

# Two-locus clines on the real line with a step environment

Reinhard Bürger

Department of Mathematics,  
University of Vienna,  
Austria

**Address for correspondence:**

Reinhard Bürger  
Fakultät für Mathematik  
Universität Wien  
Oskar-Morgenstern-Platz 1  
A-1090 Wien  
Austria  
E-mail: reinhard.buerger@univie.ac.at  
Phone: +43 1 4277 50784  
Fax: +43 1 4277 850784

## Abstract

The shape of allele-frequency clines maintained by migration-selection balance depends not only on the properties of migration and selection, but also on the dominance relations among alleles and on linkage to other loci under selection. We investigate a two-locus model in which two diallelic, recombining loci are subject to selection caused by an abrupt environmental change. The habitat is one-dimensional and unbounded, selection at each locus is modeled by step functions such that in one region one allele at each locus is advantageous and in the other deleterious. We admit an environmentally independent, intermediate degree of dominance at both loci, including complete dominance. First, we derive an explicit expression for the single-locus cline with dominance, thus generalizing classical results by Haldane (1948). We show that the slope of the cline in the center (at the step) or, equivalently, the width of the cline, is independent of the degree of dominance. Second, under the assumption of strong recombination relative to selection and migration, the first-order approximations of the allele-frequency clines at each of the loci and of the linkage disequilibrium are derived. This may be interpreted as the quasi-linkage-equilibrium approximation of the two-locus cline. Explicit asymptotic expressions for the clines are deduced as  $x \rightarrow \pm\infty$ . For equivalent loci, explicit expressions for the whole clines are derived. The influence of dominance and of linkage on the slope of the cline in the center and on a global measure of steepness are investigated. This global measure reflects the influence of dominance. Finally, the accuracy of the approximations and the dependence of the shape of the two-locus cline on the full range of recombination rates is explored by numerical integration of the underlying system of partial differential equations.

**Key words:** Selection; Migration; Recombination; Dominance; Linkage disequilibrium; Dispersal; Geographical structure; Population structure

# 1 Introduction

A cline describes a gradual change in genotypic or phenotypic frequency as a function of spatial location. Such clines are frequently observed in natural populations and are an important and active research area in evolutionary biology and ecology (e.g., Endler 1977, Hoffman et al. 2002, Lohman et al. 2017). Clines typically occur in species distributed along an environmental gradient, for instance in temperature, where alternative phenotypes or genotypes are better adapted to the different extremes of the environment. Dispersal leads to mixing, reduces local adaptation, and entails a continuous, often sigmoidal, change in type frequencies across space. The study of clines can be used to obtain insight into the relative strengths of the evolutionary and ecological forces acting on this species.

Haldane (1948) devised a model in terms of a reaction-diffusion equation which approximates migration by diffusion and assumes that there is a step environment on the real line such that one allele is advantageous if  $x > 0$  and the other if  $x < 0$ . He derived explicit expressions for the cline, the spatially non-constant stationary solution, in terms of hyperbolic functions for the two cases of no dominance and of a completely dominant allele. The slope of the cline in the center, i.e., at the environmental step, can be expressed in terms of the selection intensity and the migration variance. He used this relation to infer the strength of selection on a population of deer mouse.

The mathematical theory of clines became a very active research area in the 1970s. Various patterns of spatial variation in fitnesses were investigated (e.g., environmental pockets, periodic changes), as were variation (e.g., barriers) or asymmetry in migration (e.g., Slatkin 1973; Nagylaki 1975, 1976, 1978). These works focused on the derivation of explicit results about the shape of clines. At about the same time and motivated by this work, Conley (1975), Fleming (1975), Fife and Peletier (1977), and Henry (1981) developed and employed advanced mathematical methods to investigate existence, uniqueness, and stability of clinal solutions under a variety of assumptions about fitnesses, i.e., for quite general classes of functions that describe selection caused by the environment.

Lou and Nagylaki (2002, 2004, 2006) extended much of the previous work on spatially varying selection in several directions. In most of their analyses, migration is

modeled by general elliptic operators on bounded domains in arbitrary dimensions and for wide classes of fitness functions. Such elliptic operators arise if migration is spatially inhomogeneous or anisotropic (Nagylaki 1989). In addition, they studied the maintenance of clines at multiallelic loci, which does not only add realism but also produces new and interesting phenomena.

In the present work, we focus on the role of dominance and of linkage between loci. In particular, we investigate how the shape of a cline at one locus is affected by linkage to a second locus. We choose a step environment on the real line, such that in each of the two regions ( $x > 0$ ,  $x < 0$ ) one of the alleles at each locus is advantageous, the other deleterious. The strength of selection acting at the two loci may be different, though. The choice of a step environment on the whole real line (as opposed to a bounded interval) has the advantage that explicit results are obtained more readily.

We admit arbitrary intermediate dominance, including complete dominance, and assume that its degree is independent of the environment. The influence of dominance on the maintenance of single-locus clines has been studied before. In particular, on bounded domains the number and stability of clines depend on the degree of dominance (Henry 1981, Lou and Nagylaki 2002, Nagylaki and Lou 2008, Lou et al. 2010, Nakashima et al. 2010). Explicit results about the shape of a cline seem to be rare and confined to the two cases of no dominance and complete dominance (e.g., Haldane 1948, Nagylaki 1975).

The first study of a two-locus cline model is due to Slatkin (1975), who showed numerically that the linkage disequilibrium generated between the two loci tends to steepen the clines. Barton (1983, 1986, 1999) derived general results about the consequences of linkage and linkage disequilibrium among multiple loci. Although he derived them for and applied them to hybrid zones, they are also of relevance in our context.

One of the novel features of our work is that we derive an analytically explicit solution for the single-locus cline with dominance (Section 3). In particular, we show that the slope of the cline in its center or, equivalently, the width of the cline is independent of the degree of dominance. Our main achievement is the derivation of the first-order perturbation of the allele-frequency clines at each of the loci if recombination is strong relative to selection and diffusion (Section 4.1). In other words, we derive the quasi-linkage-equilibrium approximation for the two-locus cline. In Sections 4.2 and 4.5,

we study two measures of steepness of clines and their dependence on dominance and linkage. We derive the asymptotic properties of the allele-frequency clines of the two-locus system in Section 4.3. For equivalent loci, we obtain an analytically explicit expression for the strong-recombination approximation of the two-locus cline (Section 4.4). In Section 5, we briefly treat the case of no recombination. Finally, we provide numerical checks of the accuracy of our approximations and illustrate the dependence of two-locus clines on the full range of recombination rates by numerical integration of the system of partial differential equations (Section 6).

## 2 The model

We consider a monoecious, diploid population that occupies a linear, unbounded habitat in which it is uniformly distributed and mates locally at random. Fitness of individuals depends on location and is determined by two diallelic loci,  $\mathcal{A}$  and  $\mathcal{B}$ , which recombine at rate  $r \geq 0$ . We model dispersal by diffusion on the real line  $\mathbb{R} = (-\infty, \infty)$ , and assume it is homogeneous, isotropic, and genotype-independent, with migration variance  $\sigma^2$ .

The frequencies of the gametes  $AB$ ,  $Ab$ ,  $aB$ , and  $ab$ , at position  $x \in \mathbb{R}$  and time  $t$  are  $p_1 = p_1(x, t)$ ,  $p_2 = p_2(x, t)$ ,  $p_3 = p_3(x, t)$ , and  $p_4 = p_4(x, t)$ , respectively, where  $p_i \geq 0$  and  $\sum_{i=1}^4 p_i = 1$ . Let  $D = p_1 p_4 - p_2 p_3$  denote the usual measure of linkage disequilibrium, and let  $\mathbf{p} = (p_1, p_2, p_3, p_4)^T$ . If  $w_{ij}(x)$  is the fitness of the diploid genotype  $ij$  ( $i, j \in \{1, 2, 3, 4\}$ ) at location  $x \in \mathbb{R}$ , then  $w_i = w_i(x, \mathbf{p}) = \sum_{j=1}^4 w_{ij}(x) p_j$  is the marginal fitness of gamete  $i$ , and  $\bar{w} = \bar{w}(x, \mathbf{p}) = \sum_{i=1}^4 w_i p_i$  is the population mean fitness. For a glossary of symbols, see Table 1.

Throughout, we use primes,  $'$ , and dots,  $\dot{\cdot}$ , to indicate partial derivatives with respect to  $x$  and  $t$ , respectively. We assume that (i) the three evolutionary forces selection, migration, and recombination are of the same order of magnitude and sufficiently weak, (ii) migration is genotype independent and spatially uniform and symmetric, and (iii) Hardy-Weinberg proportions obtain locally. Defining  $\eta_1 = \eta_4 = -\eta_2 = -\eta_3 = 1$  and proceeding as in Nagylaki (1975, 1989), we derive the following diffusion approximation for the evolution of gamete frequencies:

$$\dot{p}_i = \frac{\sigma^2}{2} p_i'' + p_i(w_i - \bar{w}) - \eta_i r D, \quad i \in \{1, 2, 3, 4\}, \quad (2.1a)$$

subject to the initial conditions

$$0 \leq p_i(x, t) \leq 1, \quad \sum_{i=1}^4 p_i(x, t) = 1 \quad \text{for } t = 0 \text{ and every } x \in \mathbb{R} \quad (2.1b)$$

(cf. Slatkin 1975). Solutions satisfy the constraints (2.1b) for every  $t \geq 0$ .

Throughout, we assume absence of epistasis. Then the genotypic fitnesses can be written as

	<i>BB</i>	<i>Bb</i>	<i>bb</i>	
<i>AA</i>	$\alpha(x) + \beta(x)$	$\alpha(x) + h_B\beta(x)$	$\alpha(x) - \beta(x)$	(2.2)
<i>Aa</i>	$h_A\alpha(x) + \beta(x)$	$h_A\alpha(x) + h_B\beta(x)$	$h_A\alpha(x) - \beta(x)$	
<i>aa</i>	$-\alpha(x) + \beta(x)$	$-\alpha(x) + h_B\beta(x)$	$-\alpha(x) - \beta(x)$	

where it is easy to show that for a continuous-time model this scaling is general because absence of epistasis is assumed (Appendix A.1). We could have introduced spatially dependent dominance coefficients,  $h_A(x)$  and  $h_B(x)$ . However, in view of our applications, we refrained from doing so. In order to have unique single-locus clines (Fife and Peletier 1981), we assume throughout

$$-1 \leq h_A \leq 1 \quad \text{and} \quad -1 \leq h_B \leq 1. \quad (2.3)$$

Dominance is absent at locus  $\mathcal{A}$  ( $\mathcal{B}$ ) if  $h_A = 0$  ( $h_B = 0$ ).

For our purposes, it will be convenient to follow the evolution of the allele frequencies  $p_A = p_1 + p_2$  and  $p_B = p_1 + p_3$ , and the linkage disequilibrium  $D$ . With the abbreviations

$$\vartheta_A = 1 + h_A - 2h_A p_A, \quad \vartheta_B = 1 + h_B - 2h_B p_B, \quad (2.4)$$

straightforward calculations yield

$$w_1(x, \mathbf{p}) - \bar{w}(x, \mathbf{p}) = \alpha(x)(1 - p_A)\vartheta_A + \beta(x)(1 - p_B)\vartheta_B, \quad (2.5a)$$

$$w_2(x, \mathbf{p}) - \bar{w}(x, \mathbf{p}) = \alpha(x)(1 - p_A)\vartheta_A - \beta(x)p_B\vartheta_B, \quad (2.5b)$$

$$w_3(x, \mathbf{p}) - \bar{w}(x, \mathbf{p}) = -\alpha(x)p_A\vartheta_A + \beta(x)(1 - p_B)\vartheta_B, \quad (2.5c)$$

$$w_4(x, \mathbf{p}) - \bar{w}(x, \mathbf{p}) = -\alpha(x)p_A\vartheta_A - \beta(x)p_B\vartheta_B. \quad (2.5d)$$

Now it is easy to show that the system of differential equations (2.1a) with the fitnesses (2.5) is equivalent to

$$\dot{p}_A = p_A'' + \lambda\alpha(x)p_A(1 - p_A)\vartheta_A + \lambda\beta(x)\vartheta_B D, \quad (2.6a)$$

$$\dot{p}_B = p_B'' + \lambda\beta(x)p_B(1 - p_B)\vartheta_B + \lambda\alpha(x)\vartheta_A D, \quad (2.6b)$$

$$\dot{D} = D'' + 2p_A'p_B' + \lambda[\alpha(x)(1 - 2p_A)\vartheta_A + \beta(x)(1 - 2p_B)\vartheta_B]D - \rho D, \quad (2.6c)$$

where time has been rescaled such that

$$\lambda = 2/\sigma^2 \quad \text{and} \quad \rho = r\lambda = 2r/\sigma^2. \quad (2.7)$$

The constraints (2.1b) on the  $p_i$  are transformed to

$$0 \leq p_A \leq 1, \quad 0 \leq p_B \leq 1, \quad (2.8a)$$

and

$$-\min\{p_A p_B, (1-p_A)(1-p_B)\} \leq D \leq \min\{p_A(1-p_B), (1-p_A)p_B\}, \quad (2.8b)$$

where these inequalities hold for every  $(x, t) \in \mathbb{R} \times [0, \infty)$ . We will impose the boundary conditions

$$p'_A(\pm\infty, t) = 0, \quad p'_B(\pm\infty, t) = 0, \quad \text{and} \quad D'(\pm\infty, t) = 0, \quad \text{for every } t \geq 0. \quad (2.9)$$

Following Barton and Shpak (2000), we offer the following interpretation of the terms in (2.6c). The first term represents diffusion of linkage disequilibrium; the second term is due to migration, which mixes populations with different allele frequencies, and therefore generates associations between the loci by the Wahlund effect; the third term arises from the indirect effects on  $D$  of direct selection on each locus; the fourth term describes the decay of  $D$  caused by recombination.

**Remark 2.1.** Because we assume absence of epistasis and Hardy-Weinberg proportions, only marginal selection coefficients matter. Therefore, the above model with constant selection on genotypes and dominance is equivalent to a model with linear frequency-dependent selection on genotypes without dominance. For a model with dominance and linear frequency dependence, see Mallet and Barton (1989).

Now we specialize fitness further. For the rest of this paper, we assume a so-called step environment. More precisely, we assume that  $\alpha(x)$  and  $\beta(x)$  are the following step functions on the real line:

$$\alpha(x) = \begin{cases} \alpha_+ & \text{if } x \geq 0, \\ -\alpha_- & \text{if } x < 0, \end{cases} \quad (2.10a)$$

and

$$\beta(x) = \begin{cases} \beta_+ & \text{if } x \geq 0, \\ -\beta_- & \text{if } x < 0, \end{cases} \quad (2.10b)$$

where we posit

$$\alpha_+ > 0, \alpha_- > 0, \beta_+ > 0, \beta_- > 0. \quad (2.10c)$$

Under the assumption that both functions change sign once and at the same spatial location, this is general. For instance, the case  $\beta_+ < 0 < -\beta_-$  is obtained by relabeling the alleles  $B$  and  $b$ . Under the assumptions (2.10), selection favors alleles  $A$  and  $B$  on  $[0, \infty)$ , and  $a$  and  $b$  on  $(-\infty, 0)$ ; cf. Appendix A.1. Without loss of generality, we could set  $\alpha_+ = 1$  by rescaling  $\lambda$ , but we refrain from doing so.

By elliptic regularity (Gilbarg and Trudinger 2001), stationary solutions  $(p_A, p_B, D)$  of (2.6) with (2.10) are continuous and have continuous first derivatives. In addition,  $(p_A'', p_B'', D'')$  is continuous and uniformly bounded on every set  $(-\infty, -\delta) \cup (\delta, \infty)$ , where  $\delta > 0$ , but discontinuous at  $x = 0$ . We refer to a non-constant stationary solution  $(p_A, p_B, D)$  of (2.6) as a two-locus cline, and we call  $p_A$  and  $p_B$  the (marginal) one-locus allele-frequency clines.

Throughout, if we write  $f(\pm\infty) = f^\pm$  then both limits  $f^+ = \lim_{x \rightarrow \infty} f(x)$  and  $f^- = \lim_{x \rightarrow -\infty} f(x)$  exist and are finite. Furthermore, each equation in which plus and minus signs appear superimposed holds if either every upper or every lower sign is chosen.

For stationary solutions of (2.6) with (2.10), we impose the boundary conditions

$$p_A'(\pm\infty) = 0, p_B'(\pm\infty) = 0, D'(\pm\infty) = 0, \quad (2.11)$$

which are a consequence of (2.9). Under the assumption that the limits  $p_A'(\pm\infty)$ ,  $p_A''(\pm\infty)$ ,  $p_B'(\pm\infty)$ ,  $p_B''(\pm\infty)$ ,  $D'(\pm\infty)$ , and  $D''(\pm\infty)$  exist, (2.11) is easy to show, and a two-locus cline satisfies

$$p_A(+\infty) = 1, p_A(-\infty) = 0, p_B(+\infty) = 1, p_B(-\infty) = 0, \text{ and } D(\pm\infty) = 0 \quad (2.12)$$

(provided  $r > 0$ ); see also Remark 3.1. We conjecture that the stationary solutions of (2.6) with (2.10) obeying the constraints (2.8) automatically satisfy (2.11) and (2.12).

**Remark 2.2.** We observe from (2.6a) or (2.6b) that if one locus is fixed, the cline at the other locus is independent of which allele is fixed. Formally, the reason is that if one locus is fixed, then  $D = 0$ . This independence of the other locus is a specific property of continuous-time models. In discrete-time models, the allele frequency

at migration-selection equilibrium depends on the fitness of fixed alleles at other loci (see e.g., Aeschbacher and Bürger 2014, where both discrete-time and continuous-time continent-island models are treated). This difference results from the fast decay of higher-order terms when approximating a discrete-time model by a continuous-time model.

The following simple observation of an invariance property of (time-dependent) solutions of (2.6) with (2.10) will be useful later and provide an explanation for the qualitatively different conditions of existence and stability of clines in unbounded and bounded domains with a step environment. For arbitrary, fixed  $\alpha_{\pm}$  and  $\beta_{\pm}$ , let

$$\pi(x, t; \lambda, \rho) = (p_A(x, t; \lambda, \rho), p_B(x, t; \lambda, \rho), D(x, t; \lambda, \rho)) \quad (2.13)$$

denote the time-dependent solution of (2.6) with (2.10), where the dependence on  $\lambda$  and  $\rho$  is shown explicitly.

**Lemma 2.3.** *The following statements hold for arbitrary positive constants  $c$ ,  $c_1$ ,  $c_2$ , for arbitrary (positive) parameters  $\lambda$  and  $\rho$ , and for every  $(x, t) \in \mathbb{R} \times [0, \infty)$ .*

(i)

$$\pi(x, t; c\lambda, c\rho) = \pi(\sqrt{c}x, ct; \lambda, \rho) \quad (2.14)$$

and

$$\pi(x, t; c_1\lambda, c_2\rho) = \pi(\sqrt{c_1}x, c_1t; \lambda, \frac{c_2}{c_1}\rho) = \pi(\sqrt{c_2}x, \frac{c_2}{c_1}t; \frac{c_1}{c_2}\lambda, \rho). \quad (2.15)$$

(ii) *If  $\pi(x; \lambda, \rho)$  is a two-locus cline for  $\lambda$  and  $\rho$ , then a two-locus cline exists for  $c\lambda$  and  $c\rho$ . It is given by  $\pi(x; c\lambda, c\rho) = \pi(\sqrt{c}x; \lambda, \rho)$ .*

(iii) *If  $\pi(x; \lambda, \rho)$  is a globally asymptotically stable two-locus cline for  $\lambda$  and  $\rho$ , then  $\pi(x; c\lambda, c\rho) = \pi(\sqrt{c}x; \lambda, \rho)$  is a globally asymptotically stable two-locus cline for  $c\lambda$  and  $c\rho$ .*

*Proof.* (i) Let  $\tilde{\pi}(x, t) = \pi(\sqrt{c}x, ct; \lambda, \rho)$  and use the abbreviation  $z = (\sqrt{c}x, ct; \lambda, \rho)$ . Then, by the chain rule,  $\tilde{\pi}'(x, t) = \sqrt{c}\pi'(z)$ ,  $\tilde{\pi}''(x, t) = c\pi''(z)$ , and  $\dot{\tilde{\pi}}(x, t) = c\dot{\pi}(z)$ . From these identities, the fact that  $\alpha(x)$  and  $\beta(x)$  are step functions with step at 0,

and by applying (2.6), we obtain

$$\begin{aligned}
\dot{\tilde{D}}(x, t) &= c\dot{D}(z) \\
&= c\left\{D''(z) + 2p'_A(z)p'_B(z) \right. \\
&\quad \left. + \lambda[\alpha(\sqrt{cx})(1 - 2p_A(z))\vartheta_A(z) + \beta(\sqrt{cx})(1 - 2p_B(z))\vartheta_B(z)]D(z) - \rho D(z)\right\} \\
&= \tilde{D}''(x, t) + 2\tilde{p}'_A(x, t)\tilde{p}'_B(x, t) \\
&\quad + c\lambda[\alpha(x)(1 - 2\tilde{p}_A(x, t))\tilde{\vartheta}_A(x, t) + \beta(x)(1 - 2\tilde{p}_B(x, t))\tilde{\vartheta}_B(x, t)]\tilde{D}(x, t) \\
&\quad - c\rho\tilde{D}(x, t). \tag{2.16}
\end{aligned}$$

Analogous calculations hold for  $\tilde{p}_A$  and  $\tilde{p}_B$ . Because solutions are unique (the proof of uniqueness in Theorem 1 of Kolmogoroff et al. 1937 does not require the Lipschitz condition in  $x$ ), we have  $\tilde{\pi}(x, t) = \pi(x, t; c\lambda, c\rho)$ . This proves (2.14), and (2.15) follows immediately.

(ii) follows from (i) because clines are stationary solutions.

(iii) follows because (i) establishes a one-to-one relation between the domains of attraction of the two stationary solutions  $\pi(x; \lambda, \rho)$  and  $\pi(x; c\lambda, c\rho)$ .  $\square$

**Remark 2.4.** Assume that  $(p_A, p_B, D)$  is a two-locus cline and  $D(x) \geq 0$  holds everywhere. Then (2.6a) implies that  $p'_A$  is strictly monotone increasing on  $(-\infty, 0)$  and strictly monotone decreasing on  $[0, \infty)$ . Because the cline satisfies  $p'_A(\pm\infty) = 0$ , it follows that  $p'_A(x)$  is maximized at  $x = 0$ , and minimized at  $x = \pm\infty$ . An analogous statement holds for  $p_B$ . We conjecture that  $D(x) > 0$  on  $\mathbb{R}$  holds for every possible parameter combination satisfying our assumptions.

### 3 Single-locus clines with dominance

For a step environment on the real line, Haldane (1948) derived explicit solutions for a single-locus cline if dominance is either absent ( $h = 0$ ) or complete ( $h = -1$ ). We admit spatially constant dominance coefficients satisfying

$$-1 \leq h \leq 1 \tag{3.1}$$

and assume that  $\alpha(x)$  is given by (2.10a). Throughout this section, we write  $h$  instead of  $h_A$ . Then the single-locus cline is the solution of

$$P'' + \lambda\alpha(x)P(1 - P)(1 + h - 2hP) = 0 \tag{3.2a}$$

satisfying

$$0 < P(x) < 1, \quad (3.2b)$$

and obeying the boundary conditions

$$P(-\infty) = 0, \quad P(+\infty) = 1, \quad (3.2c)$$

and

$$P'(-\infty) = P'(+\infty) = 0. \quad (3.2d)$$

**Remark 3.1.** Solutions of (3.2a) and (3.2b) satisfy the boundary conditions (3.2c) and (3.2d) automatically: Let, e.g.,  $x > 0$ . Then (3.2a) implies  $P''(x) < 0$  for every  $x > 0$ . Because  $P(x)$  is bounded by (3.2b), we infer that  $P''(+\infty) = 0$  and that  $P'(x)$  is monotone decreasing in  $x$ . Since  $P'(x)$  must be bounded from below, the limit  $P'(+\infty)$  exists and equals 0. Since  $P''(+\infty) = 0$ , (3.2a) yields  $P(+\infty) = 0$  or  $P(+\infty) = 1$ . The properties of  $P'$ , or  $P''$ , imply the latter.

Our goal is to derive an explicit solution for this cline. We start by collecting some simple facts.

From (3.2a), (3.2b), and (3.1), we obtain

$$P''(x) \begin{cases} < 0 & \text{if } x > 0, \\ > 0 & \text{if } x < 0. \end{cases} \quad (3.3)$$

From (3.2c), (3.2b), and (3.3), we infer that  $P'(x)$  is maximized at  $x = 0$  and is strictly monotone declining to 0 as  $x \rightarrow \pm\infty$ , i.e.,

$$P'(\pm\infty) = 0. \quad (3.4)$$

To state the main result of this section in compact form, we define

$$a_+ = \sqrt{\lambda\alpha_+}, \quad a_- = \sqrt{\lambda\alpha_-}, \quad (3.5a)$$

$$X_+ = xa_+\sqrt{1-h} \quad \text{if } h < 1, \quad (3.5b)$$

$$X_- = xa_-\sqrt{1+h} \quad \text{if } h > -1, \quad (3.5c)$$

$$A_+ = F_+(a_0, h), \quad A_- = F_-(a_0, h), \quad (3.5d)$$

where

$$F_+(y, h) = \begin{cases} \frac{2 + (1 - 3h)y + \sqrt{3}\sqrt{1-h}\sqrt{\phi(1-y, -h)}}{1-y} & \text{if } h < 1, \\ \frac{\sqrt{3}\sqrt{1+3y}}{2\sqrt{1-y}} & \text{if } h = 1, \end{cases} \quad (3.6a)$$

$$F_-(y, h) = \begin{cases} 1/F_+(1-y, -h) & \text{if } h > -1, \\ F_+(1-y, 1) & \text{if } h = -1, \end{cases} \quad (3.6b)$$

$$\phi(y, h) = 3 - 2y + 3h(1-y)^2, \quad (3.7)$$

and  $a_0$  is the unique solution in  $(0, 1)$  of the quartic equation

$$a_0^2 \phi(a_0, h) = \frac{\alpha_+}{\alpha_+ + \alpha_-}. \quad (3.8)$$

The functions  $F_{\pm}$  are well defined and positive if  $0 < y < 1$ . They are strictly monotone increasing in  $y$ .

In addition, it will be convenient to write

$$Z_{\pm}(x) = A_{\pm} e^{X_{\pm}}. \quad (3.9)$$

**Theorem 3.2.** *Let  $\alpha(x)$  be given by (2.10a) and assume (3.1). Then there exists a unique continuously differentiable solution  $P(x)$  of (3.2). If  $x \geq 0$ , it is given by*

$$P(x) = \begin{cases} 1 - \frac{6(1-h)}{Z_+ + 2(1-3h) + (1+3h)Z_+^{-1}} & \text{if } h < 1, \\ 1 - \frac{12}{9 + 4(xa_+ + A_+)^2} & \text{if } h = 1. \end{cases} \quad (3.10a)$$

If  $x < 0$ , it is given by

$$P(x) = \begin{cases} \frac{6(1+h)}{(1-3h)Z_- + 2(1+3h) + Z_-^{-1}} & \text{if } h > -1, \\ \frac{12}{9 + 4(xa_- - A_-)^2} & \text{if } h = -1. \end{cases} \quad (3.10b)$$

*Proof.* Following Haldane's method, we multiply (3.2) by  $2P'$  to obtain

$$\frac{d}{dx}(P')^2 = -2\lambda\alpha(x)P(1-P)(1+h-2hP)P'. \quad (3.11)$$

From (3.7), we obtain

$$\frac{\partial}{\partial y}(y^2\phi(y, h)) = 6y(1-y)(1+h-2hy). \quad (3.12)$$

Let  $x \geq 0$ , so that  $\alpha(x) = \alpha_+$ . Separating variables in (3.11) and integrating, we find

$$[P'(x)]^2 = C_1 - \frac{\lambda\alpha_+}{3}P(x)^2\phi(P(x), h). \quad (3.13a)$$

If  $x < 0$ , then  $\alpha(x) = -\alpha_-$  and we obtain in an analogous manner

$$[P'(x)]^2 = C_2 + \frac{\lambda\alpha_-}{3}P(x)^2\phi(P(x), h). \quad (3.13b)$$

Because the boundary conditions are  $P'(x) \rightarrow 0$  and  $P(x) \rightarrow 1$  as  $x \rightarrow \infty$ , and  $P'(x) \rightarrow 0$  and  $P(x) \rightarrow 0$  as  $x \rightarrow -\infty$ , we find  $C_1 = \lambda\alpha_+/3$  and  $C_2 = 0$ .

Since we require continuity of  $P'$  at  $x = 0$ , we equate the right-hand sides of (3.13a) and (3.13b) to obtain

$$\frac{\lambda\alpha_+}{3} - \frac{\lambda\alpha_+}{3}P(0)^2\phi(P(0), h) = \frac{\lambda\alpha_-}{3}P(0-)^2\phi(P(0-), h). \quad (3.14)$$

Setting

$$a_0 = P(0) = P(0-) \quad (3.15)$$

and simplifying, we infer that  $a_0$  is the unique solution in  $(0, 1)$  of (3.8).

Substituting (3.8) into (3.13b) with  $C_2 = 0$  and  $x = 0$ , we find

$$P'(0) = \frac{\sqrt{\lambda}}{\sqrt{3}} \sqrt{\frac{\alpha_+\alpha_-}{\alpha_+ + \alpha_-}} = \frac{\sqrt{\lambda}}{\sqrt{6}} \sqrt{H(\alpha_+, \alpha_-)} = \frac{1}{\sqrt{6}} \sqrt{H(a_+^2, a_-^2)}, \quad (3.16)$$

where  $H(\alpha_+, \alpha_-)$  denotes the harmonic mean of  $\alpha_+$  and  $\alpha_-$ .

Because the constants in (3.13) are  $C_1 = \lambda\alpha_+/3 = a_+^2/3$  and  $C_2 = 0$ , we have to solve the set of differential equations

$$P' = \frac{a_+}{\sqrt{3}}(1 - P)\sqrt{\phi(1 - P, -h)} \quad \text{if } x \geq 0, \quad (3.17a)$$

$$P' = \frac{a_-}{\sqrt{3}}P\sqrt{\phi(P, h)} \quad \text{if } x < 0. \quad (3.17b)$$

By separating variables in (3.17a), multiplying by  $\sqrt{3}\sqrt{1-h}$  if  $h < 1$  and by  $\sqrt{3}$  if  $h = 1$ , integrating and observing

$$\frac{\partial}{\partial y} \ln F_+(y, h) = \frac{\sqrt{3}\sqrt{1-h}}{(1-y)\sqrt{\phi(1-y, -h)}} \quad \text{if } h < 1, \quad (3.18a)$$

$$\frac{\partial}{\partial y} F_+(y, 1) = \frac{\sqrt{3}}{(1-y)\sqrt{\phi(1-y, -1)}} \quad \text{if } h = 1, \quad (3.18b)$$

we arrive at

$$\ln F_+(P(x)) = c + xa_+\sqrt{1-h} \quad \text{if } h < 1, \quad (3.19a)$$

$$F_+(P(x)) = c + xa_+ \quad \text{if } h = 1. \quad (3.19b)$$

The constant  $c$  has to satisfy  $c = \ln F_+(a_0)$  if  $h < 1$ , and  $c = F_+(a_0)$  if  $h = 1$ . Using the abbreviations (3.5), we can rewrite (3.19) as

$$F_+(P(x)) = A_+e^{X_+} \quad \text{if } h < 1, \quad (3.20a)$$

$$F_+(P(x)) = X_+ + A_+ \quad \text{if } h = 1. \quad (3.20b)$$

Now, tedious but straightforward inversion of  $F_+$  yields (3.10a).

To derive (3.10b), we observe that if  $P(x, h, a_+)$  solves (3.17a), then  $\tilde{P}(x) = 1 - P(-x, -h, a_-)$  solves (3.17b). This follows from the fact that the right-hand side of (3.17b) is obtained from that of (3.17a) by the simultaneous substitutions  $P \rightarrow 1 - P$ ,  $h \rightarrow -h$ , and  $a_+ \rightarrow a_-$ . Therefore,

$$\frac{\partial}{\partial y} \ln F_-(y, h) = \frac{\partial}{\partial y} (-\ln F_+(1-y, -h)) = \frac{\sqrt{3}\sqrt{1+h}}{y\sqrt{\phi(y, h)}}, \quad (3.21)$$

which implies (3.10b) if  $h > -1$  because it is obtained from (3.10a) by the above substitutions. A similar argument yields (3.10b) if  $h = -1$ . The adjustment of integration constants must be such that  $P(0) = P(0-) = a_0$ ; see (3.15). This is achieved by choosing  $A_+$  and  $A_-$  according to (3.5d).

Uniqueness follows from the above derivation.  $\square$

**Corollary 3.3.** (i) *The slope of the cline in its center,  $P'(0)$ , given by (3.16), is independent of the degree of dominance.*

(ii) *For fixed step size  $\alpha_+ + \alpha_-$  and  $\lambda$ , or fixed  $a_+^2 + a_-^2$ ,  $P'(0)$  is maximized if  $\alpha_+ = \alpha_-$ , and  $P'(0)$  decays to zero as  $\alpha_+ \rightarrow 0$  (or as  $\alpha_- \rightarrow 0$ ). This is in accordance with intuition because in these limits allele  $A$  (or  $a$ ) is nowhere advantageous and the cline vanishes.*

(iii) *For fixed  $a_+$  and  $a_-$ ,  $P(0) = a_0$ , given by the solution of (3.8) in  $(0, 1)$ , is strictly monotone decreasing in  $h$ .*

(iv) *For fixed  $h$ ,  $P(0) = a_0$  is strictly monotone decreasing in  $\tilde{\alpha} = \alpha_+/(\alpha_+ + \alpha_-) = a_+^2/(a_+^2 + a_-^2)$ .*

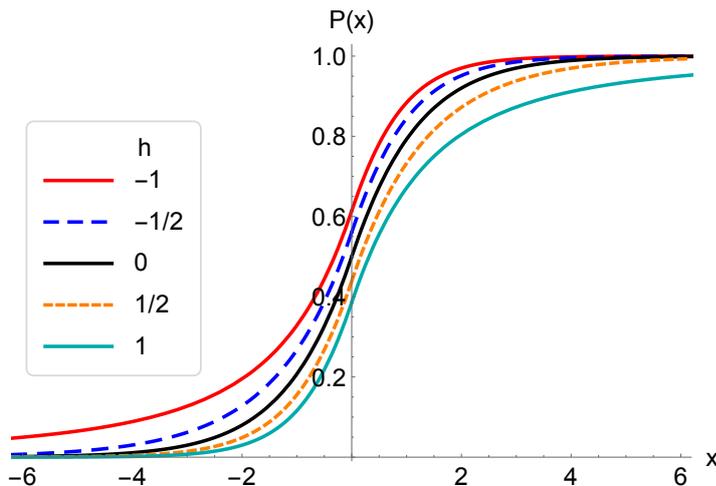


Figure 1: The single-locus cline  $P(x)$  in (3.10) as a function of  $x$  for different values of the dominance parameter  $h$ . The other parameters are  $a_+ = a_- = 1$ .

*Proof.* (i) is evident from (3.16). (ii) is also immediate. (iii) follows from implicit differentiation of (3.8): Define  $f(h, a_0) = a_0^2 \phi(a_0, h) - \frac{\alpha_+}{\alpha_+ + \alpha_-}$ . Then

$$\frac{da_0}{dh} = \frac{-\partial f / \partial h}{\partial f / \partial a_0} = \frac{-a_0(1 - a_0)}{2[1 + (1 - 2a_0)h]} < 0 \quad (3.22)$$

because  $0 < a_0 < 1$  and  $-1 \leq h \leq 1$ . (iv) is shown analogously.  $\square$

Figure 1 illustrates the dependence of the cline on the sign and degree of dominance. It shows that the slope at  $x = 0$  is independent of dominance, but it also demonstrates that dominance has a strong influence on the shape, the value  $P(0)$  in the center, and the asymptotic properties of the cline (see also Section 4.3). In particular, (environment-independent) dominance leads to remarkable asymmetries of the clines.

**Remark 3.4.** The inverse of the maximum slope, in the present model  $P'(0)$ , is often called the width of the cline. If  $\alpha_+ = \alpha_-$ , the width is proportional to Slatkin's (1973) characteristic length for allele frequency variation, which, in our notation and scaling, becomes  $\sigma / \sqrt{\alpha_+} = \sqrt{2} / \sqrt{\lambda \alpha_+}$ . To extend this concept to asymmetric selection, i.e., to  $\alpha_+ \neq \alpha_-$ , we define the characteristic length for locus  $A$  by

$$\ell_A = \frac{\sigma}{\sqrt{H(\alpha_+, \alpha_-)}} = \frac{\sqrt{2}}{\sqrt{\lambda H(\alpha_+, \alpha_-)}}. \quad (3.23)$$

By (3.16) the width of the cline is

$$\omega_A = 1/P'(0) = \sqrt{3} \ell_A. \quad (3.24)$$

Since  $a_0$  does not depend on  $\lambda$ , (3.10) informs us immediately that  $P$  satisfies the invariance property

$$P(x, c\lambda) = P(\sqrt{c}x, \lambda); \quad (3.25)$$

cf. Lemma 2.3. This invariance property lies at the heart of the definition of a characteristic length.

Although  $P$  and  $P'$  are continuous everywhere, in particular at  $x = 0$ , the second derivative of  $P$  at  $x = 0$  is not continuous because

$$P''(0+) = -\lambda\alpha_+a_0(1-a_0)(1+h-2ha_0) < 0, \quad (3.26a)$$

$$P''(0-) = \lambda\alpha_-a_0(1-a_0)(1+h-2ha_0) > 0. \quad (3.26b)$$

Finally, we note that

$$\frac{d^n P}{dx^n}(\pm\infty) = 0 \quad (3.27)$$

for derivatives of every order  $n \geq 1$ .

For later use, we note that if  $-1 < h < 1$ ,  $P'$  can be written as

$$P'(x) = \begin{cases} 6a_+(1-h)^{3/2} \frac{Z_+ - (1+3h)Z_+^{-1}}{[Z_+ + 2(1-3h) + (1+3h)Z_+^{-1}]^2} & \text{if } x \geq 0, \\ 6a_-(1+h)^{3/2} \frac{Z_-^{-1} - (1-3h)Z_-}{[(1-3h)Z_- + 2(1+3h) + Z_-^{-1}]^2} & \text{if } x < 0. \end{cases} \quad (3.28)$$

In the absence of dominance ( $h = 0$ ), we obtain the well known clinal solution (Haldane 1948, Nagylaki 1976)

$$P(x) = \begin{cases} -\frac{1}{2} + \frac{3}{2} \tanh^2\left[\frac{1}{2}(X_+ + \ln A_+)\right] & \text{if } x \geq 0, \\ \frac{3}{2} - \frac{3}{2} \tanh^2\left[\frac{1}{2}(X_- - \ln A_-)\right] & \text{if } x < 0, \end{cases} \quad (3.29)$$

where

$$\ln A_+ = 2 \tanh^{-1} \sqrt{\frac{1+2a_0}{3}}, \quad \ln A_- = 2 \tanh^{-1} \sqrt{\frac{3-2a_0}{3}}, \quad (3.30a)$$

and  $a_0$  is the unique solution in  $(0, 1)$  of the cubic equation

$$3a_0^2 - 2a_0^3 = \frac{\alpha_+}{\alpha_+ + \alpha_-} = \frac{a_+^2}{a_+^2 + a_-^2}. \quad (3.30b)$$

We observe from (3.30b) that  $a_0 = \frac{1}{2}$  if and only if  $a_+ = a_-$ , and  $a_0 > \frac{1}{2}$  if and only if  $a_+ > a_-$ .

We refer to Slatkin (1973) and Nagylaki (1975, 1976) for related results on other specific environmental functions  $\alpha(x)$ , and also on the treatment of barriers to gene flow. Conley (1975) proved existence of clines on  $\mathbb{R}$  for a large class of fitness functions, including the step function treated above. As noted by Nagylaki (1975), his proof applies to arbitrary dominance. Fife and Peletier (1977) proved existence, uniqueness, and global asymptotic stability of clines under rather general assumptions about selection. Although they assume that  $\alpha'(x)$  is uniformly bounded, the proof of their local stability result (their Theorem 3) applies without this assumption if the cline exists and is unique. This is the case in the present model, whence a form of local asymptotic stability of the single-locus cline  $P(x)$  is established if  $-1 \leq h \leq 1$ . I am grateful to Dr. Linlin Su for pointing this out. Whether their global stability result, Theorem 4, applies seems to be open.

## 4 Strong recombination

Now we return to the two-locus problem (2.6) with (2.10) and assume that

$$\epsilon = 1/\rho \tag{4.1}$$

is sufficiently small. Our aim is to derive a first-order approximation in  $\epsilon$  for the two-locus cline, i.e., for the stationary solution that satisfies (2.8) and (2.11). Under our assumptions (2.10) about the step environment, such an equilibrium solution is expected to exist whenever both single-locus problems have a cline, because linkage will lead to additional, indirect, selection on each locus, so that the loci will reinforce each other and the cline will persist. Because we assume the unbounded domain  $\mathbb{R}$ , the single-locus clines,  $P$  and  $Q$ , exist always and are uniquely determined. The main result is Theorem 4.1. In population-genetic terms, this theorem provides the quasi-linkage-equilibrium approximation to the two-locus cline.

### 4.1 Approximation of the two-locus cline

We write  $P$  and  $Q$  for the single-locus clines in the allele-frequencies of  $A$  and  $B$ , respectively. They are obtained from (3.10). For given  $\epsilon$ , we write  $(p_A, p_B, D)$  for the

cline solution of (2.6). We want to determine solutions of the form

$$p_A = p_A^{(\epsilon)} + O(\epsilon^2), \quad p_B = p_B^{(\epsilon)} + O(\epsilon^2), \quad D = D^{(\epsilon)} + O(\epsilon^2), \quad (4.2a)$$

where

$$p_A^{(\epsilon)} = P + \epsilon p, \quad p_B^{(\epsilon)} = Q + \epsilon q, \quad D^{(\epsilon)} = \epsilon d. \quad (4.2b)$$

Here, the limit  $\epsilon \rightarrow 0$  is assumed to hold uniformly in  $x \in \mathbb{R}$  for  $(p_A, p_B, D)$  and for its first and second derivatives, with the restriction that at  $x = 0$  only left and right second derivatives exist; cf. (3.26). From the boundary conditions (2.11) and the fact that  $P'(\pm\infty) = Q'(\pm\infty) = 0$ , we obtain the boundary conditions

$$p'(\pm\infty) = q'(\pm\infty) = d'(\pm\infty) = 0. \quad (4.3)$$

Before formulating our main result, we need additional notation and definitions. Let

$$I_+(x) = \int_x^\infty [P'(y)]^2 Q'(y) [1 + h_B - 2h_B Q(y)] dy, \quad (4.4a)$$

$$I_-(x) = \int_{-\infty}^x [P'(y)]^2 Q'(y) [1 + h_B - 2h_B Q(y)] dy. \quad (4.4b)$$

We recall from Section 3 that  $P'(x) > 0$  and  $Q'(x) > 0$  on  $\mathbb{R}$ . Therefore,  $I_+(x) > 0$  and  $I_-(x) > 0$  on  $\mathbb{R}$ . These integrals exist because  $P'(y)$  and  $Q'(y)$  decay rapidly as  $y \rightarrow \pm\infty$  (see Theorem 3.2, and Section 4.3 for asymptotic results). However, they can be computed explicitly only in special cases (Section 4.4).

In addition to the abbreviations (3.5), we introduce

$$b_+ = \sqrt{\lambda\beta_+}, \quad b_- = \sqrt{\lambda\beta_-}. \quad (4.5)$$

**Theorem 4.1.** *Assume that for sufficiently small  $\epsilon > 0$ , (2.6) admits a two-locus cline of the form (4.2). Then*

(i)  $d(x)$  is given by

$$d(x) = 2P'(x)Q'(x). \quad (4.6)$$

(ii)  $p(x)$  is the unique continuously differentiable solution on  $\mathbb{R}$  of the linear problem

$$\begin{aligned} p''(x) + \lambda\alpha(x)[1 - 2P(x) + h_A(1 - 6P(x) + 6P(x)^2)]p(x) \\ + 2\lambda\beta(x)P'(x)Q'(x)[1 + h_B - 2h_B Q(x)] = 0. \end{aligned} \quad (4.7)$$

(iii)  $p(x)$  is given by

$$p(x) = \begin{cases} P'(x)k_+(x) & \text{if } x \geq 0, \\ P'(x)k_-(x) & \text{if } x \leq 0, \end{cases} \quad (4.8)$$

where

$$k_+(x) = k_0 + 2b_+^2 \int_0^x \frac{I_+(y)}{[P'(y)]^2} dy \quad \text{if } x \geq 0, \quad (4.9a)$$

$$k_-(x) = k_0 - 2b_-^2 \int_x^0 \frac{I_-(y)}{[P'(y)]^2} dy \quad \text{if } x \leq 0, \quad (4.9b)$$

and

$$k_0 = \frac{2\sqrt{3}[b_+^2 I_+(0) - b_-^2 I_-(0)]}{a_0(1-a_0)(1+h_A-2h_A a_0)a_+ a_- \sqrt{a_+^2 + a_-^2}}. \quad (4.10)$$

(iv) The first-order term  $q(x)$  in the expansion of  $p_B(x)$  is of analogous form.

*Proof.* To deduce the stationary solution, we equate the right-hand sides of (2.6) to zero.

(i) If  $\rho = 1/\epsilon$  and (4.2) holds, then the right-hand side of (2.6c) becomes  $2P'(x)Q'(x) - d(x) + O(\epsilon^2)$ , which yields (4.6).

(ii) We obtain (4.7) by substituting (4.2) and (4.6) into the right-hand side of (2.6a) and observing (3.2). The solution of (4.7) is unique because the homogeneous equation

$$p'' + \lambda\alpha(x)[1 - 2P(x) + h_A(1 - 6P(x) + 6P(x)^2)]p = 0 \quad (4.11)$$

has, up to constant multiples, the unique solution  $P'(x)$ . That the latter is a solution follows immediately by differentiating (3.2) and observing that  $\alpha(x)$  is the step function (2.10a).

(iii) Because  $P'(x)$  is the appropriate solution of the homogeneous problem (4.11), the solution  $p(x)$  of the inhomogeneous problem can be obtained by variation of constants. Therefore, by substituting (4.8) into (4.7), using that  $P'$  solves (4.11), and recalling (4.5), we obtain

$$P'(x)k_{\pm}''(x) + 2P''(x)k'_{\pm}(x) \pm 2b_{\pm}^2 P'(x)Q'(x)[1 + h_B - 2h_B Q(x)] = 0. \quad (4.12)$$

If we view (4.12) as a set of two first-order equations for  $k'_{\pm}$ , one defined for  $x \geq 0$  and the other for  $x < 0$ , we find by straightforward calculations that

$$k'_{\pm}(x) = 2b_{\pm}^2 \frac{I_{\pm}(x)}{[P'(x)]^2} \quad (4.13)$$

is the only special solution that is potentially integrable (the general solution is of the form  $k'_\pm(x) + c_\pm/P'(x)$ ; it diverges if  $c_\pm \neq 0$ ). Therefore, integration yields (4.9), and  $k_+$  and  $k_-$  must have the same constant  $k_0$  because we require that  $p(x)$  be continuous at  $x = 0$  (hence, everywhere).

It remains to determine  $k_0 = k_+(0) = k_-(0)$ . Because  $p(x)$  needs to be continuously differentiable at  $x = 0$ , and because  $P'(x)$  is continuous, we require  $p'(0+) = p'(0-)$ . Differentiation of (4.8) shows that this holds if and only if

$$P''(0+)k_0 + P'(0)k'_+(0+) = P''(0-)k_0 + P'(0)k'_-(0-). \quad (4.14)$$

Using (4.13), we obtain

$$k_0 = \frac{2[b_+^2 I_+(0) - b_-^2 I_-(0)]}{P'(0)[P''(0-) - P''(0+)]} \quad (4.15)$$

and, after applying (3.16) and (3.26), (4.10).

(iv) is obvious. □

The approximation (4.6) for the linkage disequilibrium was already derived by Barton and Shpak (2000) in a multilocus context. Since  $P(x)$  and  $Q(x)$  are strictly monotone increasing, the leading term  $\epsilon d(x) = 2\epsilon P'(x)Q'(x)$  of  $D(x)$  is strictly positive on  $\mathbb{R}$ . In addition, because  $P'(x)$  is positive, maximized at  $x = 0$ , and decays monotonically to zero as  $x \rightarrow \pm\infty$ , the same properties are shared by  $d(x)$ . This does not imply that the true equilibrium solution  $D(x)$  is maximized at  $x = 0$ . If selection is not symmetric, i.e., if  $a_+ \neq a_-$  or  $b_+ \neq b_-$ , then  $D(x)$  will be maximized only in an  $\epsilon$ -neighborhood of 0 (see also Fig. 7). We note that  $D'(0)$  exists and is continuous for every  $\epsilon > 0$ , whereas  $d(x)$  has different one-sided derivatives at  $x = 0$  for every  $\epsilon > 0$ , as has  $D'(x)$ .

The integrals appearing in (4.9) can be computed explicitly only in special cases (Section 4.4). However, using integration by parts they can be reduced to expressions requiring only one-fold integration (Appendix A.2). Therefore, it is straightforward to compute  $p(x)$  numerically. Asymptotic expansions are derived in Section 4.3.

**Remark 4.2.** Below, we show that the integrals  $\int_0^\infty \frac{I_+(y)}{[P'(y)]^2} dy$  and  $\int_{-\infty}^0 \frac{I_-(y)}{[P'(y)]^2} dy$  are finite; see (4.35). Therefore, we observe from (4.9) that  $k_+$  and  $k_-$  are bounded and

strictly monotone increasing on their domain. Hence, the limits

$$\kappa_+ = \lim_{x \rightarrow \infty} k_+(x) = k_0 + 2b_+^2 \int_0^\infty \frac{I_+(y)}{[P'(y)]^2} dy, \quad (4.16a)$$

$$\kappa_- = \lim_{x \rightarrow -\infty} k_-(x) = k_0 - 2b_-^2 \int_{-\infty}^0 \frac{I_-(y)}{[P'(y)]^2} dy \quad (4.16b)$$

exist and satisfy  $\kappa_- < \kappa_+$ . Therefore, we can rewrite (4.9) as

$$k_+(x) = \kappa_+ - 2b_+^2 \int_x^\infty \frac{I_+(y)}{[P'(y)]^2} dy \quad \text{if } x \geq 0, \quad (4.17a)$$

$$k_-(x) = \kappa_- + 2b_-^2 \int_{-\infty}^x \frac{I_-(y)}{[P'(y)]^2} dy \quad \text{if } x \leq 0. \quad (4.17b)$$

Although  $p(x)$  is continuously differentiable at  $x = 0$ ,  $k'_+(0+) \neq k'_-(0-)$ . This follows directly from (4.13), but also from the structure of  $p$  and the fact that  $P'$  is not differentiable at  $x = 0$ .

## 4.2 Properties of the approximate two-locus cline

The following corollary summarizes some simple but important consequences of Theorem 4.1.

**Corollary 4.3.** (i) *We have*

$$p'(0) = \frac{2\sqrt{3}[a_-^2 b_+^2 I_+(0) + a_+^2 b_-^2 I_-(0)]}{a_+ a_- \sqrt{a_+^2 + a_-^2}} > 0. \quad (4.18)$$

*Therefore, the cline  $p_A = P + \epsilon p + O(\epsilon^2)$  gets steeper in its center as  $\epsilon$  increases from 0, i.e., if linkage between the two loci gets tighter.*

(ii) *The sign of*

$$p(0) = \frac{2[b_+^2 I_+(0) - b_-^2 I_-(0)]}{a_0(1 - a_0)(1 + h_A - 2h_A a_0)(a_+^2 + a_-^2)} \quad (4.19)$$

*can be positive, negative, or zero.*

(iii)  *$p(x)$  can be positive on  $(-\infty, \infty)$ , negative, or change sign once.*

(iv)  *$p(x)$  satisfies the invariance property*

$$p(x, c\lambda) = cp(\sqrt{cx}, \lambda). \quad (4.20)$$

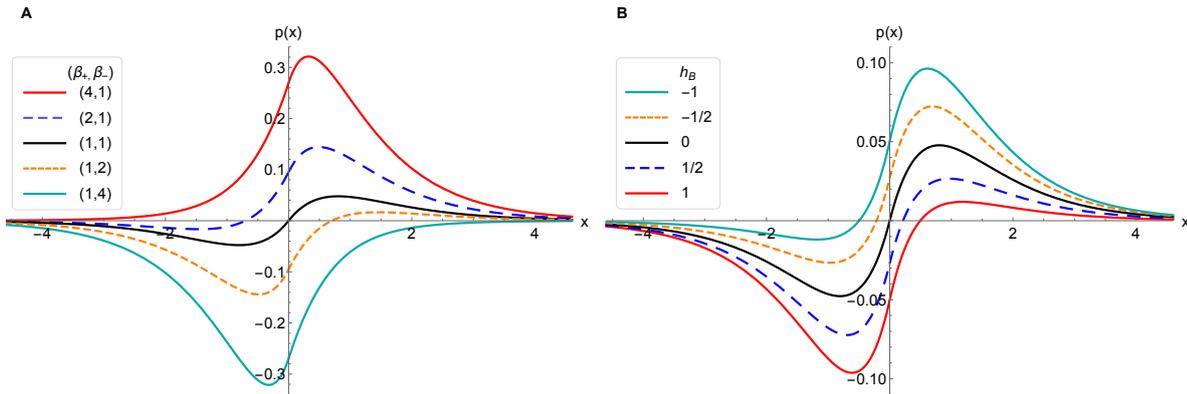


Figure 2: The first-order perturbation  $p(x)$  as a function of  $x$  for different parameter combinations. **A.** The selection intensities at locus  $\mathcal{B}$  in the two environments,  $(\beta_+, \beta_-)$ , are varied for fixed  $\alpha_+ = \alpha_- = 1$ . Dominance is absent ( $h_A = h_B = 0$ ). **B.** The selection intensities at loci  $\mathcal{A}$  and  $\mathcal{B}$  are fixed ( $\alpha_+ = \alpha_- = \beta_+ = \beta_- = 1$ ), and the strength and sign of dominance  $h_B$  at locus  $\mathcal{B}$  are varied. Dominance is absent at locus  $\mathcal{A}$  ( $h_A = 0$ ). In all cases,  $\lambda = 1$ . The perturbation  $p(x)$  is obtained from (4.8), (4.9), and (4.10) by numerical integration using Lemma A.1. We note that in panel A,  $p(x)$  does not change sign if  $(\beta_+, \beta_-) = (4, 1)$  or  $(\beta_+, \beta_-) = (1, 4)$ . In the first case, we have  $\kappa_- \approx 0.09$ ,  $\kappa_+ \approx 1.32$ ; in the second case,  $\kappa_- \approx -1.32$ ,  $\kappa_+ \approx -0.09$ ; cf. Corollary 4.3(iii).

*Proof.* The expressions for  $p(0)$  and  $p'(0)$  follow straightforwardly from Theorem 4.1, (3.16) and (3.26). Because the integrals  $I_{\pm}(0)$  are strictly positive,  $p'(0) > 0$ .

(iii) Remark 4.2 shows that  $k_{\pm}(x)$  is strictly monotone increasing and bounded. Whether  $\kappa_+$  and  $\kappa_-$  have the same or different sign depends on the parameters (see Fig. 2). Because  $P'$  is strictly positive on  $\mathbb{R}$ , statement (iii) now follows from (4.8).

(iv) The invariance property (4.20) follows easily from (4.7) and (3.25) because  $\alpha$  and  $\beta$  are step functions with step at  $x = 0$ .  $\square$

In Section 3, we have shown that the slope of the cline in its center,  $P'(0)$ , is independent of dominance. This is no longer so in the two-locus case, because  $p'(0)$  in (4.18) depends on  $I_+(0)$  and  $I_-(0)$ , and both integrals depend on  $h_A$  and  $h_B$ ; see also (A.24).

Because  $P$  and  $p$  satisfy the invariance properties (3.25) and (4.20), respectively, it follows immediately from (4.2b) that  $p_A^{(\epsilon)} = p_A^{(1/\rho)}$  satisfies the invariance property

$$p_A^{(\frac{1}{c_2\rho})}(x, c_1\lambda) = p_A^{(\frac{c_1}{c_2\rho})}(\sqrt{c_1}x, \lambda) = p_A^{(\frac{1}{\rho})}(\sqrt{c_2}x, \frac{c_1}{c_2}\lambda). \quad (4.21)$$

This parallels (2.15), as is to be expected.

For equivalent loci, an explicit representation of  $p(x)$  will be derived in Section 4.4. Figure 2 displays  $p(x)$  for various parameter choices. All cases clearly show that  $p'(0) > 0$  (4.18). However, also each of the three possibilities in statements (ii) and (iii) of Corollary 4.3 are exemplified. The accuracy of the approximation for the cline  $(p_A, p_B, D)$  derived in Theorem 4.1 will be investigated numerically in Section 6.

### 4.3 Asymptotic properties

Here, we derive asymptotic properties of  $P(x)$ ,  $p(x)$ , and  $p_A(x)$  for  $x \rightarrow \pm\infty$ . We recall that two functions  $f(x)$  and  $g(x)$  are asymptotically equal as  $x \rightarrow \infty$ , i.e.,  $f(x) \sim g(x)$  as  $x \rightarrow \infty$ , if  $\lim_{x \rightarrow \infty} f(x)/g(x) = 1$ , and analogously if  $x \rightarrow -\infty$ .

We start with the single-locus case and assume  $-1 < h_A < 1$  and  $-1 < h_B < 1$ . From (3.10), we obtain immediately the asymptotic equalities

$$1 - P(x) \sim 6(1 - h_A)A_+^{-1}e^{-X_+} \quad \text{as } x \rightarrow +\infty, \quad (4.22a)$$

$$P(x) \sim 6(1 + h_A)A_-e^{X_-} \quad \text{as } x \rightarrow -\infty, \quad (4.22b)$$

and, after differentiation of (3.10),

$$P'(x) \sim 6a_{\pm}(1 \mp h_A)^{3/2}A_{\pm}^{\mp 1}e^{\mp X_{\pm}} \quad \text{as } x \rightarrow \pm\infty. \quad (4.23)$$

As a trivial consequence, we note that

$$1 - P(x) \sim \frac{1}{a_+\sqrt{1 - h_A}}P'(x) \quad \text{as } x \rightarrow +\infty, \quad (4.24a)$$

$$P(x) \sim \frac{1}{a_-\sqrt{1 + h_A}}P'(x) \quad \text{as } x \rightarrow -\infty. \quad (4.24b)$$

In analogy to (3.5), we introduce the abbreviations

$$B_+ = F_+(b_0, h_B), \quad B_- = F_-(b_0, h_B), \quad (4.25a)$$

$$Y_+ = xb_+\sqrt{1 - h_B} \quad \text{if } h_B < 1, \quad (4.25b)$$

$$Y_- = xb_-\sqrt{1 + h_B} \quad \text{if } h_B > -1, \quad (4.25c)$$

where

$$b_0 = Q(0) \quad (4.26)$$

is the unique solution in  $(0, 1)$  of the quartic equation

$$b_0^2\phi(b_0, h_B) = \frac{\beta_+}{\beta_+ - \beta_-}; \quad (4.27)$$

cf. (3.8) and (3.7).

The main result of this section is the following.

**Proposition 4.4.** *Let  $-1 < h_A < 1$  and  $-1 < h_B < 1$ . The asymptotic rates of convergence of the marginal allele-frequency cline  $p_A$  are proportional to those of the single-locus cline  $P$ , i.e.,*

$$\lim_{x \rightarrow \infty} \frac{1 - p_A(x)}{1 - P(x)} = 1 - \epsilon a_+ \kappa_+ \sqrt{1 - h_A}, \quad (4.28a)$$

$$\lim_{x \rightarrow -\infty} \frac{p_A(x)}{P(x)} = 1 + \epsilon a_- \kappa_- \sqrt{1 + h_A}. \quad (4.28b)$$

In fact, the following stronger result holds:

$$\lim_{x \rightarrow \pm\infty} \frac{p(x) - \kappa_{\pm} P'(x)}{P'(x) Q'(x)} = \frac{\mp 2b_{\pm} \sqrt{1 \mp h_B}}{2a_{\pm} \sqrt{1 \mp h_A} + b_{\pm} \sqrt{1 \mp h_B}}. \quad (4.29)$$

Asymptotic equalities in terms of exponential functions can be obtained straightforwardly by applying (4.23).

*Proof.* The proof of (4.28) is very simple. From (4.2) and (4.8), and after invoking (4.16) and (4.24a), we obtain

$$\lim_{x \rightarrow \infty} \frac{1 - p_A(x)}{1 - P(x)} = 1 - \epsilon \lim_{x \rightarrow \infty} \frac{P'(x) k(x)}{1 - P(x)} = 1 - \epsilon \kappa_+ a_+ \sqrt{1 - h_A},$$

and analogously (4.28b).

To prove (4.29), we have to work harder. Mimicking (4.23), we introduce

$$\tilde{P}'(x) = 6a_{\pm}(1 \mp h_A)^{3/2} A_{\pm}^{\mp 1} e^{\mp X_{\pm}} \quad \text{if } x \geq 0, \quad (4.30a)$$

$$\tilde{Q}'(x) = 6b_{\pm}(1 \mp h_B)^{3/2} B_{\pm}^{\mp 1} e^{\mp Y_{\pm}} \quad \text{if } x \geq 0. \quad (4.30b)$$

Furthermore, we define

$$\tilde{I}_+(x) = \int_x^{\infty} \left[ \tilde{P}'_+(y) \right]^2 \tilde{Q}'_+(y) [1 + h_B - 2h_B \tilde{Q}_+(y)] dy, \quad (4.31a)$$

$$\tilde{I}_-(x) = \int_{-\infty}^x \left[ \tilde{P}'_-(y) \right]^2 \tilde{Q}'_-(y) [1 + h_B - 2h_B \tilde{Q}_-(y)] dy. \quad (4.31b)$$

Then straightforward integration yields

$$\tilde{I}_{\pm}(x) = \frac{6^3 a_{\pm}^2 b_{\pm} (1 \mp h_A)^3 (1 \mp h_B)^{5/2}}{2a_{\pm} \sqrt{1 \mp h_A} + b_{\pm} \sqrt{1 \mp h_B}} A_{\pm}^{\mp 2} B_{\pm}^{\mp 1} e^{\mp(2X_{\pm} + Y_{\pm})} \quad \text{as } x \rightarrow \pm\infty. \quad (4.32)$$

Because

$$I_{\pm}(x) \sim \tilde{I}_{\pm}(x) \quad \text{as } x \rightarrow \pm\infty, \quad (4.33)$$

we obtain from (4.32) and (4.30)

$$I_{\pm}(x) \sim \frac{1 \mp h_B}{2a_{\pm}\sqrt{1 \mp h_A} + b_{\pm}\sqrt{1 \mp h_B}} [P'(x)]^2 Q'(x) \quad \text{as } x \rightarrow \pm\infty. \quad (4.34)$$

Therefore,

$$\frac{I_{\pm}(x)}{[P'(x)]^2} \sim \frac{6b_{\pm}(1 \mp h_B)^{5/2}}{2a_{\pm}\sqrt{1 \mp h_A} + b_{\pm}\sqrt{1 \mp h_B}} B_{\pm}^{\mp 1} e^{\mp Y_{\pm}} \quad \text{as } x \rightarrow \pm\infty \quad (4.35)$$

and

$$\int_x^{\infty} \frac{I_+(y)}{[P'(y)]^2} dy \sim \frac{6(1 - h_B)^2}{2a_+\sqrt{1 - h_A} + b_+\sqrt{1 - h_B}} B_+^{-1} e^{-Y_+} \quad \text{as } x \rightarrow +\infty, \quad (4.36a)$$

$$\int_{-\infty}^x \frac{I_-(y)}{[P'(y)]^2} dy \sim \frac{6(1 + h_B)^2}{2a_-\sqrt{1 + h_A} + b_-\sqrt{1 + h_B}} B_- e^{Y_-} \quad \text{as } x \rightarrow -\infty. \quad (4.36b)$$

Substitution into (4.17) yields

$$k_{\pm}(x) - \kappa_{\pm} \sim \frac{\mp 12b_{\pm}^2(1 \mp h_B)^2}{2a_{\pm}\sqrt{1 \mp h_A} + b_{\pm}\sqrt{1 \mp h_B}} B_{\pm}^{\mp 1} e^{\mp Y_{\pm}} \quad \text{as } x \rightarrow \pm\infty, \quad (4.37)$$

which can be written as

$$k_{\pm}(x) - \kappa_{\pm} \sim \frac{\mp 2b_{\pm}\sqrt{1 \mp h_B}}{2a_{\pm}\sqrt{1 \mp h_A} + b_{\pm}\sqrt{1 \mp h_B}} Q'(x) \quad \text{as } x \rightarrow \pm\infty. \quad (4.38)$$

Now (4.8) establishes (4.29).  $\square$

The equations (4.28) may leave the impression that these limits are independent of  $h_B$ . This is not the case, because  $\kappa_{\pm}$  depends both on  $h_A$  and  $h_B$ , and of course on  $\alpha_{\pm}$  and  $\beta_{\pm}$ . If  $\kappa_- < 0 < \kappa_+$ , as is the case for loci of equal effects (see eqs. (4.51) and Appendix A.3.1), then (4.28) shows that the cline  $p_A$  in the two-locus system converges faster to its limits 1 or 0 as  $x \rightarrow \infty$  or  $x \rightarrow -\infty$  than the single-locus cline  $P$ . If  $\kappa_-$  and  $\kappa_+$  have the same sign (for examples, see Fig. 2), then convergence may be faster at one end but slower at the other.

Equations (4.51) and (A.21) provide explicit expressions for  $\kappa_{\pm}$  for equivalent loci without and with dominance, respectively.

In many of the above asymptotic formulas, the compound constants  $A_{\pm}^{\mp 1}$  (and the analogous  $B_{\pm}^{\mp 1}$ ) occur. From the definitions of  $A_{\pm}$  in (3.5d) and the monotonicity

of  $F_{\pm}$  in (3.6), we observe that  $A_+^{-1} = A_+^{-1}(a_0)$  is monotone decreasing in  $a_0$ , and  $A_- = A_-(a_0)$  is monotone increasing. They satisfy  $A_+^{-1}(0) = (2 + \sqrt{3}\sqrt{1-h_A})^{-1}$ ,  $A_+^{-1}(1) = 0$ ,  $A_-(0) = 0$ , and  $A_-(1) = (2 + \sqrt{3}\sqrt{1+h_A})^{-1}$ .

**Remark 4.5.** The cases where one or both of  $h_A$  and  $h_B$  equal  $\pm 1$  can be treated easily. The results, however, differ drastically from intermediate dominance. We present the case  $h_A = h_B = 1$ . For the single-locus clines, we obtain from (3.10a)

$$1 - P(x) \sim \frac{3}{a_+^2} x^{-2} \quad \text{as } x \rightarrow \pm\infty, \quad (4.39)$$

and

$$P'(x) \sim \frac{6}{a_+^2} x^{-3} \quad \text{as } x \rightarrow \pm\infty. \quad (4.40)$$

Simple calculations yield

$$I_+(x) \sim \frac{648}{5a_+^4 b_+^4} x^{-10} \quad \text{as } x \rightarrow \pm\infty, \quad (4.41)$$

and

$$\int_x^\infty \frac{I_+(y)}{[P'(y)]^2} dy \sim \frac{6}{5b_+^2} x^{-3} \quad \text{as } x \rightarrow \pm\infty. \quad (4.42)$$

Therefore, we find

$$k_+(x) - \kappa_+ \sim -\frac{12}{5b_+^2} x^{-3} \quad \text{as } x \rightarrow \pm\infty. \quad (4.43)$$

Thus, (4.28a) and (4.29) are replaced by the qualitatively different

$$\frac{1 - p_A(x)}{1 - P(x)} - 1 \sim -\frac{2\epsilon a_+ \kappa_+}{x} \quad \text{as } x \rightarrow +\infty, \quad (4.44)$$

and

$$p(x) - \kappa_+ P'(x) \sim -\frac{2}{5} P'(x) Q'(x) \quad \text{as } x \rightarrow +\infty, \quad (4.45)$$

respectively.

#### 4.4 Explicit solution for equivalent loci

We assume  $b_{\pm} = a_{\pm}$  and  $-1 < h_A = h_B < 1$ , i.e., the loci have equal effects. Then we can calculate the term  $p(x)$  in  $p_A(x) = P(x) + \epsilon p(x) + O(\epsilon^2)$  explicitly. Since the expressions with dominance are very complicated, we relegate them to Appendix A.3, which also contains the derivations. After presenting the explicit results in the absence of dominance, we describe and illustrate the influence of dominance on  $p'(0)$ .

**Theorem 4.6.** Let  $b_{\pm} = a_{\pm}$  and  $h = h_A = h_B = 0$ . Then  $p(x)$  is given by (4.8), where

$$P'(x) = \begin{cases} \frac{6a_+(Z_+ - Z_+^{-1})}{(Z_+ + 2 + Z_+^{-1})^2} & \text{if } x \geq 0, \\ \frac{6a_-(Z_-^{-1} - Z_-)}{(Z_+ - 2 + Z_-^{-1})^2} & \text{if } x < 0, \end{cases} \quad (4.46)$$

$$Z_{\pm} = A_{\pm}e^{xa_{\pm}}, \quad (4.47)$$

$$A_+ = \frac{2 + a_0 + \sqrt{3}\sqrt{1 + 2a_0}}{1 - a_0}, \quad A_- = \frac{3 - a_0 - \sqrt{3}\sqrt{3 - 2a_0}}{1 - a_0}, \quad (4.48)$$

$$k_+(x) = k_0 + 2a_+ \left[ \frac{Z_+^{-1} - 2}{Z_+ - Z_+^{-1}} - \frac{A_+^{-1} - 2}{A_+ - A_+^{-1}} \right] \quad \text{if } x \geq 0, \quad (4.49a)$$

$$k_-(x) = k_0 - 2a_- \left[ \frac{A_-^{-1} - 2}{A_- - A_-^{-1}} - \frac{Z_-^{-1} - 2}{Z_- - Z_-^{-1}} \right] \quad \text{if } x < 0, \quad (4.49b)$$

and

$$k_0 = \frac{2a_+(1 - a_0)(2a_0 - 1)}{\sqrt{3}\sqrt{1 + 2a_0}(3 - 2a_0)} = \frac{2a_-a_0(2a_0 - 1)}{\sqrt{3}\sqrt{3 - 2a_0}(1 + 2a_0)}. \quad (4.50a)$$

In addition,

$$\kappa_+ = \lim_{x \rightarrow \infty} k_+(x) = a_+ \left( 1 - \frac{2 - a_0}{3 - 2a_0} \sqrt{\frac{1 + 2a_0}{3}} \right) \quad \text{if } x \geq 0, \quad (4.51a)$$

$$\kappa_- = \lim_{x \rightarrow -\infty} k_-(x) = -a_- \left( 1 - \frac{1 + a_0}{1 + 2a_0} \sqrt{\frac{3 - 2a_0}{3}} \right) \quad \text{if } x < 0. \quad (4.51b)$$

**Remark 4.7.** It is easy to show that  $\kappa_+$  is strictly monotone decreasing in  $a_0 \in [0, 1]$  and satisfies  $1.23a_+ \approx 2a_+(1 - \frac{2}{3\sqrt{3}}) \geq \kappa_+ \geq 0$ . Similarly,  $\kappa_-$  is strictly monotone decreasing in  $a_0 \in [0, 1]$  and satisfies  $0 \geq \kappa_- \geq -2a_-(1 - \frac{2}{3\sqrt{3}}) \approx -1.23a_-$ .

The following corollary specifies several important properties of  $p(x)$ .

**Corollary 4.8.** Under the assumptions of Theorem 4.6,  $p(x)$  has the following properties:

(i)

$$p(0) = 8a_+^2 \frac{(1 - a_0)^2(2a_0 - 1)}{3(3 - 2a_0)} = \frac{8}{3}(a_+^2 + a_-^2)a_0^2(1 - a_0^2)(2a_0 - 1). \quad (4.52)$$

Therefore,  $p(0) \geq 0$  if and only if  $a_0 \geq \frac{1}{2}$ , which is the case if and only if  $a_+ \geq a_-$ .

(ii)

$$p'(0) = \frac{8a_+^3(1-a_0)^3\sqrt{3+6a_0}}{3-2a_0} = 8\sqrt{3}a_+a_-\sqrt{a_+^2+a_-^2}a_0^2(1-a_0^2), \quad (4.53)$$

which is always positive, in accordance with the general case.

(iii) The frequency  $p(x)$  changes sign once. If  $p(x_k) = 0$ , then  $p(x) > 0$  if  $x > x_k$  and  $p(x) < 0$  if  $x < x_k$ .

*Proof.* Statements (i) and (ii) follow immediately from the corresponding statements in Corollary 4.3 by applying (A.15).

(iii) From Remark 4.7, we obtain  $\kappa_- < 0 < \kappa_+$ . Because  $k_+$  and  $k_-$  are strictly monotone increasing on their domains (see eq. 4.9),  $k(x) = k_{\pm}(x)$ ,  $x \geq 0$ , changes sign once, say at  $x = x_k$ . Because  $P' > 0$  on  $\mathbb{R}$ ,  $p$  changes sign once, at  $x = x_k$ , and satisfies  $p(x) > 0$  if  $x > x_k$  and  $p(x) < 0$  if  $x < x_k$ .  $\square$

Explicit asymptotic expansions of  $p(x)$  and  $p_A(x)$  as  $x \rightarrow \pm\infty$  follow immediately from Proposition 4.4 by setting  $b_{\pm} = a_{\pm}$ ,  $h = h_A = h_B = 0$ , and using (4.51).

Corollary A.2 generalizes Corollary 4.8 to intermediate dominance. In contrast to single-locus clines, in two-locus systems the slope  $p'_A(0) \approx P'(0) + \rho^{-1}p'(0)$  of the allele-frequency clines depends on  $h$  because  $p'(0) = p'(0, h)$  depends on  $h$ . Figure 3 displays the ratio  $p'(0, h)/p'(0, 0)$  for different values of  $\tilde{\alpha}$ . That this ratio depends only on  $\tilde{\alpha} = \alpha_+/(\alpha_+ + \alpha_-)$ , and not on  $\alpha_+$  and  $\alpha_-$  separately, is a consequence of the scaling property shown in Lemma 2.3 because  $\tilde{\alpha}$  remains unchanged if and only if both  $\alpha_+$  and  $\alpha_-$  are multiplied by the same constant. The estimates (A.26) show that with intermediate dominance, the slope  $p'(0, h)$  is at most 20% larger or smaller than without dominance. Since our analysis is based on the assumption of large  $\rho$ , the influence of dominance on  $p'_A(0)$  will be small, at least for loci of equal or similar effects.

## 4.5 A global measure for the steepness of a cline

The slope of a cline in its center,  $P'(0)$  or  $p'_A(0)$ , provides a local measure for its steepness, which is the inverse of its width. Here, we study the global measure

$$s(\pi) = \int_{-\infty}^{\infty} \pi'(x)^2 dx \quad (4.54)$$

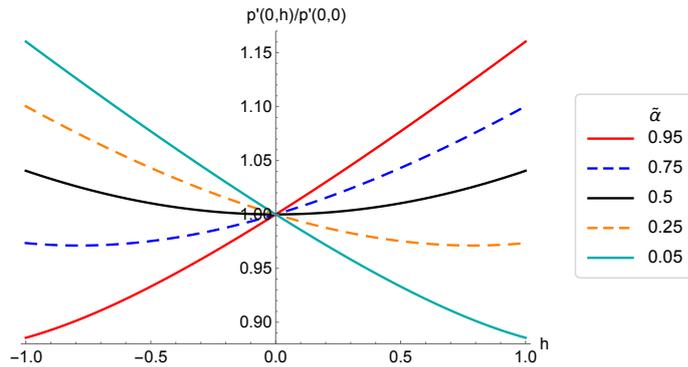


Figure 3: The effects of dominance on the slope of two-locus clines for equivalent loci. Displayed is the ratio  $p'(0, h)/p'(0, 0)$  in (A.25) as a function of  $h$  for different values of  $\tilde{\alpha} = \alpha_+ / (\alpha_+ + \alpha_-)$ . The dependence of  $p'(0) = p'(0, h)$  on  $h$  is indicated explicitly.

for the steepness of a cline  $\pi(x)$  and compare with the slope  $\pi'(0)$ . The function  $s(\pi)$  is the square of the measure introduced by Liang and Lou (2011), who proved for a more general model on a bounded domain that it is a strictly monotone increasing function of  $\lambda$ ; see also Lou et al. (2013). We mention that  $s(\pi)$  is one of the two terms in the Lyapunov function that is minimized by the solution (Fleming 1975). To shape intuition, we note that  $s(\tanh(ax)) = \frac{4a}{3}$ .

#### 4.5.1 A single locus

We start by collecting properties of the steepness  $s(P)$  for the single-locus cline (3.10). We write  $P_\lambda$  to indicate that for given  $(\alpha_+, \alpha_-)$  each  $\lambda$  defines a different function. Then the invariance property (3.25) implies  $P'_{c\lambda}(x) = cP'_\lambda(\sqrt{c}x)$ . A simple integral transformation yields

$$s(P_{c\lambda}) = \sqrt{c} s(P_\lambda). \quad (4.55)$$

Therefore, as expected, the cline becomes steeper with stronger selection, i.e., larger  $\lambda$ .

The steepness of the single-locus cline  $P(x)$  can be computed explicitly. The expression, however, is very complicated and given by equations (A.32) in Appendix A.4. Figure 4 displays  $s(P)$  as a function of  $h$  for various parameter combinations, which are chosen such that all resulting clines have the same slope  $P'(0)$ . Whereas,  $P'(0)$  does not depend on the degree of dominance, the steepness  $s(P)$  does depend on  $h$ , although not very strongly. The reason is that the region near the environmental step contributes most to  $s(P)$  because it is there that  $P'(x)$  is highest and (nearly) independent of  $h$ .

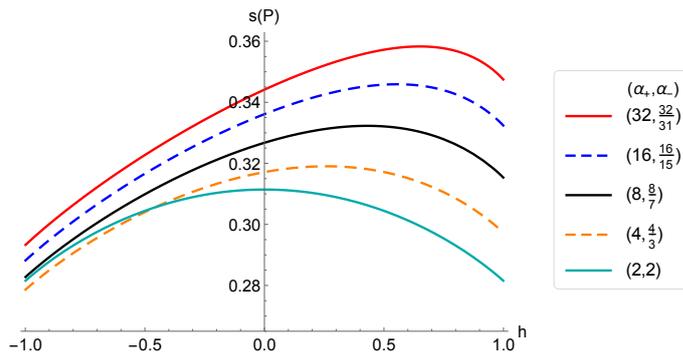


Figure 4: Steepness of single-locus clines,  $s(P)$ , as a function of the degree of dominance,  $h$ . The curves show  $s(P)$  in (A.32) as a function of  $h$  for the five clines emerging from the pairs of values  $(\alpha_+, \alpha_-)$  given in the legend, where  $\lambda = 1$ . These pairs are chosen such that they have the same harmonic mean of 2. By (3.16), the corresponding clines have the same slope in the center, i.e.,  $P'(0) = \sqrt{2/6} \approx 0.577$ .

In the absence of dominance ( $h = 0$ ), the steepness simplifies to

$$s(P) = \frac{3}{5} \left[ (a_+ + a_-) - \frac{\sqrt{a_+^2 + a_-^2}}{\sqrt{3}} \sqrt{(3 - 2a_0)(1 + 2a_0)} \right]; \quad (4.56)$$

see Appendix A.4. Evidently, (4.56) satisfies (4.55). If we fix  $\lambda$  and the step size  $\alpha_+ + \alpha_-$ , or equivalently  $a_+^2 + a_-^2$ , we can consider  $s(P)$  as a function of  $a_+$ , or of  $a_0$ . Then (4.56) shows that  $s(P)$  decreases monotonically to 0 as  $a_+ \rightarrow 0$  (because then  $a_0 \rightarrow 0$ ) or as  $a_+ \rightarrow \sqrt{a_+^2 + a_-^2}$  (because then  $a_0 \rightarrow 1$ ). In addition,  $s(P)$  is symmetric about  $a_+ = a_-$ , or  $a_0 = 1/2$ , and  $s(P)$  is maximized at this point, at which it assumes the value

$$s(P) = \frac{2}{5}(3 - \sqrt{6})a_+. \quad (4.57)$$

In Appendix A.4, we provide the simple and accurate approximation (A.38) to (4.56), and we show that in the absence of dominance the ratio  $P'(0)/s(P)$  varies between quite narrow bounds, i.e.,

$$1.564 \approx \frac{5}{23}(2 + 3\sqrt{3}) \leq \frac{P'(0)}{s(P)} \leq \frac{5}{12}(2 + \sqrt{6}) \approx 1.854. \quad (4.58)$$

#### 4.5.2 Two recombining loci

We write  $p_{A,(\lambda,\rho)}$  to indicate the dependence of the cline on  $\lambda$  and  $\rho$ . For two recombining loci the invariance property (2.15) yields, in a similar way as for a single locus,

$$s[p_{A,(c_1\lambda,c_2\rho)}] = \sqrt{c_1}s[p_{A,(\lambda,\frac{c_2}{c_1}\rho)}] = \sqrt{c_2}s[p_{A,(\frac{c_1}{c_2}\lambda,\rho)}]. \quad (4.59)$$

In particular, this shows that the steepness increases by a factor of  $\sqrt{c}$  if both  $\lambda$  and  $\rho$  are multiplied by the same constant  $c$ .

For two loosely linked loci, we obtain more detailed insight. From (4.54) and (4.2), we infer

$$\begin{aligned} s(p_A) &= \int_{-\infty}^{\infty} P'(x)^2 dx + \frac{2}{\rho} J + O(\rho^{-2}) \\ &= s(P) + \frac{2}{\rho} J + O(\rho^{-2}), \end{aligned} \quad (4.60)$$

where

$$J = \int_{-\infty}^{\infty} P'(x)p'(x)dx. \quad (4.61)$$

Therefore,  $J$  is a measure of the influence of locus  $\mathcal{B}$  on the steepness of the cline at locus  $\mathcal{A}$ . It contributes to  $s(p_A)$  in an analogous way as  $p'(0)$  to  $p'_A(0)$ . Equation (4.60) demonstrates that the allele-frequency cline (at each locus) gets steeper as  $\rho$  decreases if and only if  $J > 0$ .

To study the integral  $J$ , we write  $J(\lambda)$  to make the dependence on  $\lambda$  explicit and obtain

$$\begin{aligned} J(c\lambda) &= \int_{-\infty}^{\infty} P'(x, c\lambda)p'(x, c\lambda)dx \\ &= \int_{-\infty}^{\infty} \sqrt{c}P'(\sqrt{c}x, \lambda)c\sqrt{c}p'(\sqrt{c}x, \lambda)dx \\ &= \frac{c^2}{\sqrt{c}} \int_{-\infty}^{\infty} P'(y, \lambda)p'(y, \lambda)dy \\ &= c^{3/2}J(\lambda). \end{aligned} \quad (4.62)$$

From (4.60), (4.55) and (4.62), and again (4.60), we derive

$$\begin{aligned} s(p_{A,(c\lambda,\rho)}) &= s(P_{c\lambda}) + \frac{2}{\rho}J(c\lambda) + O(\rho^{-2}) \\ &= \sqrt{c}s(P_\lambda) + \frac{2c^{3/2}}{\rho}J(\lambda) + O(\rho^{-2}) \\ &= \sqrt{c}s(p_{A,(\lambda,\rho)}) + \frac{2\sqrt{c}}{\rho}(c-1)J(\lambda) + O(\rho^{-2}). \end{aligned} \quad (4.63)$$

This is not only in accordance with (4.59) but shows, provided  $J(\lambda) > 0$ , that stronger selection ( $c > 1$ ) causes a stronger increase in the steepness of the cline for smaller values of  $\rho$  than for larger values.

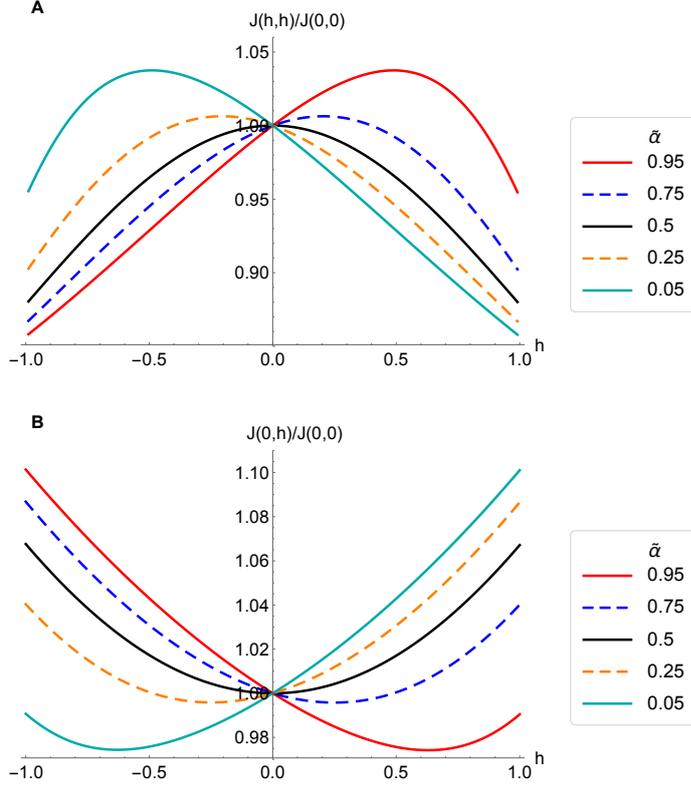


Figure 5: The effects of dominance on the steepness of two-locus clines. **A** The ratio  $J(h, h)/J(0, 0)$  as a function of  $h$  for different values of  $\tilde{\alpha} = \alpha_+ / (\alpha_+ + \alpha_-)$ . **B** The ratio  $J(0, h)/J(0, 0)$  as a function of  $h$  for different values of  $\alpha$ . The scaling property (4.62) shows that the ratios displayed in the figure do not depend on  $\alpha_+$  and  $\alpha_-$  separately, but only on  $\alpha$ . In both panels, both loci have the step sizes  $\beta_{\pm} = \alpha_{\pm}$ , so that selection on homozygotes is equally strong at both loci.

We conjecture that  $J$  is always positive. This conjecture is based on a proof for loci of equal effects without dominance (Appendix A.5.1) and comprehensive numerical evaluation of  $J$  (e.g., Figs. A.2 and 5).

Figure 5 illustrates the effects of dominance on  $J = J(h_A, h_B)$  for loci of equal effects (panel A) and for the case where only locus  $\mathcal{B}$  exhibits dominance (panel B). Panel A is analogous to Fig. 3 and shows that  $J$  depends on  $h$  in a qualitatively different way than  $p'(0)$ . If dominance is absent at locus  $\mathcal{A}$ , dominance at locus  $\mathcal{B}$  affects the steepness  $s(p_A)$  in (4.60) as depicted by Fig. 5B. For loci of equal effects and in the absence of dominance, Fig. A.1 displays the ratio  $p'_A(0)/s(p_A)$  as a function of  $\alpha_+$  for given step size  $\alpha_+ + \alpha_-$  and three (large) recombination rates.

We conclude that the global measure  $s(p_A)$  for the steepness has properties very

similar to the slope  $p'_A(0)$  if there is no dominance. In the presence of dominance, the steepness  $s(p_A)$  depends on the degree of dominance and therefore reflects its effects much better than the slope. This is most conspicuous if the ‘primary’ locus (here,  $\mathcal{A}$ ) exhibits dominance because in this case the slope in the center is independent of  $h_A$ . The indirect effects of dominance at a loosely linked locus are weak; this is true for both measures of steepness if both loci are under similarly strong selection. Of course, if the second locus is under much stronger selection than the first, then dominance at this locus may have a bigger effect on the first locus.

## 5 No recombination

If recombination is absent ( $\rho = 0$ ), the system of partial differential equations (2.1) for the gamete frequencies becomes formally equivalent to a one-locus four-allele model in which the gametes play the role of the alleles. With the fitness functions (2.10) and the assumption (2.10c), one-locus theory (Section 3) implies that on each of the edges ( $p_i + p_j = 1$ ,  $i \neq j$ ) of the state space there exists a unique cline.

By our assumptions,  $AB$  is the gamete with highest fitness ( $\alpha_+ + \beta_+$ ) if  $x \geq 0$  and that with lowest fitness ( $-\alpha_- - \beta_-$ ) if  $x < 0$ , whereas  $ab$  is the gamete with lowest fitness ( $-\alpha_+ - \beta_+$ ) if  $x \geq 0$  and that with highest fitness ( $\alpha_- + \beta_-$ ) if  $x < 0$ . The gametes  $Ab$  and  $aB$  have intermediate fitness everywhere. We conjecture that the cline formed by the gametes  $AB$  and  $ab$  (satisfying  $p_2 = p_3 = 0$ ,  $p_1 > 0$ ,  $p_4 > 0$ ) is globally asymptotically stable. The frequency  $p_1$  of gamete  $AB$  of this cline is obtained from the one-locus formula (3.29) by substituting in (3.30)  $\alpha_+ + \beta_+$  for  $\alpha_+$ , and  $\alpha_- + \beta_-$  for  $\alpha_-$ . Then the two-locus cline, denoted by  $p_{AB} = (p_A, p_B, D)$ , satisfies  $p_A = p_B = p_1$  and  $D = p_1(1 - p_1)$ .

This conjecture is supported by Corollary 4.7 of Lou and Nagylaki (2004) which implies global stability for an analogous model defined on a *bounded* domain if  $\lambda$  is sufficiently large. Because  $\rho = 0$ , Lemma 2.3(iii) shows that if our cline  $p_{AB}$  is globally asymptotically stable for one value of  $\lambda$ , then it is globally asymptotically stable for every  $\lambda > 0$ . Numerical integration of the time-dependent equation (2.6) also supports this conjecture.

In principle, a regular perturbation analysis of the equilibrium  $p_{AB}$  can be performed

to obtain an approximation of the two-locus cline for small values of  $\rho$ . This leads to a system of three coupled linear PDEs that seems to be difficult to analyze. In fact, already for a continent-island model (Bürger and Akerman 2011) and a two-deme model (Akerman and Bürger 2014), the resulting approximations were very complicated and not immediately intuitive.

## 6 Numerical results for arbitrary recombination

Here, we present numerical results to (i) test the accuracy of the approximation for the cline  $(p_A, p_B, D)$  derived in Theorem 4.1, and to (ii) illustrate the dependence of the properties of the two-locus cline on the recombination rate over the full range of recombination rates

Stationary solutions were obtained by numerical integration of the system (2.6) of partial differential equations from given initial conditions. We used the *Mathematica* function `NDSolve` with sufficiently small values of `MaxStepFrac` (mostly = 0.0005). We solved this system for  $(x, t) \in [-L, L] \times [0, T]$ , where for given parameters  $\lambda$  (usually  $\lambda = 1$ ),  $\alpha_{\pm}$ , and  $\beta_{\pm}$ , the lengths  $T$  and  $L$  of the time and spatial intervals were chosen sufficiently large in the following sense.  $T$  was chosen such that  $\max_{-L \leq x \leq L} |p_A(x, T) - p_A(x, T/2)| < 5 \times 10^{-7}$  (and analogously for  $p_B$ ). Our choice of  $L$  yields  $\max_{-2L/3 \leq x \leq 2L/3} |p_A(x, T, \rho = \infty) - P(x)| < 1.2 \times 10^{-3}$ , where  $P(x)$  is the analytically calculated cline of the one-locus case (and analogously for  $p_B$ ). We imposed zero-flux boundary conditions; cf. (2.9). Deviations of the numerically computed solution on  $[-L, L]$  from the analytical solution on  $\mathbb{R}$  close to the boundaries of this interval cannot be minimized arbitrarily because on a finite interval stationary allele frequencies are always strictly positive at the boundary. Values of  $T$  and  $L$  are given in the figure legends.

The reader may recall from (2.7) that we use scaled parameters, i.e.,  $\rho = r\lambda = 2r/\sigma^2$  is the recombination *rate* relative to the diffusion constant  $d = 1/\lambda = \frac{1}{2}\sigma^2$ , which is half the migration variance. Following (3.23), we use the harmonic mean of  $\alpha_+$  and  $\alpha_-$ ,  $H(\alpha_+, \alpha_-)$ , as a measure for the selection intensity (before scaling by  $d$ ). Then the ratio of recombination rate to selection intensity at locus  $A$  is

$$\frac{r}{H(\alpha_+, \alpha_-)} = \frac{1}{\epsilon\lambda H(\alpha_+, \alpha_-)}. \quad (6.1)$$

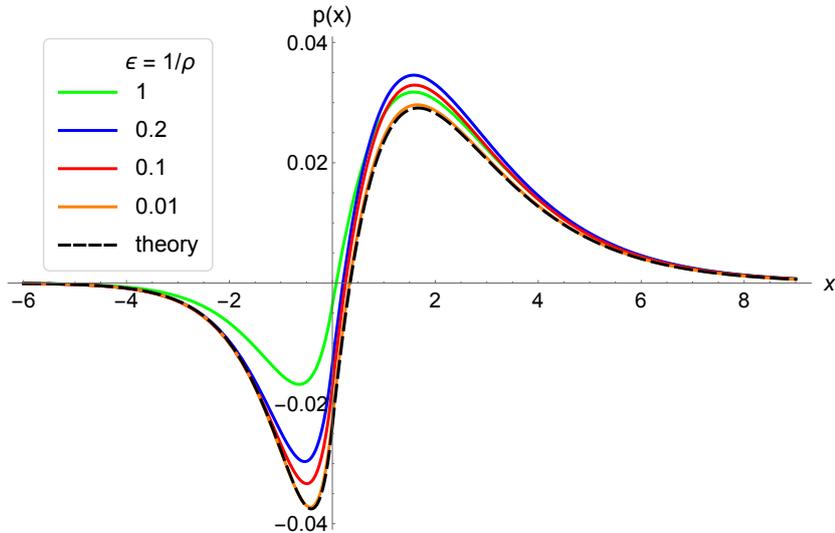


Figure 6: Accuracy of the first-order perturbation  $p(x)$  in the absence of dominance ( $h_A = h_B = 0$ ). The selection parameters are  $\lambda = 1$ ,  $\alpha_+ = \beta_+ = 0.4$ ,  $\alpha_- = \beta_- = 1.6$ . Because the loci are equivalent,  $p(x)$  (dashed black line) was computed from the explicit formulas given in Theorem 4.6. The solid lines shows  $[p_A(x) - P(x)]/\epsilon$ , where  $p_A(x)$  and  $P(x)$  were obtained by numerical integration as described in the main text. We chose  $T = 100$  and  $L = 12$ . The ordering of lines in the legend coincides with their order near  $x = -1$ .

We expect that the strong-recombination approximation derived in Section 4 is accurate if this ratio is sufficiently bigger than 1. For the parameters in Figs. 6 and 7, we obtain  $r/H(\alpha_+, \alpha_-) = 25/(16\epsilon) = 0.5625/\epsilon$ . For those in Fig. 8, we obtain  $r/H(\alpha_+, \alpha_-) = 9/(8\epsilon) = 1.125/\epsilon$  and  $r/H(\beta_+, \beta_-) = 15/(4\epsilon) = 3.75/\epsilon$ . Indeed, Fig. 9 shows that the slope at the center of the allele-frequency clines is reasonably well approximated by the weak-recombination approximation if  $r/H \gtrsim 1$ .

## 6.1 Accuracy of the strong-recombination approximation

For a range of (scaled) recombination rates  $\rho = 1/\epsilon$ , Fig. 6 compares the analytical approximation  $p(x)$  from Theorem 4.1 with  $\frac{1}{\epsilon}(p_A(x) - P(x))$ , where  $p_A(x)$  and  $P(x)$  were obtained by numerical integration of the PDE as described above. Here, the loci are equivalent and dominance is absent. Qualitatively similar results were found for a variety of parameter combinations including strong dominance and loci of different effects. If the selection strength is similar to that in Fig. 6, then the accuracy of the approximation  $p(x)$  to  $\frac{1}{\epsilon}(p_A(x) - P(x))$  is also similar to that in the figure (results not

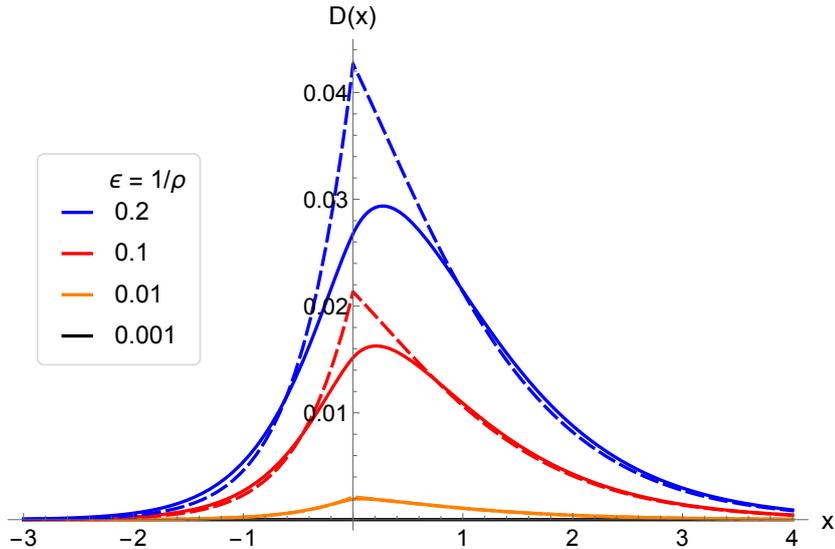


Figure 7: Accuracy of the first-order perturbation  $\epsilon d(x)$  in (4.6) of  $D(x)$ . The latter was obtained by numerical solution of (2.6). Solid lines show the true  $D(x)$ , dashed lines the corresponding approximation  $\epsilon d(x)$ . The selection parameters are as in Fig. 6. We chose  $T = 100$  and  $L = 12$ .

shown).

Figure 7 shows  $D(x)$  and its approximation  $\epsilon d(x)$  in (4.6). Whereas  $D(x)$  has a continuous first derivative at  $x = 0$ ,  $d(x)$  has not. Also recall from Sect. 2 that  $D(x)$  has different one-sided second derivatives at  $x = 0$ . Note that in contrast to its approximation  $\epsilon d(x)$ ,  $D(x)$  is in general not maximized at  $x = 0$ .

## 6.2 Dependence of the two-locus cline on the recombination rate

To investigate the dependence of the shape of two-locus clines on the recombination rate for given parameters  $\alpha_{\pm}$ ,  $\beta_{\pm}$ , it is sufficient to vary  $\rho$  and keep  $\lambda$  constant. This follows from the scaling property (2.15).

Figure 8 illustrates the dependence on the scaled recombination rate  $\rho$  of the clines in allele frequencies at each locus and of the linkage disequilibrium between the loci. As already shown by Slatkin (1975), stronger linkage has a stronger effect on the shape, in particular on the slope, of the cline at the locus under weaker selection. This is in accordance with intuition because the locus under stronger selection exerts stronger indirect selection on the weaker locus than vice versa.

If the loci are unlinked ( $\rho \rightarrow \infty$ , black lines in Fig. 8), the width (3.24) of the cline at locus  $\mathcal{A}$  ( $\mathcal{B}$ ) is  $\omega_A = \frac{3\sqrt{3}}{2} \approx 2.598$  ( $\omega_B = 3\sqrt{\frac{5}{2}} \approx 4.743$ ). For completely linked loci ( $\rho = 0$ , green lines in Fig. 8), we obtain  $\omega_A = \omega_B = \sqrt{5} \approx 2.236$  because the width can be calculated from the added step-size parameters, i.e.,  $\alpha_+ + \beta_+ = \alpha_- + \beta_- = 1.2$  (Section 5). Indeed, most gene-frequency change at loci  $\mathcal{A}$  and  $\mathcal{B}$  occurs in the intervals  $(-\omega_A, \omega_A)$  and  $(-\omega_B, \omega_B)$ , respectively.

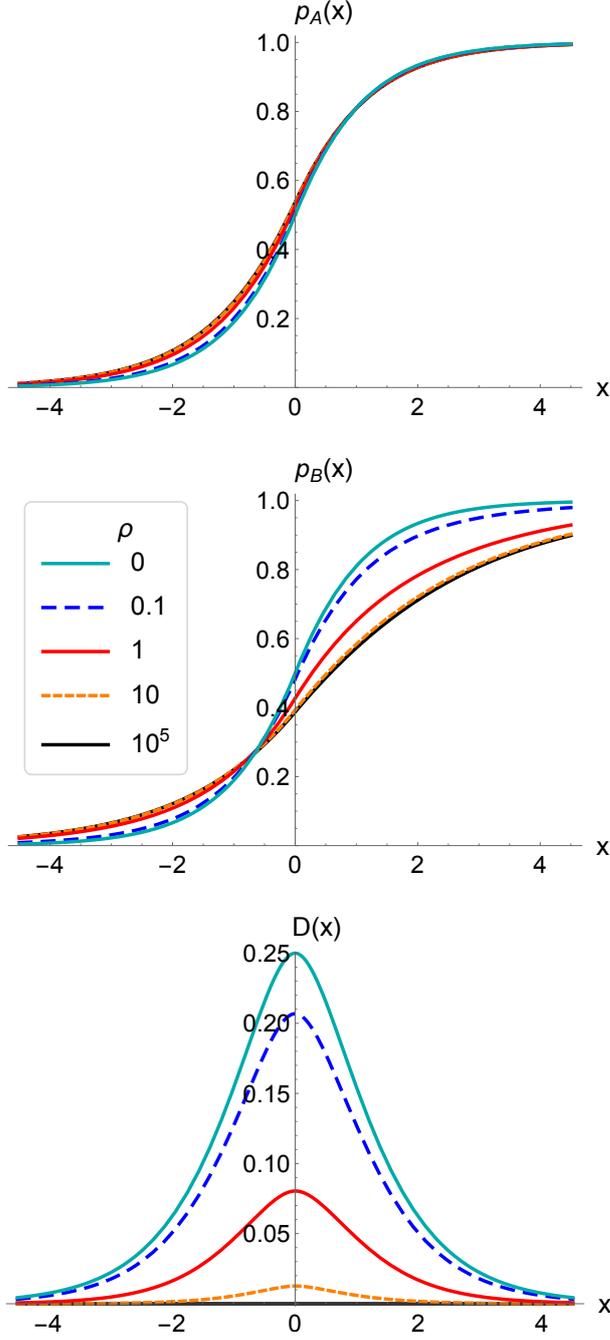


Figure 8: Dependence of clines in  $p_A$ ,  $p_B$ , and  $D$  on the scaled recombination rate  $\rho$ . The other parameters are  $\lambda = 1$ ,  $\alpha_+ = 2$ ,  $\alpha_- = 1.6$ ,  $\beta_+ = 0.4$ ,  $\beta_- = 0.8$ . For  $\rho = 0$ , the maximum relative deviation on the interval  $[-8, 8]$  of the numerically calculated  $p_A = p_B$  from the corresponding exact one-locus solution is  $< 2 \times 10^{-5}$ . For  $\rho = 10^5$ , the maximum relative deviations on the interval  $[-8, 8]$  of  $p_A$  and  $p_B$  from the corresponding exact one-locus solutions are about  $8 \times 10^{-4}$  and  $2.5 \times 10^{-3}$ , respectively. The ordering of lines in the legend coincides with their order for  $x > 0$  in the same panel and with the order in the bottom panel. We chose  $T = 200$  and  $L = 12$ .

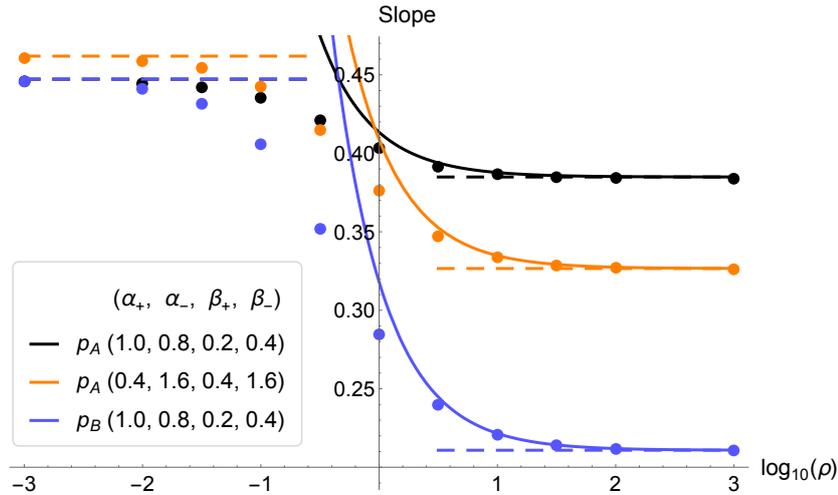


Figure 9: Slope of allele-frequency clines in the center as a function of the (decadic) logarithm of the recombination rate  $\rho$ . Black and orange solid lines show  $p'_A(0) = P'(0) + \frac{1}{\rho} p'(0)$  for the parameter combinations given in the legend (all with  $\lambda = 1$ ), where  $P'(0)$  and  $p'(0)$  are calculated from (3.16) and (A.46). The blue solid line shows  $p'_B(0)$  and was calculated analogously. For the other parameter combination (orange line) loci are equivalent, whence  $p_B = p_A$ . The ordering of lines in the legend coincides with their order for  $\log_{10}(\rho) > 0$ . The dots are the slopes obtained from numerically computed solutions of the PDE. The dashed horizontal lines on the right are for independent loci ( $\rho \rightarrow \infty$ ) and are computed from (3.16). The dashed horizontal lines on the left are for  $\rho = 0$  and are computed from (3.16) by using  $\alpha_+ + \beta_+$  and  $-(\alpha_- + \beta_-)$  as values for the step function (2.10a). The black dashed line on the left side is invisible because it coincides with the blue one since the loci are equivalent.

Figure 9 displays the slope in the center as a function of  $\log_{10}(\rho)$  for various parameter combinations. In particular, it compares analytical results from the strong-recombination approximation with results obtained from numerical integration of the PDE (2.6). The solid lines, based on the analytical approximation, always overestimate the true slope (dots), but provide an accurate approximation if, approximately,  $\rho > \sqrt{10}$ , i.e.,  $\log_{10}(\rho) > 0.5$ . For small recombination rates, the approximation diverges, whereas the true slope approaches that of one-locus clines with values  $\alpha_+ + \beta_+$  and  $-(\alpha_- + \beta_-)$  for the step function (dashed horizontal lines on the left).

## 7 Discussion

The shape of a cline in allele, genotype, or phenotype frequencies has many determinants. Among these are the migration distribution, the spatial dependencies of fitnesses,

but also dominance relations among alleles and linkage to other loci under selection. In this paper, we focus on the latter two and base our analysis on a simple but classical scenario for migration and selection (Haldane 1948; Slatkin 1973, 1975; Nagylaki 1975). We model migration by homogeneous, isotropic diffusion on the real line and let selection act additively on two loci,  $\mathcal{A}$  and  $\mathcal{B}$ , such that in one region ( $x \geq 0$ ) one allele at each locus ( $A_1, B_1$ ) is advantageous, whereas in the other region ( $x < 0$ ) the other allele ( $A_2, B_2$ ) is advantageous. We admit arbitrary intermediate, spatially independent dominance at both loci, and we assume that the two loci are recombining. Therefore, linkage disequilibrium is generated. We start with a summary of main results.

We studied the non-constant stationary, or equilibrium, solutions of the system (2.6) of partial differential equations for the allele frequencies and the linkage disequilibrium. This system is obtained from the original set of differential equations (2.1) for the gamete frequencies (Slatkin 1975) by transforming the variables and rescaling time and parameters appropriately; see (2.7). The equilibrium equations resulting from system (2.6) are a special case of a multilocus cline model developed by Barton and Shpak (2000).

Already Haldane showed that the slope of the cline in the center, i.e., at the environmental step, is proportional to the square root of the ratio of the selection coefficient and the migration variance. Therefore, the slope of the cline or its inverse, the width, can be used to infer the selection intensity if an estimate of the migration variance is available. Haldane's results reveal that the asymptotic behavior of the clines for no or complete dominance differ drastically; see also Nagylaki (1975), who concentrated on semi-infinite clines, or an environmental pocket.

In generalization of Haldane's work, we derived an explicit solution for the single-locus cline with dominance (Theorem 3.2). It shows that the slope of the cline in the center,  $P'(0)$ , is independent of the degree of dominance (eq. (3.16) and Figure 1). Since we admit non-symmetric selection, i.e., the selection strength in the two regions may differ, the appropriate measure of selection intensity is the harmonic mean of the step sizes ( $\alpha_+, \alpha_-$ ; eq. 2.10a). Slatkin's (1973) characteristic length for allele-frequency variation and the width of a cline are generalized accordingly (Remark 3.4).

The asymptotic behavior of the single-locus cline follows immediately from the explicit representation in (3.10) and is given by eqs. (4.22). The approach to constancy

(to 0 as  $x \rightarrow -\infty$ , to 1 as  $x \rightarrow +\infty$ ) is exponential, except when dominance is complete. Then the approach is inverse quadratic in the region where the advantageous allele is recessive. In the general case, the exponent of the exponential that describes convergence as  $x \rightarrow \pm\infty$  is  $a_{\pm}\sqrt{1 \mp h}$ , where  $h$  is the dominance coefficient and  $a_{\pm}$  is the square root of the ratio of step size to half the migration variance; see eqs. (3.5a) and (2.7). Therefore, strong dominance does have a strong effect on the asymptotic properties of the cline. This is also reflected by a global measure for the steepness of a cline, which is investigated in Section 4.5; cf. Figure 4.

Sections 4, 5, and 6 are primarily dedicated to the exploration of the influence of linkage on the shape of two-locus clines. Most of the analysis is performed for arbitrary intermediate dominance. Our main result is Theorem 4.1. For strong recombination relative to selection and diffusion, it provides the approximations for the clines of (marginal) one-locus allele frequencies and of the linkage disequilibrium. In other words, the quasi-linkage-equilibrium approximation of the two-locus cline is derived. The allele-frequency cline at locus  $\mathcal{A}$  is  $p_A \approx P + (1/\rho)p$ , where  $P$  is the single-locus cline at  $\mathcal{A}$ . The perturbation term  $p$  is the solution of the linear second-order differential equation (4.7) and can be obtained by integration; see (4.8), (4.9), (4.10). The linkage disequilibrium is approximately  $(1/\rho)d(x)$ , where  $d(x)$  is given by (4.6).

Corollary 4.3 summarizes simple, general properties of the perturbation term  $p(x)$ . It shows that  $p'(0)$  is always positive. Hence, the cline gets steeper in the center with tighter linkage, as is expected intuitively (see Slatkin 1975 and Barton 1983). Depending on the parameters,  $p(0)$  can be positive, negative, or zero. In fact,  $p(x)$  can be positive on the real line, negative, or change sign once (Figure 2). Thus, linkage to other loci may affect the shape of the allele-frequency clines in complex ways. We also note that, in contrast to  $P'(0)$ ,  $p'(0)$  does depend on dominance, although only weakly.

If multiple additive loci contribute to fitness, then terms resulting from all pairwise and high-order linkage disequilibria have to be added to the right-hand side of the differential equations for the allele frequencies, i.e., to (2.6a) and the corresponding equations for the other loci. If recombination is strong, only pairwise linkage disequilibria will matter and the equilibrium equations for the allele frequencies will have the same form as eqs. (24) and (25) in Barton and Shpak (2000). Also differential equations for the linkage disequilibria can be deduced. A quasi-linkage-equilibrium approximation for

the pairwise linkage disequilibria was derived by Barton and Shpak (2000). As already noted above, our equation (4.6) is a special case of their equation (15). Putting all this together, a multilocus generalization of the second-order differential equation (4.7) for the perturbation term  $p(x)$  in (4.2) can be derived. This has a more complicated inhomogeneous term, resulting from all pairwise linkage disequilibria, and is awaiting analysis. Based on the investigation of Barton (1983) of a model with hybrid inferiority (and spatially uniform selection), we expect that the cumulative effect of many loci may considerably steepen the allele-frequency clines.

In Section 4.3, we derived the asymptotic properties of the perturbation term  $p(x)$  and of the quasi-linkage-equilibrium approximation of  $p_A$ . Proposition 4.4 shows that, unless there is complete dominance,  $p_A$  approaches constancy (0 or 1) as  $x \rightarrow \pm\infty$  at the same exponential rate as the single-locus cline  $P$ . What differs is the multiplicative factor in front of the exponential, which depends on the recombination rate and the other parameters. This factor may increase or decrease with  $\rho$  (4.28).

For equivalent loci, we calculated an explicit expression for  $p(x)$ , which is relatively simple if dominance is absent (Theorem 4.6), but complicated in general (Appendix A.3). For the slope  $p'(0)$  and the value  $p(0)$  simple expressions are obtained if there is no dominance (Corollary 4.8), and somewhat more complicated expressions if there is dominance (Corollary A.2). They all are easily evaluated numerically and provide insight into the effects of dominance on the slope in the center of two-locus clines (Figure 3). These corollaries also show that for equivalent loci,  $p(x)$  changes sign once and is positive for large  $x$ , and negative for small  $x$ . Therefore,  $p_A(x) \approx P(x) + (1/\rho)p(x) > P(x)$  holds above a threshold, and  $p_A(x) < P(x)$  below that threshold.

In Section 4.5, we studied the global, cumulative measure of steepness  $s(P)$  in (4.54). It is the square of a measure investigated by Liang and Lou (2011). For a quite general one-locus cline model on a bounded domain, these authors had shown that this measure is a strictly monotone increasing function of  $\lambda$ , i.e., stronger selection or weaker migration increase the steepness of the cline. This is also true in our setting, and (4.55) quantifies the dependence of  $s(P)$  on  $\lambda$ . In contrast to the slope in the center, this cumulative measure depends on dominance, although in a complicated way (A.32). This is visualized in Figure 4.

For two recombining loci, we showed that an increase in the strength of selection

causes an increase in the steepness  $s(p_A)$  of the one-locus cline  $p_A$  (4.63). Moreover, this increase is stronger the smaller the recombination rate  $\rho$  is. We conjecture, but could not prove, that  $s(p_A)$  increases if  $\rho$  decreases. This conjecture is supported by numerical evaluations (e.g., Figures A.2 and 5) and by a proof for the case of loci of equal effects without dominance (Appendix A.5.1). In summary, the dependence of the global measure  $s(p_A)$  on the parameters is qualitatively the same as that of the slope in the center,  $p'_A(0)$ , if there is no dominance. However, in contrast to the slope, which is a local measure, the global measure reflects the influence of dominance. The advantage of the slope in the center is that it will be easier to estimate in practice.

In Section 5, we briefly treated the case of no recombination, when the model becomes formally equivalent to a one-locus four-allele system. We conjecture that the cline formed by the gametes  $AB$  and  $ab$  is globally asymptotically stable, i.e., the intermediate gametes  $Ab$  and  $aB$  are eventually lost. This cline can be computed from the one-locus formula (3.29). For weak migration, stability of this cline was proved for a finite number of demes (Akerman and Bürger 2014) and also for a bounded continuous habitat (Su and Bürger, unpublished). For strong migration, clines usually vanish in bounded discrete or continuous habitats (see below).

Finally, in Section 6, we solved the system (2.6) of PDEs numerically to (i) check the accuracy of our quasi-linkage-equilibrium approximation (Section 6.1, Figures 6 and 7) and to (ii) explore the whole range of possible recombination rates (Section 6.2, Figures 8 and 9). The latter results complement and extend findings by Slatkin (1975) for essentially the same model, and by Barton (1983, 1999), Barton and Shpak (2000), and Geroldinger and Bürger (2015) for related models, who showed numerically that tighter linkage steepens clines (in allele frequency or mean phenotype) and increases linkage disequilibrium. Unfortunately, no general theory is currently available to prove or quantify these findings for arbitrary recombination.

The model investigated in this paper assumes that the habitat is unbounded and, in particular, that each allele is advantageous in an unbounded region (of infinite measure). As a consequence, a cline exists for arbitrarily strong migration relative to selection. However, as the migration variances increases, the cline becomes increasingly flatter. In our model, this follows easily from the scaling, or invariance, properties derived in Lemma 2.3; for the single-locus case, see also Remark 3.4. For very general fitness func-

tions, but still assuming that each allele is advantageous in an unbounded region, Conley (1975) and Fife and Peletier (1977) proved that clines generally exist independently of the strength of migration. As shown by Nagylaki's (1975) analysis of an environmental pocket, an allele that is favored in a bounded region and disadvantageous in unbounded region will be lost under sufficiently strong migration. This is in line with the apparently generic property of migration-selection models on bounded domains that clinal or, more general, genetic variation can be maintained only if selection is sufficiently strong relative to migration. Otherwise, i.e., below a threshold value of the ratio of selection intensity to migration rate, the type with the highest spatially averaged fitness will swamp the whole population. For recent reviews of cline models, we refer to Nagylaki and Lou (2008) and Lou et al. (2013), and for reviews of models with discrete demes to Lenormand (2002), Nagylaki and Lou (2008), and Bürger (2014).

We conjecture that for the present step environment on the real line, the two-locus cline exists, is unique, and globally asymptotically stable for all parameter combinations satisfying our assumptions. Our conjecture is based on the invariance property derived in Lemma 2.3 and the apparent existence and stability for strong recombination and arbitrary  $\lambda$  (Theorem 4.1).

Lenormand et al. (1998) estimated the migration variance and the selection pressure from geographic gradients (clines) in allele frequencies at two loosely linked insecticide resistance loci in the mosquito *Culex pipiens pipiens*. Their simulation results showed that different hypotheses of dominance (recessiveness, no dominance, complete dominance) do not have an important effect on the estimates of migration variance, whereas the recombination rate has a significant effect. Our study provides analytical support for their finding. However, our model is not directly applicable to the case study of Lenormand et al. because their organisms are adapted to an environmental pocket. Then the slope of the cline at the transition between the environments and the maximum allele frequencies are needed to estimate the migration variance (selection intensity) if independent estimates of the selection intensity (migration variance) are available (Nagylaki 1975). It would be of interest to extend the present analysis to study clines caused by environmental pocket. In particular, the effects of dominance should be identifiable by studying the asymptotic properties of a cline (see Section 4.3) or the global measure of steepness that was investigated in Section 4.5.

## Acknowledgments

I am very grateful to Professor Tom Nagylaki and to Dr. Linlin Su for perceptive and inspiring communication. It is a pleasure to thank Professor Nick Barton, Professor Tom Nagylaki, and Dr. Swati Patel for useful comments on a previous version. Financial support by the Austrian Science Fund (FWF) through Grant P25188-N25 is gratefully acknowledged.

Symbol	Reference	Definition
$\mathcal{A}$	(2.1a)–	First locus
$A$	(2.1a)–	First allele at locus $\mathcal{A}$
$A_+, A_-$	(3.5d)	Constants associated with locus $\mathcal{A}$
$a$	(2.1a)–	Second allele at locus $\mathcal{A}$
$a_+, a_-$	(3.5a)	Scaled step sizes at locus $\mathcal{A}$
$a_0$	(3.8)	$P(0) = P(0-)$
$\mathcal{B}$	(2.1a)–	Second locus
$B$	(2.1a)–	First allele at locus $\mathcal{B}$
$B_+, B_-$	(4.25a)	Constants associated with locus $\mathcal{B}$
$b$	(2.1a)–	Second allele at locus $\mathcal{B}$
$b_+, b_-$	(4.5)	Scaled step sizes at locus $\mathcal{B}$
$b_0$	(4.26)	$Q(0) = Q(0-)$
$c, c_1, c_2$	(2.14)–	Constants
$D$	(2.1a)–	Linkage disequilibrium
$d$	(4.2)	Approximation to $D$
$F_+, F_-$	(3.6)	Functions occurring in $P(x), Q(x)$
$H$	(3.16)	Harmonic mean
$h_A, h_B$	(2.2)	Dominance coefficient at locus $\mathcal{A}, \mathcal{B}$
$h$	(3.1)	Dominance coefficient in Section 3
$I_+, I_-$	(4.4)	Integrals occurring in $p(x)$
$i$	(2.1a)–	Index for gamete
$J$	(4.61)	Integral occurring in approximation for $s$
$j$	(2.1a)–	Index for gamete
$k_+, k_-$	(4.9)	Functions occurring in $p(x)$
$k_0$	(4.10)	Constant occurring in $k_{\pm}(x)$
$\ell_A$	(3.23)	Characteristic length of cline at $\mathcal{A}$
$P$	(3.2), (3.10)	Single-locus cline at locus $\mathcal{A}$
$P_{\lambda}$	(4.55)–	$P$ , with dependence on $\lambda$ made explicit
$\mathbf{p}$	(2.1a)–	Vector of gamete frequencies
$p_i$	(2.1a)–	Frequency of gamete $i$
$p_A$	(2.4)–	Frequency of allele $A$
$p_A^{(\epsilon)}$	(4.2)	Approximation to allele-frequency cline at $\mathcal{A}$

$p$	(4.2b)	Perturbation term in $p_A^{(e)}$
$p_B$	(2.4)–	Frequency of allele $B$
$p_B^{(e)}$	(4.2)	Approximation to allele-frequency cline at $\mathcal{B}$
$p_{A,(\lambda,\rho)}$	(4.59)–	$p_A$ , with dependence on $\lambda$ and $\rho$ made explicit
$Q$	(4.2a)–	Single-locus cline at locus $\mathcal{B}$
$q$	(4.2b)	Perturbation term in $p_B^{(e)}$
$\mathbb{R}$	(2.1a)–	Real line $(-\infty, \infty)$
$r$	(2.1a)–	Recombination rate
$s$	(4.54)	Measure of global steepness
$t$	(2.1a)–	Time
$w_{ij}$	(2.1a)–	Fitness of genotype $ij$
$w_i$	(2.1a)–	Fitness of gamete $i$
$\bar{w}$	(2.1a)–	Mean fitness of population
$X_+, X_-$	(3.5b), (3.5c)	Scaled spatial variable for locus $\mathcal{A}$
$x$	(2.1a)–	Spatial position
$Y_+, Y_-$	(4.25b), (4.25c)	Scaled spatial variable for locus $\mathcal{B}$
$y$	(3.6)	Spatial position
$Z_+, Z_-$	(3.9)	Important term in single-locus solution $P$
<hr/>		
$\alpha(x)$	(2.2), (2.10a)	Fitness function for locus $\mathcal{A}$
$\alpha_+, \alpha_-$	(2.10a)	Step sizes at locus $\mathcal{A}$
$\tilde{\alpha}$	(3.22)–	$\alpha_+ / (\alpha_+ + \alpha_-)$
$\beta(x)$	(2.2), (2.10b)	Fitness function for locus $\mathcal{B}$
$\beta_+, \beta_-$	(2.10b)	Step sizes at locus $\mathcal{B}$
$\epsilon$	(4.1)	$1/\rho$
$\eta_i$	(2.1a)–	Constants
$\vartheta_A, \vartheta_B$	(2.4)	Dominance effect at locus $\mathcal{A}, \mathcal{B}$
$\kappa_+, \kappa_-$	(4.16)	Limits of $k_+, k_-$
$\lambda$	(2.7)	$2/\sigma^2$
$\pi$	(2.13)	$(p_A, p_B, D)$
$\rho$	(2.7)	Scaled recombination rate, $2r/\sigma^2$
$\sigma^2$	(2.1a)–	Migration variance
$\phi$	(3.7)	Function occurring in definition of $F_{\pm}$
$\omega_A$	(3.24)	Width of cline at locus $\mathcal{A}$
<hr/>		
$'$	(2.1a)–	Partial derivative with respect to $x$
$\cdot$	(2.1a)–	Partial derivative with respect to $t$
$\sim$	(4.22)–	Asymptotic equivalence
<hr/>		

Table 1: Glossary of symbols. We list and briefly define the symbols in the main text that occur in more than one paragraph. Roman, Greek, and Script letters are listed separately. Uppercase letters precede lower case ones and for each letter listing is in order of appearance in the text. The references are to the position of first appearance in the text. Reference (2.1)– refers to the text above Equation (2.1), whereas (2.1)+ refers to the text below Equation (2.1).

# A Appendix

## A.1 Generality of the fitness scaling (2.2)

We assume absence of epistasis and assign spatially dependent fitnesses to one-locus genotypes as follows:

$$\frac{AA}{2\alpha_1(x)} \quad \frac{Aa}{\alpha_1(x) + \alpha_2(x) + h_A(\alpha_1(x) - \alpha_2(x))} \quad \frac{aa}{2\alpha_2(x)}, \quad (\text{A.1})$$

and analogously for the other locus. The dominance coefficient  $h_A$  could also depend on  $x$ , but in view of our applications, we assume constant dominance coefficients. Therefore, suppressing the variable  $x$ , we obtain the genotypic fitnesses

	<i>BB</i>	<i>Bb</i>	<i>bb</i>
<i>AA</i>	$2\alpha_1 + 2\beta_1$	$2\alpha_1 + \beta_1 + \beta_2 + h_B(\beta_1 - \beta_2)$	$2\alpha_1 + 2\beta_2$
<i>Aa</i>	$\left\{ \begin{array}{l} \alpha_1 + \alpha_2 + 2\beta_1 \\ +h_A(\alpha_1 - \alpha_2) \end{array} \right.$	$\left\{ \begin{array}{l} \alpha_1 + \alpha_2 + \beta_1 + \beta_2 \\ +h_A(\alpha_1 - \alpha_2) + h_B(\beta_1 - \beta_2) \end{array} \right.$	$\left\{ \begin{array}{l} \alpha_1 + \alpha_2 + 2\beta_2 \\ +h_A(\alpha_1 - \alpha_2) \end{array} \right.$
<i>aa</i>	$2\alpha_2 + 2\beta_1$	$2\alpha_2 + \beta_1 + \beta_2 + h_B(\beta_1 - \beta_2)$	$2\alpha_2 + 2\beta_2$

(A.2)

Because in the continuous-time model (2.1a), the same function of  $x$  can be added to all genotypic fitness functions  $w_{ij}(x)$  without changing the dynamics, we subtract  $\alpha_1 + \alpha_2 + \beta_1 + \beta_2$  from all entries in (A.2) and obtain

	<i>BB</i>	<i>Bb</i>	<i>bb</i>
<i>AA</i>	$\alpha_1 - \alpha_2 + \beta_1 - \beta_2$	$\alpha_1 - \alpha_2 + h_B(\beta_1 - \beta_2)$	$\alpha_1 - \alpha_2 - \beta_1 + \beta_2$
<i>Aa</i>	$h_A(\alpha_1 - \alpha_2) + \beta_1 - \beta_2$	$h_A(\alpha_1 - \alpha_2) + h_B(\beta_1 - \beta_2)$	$h_A(\alpha_1 - \alpha_2) - \beta_1 + \beta_2$
<i>aa</i>	$-\alpha_1 + \alpha_2 + \beta_1 - \beta_2$	$-\alpha_1 + \alpha_2 + h_B(\beta_1 - \beta_2)$	$-\alpha_1 + \alpha_2 - \beta_1 + \beta_2$

(A.3)

Defining

$$\alpha(x) = \alpha_1(x) - \alpha_2(x) \quad \text{and} \quad \beta(x) = \beta_1(x) - \beta_2(x), \quad (\text{A.4})$$

we arrive at (2.2).

Assuming that  $\alpha_i(x)$  and  $\beta_i(x)$  are step functions with a single step at  $x = x_0$ , and that  $A$  and  $B$  are advantageous on  $[x_0, \infty)$  and disadvantageous on  $(-\infty, x_0)$ , we obtain the assumptions (2.10) and (2.10c). Without loss of generality, we choose  $x_0 = 0$ .

## A.2 The integrals in (4.9)

We show that the integrals occurring in the definition (4.9) of  $k_{\pm}(x)$  can be obtained by one-fold integration. We define

$$\psi_+(x) = \int_0^x \frac{1}{[P'(y)]^2} dy \quad \text{and} \quad \psi_-(x) = - \int_x^0 \frac{1}{[P'(y)]^2} dy. \quad (\text{A.5})$$

**Lemma A.1.**

$$\int_0^x \frac{I_+(y)}{[P'(y)]^2} dy = \psi_+(x)I_+(x) + \int_0^x \psi_+(y)[P'(y)]^2 Q'(y)[1 + h_B - 2h_B Q(Y)] dy \quad \text{if } x \geq 0, \quad (\text{A.6a})$$

$$\int_x^0 \frac{I_-(y)}{[P'(y)]^2} dy = -\psi_-(x)I_-(x) - \int_x^0 \psi_-(y)[P'(y)]^2 Q'(y)[1 + h_B - 2h_B Q(Y)] dy \quad \text{if } x \leq 0, \quad (\text{A.6b})$$

where

$$\psi_{\pm}(x) = \tilde{\psi}_{\pm}(x) - \tilde{\psi}_{\pm}(0), \quad (\text{A.7})$$

and

$$\begin{aligned} \tilde{\psi}_+(x) = & \frac{1}{72a_+^3(1-h_A)^{7/2}} \left\{ 12xa_+ \sqrt{1-h_A}(5-21h_A+36h_A)^2 \right. \\ & + \frac{1}{Z_+ - (1+3h_A)Z_+^{-1}} \left[ Z_+^3 + 16(1-3h_A)Z_+^2 \right. \\ & - 3(1+3h_A)^{-1}(43-126h_A+51h_A^2+288h_A^3+432h_A^4)Z_+ \\ & - 32(1-3h_A)(5-3h_A+18h_A^2) - (1+3h_A)^2 Z_+^{-1} + 16(1-3h_A)(1+3h_A)^2 Z_+^{-2} \\ & \left. \left. + (1+3h_A)^3 Z_+^{-3} \right] \right\} \quad \text{if } x \geq 0, \quad h_A < 1, \quad (\text{A.8a}) \end{aligned}$$

$$\begin{aligned} \tilde{\psi}_-(x) = & \frac{1}{72a_-^3(1+h_A)^{7/2}} \left\{ 12a_- \sqrt{1+h_A}(5+21h_A+36h_A)^2 \right. \\ & + \frac{1}{(1-3h_A)Z_- - Z_-^{-1}} \left[ (1-3h_A)^3 Z_-^3 + 16(1+3h_A)(1-3h_A)^2 Z_-^2 \right. \\ & - 3(43+126h_A+51h_A^2-288h_A^3+432h_A^4)Z_- \\ & - 32(1+3h_A)(5+3h_A+18h_A^2) - (1-3h_A)Z_-^{-1} + 16(1+3h_A)Z_-^{-2} \\ & \left. \left. + Z_-^{-3} \right] \right\} \quad \text{if } x \leq 0, \quad h_A > -1. \quad (\text{A.8b}) \end{aligned}$$

We recall that  $Z_{\pm}$  is defined in (3.9).

*Proof.* The equations (A.6) follow immediately by partial integration because  $\psi_{\pm}(0) = 0$  and  $I_{\pm}(0)$  are finite. Equations (A.8) were derived with the help of *Mathematica*. They can be checked by differentiation.  $\square$

Expressions for  $h_A = \pm 1$  are available in a *Mathematica* notebook.

### A.3 Explicit solution for equivalent loci with dominance

We assume  $b_{\pm} = a_{\pm}$  and  $-1 < h = h_A = h_B < 1$ . Under these assumptions, we can calculate the first-order term  $p(x)$  of  $p_A(x)$  explicitly. In the following, we provide the most important steps of the derivation. Our main result are equations (A.20), which present  $k_{\pm}(x)$ . Together with (4.8), they provide  $p(x)$ . Corollary A.2 gives explicit expressions for  $p(0)$  and  $p'(0)$ . The dependence of  $p'(0)$  on the degree of dominance is analyzed thereafter.

#### A.3.1 Derivation of $k_{\pm}(x)$

First, we compute  $I_+(x)$  in (4.4). Therefore we need to evaluate integrals of the form  $\int_x^{\infty} [f'(y)]^3 [1 + h - 2hf(y)] dy$ , where

$$f''(x) = a_1 f(x) + a_2 f(x)^2 + a_3 f(x)^3. \quad (\text{A.9})$$

With  $f(x) = P(x)$ , (3.2a) yields

$$a_1 = -a_+^2(1 + h), \quad a_2 = a_+^2(1 + 3h), \quad a_3 = -2a_+^2 h. \quad (\text{A.10})$$

We start with the following simple observations:

$$[(f')^2]' = 2f'(a_1 f + a_2 f^2 + a_3 f^3), \quad (\text{A.11a})$$

$$[f(f')^2]' = 2f'(a_1 f^2 + a_2 f^3 + a_3 f^4) + (f')^3, \quad (\text{A.11b})$$

$$\int_x^{\infty} f(y)^n f'(y) dy = \frac{1}{n+1} (1 - f(x)^{n+1}). \quad (\text{A.11c})$$

The first two identities follow from (A.9), and the last from one-fold partial integration and the boundary condition  $f(\infty) = 1$ .

Therefore, we obtain by one-fold partial integration and the boundary condition

$f'(\infty) = 0$ :

$$\int_x^\infty [f'(y)]^3 dy = -f(x)[f'(x)]^2 - 2\left(\frac{a_1}{3} + \frac{a_2}{4} + \frac{a_3}{5}\right) + 2f(x)^3 \left(\frac{a_1}{3} + \frac{a_2 f(x)}{4} + \frac{a_3 f(x)}{5}\right), \quad (\text{A.12a})$$

$$\int_x^\infty [f'(y)]^3 f(y) dy = -f(x)^2 [f'(x)]^2 - 2\left(\frac{a_1}{4} + \frac{a_2}{5} + \frac{a_3}{6}\right) + 2f(x)^3 \left(\frac{a_1}{4} + \frac{a_2 f(x)}{5} + \frac{a_3 f(x)}{6}\right) - \int_x^\infty [f'(y)]^3 f(y) dy. \quad (\text{A.12b})$$

By collecting terms and substituting (A.10), we arrive at

$$\int_x^\infty [f'(y)]^3 [1 + h - 2hf(y)] dy = \frac{a_+^2}{30} (1 - f)^2 [5(1 + 2f + 3f^2) + h(1 + 2f + 3f^2 - 36f^3) - 20h^2(1 - f)f^3] - f(1 + h - hf)(f')^2, \quad (\text{A.13})$$

where on the right-hand side we abbreviated  $f(x)$  by  $f$ . With  $f(x) = P(x)$  and (3.10), (A.13) yields

$$I_+(x) = \frac{72a_+^2(1-h)^5}{5[Z_+ + 2(1-3h) + (1+3h)Z_+^{-1}]^6} \left[ 5Z_+^3 + 15Z_+^2 + (15-27h)Z_+ + 2(5-27h-54h^2) + 3(1+3h)(5-9h)Z_+^{-1} + 15(1+3h)^2Z_+^{-2} + 5(1+3h)^3Z_+^{-3} \right] \quad \text{if } x \geq 0, \quad (\text{A.14a})$$

$$I_-(x) = \frac{72a_-^2(1+h)^5}{5[(1-3h)Z_- + 2(1+3h) + Z_-^{-1}]^6} \left[ 5(1-3h)^3Z_-^3 + 15(1-3h)^2Z_-^2 + 3(1-3h)(5+9h)Z_- + 2(5+27h-54h^2) + 3(5+9h)Z_-^{-1} + 15Z_-^{-2} + 5Z_-^{-3} \right] \quad \text{if } x < 0, \quad (\text{A.14b})$$

where the expression for  $I_-(x)$  follows from an analogous computation using the boundary conditions at  $-\infty$ .

At  $x = 0$ , these expressions simplify to

$$I_+(0) = \frac{a_+^2}{30} (1 - a_0)^3 [5(1 + a_0) + h(1 - 7a_0 - 14a_0^2) + 10h^2a_0^3], \quad (\text{A.15a})$$

$$I_-(0) = \frac{a_-^2}{30} a_0^3 [5(2 - a_0) + h(20 - 35a_0 + 14a_0^2) + 10h^2(1 - a_0)^3]. \quad (\text{A.15b})$$

With (A.15), we obtain from (4.10),

$$k_0 = \frac{1}{5\sqrt{3}\sqrt{a_+^2 + a_-^2}[1 - h(2a_0 - 1)]} \times \left( \frac{a_+^3(1 - a_0)^2}{a_- a_0} [5(1 + a_0) + h(1 - 7a_0 - 14a_0^2) + 10h^2 a_0^3] - \frac{a_-^3 a_0^2}{a_+(1 - a_0)} [5(2 - a_0) + h(20 - 35a_0 + 14a_0^2) + 10h^2(1 - a_0)^3] \right) \quad (\text{A.16a})$$

$$= \frac{a_+(1 - a_0)}{5\sqrt{3}\sqrt{1 + 2a_0 - 3ha_0^2}[3 - 2a_0 + 3h(1 - a_0)^2][1 - h(2a_0 - 1)]} \times \left[ 10(2a_0 - 1) - 2h(10 - 37a_0 + 37a_0^2) + h^2(2a_0 - 1)(10 - 43a_0 + 43a_0^2) + 3h^3 a_0(1 - a_0)(3 - 10a_0 + 10a_0^2) \right], \quad (\text{A.16b})$$

where the second equality is obtained by using (3.8) to express  $a_-$  in terms of  $a_+$  and  $a_0$ .

We note that

$$\left. \frac{\partial k_0}{\partial h} \right|_{h=0} = a_+ \frac{a_0(1 - a_0)(2 - 61a_0 + 88a_0^2 - 44a_0^3)}{5\sqrt{3}(3 - 2a_0)^2(1 + 2a_0)^{3/2}}. \quad (\text{A.17})$$

This is negative if  $a_0 \gtrsim 0.0345$ .

Defining

$$\varphi_+(z) = \sqrt{1 - h} \left[ \frac{(7 + 9h)z^{-1} + 8}{(1 + 3h)z^{-1} - z} - \frac{10(1 - 3h + (1 + 3h)z^{-1})}{z + 2(1 - 3h) + (1 + 3h)z^{-1}} \right], \quad (\text{A.18a})$$

$$\varphi_-(z) = \sqrt{1 + h} \left[ \frac{(7 - 9h)z + 8}{z^{-1} - (1 - 3h)z} + \frac{10(1 + 3h + (1 - 3h)z)}{z^{-1} + 2(1 + 3h) + (1 - 3h)z} \right], \quad (\text{A.18b})$$

the following indefinite integrals can be written as

$$15a_+ \int^x \frac{I_+(y)}{[P'(y)]^2} dy = \varphi_+(Z_+) + \frac{4}{\sqrt{h}} \arctan \left( \frac{Z_+ + 1 - 3h}{3\sqrt{1 - h}\sqrt{h}} \right) - \frac{2\pi \operatorname{sgn} h}{\sqrt{h}} \quad \text{if } x \geq 0, \quad (\text{A.19a})$$

$$15a_- \int^x \frac{I_-(y)}{[P'(y)]^2} dy = \varphi_-(Z_-) + \frac{4}{\sqrt{h}} \operatorname{arctanh} \left( \frac{Z_-^{-1} + 1 + 3h}{3\sqrt{1 + h}\sqrt{h}} \right) + \frac{2i\pi}{\sqrt{h}} \quad \text{if } x < 0. \quad (\text{A.19b})$$

These integrals can be found with *Mathematica* after rearranging terms. Checking by differentiation is easily done in *Mathematica*. We note that (A.19b) is obtained

from (A.19a) by the joint substitutions  $Z_+ \rightarrow Z_-^{-1}$ ,  $a_+ \rightarrow a_-$ , and  $h \rightarrow -h$ . The constants  $\frac{2\pi \operatorname{sgn} h}{\sqrt{h}}$  and  $\frac{2i\pi}{\sqrt{h}}$  were added only to have real-valued right-hand sides for every  $h \in (-1, 1)$ . (Note that the argument of  $\operatorname{arctanh}$  is greater than 1 if  $h > 0$ , and imaginary if  $h < 0$ .)

Then we obtain from (4.9),

$$k_+(x) = k_0 + \frac{2a_+}{15}[\varphi_+(Z_+) - \varphi_+(A_+)] + \frac{8a_+}{15\sqrt{h}} \left[ \arctan \left( \frac{Z_+ + 1 - 3h}{3\sqrt{1-h}\sqrt{h}} \right) - \arctan \left( \frac{A_+ + 1 - 3h}{3\sqrt{1-h}\sqrt{h}} \right) \right] \quad \text{if } x \geq 0, \quad (\text{A.20a})$$

$$k_-(x) = k_0 - \frac{2a_-}{15}[\varphi_-(A_-) - \varphi_-(Z_-)] - \frac{8a_-}{15\sqrt{h}} \left[ \operatorname{arctanh} \left( \frac{A_-^{-1} + 1 + 3h}{3\sqrt{1+h}\sqrt{h}} \right) - \operatorname{arctanh} \left( \frac{Z_-^{-1} + 1 + 3h}{3\sqrt{1+h}\sqrt{h}} \right) \right] \quad \text{if } x < 0. \quad (\text{A.20b})$$

Although the terms with  $\arctan$  and  $\operatorname{arctanh}$  are non-real for certain ranges of values of  $h$ , their respective differences divided by  $\sqrt{h}$  are always real and finite. This follows from the above comment or from the identity  $\arctan z = -i \operatorname{arctanh}(iz)$ .

Combining (A.20) with (3.28), we find from (4.8) explicit, but complicated, expressions for  $p(x)$ .

For the limits  $\kappa_{\pm} = \lim_{x \rightarrow \pm\infty} k_{\pm}(x)$ , we obtain

$$\kappa_+ = k_0 - \frac{2a_+}{15} \varphi_+(A_+) + \frac{8a_+}{15\sqrt{h}} \left[ \frac{\pi \operatorname{sgn} h}{2} - \arctan \left( \frac{A_+ + 1 - 3h}{3\sqrt{1-h}\sqrt{h}} \right) \right], \quad (\text{A.21a})$$

$$\kappa_- = k_0 - \frac{2a_-}{15} \varphi_-(A_-) - \frac{8a_-}{15\sqrt{h}} \left[ \operatorname{arctanh} \left( \frac{A_-^{-1} + 1 + 3h}{3\sqrt{1+h}\sqrt{h}} \right) + \frac{i\pi}{2} \right]. \quad (\text{A.21b})$$

We note that  $(\kappa_{\pm})/a_{\pm}$  depends only on  $a_0$  and  $h$ . It can be shown that  $\lim_{a_0 \uparrow 1} \frac{\kappa_{\pm}}{a_{\pm}} = 0$  and

$$\lim_{a_0 \downarrow 0} \frac{\kappa_+}{a_+} = \frac{2\sqrt{3}(1-9h) + 9\sqrt{1-h}(1+7h)}{45(1+3h)} - \frac{8}{15\sqrt{h}} \left[ \arctan \left( \frac{\sqrt{3(1-h)} + 3(1-h)}{3\sqrt{1-h}\sqrt{h}} \right) - \frac{2\pi \operatorname{sgn} h}{\sqrt{h}} \right] \geq 0. \quad (\text{A.22})$$

We have shown numerically that  $\kappa_+ > 0$  whenever  $0 < a_0 < 1$ , and that  $\kappa_+/a_+$  is a decreasing function of  $a_0$  and of  $h$ . Similarly,  $\kappa_- < 0$  if  $0 < a_0 < 1$ , and  $\kappa_-/a_-$  is a decreasing function of  $a_0$  and of  $h$ .

### A.3.2 Properties of $p(x)$

Corollary 4.8 can be generalized as follows.

**Corollary A.2.** *Let  $b_{\pm} = a_{\pm}$  and  $-1 < h = h_A = h_B < 1$ . Then  $p(x)$  has the following properties:*

(i)

$$p(0) = \frac{a_0^2(1 - a_0^2)(a_+^2 + a_-^2)}{15[1 + (1 - 2a_0)h]} [10(2a_0 - 1) - 2(10 - 37a_0 + 37a_0^2)h + (2a_0 - 1)(10 - 43a_0 + 43a_0^2)h^2 + 3a_0(1 - a_0)(3 - 10a_0 + 10a_0^2)h^3]. \quad (\text{A.23})$$

If  $h = 0$ , then  $p(0) \geq 0$  if and only if  $a_0 \geq \frac{1}{2}$ , which is the case if and only if  $a_+ \geq a_-$ .

(ii)

$$p'(0) = \frac{a_+ a_- \sqrt{a_+^2 + a_-^2}}{5\sqrt{3}} a_0^2(1 - a_0^2)[15 + 18(1 - 2a_0)h + (3 - 20a_0 + 20a_0^2)h^2], \quad (\text{A.24})$$

which is always positive, as in the general case.

(iii) *The frequency  $p(x)$  changes sign once. If  $p(x_k) = 0$ , then  $p(x) > 0$  if  $x > x_k$  and  $p(x) < 0$  if  $x < x_k$ .*

The proof is analogous to that of Corollary 4.3 and based on the expressions derived in Section A.3.1.

### A.3.3 Dependence of the slope $p'(0)$ on dominance

To indicate the dependence of the slope  $p'(0)$  on the dominance parameter  $h$ , we write  $p'(0, h)$ . Our aim is to investigate the ratio  $p'(0, h)/p'(0, 0)$ . This has the advantage that this ratio depends only on  $\tilde{\alpha} = \alpha_+ / (\alpha_+ + \alpha_-)$ , and not on  $\alpha_+$  and  $\alpha_-$  separately. Indeed, (A.24) implies

$$\begin{aligned} & \frac{p'(0, h)}{p'(0, 0)} \\ &= \frac{a_0(h, \tilde{\alpha})^2(1 - a_0(h, \tilde{\alpha}))^2[15 + 18h(1 - 2a_0(h, \tilde{\alpha})) + h^2(3 - 20a_0(h, \tilde{\alpha}) + 20a_0(h, \tilde{\alpha})^2)]}{a_0(0, \tilde{\alpha})^2(1 - a_0(0, \tilde{\alpha}))^2}, \end{aligned} \quad (\text{A.25})$$

where  $a_0$  depends on  $h$  and  $\tilde{\alpha}$ ; see (3.8).

Figure 3 displays  $p'(0, h)/p'(0, 0)$  as a function of  $h$  for different values of  $\tilde{\alpha}$ . It suggests that the minimum is obtained as either  $(h, \tilde{\alpha}) \rightarrow (-1, 0)$  or  $(h, \tilde{\alpha}) \rightarrow (1, 1)$ ,

and the maximum is obtained as  $(h, \tilde{\alpha}) \rightarrow (1, 0)$  or  $(h, \tilde{\alpha}) \rightarrow (-1, 1)$ . We will show that for every  $h \in (-1, 1)$ ,

$$\frac{4}{5} < \frac{p'(0, h)}{p'(0, 0)} < \frac{6}{5} \quad (\text{A.26})$$

holds for every  $\tilde{\alpha}$ .

First, we determine the limits  $\tilde{\alpha} \rightarrow 0$  and  $\tilde{\alpha} \rightarrow 1$ , from which we obtain the minimum and maximum values. Corollary 3.3(iv) informs us that for each given  $h$ ,  $a_0(h, \tilde{\alpha})$  is a monotone increasing function of  $\tilde{\alpha}$ ; in addition,  $a_0(h, \tilde{\alpha}) \rightarrow 0$  as  $\tilde{\alpha} \rightarrow 0$  and  $a_0(h, \tilde{\alpha}) \rightarrow 1$  as  $\tilde{\alpha} \rightarrow 1$ . It is easy to show that if  $\tilde{\alpha}$  is sufficiently close to 0, then  $a_0(h, \tilde{\alpha}) = \sqrt{\tilde{\alpha}/(3+3h)}$  to leading order in  $\tilde{\alpha}$ . Therefore, taking the limit  $\tilde{\alpha} \rightarrow 0$  in (A.25), we obtain

$$\frac{p'(0, h)}{p'(0, 0)} \xrightarrow{\tilde{\alpha} \rightarrow 0} \frac{5+h}{5}. \quad (\text{A.27})$$

If  $\tilde{\alpha}$  is sufficiently close to 1, then  $a_0(h, \tilde{\alpha}) = 1 - \sqrt{\tilde{\alpha}/(3-3h)}$  to leading order. Then, taking the limit  $\tilde{\alpha} \rightarrow 1$  in (A.25), we obtain

$$\frac{p'(0, h)}{p'(0, 0)} \xrightarrow{\tilde{\alpha} \rightarrow 1} \frac{5-h}{5}. \quad (\text{A.28})$$

We have not yet proved that  $p'(0, h)/p'(0, 0)$  is always between the bounds given by (A.26). Figure 3 suggests that for every given  $h$ ,  $p'(0, h)/p'(0, 0)$  is monotone in  $\tilde{\alpha}$ , and for every given  $\tilde{\alpha}$ ,  $p'(0, h)/p'(0, 0)$  is monotone in  $h$ . The first statements seems to be true, but we cannot prove it. The second statement is valid if  $a_0(h, \tilde{\alpha}) < \frac{1}{35}(34 - \sqrt{421}) \approx 0.385$ ; then  $p'(0, h)/p'(0, 0)$  is strictly monotone increasing in  $h$ . It is also valid if  $a_0(h, \tilde{\alpha}) > \frac{1}{35}(1 + \sqrt{421}) \approx 0.615$ ; then  $p'(0, h)/p'(0, 0)$  is strictly monotone decreasing in  $h$ . If  $\frac{1}{35}(34 - \sqrt{421}) < a_0(h, \tilde{\alpha}) < \frac{1}{35}(1 + \sqrt{421})$ , then  $p'(0, h)/p'(0, 0)$  is concave in  $h$ . This follows from differentiation of (A.25) with respect to  $h$ , which yields

$$\begin{aligned} \frac{\partial}{\partial h} \frac{p'(0, h, \tilde{\alpha})}{p'(0, 0, \tilde{\alpha})} &= \frac{a_0(h, \tilde{\alpha})^2(1 - a_0(h, \tilde{\alpha}))^2}{a_0(0, \tilde{\alpha})^2(1 - a_0(0, \tilde{\alpha}))^2} \\ &\quad \times [33(1 - 2a_0(h, \tilde{\alpha})) + 9h - 70ha_0(h, \tilde{\alpha})(1 - a_0(h, \tilde{\alpha}))], \end{aligned} \quad (\text{A.29})$$

and straightforward analysis of the term in brackets. Therefore, critical points of  $p'(0, h, \tilde{\alpha})/p'(0, 0, \tilde{\alpha})$  are on the curve given by equating this bracket to 0. Substituting this curve into  $\frac{\partial}{\partial \tilde{\alpha}} \frac{p'(0, h, \tilde{\alpha})}{p'(0, 0, \tilde{\alpha})}$  shows that the only critical point in the interior is at  $h = 0$  and  $a_0 = 1/2$ . The minima of the concave function  $h \rightarrow p'(0, h, \tilde{\alpha})/p'(0, 0, \tilde{\alpha})$  are assumed at  $h = -1$  or  $h = 1$ , and these values lie within the bounds given by (A.26). Therefore, we have shown (A.26).

### A.3.4 No dominance

Let  $h = 0$ . Then, instead of (A.19), we obtain the much simpler expressions:

$$\int^x \frac{I_+(y)}{[P'(y)]^2} dy = \frac{1}{a_+} \frac{Z_+^{-1} - 2}{Z_+ - Z_+^{-1}} \quad \text{if } x \geq 0, \quad (\text{A.30a})$$

$$\int^x \frac{I_-(y)}{[P'(y)]^2} dy = \frac{1}{a_-} \frac{Z_-^{-1} - 2}{Z_- - Z_-^{-1}} \quad \text{if } x < 0. \quad (\text{A.30b})$$

These integrals can be obtained either by direct integration of  $I_{\pm}(y)/[P'(y)]^2$  if  $h = 0$ , or from (A.19) by taking the limit  $h \rightarrow 0$ . Moreover,  $k_0$  (A.16a) simplifies to

$$k_0 = \frac{a_+^4(1 - a_0)^3(1 + a_0) - a_-^4 a_0^3(2 - a_0)}{\sqrt{3}a_+a_- \sqrt{a_+^2 + a_-^2} a_0(1 - a_0)} \quad (\text{A.31a})$$

$$= \frac{2a_+(1 - a_0)(2a_0 - 1)}{\sqrt{3}\sqrt{1 + 2a_0}(3 - 2a_0)} = \frac{2a_-a_0(2a_0 - 1)}{\sqrt{3}\sqrt{3 - 2a_0}(1 + 2a_0)}. \quad (\text{A.31b})$$

The expressions for  $k_{\pm}(x)$  and  $\kappa_{\pm}$  follow easily and are given in (4.49) and (4.51), respectively.

## A.4 Steepness of single-locus clines

For the single-locus cline  $P(x)$  in (3.10), the integral (4.54) can be computed with *Mathematica*. We assume  $-1 < h < 1$ . The steepness  $s(P)$  is

$$s(P) = \int_0^{\infty} [P'(y)]^2 dy + \int_{-\infty}^0 [P'(y)]^2 dy, \quad (\text{A.32a})$$

where, after rearrangement,

$$\begin{aligned} \int_0^{\infty} [P'(y)]^2 dy &= a_+ \left\{ \frac{\sqrt{1-h}}{9h^2 [A_+ + 2(1-3h) + (1+3h)A_+^{-1}]^3} \right. \\ &\quad \times \left[ (1-9h^2)A_+^2 + (5-15h+117h^2-189h^3+162h^4)A_+ \right. \\ &\quad + 2(1-3h)(1+3h^2)(5-3h+18h^2) + 2(1-3h)^2(1+3h)(5+6h+9h^2)A_+^{-1} \\ &\quad \left. \left. + (1-3h)(1+3h)^2(5-3h+18h^2)A_+^{-2} + (1+3h)^3(1+3h^2)A_+^{-3} \right] \right. \\ &\quad \left. + \frac{1-9h^2}{27h^{5/2}} \left[ \arctan \left( \frac{A_+ + 1 - 3h}{3\sqrt{1-h}\sqrt{h}} \right) - \frac{\pi \operatorname{sgn} h}{2} \right] \right\}, \quad (\text{A.32b}) \end{aligned}$$

$$\int_{-\infty}^0 [P'(y)]^2 dy = a_- \left\{ \frac{\sqrt{1+h}}{9h^2 [(1-3h)A_- + 2(1+3h) + A_-^{-1}]^3} \right.$$

$$\begin{aligned}
& \times \left[ (1-3h)^3(1+3h^2)A_-^3 + (1-3h)^2(1+3h)(5+3h+18h^2)A_-^2 \right. \\
& + 2(1-3h)(1+3h)^2(5-6h+9h^2)A_- + 2(1+3h)(1+3h^2)(5+3h+18h^2) \\
& \left. + (5+15h+117h^2+189h^3+162h^4)A_-^{-1} + (1-9h^2)A_-^{-2} \right] \\
& + \frac{1-9h^2}{27h^{5/2}} \left[ \operatorname{arctanh} \left( \frac{A_-(1-3h)+1+3h}{3\sqrt{1+h}\sqrt{h}} \right) - \operatorname{arctanh} \left( \frac{1+3h}{3\sqrt{1+h}\sqrt{h}} \right) \right] \Bigg\}. \tag{A.32c}
\end{aligned}$$

Although these expressions are too complicated to yield analytical insight, they can be computed readily. In the limit  $h \rightarrow 0$ , they simplify to

$$\int_0^\infty [P'(y)]^2 dy = \frac{6a_+(1+5A_+ - 5A_+^2 + 15A_+^3)}{5(1+A_+)^5}, \tag{A.33a}$$

$$\int_{-\infty}^0 [P'(y)]^2 dy = \frac{6a_-A_-^2(15-5A_- + 5A_-^2 + 15A_-^3)}{5(1+A_-)^5}. \tag{A.33b}$$

Employing the substitutions (3.5d), we obtain

$$\int_0^\infty [P'(y)]^2 dy = \frac{a_+}{15} [9 - (2-a_0)(1+2a_0)\sqrt{3}\sqrt{1+2a_0}], \tag{A.34a}$$

$$\int_{-\infty}^0 [P'(y)]^2 dy = \frac{a_-}{15} [9 - (1+a_0)(3-2a_0)\sqrt{3}\sqrt{3-2a_0}]. \tag{A.34b}$$

We emphasize that taking the limits  $h \rightarrow 1$  in (A.32b) or  $h \rightarrow -1$  in (A.32c) does not yield valid results. In fact, this is obvious from (3.5d) and (3.6).

For the rest of this subsection, we assume  $h = 0$ . To derive (4.56), we observe from (3.30b) that

$$a_+ = a_0\sqrt{3-2a_0}\sqrt{a_+^2 + a_-^2}. \tag{A.35}$$

Because (3.30b) is equivalent to

$$(1-a_0)^2(1+2a_0) = \frac{a_-^2}{a_+^2 + a_-^2}, \tag{A.36}$$

we obtain

$$a_- = (1-a_0)\sqrt{1+2a_0}\sqrt{a_+^2 + a_-^2}. \tag{A.37}$$

Substituting (A.35) and (A.37) into (A.34), we find (4.56).

There seems to be no simple representation of  $s(P)$  in (4.56) solely in terms of  $a_+$  and  $a_-$ . However, we found the following accurate approximation:

$$s(P) \approx \frac{6-2\sqrt{6}}{5} \sqrt{H(a_+^2, a_-^2)}, \tag{A.38}$$

where  $(6 - 2\sqrt{6})/5 \approx 0.220$  (results not shown). It is proportional to  $P'(0)$  in (3.16).

To compare the global and the local measure of steepness of a cline, we assume that  $\lambda$  and  $\alpha_+ + \alpha_-$ , or equivalently  $a_+^2 + a_-^2$ , are fixed but arbitrary. We already know that then both  $P'(0)$  and  $s(P)$  are maximized at  $a_+ = a_-$ , or  $a_0 = \frac{1}{2}$ , and are symmetric about this value. From (3.16) we have  $P'(0) = \frac{1}{\sqrt{6}}a_+$  if  $a_+ = a_-$ . Together with (4.57), this yields

$$\frac{P'(0)}{s(P)} = \frac{5}{12}(2 + \sqrt{6}) \approx 1.854. \quad (\text{A.39})$$

Furthermore, starting from (3.16) and (4.56), straightforward calculations yield

$$\lim_{\alpha_+ \rightarrow 0} \frac{P'(0)}{s(P)} = \frac{5}{23}(2 + 3\sqrt{3}) \approx 1.564 \quad (\text{A.40})$$

and (4.58).

## A.5 Steepness of two-locus clines

In the first subsection, we derive an explicit expression for the integral  $J$  in (4.61) for loci of equal effects without dominance and show that  $J > 0$ . We also compare the two measures of steepness of  $p_A$ ,  $p'_A(0) \approx P'(0) + \frac{1}{\rho}p'(0)$  and  $s(p_A) \approx s(P) + \frac{2}{\rho}J$ , and show that their ratio varies only between narrow bounds.

In the second subsection, we provide a method for numerical computation of  $J$  in the general case and apply it to support our conjecture that  $J$  is always positive.

### A.5.1 Properties of $J$ for loci of equal effects without dominance

We assume  $a_{\pm} = b_{\pm}$  and  $h_A = h_B = 0$ . We use partial integration and (3.2a) to obtain

$$\begin{aligned} J &= - \int_{-\infty}^{\infty} P''(x)p(x)dx \\ &= \lambda\alpha_+ \int_0^{\infty} P(x)(1 - P(x))p(x)dx - \lambda\alpha_- \int_{-\infty}^0 P(x)(1 - P(x))p(x)dx. \end{aligned} \quad (\text{A.41})$$

With the explicit expression for  $P(x)$  and  $p(x)$  given in Sects. 3 and 4, these integrals can be computed:

$$\begin{aligned} &\int_0^{\infty} P(x)(1 - P(x))p(x)dx \\ &= \frac{54(3 - 2a_0) - \sqrt{3}\sqrt{1 + 2a_0}(86 - 127a_0 + 114a_0^2 - 77a_0^3 + 22a_0^4)}{63(1 + 2a_0)} \end{aligned} \quad (\text{A.42})$$

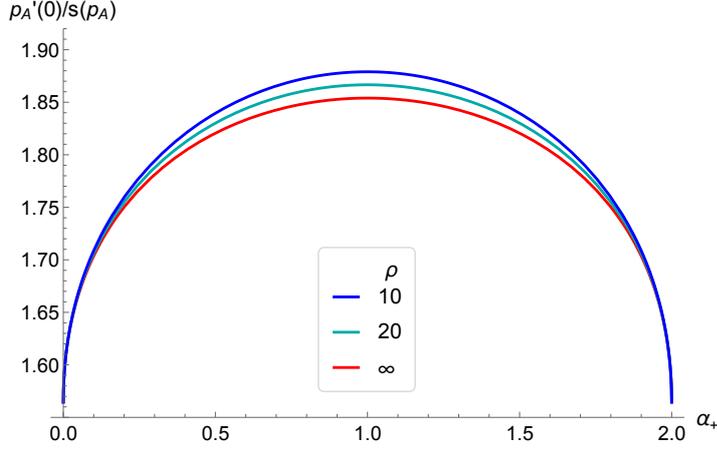


Figure A.1: Comparison of measures of steepness of a cline. This figure shows  $p'_A(0)/s(p_A)$  as a function of  $\alpha_+$  for  $\rho = \infty$  (independent loci),  $\rho = 20$ , and  $\rho = 10$ . The loci have equal effects,  $\lambda = 1$ , and  $\alpha_+ + \alpha_- = 2$ . The slope  $p'_A(0) \approx P'(0) + \frac{1}{\rho} p'(0)$  is computed from (3.16) and (A.46), and the steepness  $s(p_A)$  in (4.60) from (4.56) and (A.44).

and

$$\int_{-\infty}^0 P(x)(1-P(x))p(x)dx = \frac{54(1+2a_0) - \sqrt{3}\sqrt{3-2a_0}(18+42a_0+15a_0^2-11a_0^3+22a_0^4)}{63(3-2a_0)}. \quad (\text{A.43})$$

Straightforward algebra using (3.30b), (A.36), and  $A_0 = a_0^2(3-2a_0) = 1 - (1-a_0)^2(1+2a_0) = 1 - (1-A_0)$ , yields

$$J = \frac{\sqrt{2}}{21} \lambda^{3/2} \left( \frac{\alpha_+ + \alpha_-}{2} \right)^{3/2} \left[ 18(A_0^{3/2} + (1-A_0)^{3/2}) - \sqrt{3}\sqrt{3-2a_0}\sqrt{1+2a_0}(6-4a_0-19a_0^2+46a_0^3-23a_0^4) \right]. \quad (\text{A.44})$$

Obviously,  $J$  is proportional to  $\lambda^{3/2}$ . Moreover, it is easily shown that  $J > 0$  because  $\alpha_+ > 0$  and  $\alpha_- > 0$ . Indeed, as a function of  $a_0$ ,  $J$  is symmetric about  $a_0 = \frac{1}{2}$  and is maximized at  $a_0 = 1/2$ , i.e., for symmetric selection ( $\alpha_+ = \alpha_-$ ). Then it simplifies to

$$J = \frac{1}{56}(48 - 19\sqrt{6})(\lambda\alpha_+)^{3/2}, \quad (\text{A.45})$$

where  $\frac{1}{56}(48 - 19\sqrt{6}) \approx 0.0261$ .  $J$  is minimized and equals 0 if either  $a_0 = 0$  or  $a_0 = 1$ , i.e., if  $\alpha_+ = 0$  or  $\alpha_- = 0$ , respectively (see also the more general Figure A.2).

Now, we compare the two measures of steepness of  $p_A$ ,  $p'_A(0) \approx P'(0) + \frac{1}{\rho} p'(0)$  and  $s(p_A) \approx s(P) + \frac{2}{\rho} J$ , for two loosely linked loci.

We assume that the loci are equivalent. Then we obtain from the first equality in (4.53), by substituting (A.36) raised to the power 3/2,

$$p'(0) = \frac{\sqrt{3} \lambda^{3/2} H(\alpha_+, \alpha_-)^{3/2}}{\sqrt{8} (3 - 2a_0)(1 + 2a_0)}. \quad (\text{A.46})$$

For fixed  $\lambda$  and  $\alpha_+ + \alpha_-$ , and considered as a function of  $a_0$  (or  $\alpha_+$ ),  $p'(0)$  is symmetric about  $a_0 = \frac{1}{2}$  and maximized at  $a_0 = 0$ . There it assumes the value

$$p'(0) = \frac{1}{8} \frac{\sqrt{3}}{\sqrt{2}} (\lambda \alpha_+)^{3/2} \approx 0.153 (\lambda \alpha_+)^{3/2}, \quad (\text{A.47})$$

and it decays to zero as  $a_0 \rightarrow 0$  or  $a_0 \rightarrow 1$ . From (A.45) we find that at  $a_0 = \frac{1}{2}$ , or  $\alpha_+ = \alpha_-$ ,

$$\frac{p'(0)}{2J} \approx 2.937. \quad (\text{A.48})$$

However, we find that  $\lim_{\alpha_+ \rightarrow 0} \frac{p'(0)}{2J} = 0$ . Nevertheless, if  $10^{-5} \leq \alpha_+ \leq 1 - 10^{-5}$ , then

$$2.406 < \frac{p'(0)}{2J} < 2.937. \quad (\text{A.49})$$

Therefore, unless selection against one of the alleles essentially vanishes,  $p'(0)/(2J)$  varies only between quite narrow bounds.

Figure A.1 displays the ratio  $p'_A(0)/s(p_A)$  as a function of  $\alpha_+$  for given step size  $\alpha_+ + \alpha_-$  and three values of recombination. For the dependence of the slope of a cline on the recombination rate in its full range, we refer to Figure 9.

## A.6 Computation and properties of $J$ in the general case

To compute the steepness of two-locus clines in the general case, we use (4.60) and evaluate  $J$  as follows:

$$\begin{aligned} J &= \int_{-\infty}^{\infty} P'(x)p'(x) dx \\ &= P(+\infty)p'(+\infty) - P(-\infty)p'(-\infty) - \int_{-\infty}^{\infty} P(x)p''(x) dx \\ &= - \int_{-\infty}^{\infty} P(x)p''(x) dx, \end{aligned} \quad (\text{A.50})$$

where we used the boundary conditions (4.3). Since  $p''(x)$  satisfies (4.7), we obtain

$$\begin{aligned} - \int_0^{\infty} P(x)p''(x) dx &= a_+^2 \int_0^{\infty} [1 - 2P(x) + h_A(1 - 6P(x) + 6P(x)^2)]P(x)p(x) dx \\ &\quad + 2b_+^2 \int_0^{\infty} P(x)P'(x)Q'(x)(1 + h_B - 2h_BQ(x)) dx \end{aligned} \quad (\text{A.51a})$$

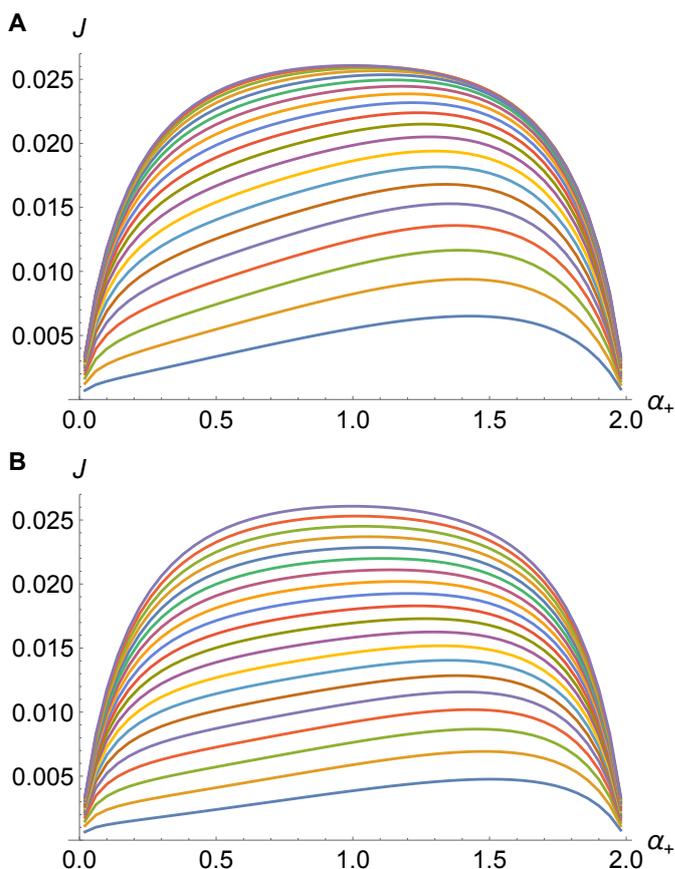


Figure A.2: The integral  $J$  in (4.61) as a function of various parameters. In both panels,  $J$  is shown as a function of  $\alpha_+$  for the constant step size  $\alpha_+ + \alpha_- = 2$  and different combinations of  $\beta_+$  and  $\beta_-$ . In **A**,  $\beta_+ + \beta_- = 2$  and  $\beta_+$  is varied from 0.05 to 1 in steps of 0.05 (curves from bottom to top). In **B**,  $\beta_- = 1$  and  $\beta_+$  is varied from 0.05 to 1 in steps of 0.05 (curves from bottom to top). In both panels,  $\lambda = 1$ , and there is no dominance,  $h_A = h_B = 0$ . Note that by (4.62), a change in  $\lambda$  affects all curves in the same way.

and

$$\begin{aligned}
 - \int_{-\infty}^0 P(x)p''(x) dx &= -a_-^2 \int_0^{\infty} [1 - 2P(x) + h_A(1 - 6P(x) + 6P(x)^2)]P(x)p(x) dx \\
 &\quad - 2b_-^2 \int_0^{\infty} P(x)P'(x)Q'(x)(1 + h_B - 2h_BQ(x)) dx. \quad (\text{A.51b})
 \end{aligned}$$

Both expressions can be computed numerically from the formulas given for  $P(x)$ ,  $P'(x)$ , and  $p(x)$  in the main text. We used *Mathematica* for these evaluations.

In general, we cannot prove that  $J$  is always positive. In the absence of dominance, multiplication of all four fitness effects ( $\alpha_{\pm}, \beta_{\pm}$ ) by the same positive constant  $c$  can be compensated for by multiplying  $\lambda$  by  $c$ . Therefore, (4.62) shows that, up to a

multiplicative factor,  $J$  depends only on three independent parameters. In Figure A.2, we fixed the step size  $\alpha_+ + \alpha_-$  and varied the other parameters as explained in the legend. The figure suggests convincingly that  $J$  is always positive. Thus, as expected from intuition, tighter linkage apparently always increases the measures  $s(p_A)$  and  $s(p_B)$  of steepness of the allele-frequency clines. Based on additional numerical computations for a wide variety of parameter combinations (not shown), we conjecture that  $J$  is always positive.

## References

- Aeschbacher, S., Bürger, R. 2014. The effect of linkage on establishment and survival of locally beneficial mutations. *Genetics* 197, 317-336.
- Akerman, A., Bürger, R. 2014. The consequences of gene flow for local adaptation and differentiation: a two-locus two-deme model. *J. Math. Biol.* 68, 1135-1198.
- Barton, N.H. 1983. Multilocus clines. *Evolution* 37, 454-471.
- Barton, N.H. 1986. The effects of linkage and density-dependent regulation on gene flow. *Heredity* 57, 415-426.
- Barton, N.H. 1999. Clines in polygenic traits. *Genet. Res. Camb.* 74, 223-236.
- Barton, N.H., Shpak, M. 2000. The effect of epistasis on the structure of hybrid zones. *Genet. Res. Camb.* 75, 179-198.
- Bürger, R. 2000. *The Mathematical Theory of Selection, Recombination, and Mutation*. Chichester: Wiley.
- Bürger, R. 2014. A survey of migration-selection models in population genetics. *Discrete Cont. Dyn. Syst. B* 19, 883 - 959.
- Bürger, R., Akerman, A. 2011. The effects of linkage and gene flow on local adaptation: A two-locus continent-island model. *Theor. Popul. Biol.*, 80, 272-288.
- Conley, C.C. 1975. An application of Wazewski's method to a non-linear boundary value problem which arises in population genetics. *J. Math. Biol.* 2, 241-249.
- Endler, J.A. 1977. *Geographic Variation, Speciation, and Clines*. Princeton Univ. Press, Princeton, New Jersey.
- Fife, P.C., Peletier, L.A. 1977. Nonlinear diffusion in population genetics. *Arch. Rat. Mech. Anal.* 64, 93-109.
- Fife, P.C., Peletier, L.A. 1981. Clines induced by variable selection and migration. *Proc. R. Soc. Lond. B* 214, 99-123.
- Fleming, W.H., 1975. A selection-migration model in population genetics. *J. Math. Biol.* 2, 219-233.
- Geroldinger, L., Bürger, R. 2015. Clines in quantitative traits: The role of migration patterns and selection scenarios. *Theor. Popul. Biol.* 90, 43-66.

- Gilbarg, D., Trudinger, N.S. 2001. Elliptic Partial Differential Equations of Second Order. Springer, Berlin.
- Haldane, J.B.S. 1948. The theory of a cline. *J. Genetics* 48, 277-284.
- Henry, D. 1981. Geometric Theory of Semilinear Parabolic Equations. In: Lecture Notes in Mathematics, vol. 840. Springer, Berlin.
- Hoffmann, A.A., Anderson A., Hallas R. 2002. Opposing clines for high and low temperature resistance in *Drosophila melanogaster*. *Ecology Letters* 5, 614-618.
- Kolmogoroff, A., Petrovsky, I., Piscounoff, N. 1937. Étude de l'équation de la diffusion avec croissance de la quantité de matière et son application à un problème biologique. *Bull. Univ. Etat Moscou, Ser. Int., Sect. A, Math. et Mecan.* 1, Fasc. 6, 1-25.
- Lenormand, T., 2002. Gene flow and the limits to natural selection. *Trends Ecol. Evol.* 17, 183-189.
- Lenormand, T., Guillemaud, T., Bourguet, D., Raymond, R. 1998. Evaluating gene flow using selected markers: A case study. *Genetics* 149, 1383-1392.
- Liang, S., Lou, Y., 2011. On the dependence of the population size on the dispersal rate. *Disc. Cont. Dynam. Sys. Series B* 17, 2771-2788.
- Lohman B.K., Berner D., Bolnick D.I. 2017. Clines arc through multivariate morphospace. *Amer. Nat.* 189, 354-367.
- Lou, Y., Nagylaki, T., 2002. A semilinear parabolic system for migration and selection in population genetics. *J. Differential Equations* 181, 388-418.
- Lou, Y., Nagylaki, T., 2004. Evolution of a semilinear parabolic system for migration and selection in population genetics. *J. Differential Equations* 204, 292-322.
- Lou, Y., Nagylaki, T., 2006. Evolution of a semilinear parabolic system for migration and selection without dominance.. *J. Differential Equations* 225, 624-665.
- Lou, Y., Nagylaki, T., Ni, W.-M. 2013. An introduction to migration-selection PDE models. *Disc. Cont. Dyn. Syst. A* 33, 4349-4373.
- Lou, Y., Ni, W.-M., Su, S. 2010. An indefinite nonlinear diffusion problem in population genetics, II: Stability and multiplicity. *Disc. Cont. Dyn. Syst. A* 27, 643-655.
- Mallet, J., Barton, N. 1989. Inference from clines stabilized by frequency-dependent selection. *Genetics* 122, 967-976.
- Nagylaki, T., 1975. Conditions for the existence of clines. *Genetics* 80, 595-615.
- Nagylaki, T., 1976. Clines with variable migration. *Genetics* 83, 867-886.
- Nagylaki, T., 1978. Clines with asymmetric migration. *Genetics* 88, 813-827.
- Nagylaki, T. 1989. The diffusion model for migration and selection. Pp. 55-75 in: Hastings, A. (Ed.), *Some Mathematical Questions in Biology*. Lecture Notes on Mathematics in the Life Sciences, vol. 20. American Mathematical Society, Providence, RI.
- Nagylaki, T., Lou, Y. 2008. The dynamics of migration-selection models. In: Friedman, A. (ed) *Tutorials in Mathematical Biosciences IV*. *Lect. Notes Math.* 1922, pp. 119

- 172. Berlin Heidelberg New York: Springer.

Nakashima, K., Ni, W.-M., Su, S. 2010. An indefinite nonlinear diffusion problem in population genetics, I: Existence and limiting profiles. *Disc. Cont. Dyn. Syst. A* 27, 617-641.

Slatkin, M. 1973. Gene flow and selection in a cline. *Genetics* 75, 773-756.

Slatkin, M. 1975. Gene flow and selection on a two-locus system. *Genetics* 81: 787802.