

Collegio Carlo Alberto



Non-monotonic effects of migration in populations with balancing selection

Pierangelo Lombardo
Andrea Gambassi
Luca Dall'Asta

No. 323
December 2013

Carlo Alberto Notebooks

www.carloalberto.org/research/working-papers

Non-monotonic effects of migration in populations with balancing selection

Pierangelo Lombardo,¹ Andrea Gambassi,¹ and Luca Dall'Asta^{2,3}

¹SISSA – International School for Advanced Studies and INFN, via Bonomea 265, 34136 Trieste, Italy

²Department of Applied Science and Technology – DISAT,

Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

³Collegio Carlo Alberto, Via Real Collegio 30, 10024 Moncalieri, Italy

(Dated: October 21, 2013)

Balancing selection is recognized as a prominent evolutionary force responsible for the maintenance of genetic diversity in natural populations. We quantify its influence on the evolution of a subdivided population, investigating how the mean-fixation time (MFT) depends on the migration rate among subpopulations. We identify a threshold in the strength of the balancing selection above which the MFT changes its qualitative behavior compared to that of neutral populations, developing an unexpected non-monotonic dependence on the migration rate. This feature carries over into an analogous behavior of the heterozygosity, which is an index of the biodiversity of the population.

PACS numbers: 87.23.Kg, 87.23.Cc, 05.40.-a

I. INTRODUCTION

A central problem in population biology is that of identifying the mechanisms which might be responsible for the conservation of biodiversity (i.e., for the presence of different genotypes within a population) under evolution. Balancing selection [1] in various forms, such as frequency-dependent selection, over-dominance, and heterozygote advantage, was demonstrated to maintain high biodiversity in several contexts, most notably in mammalian major histocompatibility complex genes [2] and in some genetic self-incompatibility loci in plants [3]. The maintenance of some genetic diseases in human populations, such as sickle-cell anemia [4], thalassemia [5] and cystic fibrosis [6] is also ascribed to balancing selection. Analogous mechanisms are responsible for cooperation and mutualistic behaviors in ecology [7, 8] — such as those recently observed in microbial communities [9] — and in game theory [10].

In the absence of mutations, the fate of a finite population is inescapably determined by fixation events induced by the intrinsic noise (*genetic drift*). However, their statistical properties may depend on several features such as the form of the selective pressure and the geographical structure [8, 11–13]. The role of balancing selection on the fixation properties of well-mixed (panmictic) populations has been known for long time [14], but natural populations are often subdivided or fragmented in space, with the possibility of internal migration. Here we investigate the interplay between balancing selection and subdivision, focusing on their combined influence on relevant features of the evolution, such as the fixation probability and the mean fixation time (MFT), for subpopulations of equal and constant sizes.

In population genetics, it is widely accepted that fixation is not significantly affected by subdivision [15], because a subdivided population is approximately equivalent to a well-mixed one with *effective* selection coefficient s_e and population size N_e [16, 17]. However, N_e and s_e depend on the migration rate m among subpopulations,

and therefore also the fixation probability and the MFT depend on it. For neutral evolution (i.e., in the absence of genetic selection) or constant selection, the MFT monotonically increases upon decreasing m [12, 13, 18]. In the presence of balancing selection a similar behavior is expected, with a fixation probability and MFT which satisfy certain bounds for slow and fast migration [17].

Here we develop an approximate self-consistent mean-field-like approach to evaluate the MFT for a subdivided population that undergoes balancing selection. For weak selection, our approximation renders the one of Ref. [12]. However, we show that, contrary to what implicitly assumed in Ref. [17], the MFT can develop a minimum as a function of the migration rate for sufficiently strong selection. The existence of this minimum depends, *inter alia*, on the optimal frequency, i.e., on the amount of biodiversity, promoted by balancing selection alone. The non-monotonicity of MFT is reflected in the behavior of the time-dependent heterozygosity, which quantifies the biodiversity within the subdivided population.

II. THE MODEL

Inspired by common models in population genetics, we consider $\Omega \gg 1$ haploid individuals and focus on a single locus with two competing alleles A and B . The evolution of a large but finite diallelic population can be effectively described by using a *diffusion approximation* [19], i.e., a Langevin equation for the frequency x of, e.g., allele A . The mean change of x in a panmictic population is $\mu(x) = \tilde{s}x(1-x)$, where \tilde{s} is the selection rate, while the variance is approximately given by $v(x) = x(1-x)/(\Omega\tau_g)$, where τ_g is the generation time (see Ref. [20] for a derivation of these expressions from microscopic models). Hereafter, time is measured in units of generations, so that $\tau_g = 1$ and the rates become dimensionless quantities. Balancing selection is characterized by $\tilde{s} = s(x_* - x)$, where $s > 0$ is a constant selection coefficient and x_* represents the internal optimal frequency which is pro-

moted by balancing effects in an infinite population.

In order to investigate the influence of migration on structured populations with balancing selection in the simplest possible setting, we consider the celebrated Island model, originally proposed by Wright [21] for neutral evolution. It consists of N subpopulations (demes), each forming a panmictic population of Ω identical haploid individuals. Each deme is subject to balancing selection with selection rate s and optimal frequency x_* , while being allowed to *exchange* a randomly picked individual with any other deme at a rate m/N , such that the sizes of the involved subpopulations are unchanged. For sufficiently large values of Ω , and small m and s , the evolution of the allele frequency $x_i \in [0, 1]$ in the i -th deme is described, within the diffusion approximation [19, 20], by the Langevin equation (with Itô prescription)

$$\dot{x}_i = \mu(x_i) + m(\bar{x} - x_i) + \sqrt{v(x_i)} \eta_i, \quad (1)$$

where η_i are independent Gaussian noises with $\langle \eta_i(t)\eta_j(t') \rangle = \delta_{i,j}\delta(t-t')$; hereafter the overbar denotes inter-deme averages, e.g., $\bar{x}^k = \sum_i x_i^k/N$, and thus \bar{x} is the inter-deme mean frequency (IDMF). For $m=0$, the demes are independent: the deterministic selection term μ in Eq. (1) drives x_i towards x_* , while the random genetic drift eventually drives x_i towards one of the two possible absorbing states $x_i = 0$ and 1 , corresponding to fixation of allele B and A , respectively (see Fig. 1(a)). For $m > 0$, migration acts as a source of biodiversity on the subpopulations, preventing the occurrence of independent fixation events (see Fig. 1(b)-(c)) and favoring a coordinate evolution of the interacting demes. While for $\Omega m \gg 1$ and x_* sufficiently close to 0 or 1, the collective evolution rapidly drives all demes into the same absorbing state, within the range of parameters considered here the IDMF \bar{x} fluctuates for a long time around a value \hat{x} — characterized by the vanishing of the deterministic force in the dynamics of \bar{x} — until fixation eventually occurs through a rare (for large N) fluctuation [20]. This coordinated behavior around \hat{x} becomes effectively a *metastable* state if the typical time T_{rel} required to reach it from the initial condition is significantly smaller than the typical time T_{fluct} for fixation to occur. This condition is satisfied for $ms\Omega^2 N \gg 1$ [20]. The fixation properties of the population can be studied by considering the evolution equation of \bar{x} , which follows from Eq. (1) [20],

$$\dot{\bar{x}} = s[x_*\bar{x} - (1+x_*)\bar{x}^2 + \bar{x}^3] + \sqrt{(\bar{x} - \bar{x}^2)/(\Omega N)} \eta, \quad (2)$$

where η is a Gaussian noise with $\langle \eta(t)\eta(t') \rangle = \delta(t-t')$. This equation involves higher-order moments, and the hierarchy does not close; however, we can proceed by introducing a moment closure scheme based on a self-consistent mean-field-like approximation.

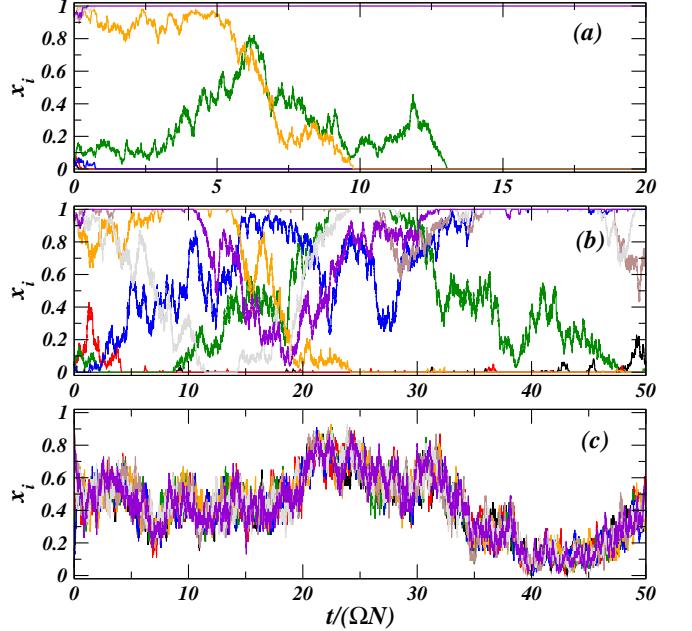


Figure 1. Time evolution of the frequency x_i of allele A in the various demes of a fully-connected population consisting of $N = 8$ demes with $\Omega = 100$ individuals each (a) in the absence of migration ($m = 0$) or (b) for small ($\Omega m = 0.05$) and (c) large ($\Omega m = 50$) migration rate. The balancing selection is characterized here by $x_* = 0.5$ and $\Omega s = 5$. At time $t = 0$, half of the demes have $x_i = 0.05$, while the remaining ones $x_i = 0.95$. Upon increasing N , the fluctuations of \bar{x} around x_* reduce significantly in panel (c).

III. THE APPROXIMATION

In the absence of selection ($s = 0$) Eq. (2) is driven only by the genetic drift, and therefore, for sufficiently large N ($N\Omega m \gg 1$ [20]), a separation of time scales emerges between this global dynamics of \bar{x} and that of the local allele frequencies $\{x_i\}$. Being coupled only via the slowly varying quantity \bar{x} , the frequencies $\{x_i\}$ can be considered as almost independent random variables, each one described by a conditional quasi-stationary distribution $P_{\text{qs}}(x_i|\bar{x})$. The latter can be obtained by solving the stationary Fokker-Planck equation associated with Eq. (1), in which \bar{x} is treated as a constant parameter. Under these assumptions the population average $\bar{x}^k(t)$ can be approximated, for $N \gg 1$, by the corresponding mean $\int dx_i x_i^k P_{\text{qs}}(x_i|\bar{x})$. For $s = 0$ one obtains $P_{\text{qs}}(x|\bar{x}) \propto x^{2m'\bar{x}-1}(1-x)^{2m'(1-\bar{x})-1}$, where $m' = \Omega m$ is a rescaled rate introduced for convenience and $P_{\text{qs}}(x|\bar{x})$ satisfies the consistency condition $\bar{x} = \int_0^1 dx x P_{\text{qs}}(x|\bar{x})$. The distribution $P_{\text{qs}}(x|\bar{x})$ can then be used for evaluating the higher-order moments contained in Eq. (2) and for calculating the mean drift $M(\bar{x})$ and variance $V(\bar{x})$ of the (stochastic) variable \bar{x} [12]:

$$M(\bar{x}) = s_e \bar{x}(1-\bar{x})(x_* - \bar{x}) \text{ and } V(\bar{x}) = \frac{\bar{x}(1-\bar{x})}{N_e}. \quad (3)$$

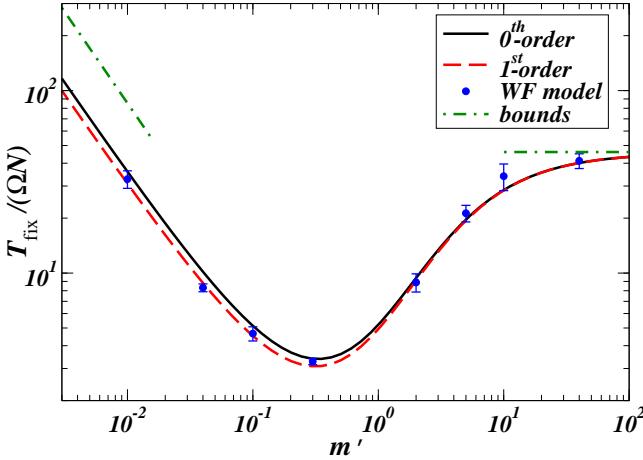


Figure 2. Mean fixation time as a function of the migration rate m' with $N = 30$, $\Omega = 100$, $s' = 1$, and $x_* = 0.5$. The solid line corresponds to Eq. (5), while the dashed line accounts for the first-order correction in s_e/m ; symbols with errorbars are the results of numerical simulations of the Wright-Fisher (WF) model with selection. The dash-dotted lines indicate the upper bounds for small and large migration, found in Ref. [17] (see also Ref. [20]).

This means that at the lowest, non-vanishing order in s , the structured population behaves like a well-mixed one with an *effective selection coefficient* $s_e = s / [(1 + \frac{1}{m'}) (1 + \frac{1}{2m'})]$ and an *effective population size* $N_e = N\Omega (1 + \frac{1}{2m'})$. The time scale T_{migr} associated with the response of x_i to a variation in \bar{x} can be read from Eq. (1) and it is $T_{\text{migr}} \simeq 1/m$. The typical time scale of the dynamics of \bar{x} is determined, instead, either by the time scale $T_{\text{rel}} \simeq 1/s_e$ of the drift or by the time scale $T_{\text{fluct}} \simeq N_e$ of the stochastic term in Eq. (3). When $T_{\text{rel}} < T_{\text{fluct}}$, i.e., $N\Omega s > 1 + 1/m'$, our approximation requires $T_{\text{rel}} \gg T_{\text{migr}}$, i.e., $s_e \ll m$, while in the opposite case, it is accurate whenever $N \gg 1$ (see Ref. [20] for a detailed discussion). This approximation can be generalized to small but non-vanishing values of s_e/m by accounting (a) for $s \neq 0$ in the quasi-stationary distribution P_{qs} and (b) for the fact that \bar{x} slowly changes during the fast evolution of x_i , which results in a distribution $P_{\text{qs}}(x_i|y(t))$ where the *effective field* $y(t) \simeq \bar{x}(t)$ has to be determined self-consistently. The single-deme quasi-stationary distribution for $s' \equiv \Omega s \neq 0$ is

$$P_{\text{qs}}(x|y) \propto x^{2m'y-1} (1-x)^{2m'(1-y)-1} e^{s'x(2x_*-x)}. \quad (4)$$

The consistency condition $\bar{x} = \int_0^1 dx x P_{\text{qs}}(x|y)$ gives $y = \bar{x} - (s_e/m)\bar{x}(1-\bar{x})(x_* - \bar{x}) + O((s_e/m)^2)$, which can be used together with Eq. (4) in order to calculate higher-order corrections in s to $M(\bar{x})$ and $V(\bar{x})$ [20].

IV. MEAN FIXATION TIME

On the basis of $M(\bar{x})$ and $V(\bar{x})$ calculated as discussed above, the MFT $T_{\text{fix}}(\bar{x})$ for the whole population with an initial IDMF \bar{x} is determined within the diffusion approximation by $V(\bar{x})T_{\text{fix}}''(\bar{x})/2 + M(\bar{x})T_{\text{fix}}'(\bar{x}) = -1$ [22]. For $x_* = 1/2$, by using the lowest-order approximations ${}^{(0)}$ for M and V in Eq. (3) and choosing the state $\bar{x} = 1/2$ (corresponding to the metastable state) as initial condition, we get

$$T_{\text{fix}}^{(0)} = \frac{N_e}{2} \int_0^1 dy \int_0^1 dz \frac{e^{s_e N_e y(1-z^2)/4}}{1 - yz^2}, \quad (5)$$

which reaches a constant value for $m' \gg 1$, while $T_{\text{fix}}^{(0)}/(N\Omega) \simeq (\log 2)/(2m')$ for $m' \ll 1$. Figure 2 shows $T_{\text{fix}}^{(0)}$ (solid line) as a function of m' for the population specified in the caption, together with the prediction (dashed line) obtained by including the first-order correction in s_e/m to the mean drift $M(\bar{x})$ and variance $V(\bar{x})$ [20]. The MFT shows a marked non-monotonic dependence on the migration rate m' , while complying with the bounds of Ref. [17] for small and large m' (dash-dotted lines). In fact, $T_{\text{fix}}^{(0)}(m' \gg 1)$ approaches the value it would have in a well-mixed population of ΩN individuals, whereas for $m' \ll 1$ fixation — and thus T_{fix} — is controlled by the growing time scale $T_{\text{migr}} \propto 1/m'$ associated with migration. In this respect, the limit $m' \rightarrow 0$ differs essentially from the case $m' = 0$, in which T_{fix} is governed by the single-deme fixation times, is finite, and scales $\propto \log N$ for large N [20].

In order to demonstrate the accuracy of our analytical predictions, Fig. 2 reports the results (symbols with errorbars) of numerical simulations of the Wright-Fisher (WF) microscopic model with balancing selection [20]. Their agreement with the analytical prediction of Eq. (5) is very good and further improves upon including the first-order corrections in s_e/m (dashed line).

Figure 3(a) shows that the non-monotonicity displayed in Fig. 2 is enhanced upon increasing $\sigma \equiv s'N$, while it disappears for $\sigma < \sigma_c$, where σ_c is a critical threshold below which the MFT behaves qualitatively as in a neutral population with $s = 0$. The value m'_{\min} of m' at which T_{fix} is minimum diverges for $\sigma \rightarrow \sigma_c \simeq 5.2$ [20] and decreases upon increasing $\sigma > \sigma_c$, as shown in Fig. 3(b). The value σ_c slightly depends on s_e/m if the corrections to Eq. (3) are included. Figure 3(c) shows that the non-monotonicity of $T_{\text{fix}}^{(0)}$ persists also for $x_* \neq 1/2$, but only within an interval of values of x_* which depends on σ — as indicated by the shaded area in Fig. 3(d) — and which stretches over the entire range for very large σ .

V. BIODIVERSITY

Migration is expected to affect the level of biodiversity observed in a population. In diallelic models, this effect

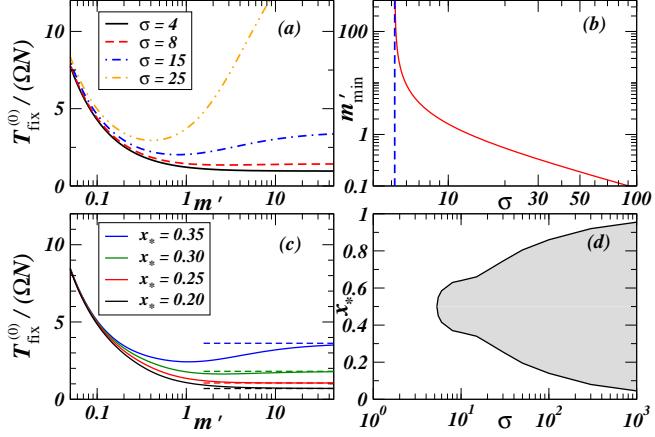


Figure 3. Features of the mean fixation time $T_{\text{fix}}^{(0)}$ in Eq. (5) for a population of $N = 30$ demes with $\Omega = 100$ individuals each: (a) dependence of $T_{\text{fix}}^{(0)}$ on m' for $x_* = 0.5$ and various values of σ ; (b) m'_min as a function of σ ; (c) $T_{\text{fix}}^{(0)}$ as a function of m' for $\sigma = 30$ and various $x_* \neq 0.5$; (d) region of the parameter space (σ, x_*) within which $T_{\text{fix}}^{(0)}$ is a non-monotonic function of m' .

is usually studied in terms of (i) the *global heterozygosity* $H = 2\bar{x}(1 - \bar{x})$, which quantifies the diversification of the global population but neglects the possible subdivision in demes, and of (ii) the *intra-deme heterozygosity* $h = (2/N) \sum_{i=1}^N x_i(1 - x_i) = 2\bar{x}(1 - \bar{x})$, which measures the average level of diversification inside each deme. Note that $0 \leq h \leq H \leq 1/2$. $H = 0$ corresponds to the loss of global biodiversity, namely all individuals within the population have the same genotype; $H = 1/2$, instead, corresponds to the maximal possible global biodiversity in which the two genotypes are equally present within the whole population. Analogous interpretation holds for $h = 0$ and $h = 1/2$ at the intra-deme level. As depicted in Fig. 1(c) the local allele frequencies $\{x_i\}$ approach each other for $m' \gg 1$, with $x_i \simeq x_j$ and therefore $h \simeq H$. In the case of moderate migration rate $m' \lesssim 1$ of Fig. 1(b), instead, different demes fix different alleles, causing $h \simeq 0$, while H is maintained positive by migration which acts as a constant source of biodiversity.

In order to understand how migration affects biodiversity before the eventual fixation $H = h = 0$, we assume that the population at time $t = 0$ is in the metastable state $\bar{x} = \hat{x}$ [20], such that $H(0) = 2\hat{x}(1 - \hat{x})$ and that it persists in this state until fixation occurs. Under this heuristic assumption, one can approximate $H(t) \simeq [1 - p_{\text{fix}}(\hat{x}, t)]H(0)$, where $p_{\text{fix}}(x_0, t)$ is the probability that a population prepared with $\bar{x} = x_0$ at time $t = 0$ has already fixed at time t . This probability satisfies the backward Fokker-Planck equation $\partial_t p_{\text{fix}} = M(x_0) \partial_{x_0} p_{\text{fix}} + V(x_0) \partial_{x_0}^2 p_{\text{fix}}/2$, which can be integrated numerically [20]. By using the expressions of M and V in Eq. (3), the results of this approximation for H are presented in Fig. 4 as functions of m' for some values of the time t and they are compared with the results of numer-

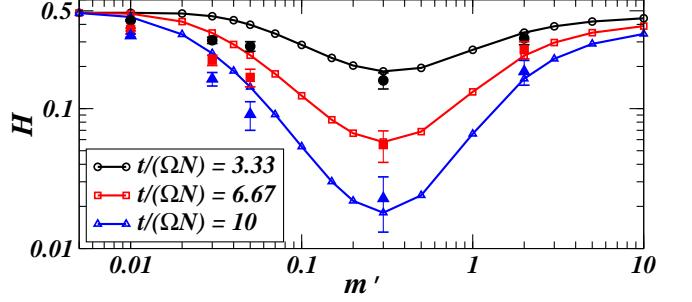


Figure 4. Dependence of the global heterozygosity H on the migration rate m' at various times, for a subdivided population with $\Omega = 100$, $N = 30$, $s' = 1$, and $x_* = 1/2$: the prediction of the approximation described in the text (solid lines) is compared with the results of simulations of the WF model (symbols with error-bars).

ical simulations of the WF model (symbols with error-bars) [20]. Note that the resulting estimate of $H(t)$ is expected to become less accurate as $m'\sigma$ exceeds 1 because, correspondingly, the state $\bar{x} \simeq \hat{x}$ is no longer metastable [20]. For slow and fast migration $H(t) \simeq H(0)$ for a rather long time whereas $H(t)$ rapidly decreases in time for intermediate values of the migration rate. For a fixed time and as a function of m' , instead, H has a minimum at $m' \simeq m'_\text{min}$, indicating — in contrast to the heuristic expectation — that the global biodiversity can be enhanced upon increasing migration [23]. Our predictions agree rather well with the results of simulations, apart, as expected, from $m' \lesssim 1/\sigma \simeq 0.03$. A similar study of both $H(t)$ for different values of the parameters and $h(t)$ [20] highlights a non-monotonic dependence on m' whenever the corresponding T_{fix} develops a minimum.

VI. CONCLUSIONS AND OUTLOOK

We have shown that, contrary to a common belief in population biology, the mean fixation time of a subdivided population can develop a minimum, becoming a non-monotonic function of the migration rate. This occurs in the presence of balancing selection, an evolutionary mechanism which promotes the coexistence of different genetic traits within the same populations. The emergence of the minimum depends on both the strength $\sigma \equiv s\Omega N$ of the balancing selection exceeding a threshold and on the frequency x_* of coexistence which is promoted by the selection. We describe the dynamics of the entire population via an effective Langevin equation, which hinges on the separation of time scales between the intra- and the inter-deme dynamics and on a mean-field-like approximation. The perturbative scheme which we used in order to calculate the mean fixation time can be extended to additional dynamical features of the population, while the accuracy of our predictions is demonstrated by the comparison with the numerical simulation of the Wright-Fisher model. Our approach carries over to

those problems of evolutionary dynamics in subdivided populations, in which fast local variables are influenced by slow global "mean-field-like" quantities. The observed non-monotonic behavior of the MFT should appear also in subdivided populations with more complex migration structure, possibly even in spatially embedded popula-

tions, provided that balancing selection is present.

ACKNOWLEDGMENTS

L.D. acknowledges the Italian FIRB Project No. RBFR10QUW4.

-
- [1] M. Hamilton, *Population Genetics*, Wiley-Blackwell (2009).
 - [2] A. L. Hughes and M. Nei, *Nature* **335**, 167 (1998); N. Takahata, *Proc. Natl. Acad. Sci. USA* **87**, 2419 (1990).
 - [3] S. Wright, *Genetics* **24**, 538 (1939); A. D. Richman, M. K. Uyenoyama, and J. R. Kohn, *Science* **273**, 1212 (1996).
 - [4] M. Aidoo, D. J. Terlouw, M. S. Kolczak, P. D. McElroy, F. O. ter Kuile, S. Kariuki, B. L. Nahlen, A. A. Lal, and V. Udhayakumar, *Lancet* **359**, 1311 (2002).
 - [5] D. J. Weatherall, *Ann. Trop. Med. Parasitol.* **91**, 885 (1997).
 - [6] S. A. Schroeder, D. M. Gaughan, and M. Swift, *Nature Med.* **1**, 703 (1995).
 - [7] M. A. Nowak and K. Sigmund, *Science* **303**, 793 (2004).
 - [8] K. S. Korolev and D. R. Nelson, *Phys. Rev. Lett.* **107**, 088103 (2011); L. Dall'Asta, F. Caccioli, and D. Beghè, *EPL* **101**, 18003 (2013).
 - [9] J.B. Xavier, *Mol. Sys. Biol.* **7**, 483 (2011).
 - [10] A. Traulsen, J. C. Claussen, and C. Hauert, *Phys. Rev. Lett.* **95**, 238701 (2005).
 - [11] M. Khasin, B. Meerson, E. Khain, and L. M. Sander, *Phys. Rev. Lett.* **109**, 138104 (2012); M. Khasin, E. Khain, and L. M. Sander, *Phys. Rev. Lett.* **109**, 248102 (2012); B. Waclaw, R. J. Allen, and M. R. Evans, *Phys. Rev. Lett.* **105**, 268101 (2010).
 - [12] J. L. Cherry and J. Wakeley, *Genetics* **163**, 421 (2003); J. L. Cherry, *Genetics* **163**, 1511 (2003).
 - [13] G. J. Baxter, R. A. Blythe, and A. J. McKane, *Phys. Rev. Lett.* **101**, 258701 (2008).
 - [14] A. Robertson, *Genetics* **47**, 1291 (1962); M. Nei and A. K. Roychoudhury, *Genetics* **74**, 371 (1973).
 - [15] With the notable exception of unequal or non-constant subpopulation sizes [11].
 - [16] T. Maruyama, *Genet. Res. Camb.* **15**, 221 (1970).
 - [17] M. Slatkin, *Evolution* **35**, 477 (1981).
 - [18] M. C. Whitlock, *Genetics* **164**, 767 (2003).
 - [19] M. Kimura, *J. Appl. Prob.* **1**, 177 (1964).
 - [20] P. Lombardo, A. Gambassi, and L. Dall'Asta, to appear (2013).
 - [21] S. Wright, *Genetics* **16**, 97 (1931).
 - [22] M. Kimura, and T. Ohta, *Genetics* **61**, 763 (1969).
 - [23] A non-monotonic heterozygosity as a function of migration has been already reported in a different class of models: see M. H. Schierup, *Genetics* **149**, 1153 (1998); M. H. Schierup, X. Vekemans, and D. Charlesworth, *Genet. Res. Camb.* **76**, 51 (2000); C. A. Muirhead, *Evolution* **55**, 1532 (2001); J. Nishino and F. Tajima, *Genes Genet. Syst.* **80**, 287 (2005). In these works, however, biodiversity is sustained by mutations, no fixation occurs, and the population reaches a non trivial stationary state.