutation does not “overwhelm” selection. Within those parameter
 178 constraints, the fraction of mutants at mutation-selection balance is

$$179 \quad \Phi = \frac{\mu \delta_W}{(1 - \mu)(\delta_M - \delta_W)}. \quad (4)$$

180 Like the classical result for a panmictic haploid population, for which $\Phi = \frac{\mu}{s}$ where $1 - s$ is the fitness
 181 of a mutant relative to a wild-type (Crow and Kimura, 1970), our expression is proportional to μ (for
 182 small μ) and inversely proportional to a measure of the selective disadvantage of the mutant ($\frac{\delta_M}{\delta_W} - 1$).

183 How does this fraction change as spatial structure is introduced; that is, what happens to Φ as
 184 m is lowered from unity? Since Equations (2) and (3) are exact for $m = 1$, the partial derivative of Φ
 185 with respect to m can be computed exactly at $m = 1$ (Neuhauser, 2002). In Appendix 3, we derive the
 186 following:

$$\left. \frac{\partial \Phi}{\partial m} \right|_{m=1} = - \frac{\beta \mu}{(1-\mu)^2 \delta_M \{(\delta_M / \delta_W)^2 - 1\}} . \quad (5)$$

This expression demonstrates that $\left. \frac{\partial \Phi}{\partial m} \right|_{m=1} < 0$ for $0 < \mu < 1$, so the fraction of deleterious mutants at equilibrium always increases when a small amount of structure is introduced into the model.

190

191 **Simulation of the Spatial Model**

Our analysis is exact when all offspring migrate, but becomes approximate as soon as some offspring remain in their natal patches. How well do the approximations capture actual dynamics? Here, we explore the model via simulation.

In the simulation, we seed a finite (but large) number of patches P with wild type and mutant individuals, and simulate evolution using a Gillespie algorithm in which birth and death events occur stochastically (see Appendix 4 for details). In simulation runs with absolute migration, all first order moments and second order moments approach our analytic predictions as equilibrium is reached, even when initialized far from the calculated equilibrium (Sup. Fig. 1). This is expected, as our analysis is exact when migration is absolute. As the probability of migration is lowered from unity, our analysis becomes approximate. Figure 1 shows simulation results across a range of migration probabilities, and the analytical prediction extrapolated from Equation (5). At high levels of migration, the simulation corresponds well with the analysis, with Φ following its derivative calculated at $m = 1$ (Fig. 1, inset). As migration drops further, the mutant frequency rises faster than the linear extrapolation from our analytical model (Fig. 1). The correspondence between our finite simulation and our deterministic analysis for high migration indicates that the number of patches P is large enough for the metapopulation to behave deterministically. Moreover, results are not appreciably affected when fewer patches are used (Sup. Fig. 2B). Results are also not appreciably affected when back mutation is allowed (Sup. Fig. 2A).

Parameters for Figure 1 were chosen to illustrate a large effect of limited migration on the mutant fraction at mutation-selection balance. When the competition parameter β is decreased, $\text{abs}\left(\frac{\partial \Phi}{\partial m}\right)_{m=1}$ is proportionately decreased (see Eq. 5) and the per-patch carrying capacity is increased, but the simulated mutant frequency still rises faster than the linear extrapolation from our model (Sup. Fig. 2C). Similar results to those shown in Figure 1 occur when both the mutation rate and the selective disadvantage of mutants are decreased (Sup. Fig. 2D).

To see why limited migration increases the fraction of deleterious mutants in a population, we follow a simulation (Fig. 2) as it transitions from absolute migration to limited migration (i.e., $m = 1$ to $m = 0.5$). We see that when limited migration is introduced, the mutant frequency increases (Fig. 2A), the variance in mutant density increases (i.e., the mutants become more clumped), and the covariance between the densities of the two genotypes becomes negative (i.e., patches with many wild-type genotypes tend to have fewer mutant genotypes, and *vice versa*) (Fig. 2B).

Such spatial segregation leads to an increase in the fraction of inhabited patches that house mutant-only populations (Sup. Fig. 3A). The fraction of mutants in mutant-only patches also increases as the migration rate decreases (Sup. Fig. 3B). Notably, the fraction of mutants in patches that also house wild type genotypes does not increase as the migration rate decreases (Sup. Fig. 3B). Thus, the increase of mutants in mutant-only patches may suffice to explain the overall increase in mutants at limited migration rates.

We conclude that limited migration leads to a higher mutant frequency at mutation-selection balance because the less fit mutant is able to escape competition with the wild type due to spatial segregation. Thus, mutant-rich patches are competitive refugia that allow the mutant genotype to persist in relative isolation from the competitively superior wild type. If this explanation is correct, limited migration should safeguard deleterious mutants regardless of whether selection occurs via

233 differences in viability or fecundity, whether space is explicit or implicit, and whether the spatially
234 distributed units are populations or individuals.

235

236 **A LATTICE-BASED APPROACH**

237 Our next approach considers individuals that are embedded in a lattice. Here, unlike our first approach,
238 (a) space is explicit, (b) population structure varies with dispersal distance, (c) the “patches” house
239 individuals rather than subpopulations, and (d) we consider both viability and fecundity selection. A
240 similar lattice-based approach has been used to explore many eco-evolutionary aspects of spatially
241 structured populations, including the invasion of rare types, species coexistence, host-parasite
242 evolution, spatial structuring of communities, and evolutionary trajectories (Débarre et al., 2012;
243 Durrett and Levin, 1997; Hauert and Doebeli, 2004; Kerr et al., 2002).

244 In our simulation, we consider two haploid asexual genotypes: wild type (W) and mutant (M).
245 These genotypes occupy an $L \times L$ regular square lattice with periodic boundaries (i.e., a toroidal
246 geometry). Each lattice point may take one of three states: empty, wild type, or mutant. At each update,
247 a point is chosen at random. If this focal point is “filled” with a wild type, the wild type dies with
248 probability δ_W^* , giving a transformation to the empty state. Likewise, a mutant that is chosen will die
249 with probability δ_M^* (where $\delta_M^* \geq \delta_W^* > 0$).

250 If the focal point is already empty, then a birth event can occur, where an individual in a pre-
251 defined neighborhood of the focal point produces an offspring that fills the focal point (giving a
252 transformation to a filled state). Let x_W and x_M be the fraction of the focal point’s neighborhood
253 occupied by wild type and mutant lattice points, respectively. Then with probabilities $f_W^* x_W$ and $f_M^* x_M$
254 the parent of the individual “born into” the focal point is wild type and mutant, respectively. The
255 parameters f_W^* and f_M^* represent the fecundities of wild-type and mutant individuals (where $0 \leq f_M^* \leq$

256 $f_W^* \leq 1$). The focal point stays empty with probability $1 - f_W^* x_W - f_M^* x_M$. Mutation occurs at birth:
 257 from wild type to mutant with probability $\mu_{W \rightarrow M}^*$, and from mutant to wild type with probability $\mu_{M \rightarrow W}^*$.
 258 The degree of population structure is controlled by adjusting the size of the neighborhood around any
 259 focal point (effectively altering the distribution of distance at dispersal). We focus on three cases: a
 260 von Neumann neighborhood (where the lattice points immediately to the north, east, south and west of
 261 the focal point comprise the neighborhood), a Moore neighborhood (where the eight lattice points
 262 nearest the focal point constitute the neighborhood), and a Global neighborhood (where the entire
 263 lattice, minus the focal point, comprises the neighborhood). Thus, the evolving population can range
 264 from highly structured (von Neumann neighborhood) to effectively well mixed (Global neighborhood).

265 Figure 3 shows that smaller dispersal neighborhoods lead to higher mutant frequencies at
 266 equilibrium, corroborating our prior analysis. This pattern holds under both pure viability selection
 267 ($\delta_M^* > \delta_W^*$ and $f_M^* = f_W^*$) and pure fecundity selection ($\delta_M^* = \delta_W^*$ and $f_M^* < f_W^*$).
 268

269 DISCUSSION

270 We find analytically and computationally that limited migration increases the frequency of deleterious
 271 mutants, and this increase is not restricted to a specific form of space or mode of selection. Prior
 272 models of selection in metapopulations have shown effects of limited migration when local
 273 interactions are defined by mating, or when mutant-rich demes have similar productivity to wild type
 274 enriched demes, and can thus act as competitive refugia for mutants (Glémin et al., 2003; Roze and
 275 Rousset, 2004; Whitlock, 2002). In this paper, we embedded competition in an explicitly ecological
 276 framework, which allows us to manifest local interactions explicitly as density-dependent fecundity.
 277 We used a moment closure approach that expresses higher-order moments in terms of lower-order
 278 moments, therefore allowing those higher-order moments to vary dynamically as we began to limit

279 migration. We showed that limited migration can affect asexual populations by segregating types.
280 Essentially, limiting migration has no effect on the generation of mutants, but hampers the effective
281 strength of selection (Cherry and Wakeley, 2003) by sheltering alleles from global competition, and so
282 tips the mutation-selection balance in favor of deleterious mutations. Generally, whenever there is both
283 variation in localities and local interaction, migration rate will be a salient factor in determining the
284 frequency of deleterious mutants.

285 Mutant frequency is sometimes used to estimate mutation rates of microbes. Using such a
286 method, a structured environment may appear mutagenic because a higher frequency of mutants is
287 found. For example, Bjedov *et al.* (2003) find a disparity in mutant frequencies between liquid and
288 agar bacterial cultures, and attribute it to oxidative stress incurred during colonial growth on agar. This
289 explanation is certainly plausible, but the colony structure itself may contribute to the increased mutant
290 frequency. When going from an unstructured to a structured environment (e.g., a flask to an agar
291 plate), the frequency of deleterious mutants may increase even if the mutation rate is constant.

292
293 **Adaptive valley crossing:** If a single deleterious mutation is complemented by a second mutation that
294 improves the fitness of the organism above the wild type, the frequency (and number) of the original
295 mutant may be relevant for crossing “adaptive valleys”. Recent theoretical studies have elucidated how
296 well-mixed populations cross adaptive valleys, and at what rate (Weissman et al., 2010, 2009).
297 However, Sewall Wright’s shifting balance process is predicated on the idea that, collectively, semi-
298 isolated subpopulations would explore a landscape in a way unavailable to well-mixed populations
299 (Pigliucci, 2008; Wright, 1988, 1932). To cross a valley, a population must first discover a new peak,
300 and then have the peak genotype spread through the population. Increasing migration between
301 separated patches hinders exploration of novel genotypes (Whitlock, 2003), but, once a beneficial

302 genotype is discovered, the increased migration facilitates its spread (Jain et al., 2011; Rozen et al.,
303 2008). In the rugged landscapes that were the focus of Wright’s shifting balance, the rate-limiting step
304 in adaptation may be the discovery of novel genotypes (i.e., finding new peaks) rather than their spread
305 through a population. If this is the case, limited migration may speed the rate of adaptation.

306 If valley crossing requires multiple “downward” steps, the facilitating effect of limited
307 migration is amplified. Limited migration protects not only deleterious single mutants from
308 competition with wild types, but also relatively deleterious double mutants from competition with
309 single mutants (and wild types). When deleterious double mutants are added to our metapopulation
310 simulation, we see the amplified effect of limited migration on the double-mutant frequency (Fig. 4A);
311 when we add triple mutants, the effect of limited migration amplifies further (Fig. 4B). This effect on
312 double and triple mutants is also observed in our lattice-based approach (Sup. Fig. 4).

313 For sufficiently wide valleys, a population starting with only wild type individuals may
314 discover the peak genotype faster when its migration is limited (Fig. 5). For certain parameter values, a
315 population whose migration is limited may cross even the narrowest valley—one deleterious mutant
316 between two peak genotypes—faster than an unstructured population (Bitbol and Schwab, 2014). Note
317 that when only upward steps are required for adaptation (e.g., a smooth landscape) then the rate-
318 limiting step in adaptation is the spread of beneficial genotypes, and thus limited migration will inhibit
319 adaptation (Kryazhimskiy et al., 2012).

320 The limited migration of individuals generally slows the spread of advantageous traits.
321 However, it is precisely this dampening of competition that can allow spatially structured populations
322 to safeguard deleterious mutants. By harboring this diversity, it may be possible for structured
323 populations to discover novel genotypes faster, even if the benefit spreads more slowly.

324

- 326 Bitbol, A.-F., Schwab, D.J., 2014. Quantifying the Role of Population Subdivision in Evolution on
 327 Rugged Fitness Landscapes. *PLoS Comput Biol* 10, e1003778.
 328 doi:10.1371/journal.pcbi.1003778
- 329 Bjedov, I., Tenaillon, O., Gérard, B., Souza, V., Denamur, E., Radman, M., Taddei, F., Matic, I., 2003.
 330 Stress-induced mutagenesis in bacteria. *Science* 300, 1404–1409. doi:10.1126/science.1082240
- 331 Bolker, B., Pacala, S.W., 1997. Using moment equations to understand stochastically driven spatial
 332 pattern formation in ecological systems. *Theoretical population biology* 52, 179–197.
- 333 Cherry, J.L., Wakeley, J., 2003. A Diffusion Approximation for Selection and Drift in a Subdivided
 334 Population. *Genetics* 163, 421–428.
- 335 Crow, J.F., Kimura, M., 1970. An introduction to population genetics theory.
- 336 Débarre, F., Lion, S., Baalen, Gandon, S., 2012. Evolution of Host Life-History Traits in a Spatially
 337 Structured Host-Parasite System. *The American Naturalist* 179, 52–63. doi:10.1086/663199
- 338 Durrett, R., Levin, S., 1997. Allelopathy in spatially distributed populations. *Journal of Theoretical*
 339 *Biology* 185, 165–171.
- 340 Evans, K.M., Chepurinov, V.A., Sluiman, H.J., Thomas, S.J., Spears, B.M., Mann, D.G., 2009. Highly
 341 Differentiated Populations of the Freshwater Diatom *Sellaphora capitata* Suggest Limited
 342 Dispersal and Opportunities for Allopatric Speciation. *Protist* 160, 386–396.
 343 doi:10.1016/j.protis.2009.02.001
- 344 Eyre-Walker, A., Keightley, P.D., 2007. The distribution of fitness effects of new mutations. *Nat. Rev.*
 345 *Genet.* 8, 610–618. doi:10.1038/nrg2146
- 346 Fisher, R.A., 1930. The Distribution of Gene Ratios for Rare Mutations. *Proc. Roy. Soc. Edinburgh* 50,
 347 205–220.
- 348 Gillespie, D.T., 1977. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81,
 349 2340–2361. doi:10.1021/j100540a008
- 350 Glémin, S., Ronfort, J., Bataillon, T., 2003. Patterns of inbreeding depression and architecture of the
 351 load in subdivided populations. *Genetics* 165, 2193–2212.
- 352 Haegeman, B., Loreau, M., 2011. A mathematical synthesis of niche and neutral theories in
 353 community ecology. *Journal of Theoretical Biology* 269, 150–165.
- 354 Haldane, J.B.S., 1927. A Mathematical Theory of Natural and Artificial Selection, Part V: Selection
 355 and Mutation. *Mathematical Proceedings of the Cambridge Philosophical Society* 23, 838–844.
 356 doi:10.1017/S0305004100015644
- 357 Hauert, C., Doebeli, M., 2004. Spatial structure often inhibits the evolution of cooperation in the
 358 snowdrift game. *Nature* 428, 643–646.
- 359 Howells, E.J., Willis, B.L., Bay, L.K., van Oppen, M.J.H., 2013. Spatial and temporal genetic structure
 360 of *Symbiodinium* populations within a common reef-building coral on the Great Barrier Reef.
 361 *Molecular Ecology* 22, 3693–3708. doi:10.1111/mec.12342
- 362 Jain, K., Krug, J., Park, S.-C., 2011. Evolutionary Advantage of Small Populations on Complex Fitness
 363 Landscapes. *Evolution* 65, 1945–1955. doi:10.1111/j.1558-5646.2011.01280.x
- 364 Kerr, B., Riley, M.A., Feldman, M.W., Bohannan, B.J.M., 2002. Local dispersal promotes biodiversity
 365 in a real-life game of rock-paper-scissors. *Nature* 418, 171–174.
- 366 Kryazhimskiy, S., Rice, D.P., Desai, M.M., 2012. Population Subdivision and Adaptation in Asexual
 367 Populations of *Saccharomyces Cerevisiae*. *Evolution* 66, 1931–1941. doi:10.1111/j.1558-
 368 5646.2012.01569.x

- Martin, P.H., Canham, C.D., 2010. Dispersal and recruitment limitation in native versus exotic tree species: life-history strategies and Janzen-Connell effects. *Oikos* 119, 807–824.
doi:10.1111/j.1600-0706.2009.17941.x
- Neuhauser, C., 2002. Effects of local interactions and local migration on stability. *Theoretical population biology* 62, 297–308.
- Pacala, S.W., Levin, S.A., 1997. Biologically generated spatial pattern and the coexistence of competing species. *Spatial ecology*. Princeton University Press, Princeton, New Jersey, USA 204–232.
- Pigliucci, M., 2008. Sewall Wright’s adaptive landscapes: 1932 vs. 1988. *Biology and Philosophy* 23, 591–603.
- Roze, D., Rousset, F., 2004. Joint effects of self-fertilization and population structure on mutation load, inbreeding depression and heterosis. *Genetics* 167, 1001–1015.
doi:10.1534/genetics.103.025148
- Rozen, D.E., Habets, M.G.J.L., Handel, A., de Visser, J.A.G.M., 2008. Heterogeneous Adaptive Trajectories of Small Populations on Complex Fitness Landscapes. *PLoS ONE* 3, e1715.
doi:10.1371/journal.pone.0001715
- Vanpeteghem, D., Haegeman, B., 2010. An analytical approach to spatio-temporal dynamics of neutral community models. *Journal of Mathematical Biology* 61, 323–357.
- Weissman, D.B., Desai, M.M., Fisher, D.S., Feldman, M.W., 2009. The rate at which asexual populations cross fitness valleys. *Theoretical population biology* 75, 286–300.
- Weissman, D.B., Feldman, M.W., Fisher, D.S., 2010. The rate of fitness-valley crossing in sexual populations. *Genetics* 186, 1389–1410.
- Whitlock, M.C., 2002. Selection, load and inbreeding depression in a large metapopulation. *Genetics* 160, 1191–1202.
- Whitlock, M.C., 2003. Fixation Probability and Time in Subdivided Populations. *Genetics* 164, 767–779.
- Wright, S., 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution, in: *Proceedings of the Sixth International Congress on Genetics*. pp. 356–366.
- Wright, S., 1988. Surfaces of selective value revisited. *The American Naturalist* 131, 115–123.

Figure Legends

Figure 1. Stochastic simulation under various migration rates. Frequency of deleterious mutants at mutation-selection balance across various probabilities of migration as found by simulation (red circles), compared to the $m=1$ derivative of our analytical model (black dashed line). The analytically calculated mutant fraction at full migration is given as a gray dotted horizontal line for comparison. At high migration probabilities, the simulation results agree well with our analytical model (see inset). As the probability of migration decreases further from unity, the fraction of deleterious mutants increases faster than the analytical extrapolation. Large data points and shading represent mean values and standard deviation of 20-40 replicate simulations (small data points) using parameter values $P = 10^4$, $f = 0.5$, $\beta = 0.2$, $\mu = 0.1$, $\delta_W = 0.05$, $\delta_M = 0.1$, corresponding to a per-patch carrying capacity of approximately $\frac{f-\delta_W}{\beta} = 2.25$.

Figure 2. A simulated shift in the probability of migration. When migration becomes limited (grey-shaded portion of plots), the increase in deleterious mutant frequency (A) coincides with an increase in the variance in mutant density and a decrease in the covariance between mutant and wild type densities (B, variances are divided by the means of their corresponding variables, and covariance is divided by the product of the roots of the two means). Time units are relative, and defined by a Gillespie algorithm described in Appendix 4. Solid lines and colored shading represent the mean $\pm$ SD of 16 replicate simulations using the parameter values listed in Figure 1.

Figure 3. Lattice-based simulation results. The frequency of deleterious mutants at mutation-selection balance across various neighborhood sizes in lattice-based simulations with viability selection (A) or fecundity selection (B). As dispersal is limited to smaller neighborhoods, the frequency of deleterious mutants increases. Large data points represent mean values of 24 replicate simulations (small data points) using parameter values $L = 200$, $\mu_{W \rightarrow M}^* = 0.1$, $\mu_{M \rightarrow W}^* = 0.02$, and either viability selection (A, with $\delta_W^* = 0.1$, $\delta_M^* = 0.2$, $f_W^* = f_M^* = 1$) or fecundity selection (B, with $\delta_W^* = \delta_M^* = 0.1$, $f_W^* = 1$, $f_M^* = 0.5$).

Figure 4. Stochastic simulations with a chain of sequential deleterious mutants in a metapopulation approach. The wild-type genotype yields the first mutant via mutation; the first mutant yields the relatively deleterious double mutant via mutation; and so on. The density of each genotype is shown relative to its density in a well-mixed population (given by the dotted line at unity). The effect of migration probability on relative density in a community with two (A) and three (B) mutants are shown. Points and colored shading represent the mean $\pm$ SD of 12 replicate simulations, while dashed lines roughly matching the double-mutant frequency represent the square of the mutant to wild-type ratio, multiplied by the wild-type density, for each m . In (B), the cubes of the ratios are also shown, roughly matching the triple-mutant density. Parameter values used in this simulation are: $P = 10^4$, $f = 0.5$, $\beta = 0.2$,

$\mu_{W \rightarrow M} = \mu_{M \rightarrow M2} = \mu_{M2 \rightarrow M3} = 0.1$, $\mu_{M3 \rightarrow M2} = \mu_{M2 \rightarrow M} = \mu_{M \rightarrow W} = 0.01$, $\delta_W = 0.05$, $\delta_M = 0.1$,
 $\delta_{M2} = 0.2$, $\delta_{M3} = 0.4$.

Figure 5. Discovery versus spread in simulated populations. Both metapopulation (A) and lattice (B) simulations were initialized with wild type genotypes only. There are three successively more deleterious mutants comprising a valley between the wild type genotype and a highly beneficial mutant (accessible from the third deleterious mutant). The waiting time before the beneficial mutant reaches a given frequency is shown. As migration becomes more limited, the waiting time to the discovery of the beneficial mutant decreases (as m decreases from 1, the 0.01% profile drops; as the dispersal neighborhood shrinks, the 0.01% profile drops). However, a greater degree of structure inhibits the spread of these beneficial mutants (and thus the time to reach a substantial frequency of the beneficial mutant can increase under initial limitations to migration—see 1% and 50% trajectories). Extremely limited migration decreases total population size, facilitating spread. Fixation does not occur as back mutation is allowed and the numbers of patches are large. Points and shaded regions represent the mean $\pm$ SEM of 24-36 (A) or 16 (B) replicate simulations using parameter values $f = 0.5$, $\beta = 0.2$, $\mu_{W \rightarrow M} = \mu_{M \rightarrow M2} = \mu_{M2 \rightarrow M3} = 0.1$, $\mu_{M3 \rightarrow M2} = \mu_{M2 \rightarrow M} = \mu_{M \rightarrow W} = 0.01$, $\delta_W = 0.05$, $\delta_M = 0.1$, $\delta_{M2} = 0.2$, $\delta_{M3} = 0.4$, $\delta_{M4} = 0.025$. For (A), $P = 104$. For (B) $L = 200$.

505
506
507
508
509
510
511
512
513
514
515
516
517
518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

Appendices

535 *Appendix 1: Moment Equations*

536 *Mean Density Dynamics*

537 In this appendix, we derive the dynamical equations for our first and second-order moments.
 538 We assume our population inhabits a metapopulation of infinite patches (each of which houses a finite
 539 population), allowing us to use expectation values for our patch dynamics. We start with the dynamics
 540 of the mean genotype abundances. Let $q_i(t)$ be a random variable giving the number of individuals of
 541 genotype $i \in \{W, M\}$ within a randomly selected patch at time t . If we consider a period of time, Δt ,
 542 small enough that the probability of more than one event occurring during that interval is vanishing
 543 small, we have the following:

$$544 \quad q_i(t + \Delta t) = \begin{cases} q_i(t) - 1 & \text{with probability } P_i^- \\ q_i(t) + 1 & \text{with probability } P_i^+ \end{cases}, \quad (\text{A1.1})$$

545 where

$$546 \quad P_i^- = \delta_i q_i(t) \Delta t, \quad (\text{A1.2})$$

547 and

$$548 \quad P_i^+ = (1 - m_i)(1 - \mu_{i \rightarrow j})F_i(q_i(t), q_j(t))q_i(t)\Delta t + (1 - m_j)\mu_{j \rightarrow i}F_j(q_j(t), q_i(t))q_j(t)\Delta t \\ + m_i(1 - \mu_{i \rightarrow j})E[F_i(q_i(t), q_j(t))q_i(t)]\Delta t + m_j\mu_{j \rightarrow i}E[F_j(q_j(t), q_i(t))q_j(t)]\Delta t. \quad (\text{A1.3})$$

549 where E is the expectation value over all patches. Thus, the expected change in q_i over our small
 550 interval of time is given by:

$$551 \quad E[\Delta q_i] = E[P_i^+ - P_i^-]. \quad (\text{A1.4})$$

552 For typographical convenience, we drop the explicit time dependence in our notation for the terms and
 553 equations that follow. We use the following notations

$$554 \quad N_i = E[q_i],$$

$$555 \quad \sigma_i^2 = \text{var}[q_i],$$

$$C = \text{cov}[q_i, q_j] = \text{cov}[q_j, q_i],$$

and we have the following relations

$$E[q_i^2] = N_i^2 + \sigma_i^2, \quad (\text{A1.5})$$

$$E[q_i q_j] = N_i N_j + C. \quad (\text{A1.6})$$

Using (A1.5), (A1.6), and our per capita birth rate of $F_i(n_i, n_j) = f_i - \beta_i(n_i + \alpha_{ij}n_j)$ we can rewrite

(A1.4) as follows:

$$\begin{aligned} \frac{E[\Delta q_i]}{\Delta t} = & -\delta_i N_i + (1 - \mu_{i \rightarrow j}) \left\{ f_i N_i - \beta_i \left((N_i^2 + \sigma_i^2) + \alpha_{ij} (N_i N_j + C) \right) \right\} \\ & + \mu_{j \rightarrow i} \left\{ f_j N_j - \beta_j \left((N_j^2 + \sigma_j^2) + \alpha_{ji} (N_i N_j + C) \right) \right\}. \end{aligned} \quad (\text{A1.7})$$

Taking the limit $\Delta t \rightarrow 0$, and factoring N_i from the second term and N_j from the third term, we have

$$\begin{aligned} \frac{dN_i}{dt} = & -\delta_i N_i + (1 - \mu_{i \rightarrow j}) \left\{ f_i - \beta_i \left(\left(N_i + \frac{\sigma_i^2}{N_i} \right) + \alpha_{ij} \left(N_j + \frac{C}{N_i} \right) \right) \right\} N_i \\ & + \mu_{j \rightarrow i} \left\{ f_j - \beta_j \left(\left(N_j + \frac{\sigma_j^2}{N_j} \right) + \alpha_{ji} \left(N_i + \frac{C}{N_j} \right) \right) \right\} N_j. \end{aligned} \quad (\text{A1.8})$$

The terms $N_i + \sigma_i^2/N_i$ and $N_j + C/N_i$ (and the two other similar terms) are more approachable if we allow $N_{j|i}$ to represent the expected number of individuals of genotype j in the patch of a randomly chosen individual of genotype i (rather than a randomly chosen patch). Eq. A1.6 can now be rewritten as $N_i N_{j|i} = E[q_i q_j] = N_i N_j + C$, and therefore $N_{j|i} = N_j + C/N_i$. Similarly, Eq. A1.5 yields $N_{i|i} = N_i + \sigma_i^2/N_i$. Using this notation, (A1.8) can be simplified to

$$\frac{dN_i}{dt} = -\delta_i N_i + (1 - \mu_{i \rightarrow j}) F_i(N_{i|i}, N_{j|i}) N_i + \mu_{j \rightarrow i} F_j(N_{j|j}, N_{i|j}) N_j. \quad (\text{A1.9})$$

From these equations we see that change in the first order moment (the expected density of genotype i) depends on second order moments (the variances and covariance of genotype densities). Thus, we now derive the dynamical equations for the change in these second order moments.

575 *Variance Dynamics*

576 Again, we consider a very small interval of time, Δt . The following holds:

577
$$\Delta q_i^2(t) = q_i^2(t + \Delta t) - q_i^2(t). \quad (\text{A1.10})$$

578 Using (A1.1), (A1.2), (A1.3) and (A1.10), and again dropping the explicit time dependence in our
579 notation, we see that

580
$$\Delta q_i^2 = (q_i - 1)^2 - q_i^2 = -2q_i + 1 \text{ with probability } P_i^-, \text{ and} \quad (\text{A1.11})$$

581
$$\Delta q_i^2 = (q_i + 1)^2 - q_i^2 = 2q_i + 1 \text{ with probability } P_i^+. \quad (\text{A1.12})$$

582 Thus, the expected change in q_i^2 is given by:

583
$$E[\Delta q_i^2] = E[(2q_i + 1)P_i^+ + (-2q_i + 1)P_i^-] \quad (\text{A1.13})$$

584 We have the following relations:

585
$$E[q_i^3] = T_{iii} + 3N_i\sigma_i^2 + N_i^3, \quad (\text{A1.14})$$

586
$$E[q_i^2 q_j] = T_{iij} + 2N_i C + N_j \sigma_i^2 + N_i^2 N_j, \quad (\text{A1.15})$$

587 where T_{iii} and T_{iij} are the central third-order moments. Because of (A1.5), we also have

588
$$\frac{E[\Delta q_i^2]}{\Delta t} = \frac{\Delta N_i^2}{\Delta t} + \frac{\Delta \sigma_i^2}{\Delta t} \quad (\text{A1.16})$$

589 Using (A1.5), (A1.6), (A1.14), (A1.15) and (A1.16), taking the limit $\Delta t \rightarrow 0$ and using the chain rule

590 (i.e., $\frac{dN_i^2}{dt} = 2N_i \frac{dN_i}{dt}$), we have

$$\begin{aligned} \frac{d\sigma_i^2}{dt} = & \frac{dN_i}{dt} + 2(N_i - \sigma_i^2)\delta_i \\ & + 2(1 - m_i)(1 - \mu_{i \rightarrow j})\{f_i \sigma_i^2 - \beta_i(T_{iii} + 2N_i \sigma_i^2 + \alpha_{ij}(T_{iij} + N_i C + N_j \sigma_i^2))\} \\ 591 & + 2(1 - m_j)\mu_{j \rightarrow i}\{f_j C - \beta_j(T_{jji} + 2N_j C + \alpha_{ji}(T_{iij} + N_i C + N_j \sigma_i^2))\}. \end{aligned} \quad (\text{A1.17})$$

592 If we describe the third order moments exactly, we will find ourselves needing to describe fourth order
593 moments, which will in turn require fifth order moments, and so on. Here we use our moment closure
594 technique.

595

596 *Closing the Moments*

597 When migration is absolute (i.e., $m_i = m_j = 1$), the random variables q_i and q_j are
 598 independently Poisson distributed among the patches with means equal to N_i and N_j , respectively (see
 599 Neuhauser, 2002). For any independent Poisson-distributed random variables, their third order
 600 moments can be described exactly in terms of lower-order moments; the homogeneous third central
 601 moment is the corresponding first-order moment, while all mixed third central moments are zero:

$$T_{iii} = N_i$$

$$T_{ijj} = 0$$

602 By using these substitutions as approximations when $m_i \approx m_j \approx 1$, we obviate the need to describe
 603 higher order moments. This moment closure technique is exact when $m_i = m_j = 1$, and approximate
 604 when $m_i \approx m_j \approx 1$.

605 Substituting our approximations for the third central moments into equation (A1.17), we have

$$\begin{aligned} \frac{d\sigma_i^2}{dt} = & \frac{dN_i}{dt} + 2(N_i - \sigma_i^2)\delta_i \\ & + 2(1 - m_i)(1 - \mu_{i \rightarrow j})\{f_i\sigma_i^2 - \beta_i(N_i + 2N_i\sigma_i^2 + \alpha_{ij}(N_iC + N_j\sigma_i^2))\} \\ & + 2(1 - m_j)\mu_{j \rightarrow i}\{f_jC - \beta_j(2N_jC + \alpha_{ji}(N_iC + N_j\sigma_i^2))\}. \end{aligned} \quad (\text{A1.18})$$

607

608

609 *Covariance Dynamics*

610 Again, we consider a very small interval of time, Δt . The following holds:

$$\Delta(q_i(t)q_j(t)) = q_i(t + \Delta t)q_j(t + \Delta t) - q_i(t)q_j(t) \quad (\text{A1.19})$$

612 Using, (A1.1), (A1.2), (A1.3) and (A1.16), and again dropping the explicit time dependence in our
 613 notations, we see that

$$\Delta(q_i q_j) = (q_i - 1)q_j - q_i q_j = -q_j \text{ with probability } P_i^-, \text{ and} \quad (\text{A1.20})$$

$$\Delta(q_i q_j) = (q_i + 1)q_j - q_i q_j = q_j \text{ with probability } P_i^+. \quad (\text{A1.21})$$

Thus, the expected change in the quantity $q_i q_j$ is:

$$E[\Delta(q_i q_j)] = E[-q_j P_i^- - q_i P_j^- + q_j P_i^+ + q_i P_j^+]. \quad (\text{A1.22})$$

From (A1.6), we have the following relation:

$$\frac{E[\Delta(q_i q_j)]}{\Delta t} = \frac{\Delta(N_i N_j)}{\Delta t} + \frac{\Delta C}{\Delta t}. \quad (\text{A1.23})$$

Using (A1.5), (A1.6), (A1.14), (A1.15) and (A1.23), taking the limit $\Delta t \rightarrow 0$, and using the product

rule (i.e., $\frac{d(N_i N_j)}{dt} = N_i \frac{dN_j}{dt} + N_j \frac{dN_i}{dt}$), we have

$$\begin{aligned} \frac{dC}{dt} = & -(\delta_i + \delta_j)C + (1 - m_i)(1 - \mu_{i \rightarrow j}) \left\{ f_i C - \beta_i \left((T_{iij} + 2N_i C) + \alpha_{ij}(T_{jji} + N_j C + N_i \sigma_j^2) \right) \right\} \\ & + (1 - m_j)\mu_{j \rightarrow i} \left\{ f_j \sigma_j^2 - \beta_j \left((T_{jjj} + 2N_j \sigma_j^2) + \alpha_{ji}(T_{jji} + N_j C + N_i \sigma_j^2) \right) \right\} \\ & + (1 - m_j)(1 - \mu_{j \rightarrow i}) \left\{ f_j C - \beta_j \left((T_{jji} + 2N_j C) + \alpha_{ji}(T_{iij} + N_i C + N_j \sigma_i^2) \right) \right\} \\ & + (1 - m_i)\mu_{i \rightarrow j} \left\{ f_i \sigma_i^2 - \beta_i \left((T_{iii} + 2N_i \sigma_i^2) + \alpha_{ij}(T_{iij} + N_i C + N_j \sigma_i^2) \right) \right\}. \end{aligned}$$

Substituting our approximations for the third central moments yields

$$\begin{aligned} \frac{dC}{dt} = & -(\delta_i + \delta_j)C + (1 - m_i)(1 - \mu_{i \rightarrow j}) \left\{ f_i C - \beta_i(2N_i C + \alpha_{ij}(N_j C + N_i \sigma_j^2)) \right\} \\ & + (1 - m_j)\mu_{j \rightarrow i} \left\{ f_j \sigma_j^2 - \beta_j((N_j + 2N_j \sigma_j^2) + \alpha_{ji}(N_j C + N_i \sigma_j^2)) \right\} \\ & + (1 - m_j)(1 - \mu_{j \rightarrow i}) \left\{ f_j C - \beta_j(2N_j C + \alpha_{ji}(N_i C + N_j \sigma_i^2)) \right\} \\ & + (1 - m_i)\mu_{i \rightarrow j} \left\{ f_i \sigma_i^2 - \beta_i((N_i + 2N_i \sigma_i^2) + \alpha_{ij}(N_i C + N_j \sigma_i^2)) \right\}. \end{aligned} \quad (\text{A1.24})$$

With equations (A1.8), (A1.18) and (A1.24), we have a closed system of five differential equations

describing the dynamics of N_W , N_M , σ_W^2 , σ_M^2 and C .

Appendix 2: Equilibrium Densities

629 At equilibrium, $\frac{dN_W}{dt} = \frac{dN_M}{dt} = 0$. In this appendix, we assume $m_W = m_M = m$, $f_W = f_M = f$,

630 $\beta_W = \beta_M = \beta$, $\alpha_{WM} = \alpha_{MW} = 1$, $\mu_{W \rightarrow M} = \mu$ and $\mu_{M \rightarrow W} = 0$. Using these assumptions and equation

631 (A1.8), the equilibrium value $\hat{N}_W$ must satisfy the following:

$$632 \quad 0 = -\delta_W \hat{N}_W + (1 - \mu) \{f \hat{N}_W - \beta (\hat{N}_W^2 + \hat{\sigma}_W^2 + \hat{N}_W \hat{N}_M + \hat{C})\}. \quad (\text{A2.1})$$

633 If we assume that $m = 1$, then q_i and q_j are independently Poisson distributed, and therefore:

$$634 \quad \hat{\sigma}_W^2 = \hat{N}_W, \quad (\text{A2.2})$$

$$635 \quad \hat{C} = 0. \quad (\text{A2.3})$$

636 Using (A2.2) and (A2.3), the non-zero equilibrium in (A2.1) is

$$637 \quad \hat{N}_W = \frac{(1-\mu)(f-\beta)-\delta_W}{(1-\mu)\beta} - \hat{N}_M. \quad (\text{A2.4})$$

638 We denote the total density at equilibrium $\hat{T} = \hat{N}_W + \hat{N}_M$. So, we have

$$639 \quad \hat{T} = \frac{(1-\mu)(f-\beta)-\delta_W}{(1-\mu)\beta}, \quad (\text{A2.5})$$

640 and

$$641 \quad \hat{N}_W = \hat{T} - \hat{N}_M. \quad (\text{A2.6})$$

642 Now we turn to the equilibrial density of the mutant genotype, $\hat{N}_M$, again using (A1.8):

$$643 \quad 0 = -\delta_M \hat{N}_M + \{f \hat{N}_M - \beta (\hat{N}_M^2 + \hat{\sigma}_M^2 + \hat{N}_M \hat{N}_W + \hat{C})\} + \mu \{f \hat{N}_W - \beta (\hat{N}_W^2 + \hat{\sigma}_W^2 + \hat{N}_M \hat{N}_W + \hat{C})\}. \quad (\text{A2.7})$$

644 If we are assuming $m = 1$, the resulting Poisson distribution yields

$$645 \quad \hat{\sigma}_M^2 = \hat{N}_M, \quad (\text{A2.8})$$

646 Using (A2.3), (A2.6), and (A2.8), the non-zero mutant equilibrium in (A2.7) is

$$647 \quad \hat{N}_M = \frac{-\mu \hat{T} (f - \beta (\hat{T} + 1))}{-\delta_M + (1 - \mu) (f - \beta (\hat{T} + 1))}. \quad (\text{A2.9})$$

648 After substituting, using (A2.5), and simplifying, equations (A2.4) and (A2.9) simplify to the

649 following:

$$\hat{N}_W = \frac{\{(1-\mu)\delta_M - \delta_W\}\{(f-\beta)(1-\mu) - \delta_W\}}{\beta(\delta_M - \delta_W)(1-\mu)^2}, \quad (\text{A2.10})$$

$$\hat{N}_M = \frac{\mu\delta_W\{(f-\beta)(1-\mu) - \delta_W\}}{\beta(\delta_M - \delta_W)(1-\mu)^2}. \quad (\text{A2.11})$$

In order for $\hat{N}_W$ and $\hat{N}_M$ to be positive, we must have the following two conditions:

$$(1 - \mu)(f - \beta) > \delta_W, \quad (\text{A2.12})$$

$$(1 - \mu)\delta_M > \delta_W. \quad (\text{A2.13})$$

Note that (A2.13) is more stringent than the already assumed $\delta_M > \delta_W$. In all of what follows, we will assume conditions (A2.12) and (A2.13), except where explicitly mentioned. When $\mu = 0$, equations (A2.10) and (A2.11) simplify to:

$$\hat{N}_W = \frac{f - \beta - \delta_W}{\beta}, \quad \hat{N}_M = 0, \quad (\text{A2.14})$$

which gives a positive density of the wild type (by condition (A2.12)) and no mutant density. When $(1 - \mu)\delta_M = \delta_W$ (i.e., right where equation (A2.13) starts to be violated), equations (A2.10) and (A2.11) simplify to:

$$\hat{N}_W = 0, \quad \hat{N}_M = \frac{f - \beta - \delta_M}{\beta}, \quad (\text{A2.15})$$

which gives a positive density of the mutant (by condition (A2.12), replacing δ_W with $(1 - \mu)\delta_M$) and no wild-type density. Equilibria in (A2.14) and (A2.15) agree with single species equilibria from ecological models (Neuhauser, 2002; Pacala and Levin, 1997).

We let the fraction of mutants in the population be given by $\Phi(t)$, where

$$\Phi(t) = \frac{N_M(t)}{N_W(t) + N_M(t)}. \quad (\text{A2.16})$$

Using equations (A2.10) and (A2.11), the mutation-selection balance under full migration is:

$$\hat{\Phi} = \frac{\mu\delta_W}{(1-\mu)(\delta_M - \delta_W)}. \quad (\text{A2.17})$$

670 Note that if $\mu = 0$, then $\hat{\Phi} = 0$. That is, when there is no supply of new mutants through mutation,
 671 selection “wins” and no mutants remain at equilibrium; this corresponds to the special case of (A2.14).
 672 If $\mu = (\delta_M - \delta_W)/\delta_M$, then $\hat{\Phi} = 1$. That is, as $\mu \rightarrow (\delta_M - \delta_W)/\delta_M$, mutation “wins” by overwhelming
 673 selection and only mutants remain at equilibrium; this corresponds to the special case of (A2.15).
 674

675 *Appendix 3: The Effect of Structure on Mutation-Selection Balance*

676 In order to explore the role of structure on the mutant frequency, we look at

$$677 \quad \frac{\partial \Phi}{\partial m} = \frac{\frac{\partial N_M}{\partial m} N_W - \frac{\partial N_W}{\partial m} N_M}{(N_W + N_M)^2}. \quad (\text{A3.1})$$

678 Here we will evaluate $\frac{\partial \Phi}{\partial m} \Big|_{m=1}$. In order to do so, we must find $\frac{\partial N_W}{\partial m} \Big|_{m=1}$ and $\frac{\partial N_M}{\partial m} \Big|_{m=1}$, which we
 679 abbreviate with $\frac{\partial N_W}{\partial m} \Big|_1$ and $\frac{\partial N_M}{\partial m} \Big|_1$. To do this we differentiate (A1.8) with respect to m and evaluate at
 680 the $m = 1$ equilibrium. We start with equation (A1.8) where $i = W$.

$$681 \quad 0 = \{(1 - \mu)[f - \beta(2\hat{N}_W + \hat{N}_M)] - \delta_W\} \frac{\partial N_W}{\partial m} \Big|_1 - \{\beta(1 - \mu)\hat{N}_W\} \frac{\partial N_M}{\partial m} \Big|_1 - \{\beta(1 - \mu)\} \frac{\partial \sigma_W^2}{\partial m} \Big|_1 - \{\beta(1 - \mu)\} \frac{\partial c}{\partial m} \Big|_1. \quad (\text{A3.2})$$

682 Again, we see that we will need to consider partial derivatives of higher-order moments with respect to
 683 m to solve (A3.1). By differentiating equations (A1.8) with $i = M$, (A1.18) with $i = W$, (A1.18) with
 684 $i = M$, and (A1.24), all with respect to m and making the appropriate substitutions for when $m = 1$,
 685 we obtain other equalities involving partial derivatives (similar to (A3.2)). This leads to the following
 686 linear system:

$$687 \quad \mathbf{A} \vec{\partial}_1 = \vec{c}, \quad (\text{A3.3})$$

688 where,

$$689 \quad \mathbf{A} = \begin{bmatrix} (1 - \mu)\{f - \beta(2\hat{N}_W + \hat{N}_M)\} - \delta_W & -\beta(1 - \mu)\hat{N}_W & -\beta(1 - \mu) & 0 & -\beta(1 - \mu) \\ -\beta(1 + \mu)\hat{N}_M + \mu(f - 2\beta\hat{N}_W) & f - 2\beta\hat{N}_M - \beta(1 + \mu)\hat{N}_W - \delta_M & -\beta\mu & -\beta & -\beta(1 + \mu) \\ \delta_W & 0 & -\delta_W & 0 & 0 \\ 0 & \delta_M & 0 & -\delta_M & 0 \\ 0 & 0 & 0 & 0 & -\delta_W - \delta_M \end{bmatrix},$$

690

$$\vec{\partial}_1 = \begin{bmatrix} \left. \frac{\partial N_W}{\partial m} \right|_1 \\ \left. \frac{\partial N_M}{\partial m} \right|_1 \\ \left. \frac{\partial \sigma_W^2}{\partial m} \right|_1 \\ \left. \frac{\partial \sigma_M^2}{\partial m} \right|_1 \\ \left. \frac{\partial C}{\partial m} \right|_1 \end{bmatrix}, \text{ and } \vec{c} = \begin{bmatrix} 0 \\ 0 \\ \hat{N}_W(1-\mu)\{f - \beta(1 + 2\hat{N}_W + \hat{N}_M)\} \\ \hat{N}_M\{f - \beta(1 + 2\hat{N}_M + (1+\mu)\hat{N}_W)\} \\ \hat{N}_W\{\mu(f - \beta(1 + 2\hat{N}_W)) - 2\beta\hat{N}_M\} \end{bmatrix}.$$

691

Solving system (A3.3) and using $\left. \frac{\partial N_W}{\partial m} \right|_1$ and $\left. \frac{\partial N_M}{\partial m} \right|_1$ for equation (A3.1) gives the following:

692

$$\left. \frac{\partial \Phi}{\partial m} \right|_{m=1} = - \frac{\beta\mu}{(1-\mu)^2\delta_M\{(\delta_M/\delta_W)^2-1\}}. \quad (\text{A3.4})$$

693

We abbreviate $\left. \frac{\partial \Phi}{\partial m} \right|_{m=1}$ as $\partial_m \Phi$. From equation (A3.4), it is not difficult to show that $\frac{\partial |\partial_m \Phi|}{\partial \beta} > 0$,

694

$\frac{\partial |\partial_m \Phi|}{\partial \mu} > 0$, $\frac{\partial |\partial_m \Phi|}{\partial \delta_M} < 0$, and $\frac{\partial |\partial_m \Phi|}{\partial \delta_W} > 0$. That is, as the competition coefficient β , the mutation rate μ ,

695

or the death rate of the wild type genotype increase, the addition of structure to an unstructured system

696

leads to a greater increase in the mutant class frequency. As the death rate of the mutant is increased,

697

the addition of structure to an unstructured system leads to a smaller increase in the mutant class

698

frequency.

699

700

Appendix 4: Gillespie Algorithm

701

Our simulation is based on a Gillespie algorithm (Gillespie, 1977) that we coded in the Python

702

2.7 scripting language. The Gillespie algorithm simulates a possible trajectory of a continuous time

703

stochastic system.

704

In our system of P connected patches, patches must be initialized before simulating evolution.

705

Unless otherwise indicated, we seeded our patches with wild-type and mutant individuals by

706

repeatedly drawing from independent Poisson distributions whose parameters are the full migration

707

equilibria $\hat{N}_W$ and $\hat{N}_M$ from (A2.10) and (A2.11), respectively. The initial population defines update

708 zero, for which the time variable t is also zero. As the populations are seeded from their corresponding
 709 $m = 1$ equilibrium distributions, structure is introduced as any limited migration simulation begins.

710 Evolution of the population occurs over “update” steps. First, for update u each patch p
 711 receives four “weights”, corresponding to the four possible events in that patch: a wild-type birth, a
 712 mutant birth, a wild-type death, and a mutant death. Each event’s weight is proportional to its rate. If
 713 we let the number of genotypes W and M in a patch p at update step u be given by $n_W(u, p)$ and
 714 $n_M(u, p)$, respectively, then the weights are defined as follows:

$$715 \quad k_1(u, p) = f n_W(u, p) - \beta n_W(u, p)(n_W(u, p) + n_M(u, p)), \quad (\text{A4.1})$$

$$716 \quad k_2(u, p) = f n_M(u, p) - \beta n_M(u, p)(n_W(u, p) + n_M(u, p)), \quad (\text{A4.2})$$

$$717 \quad k_3(u, p) = \delta_W n_W(u, p), \quad (\text{A4.3})$$

$$718 \quad k_4(u, p) = \delta_M n_M(u, p). \quad (\text{A4.4})$$

719 The event that is attempted at update step u is either a death, a birth with migration, or a birth without
 720 migration, and the decision is made stochastically using the following weights:

$$721 \quad K_{death}(u) = \sum_{p=1}^P k_3(u, p) + \sum_{p=1}^P k_4(u, p), \quad (\text{A4.5})$$

$$722 \quad K_{birth_mig}(u) = m \{ \sum_{p=1}^P k_1(u, p) + \sum_{p=1}^P k_2(u, p) \}, \quad (\text{A4.6})$$

$$723 \quad K_{birth_natal}(u) = (1 - m) \{ \sum_{p=1}^P |k_1(u, p)| + \sum_{p=1}^P |k_2(u, p)| \}. \quad (\text{A4.7})$$

724 Time increment Δt is drawn from an exponential distribution whose rate is equal to $K_{death}(u) +$
 725 $K_{birth_mig}(u) + K_{birth_natal}(u)$, and time parameter t is incremented to $t + \Delta t$. For Figure 2 and
 726 Supplementary Figure 1, the time increment’s exponential distribution rate is equal to $k_1 + k_2 + k_3 +$
 727 k_4 , which does not noticeably affect the resulting time steps.

728 The $K_{birth_natal}(u)$ terms are summed using absolute values due to potentially negative intra-
 729 patch birth rates. Since birth rates decrease linearly with density, $k_1(u, p)$ and $k_2(u, p)$ may be
 730 negative occasionally in particularly crowded patches. For migrating events, the negative births reduce

731 the mean birth rate, which was never negative for the conditions of our simulations. For non-migrating
 732 births, a negative birth event decrements, rather than increments, the chosen genotype population in a
 733 patch. Thus, for any patch r , if a non-migrating birth event is chosen:

$$734 \quad n_W(u + 1, r) = n_W(u, r) + \text{sgn}(k_1(u, r)) \text{ with probability } (1 - \mu) \frac{|k_1(u, r)|}{K_{\text{birth_natal}}(u)}, \quad (\text{A4.8})$$

$$735 \quad n_M(u + 1, r) = n_M(u, r) + \text{sgn}(k_2(u, r)) \text{ with probability } \frac{|k_2(u, r)|}{K_{\text{birth_natal}}(u)} + \mu \frac{|k_1(u, r)|}{K_{\text{birth_natal}}(u)}, \quad (\text{A4.9})$$

736 If a migrating birth event is chosen (i.e., a migrant will ‘land’ on a random patch):

$$737 \quad n_W(u + 1, r) = n_W(u, r) + 1 \text{ with probability } \frac{1-\mu}{P} \sum_{p=1}^P \frac{k_1(u, p)}{K_{\text{birth_mig}}(u)}, \quad (\text{A4.10})$$

$$738 \quad n_M(u + 1, r) = n_M(u, r) + 1 \text{ with probability } \frac{1}{P} \left\{ \sum_{p=1}^P \frac{k_2(u, p)}{K_{\text{birth_mig}}(u)} + \mu \sum_{p=1}^P \frac{k_1(u, p)}{K_{\text{birth_mig}}(u)} \right\}, \quad (\text{A4.11})$$

739 If a death event is chosen:

$$740 \quad n_W(u + 1, r) = n_W(u, r) - 1 \text{ with probability } \frac{k_3(u, r)}{K_{\text{death}}(u)}, \quad (\text{A4.12})$$

$$741 \quad n_M(u + 1, r) = n_M(u, r) - 1 \text{ with probability } \frac{k_4(u, r)}{K_{\text{death}}(u)} \quad (\text{A4.13})$$

742 Simulations were run for 200,000 updates (for $1 \geq m \geq 0.95$), 500,000 updates (for $0.95 > m \geq 0.7$),
 743 or 1,000,000 updates (for $0.7 > m > 0$). Data was recorded every 5000 updates, and equilibril values
 744 represent a simulation’s average state over its final 10,000 updates.

745

746

747 **Acknowledgements**

748 The authors thank Michael Whitlock and Ruth Shaw for their helpful comments, and Jacob Speidel,
 749 Brian Connelly, and Joshua Nahum for their coding help. The authors are also grateful to the
 750 anonymous reviewers for their many useful suggestions. This work was supported in part by the
 751 National Science Foundation Graduate Research Fellowship under grants DGE-0718124 and DGE-

752 1256082, as well as NSF Cooperative Agreement DBI-0939454 and an NSF CAREER Award
753 (DEB0952825).

Figures

Figure 1

Figure 2

A.

B.

Figure 3

A.

B.

Figure 4

A.

B.

Figure 5

A.

B.

TERM	EXPLANATION	ASSUMPTIONS
W, M	wild type, mutant genotype labels	
i, j	arbitrary genotype indices	
n_i	number of genotype i in a patch	
f_i	intrinsic growth rate of genotype i	$f_W = f_M = f$
β_i	competition parameter of genotype i	$\beta_W = \beta_M = \beta$
α_{ij}	competitive effect of genotype j on genotype i	$\alpha_{WM} = \alpha_{MW} = 1$
δ_i	death rate of genotype i	$\delta_M > \delta_W > 0$
$\mu_{i \rightarrow j}$	mutation probability from genotype i to j	$\mu_{W \rightarrow M} = \mu, \mu_{M \rightarrow W} = 0$
m_i	migration probability of genotype i	$m_W = m_M = m$
N_i	expected number of genotype i over all patches	
$N_{i j}$	expected number of genotype i in a patch of a randomly chosen genotype j	
σ_i^2	variance of genotype i over all patches	
C	covariance between genotypes i and j over all patches	
Φ	fraction of mutants at mutation-selection balance	