The shapes of clines and wavefronts

Stuart Baird[1] and Nina Daley[2]

[1]The Academy of Sciences of the Czech Republic
[2]Masaryk University

February 14, 2025

Abstract

Cline theory has a central place in speciation studies. Cline locations delimit taxon boundaries, cline widths scale with barrier strength, and the shapes of clines (smooth or stepped) suggest whether species barriers are mono- or polygenic. How cline shapes vary along chromosomes therefore forms part of the genome species barrier landscape. Further, asymmetric moving clines (wave fronts) can mark adaptive introgression puncturing species barriers, potentially leading to their collapse or decay. Here we review the development of cline and wavefront models and relate this to the use of dispersal kernels in epidemiology and ecology. We contrast classical results to those for a thick-tailed kernel, showing how cline shape affects the speed of spatial process, including the widening of neutral clines and the spatial coalescent. We critique current cline models used for inference (both spatial and genomic clines) and address Barton's question: Why (after decades of cline fitting) is there so little evidence of stepped clines? We suggest evidence is weak because stepped cline models are over-parameterised. We propose minimum parameter stepped cline models, and discuss non-parametric approaches, that may help resolve the issue. This broadens to a discussion of the future of, and alternatives to, cline fitting.

The shapes of clines and wavefronts

Stuart J.E. Baird[1]
Nina Daley[2]

1- Corresponding author. StuartJ.E.Baird@gmail.com
Institute of Vertebrate Biology, Czech Academy of Sciences, Brno, Czech Republic
2- RECETOX, Masaryk University, Brno, Czech Republic

Abstract

Cline theory has a central place in speciation studies. Cline locations delimit taxon boundaries, cline widths scale with barrier strength, and the shapes of clines (smooth or stepped) suggest whether species barriers are mono- or polygenic. How cline shapes vary along chromosomes therefore forms part of the genome species barrier landscape. Further, asymmetric moving clines (wave fronts) can mark adaptive introgression puncturing species barriers, potentially leading to their collapse or decay. Here we review the development of cline and wavefront models and relate this to the use of dispersal kernels in epidemiology and ecology. We contrast classical results to those for a thick-tailed kernel, showing how cline shape affects the speed of spatial process, including the widening of neutral clines and the spatial coalescent. We critique current cline models used for inference (both spatial and genomic clines) and address Barton's question: Why (after decades of cline fitting) is there so little evidence of stepped clines? We suggest evidence is weak because stepped cline models are over-parameterised. We propose minimum parameter stepped cline models, and discuss non-parametric approaches, that may help resolve the issue. This broadens to a discussion of the future of, and alternatives to, cline fitting.

Introduction

> "Those forms which possess in some considerable degree the character of species, but which are so closely similar to some other forms, or are so closely linked to them by intermediate gradations, that naturalists do not like to rank them as distinct species are in several respects the most important to us." (Darwin 1859)

> "Although the early writers thought a good deal about the effects of geography and dispersal […], intense geographic differentiation and speciation was thought to require some kind of island or complete isolation situation. Only in the last few years have population geneticists become seriously concerned with the effects of gene flow in continuous populations." Endler ((Endler 1977), p3)

> "How many genomic regions differentiate during speciation? How small are regions where divergence significantly exceeds the genomic average ([…])? How are regions of exceptional divergence dispersed around the genome? We suggest that recent discussions of these issues in the context of ecological speciation would benefit from closer attention to well-established cline theory." (Abbott, Albach et al. 2013)

The first mathematical treatment of migration and selection in continuous populations was by R.A. Fisher (Fisher 1937), who studied the wave of advance of favourable genes. The next year Huxley coined the term 'cline' (Huxley 1938), and a decade later, by allowing selection coefficients to depend on location, Haldane (Haldane 1948) developed an equilibrium model of gene flow and selection in a cline, closely related to Fisher's model (Nagylaki 1975), and used it to estimate the intensity of natural selection in deer mice. Thus wavefronts, clines, and inference from them, lie at the roots of the modern evolutionary synthesis, and here we will use cline theory as a catch-all for this body of work. While (Nagylaki 1975) was setting out to "relate clearly the theory of clines to the diffusion methods [...] which have been very productive in population genetics", tropical ecologists were formulating a very different treatment of migration and selection: the Janzen-Connell model (See(Terborgh 2020)) arose from the empirical observation that seedfall is most concentrated around fruiting trees, whereas sapling recruitment fails close to parent trees and succeeds at a distance. These and analogous empirical observations inspired the use of dispersal kernels to summarise dispersal probabilities in ecology (Nathan, Klein et al. 2012) epidemiology (Pybus, Suchard et al. 2012) and invasion biology (Kot, Lewis et al. 1996, Lindström, Håkansson et al. 2011).

It would seem then that dispersal kernels should also appear at the heart of cline theory, but this was not to be. This is because the diffusion method made famous by Einstein's description of Brownian motion (Einstein 1905), has an *implicit* dispersal kernel: the Normal distribution. The first part of Einstein's argument was to determine how far a Brownian particle travels in a given time interval. He found the density of particles at a given time satisfies a diffusion equation, the solution of which is the Normal distribution, a stable distribution which widens over time, changing scale without changing shape (Figure1a). In this way Einstein demonstrated that the displacement of a Brownian particle increases with the square root of time (Einstein 1905). The mission of Nagylaki and others to "relate clearly the theory of clines to the diffusion methods" (Nagylaki 1975) tied early spatial genetics to the implicit Normal dispersal kernel of the Brownian particle at the same time as ecologists were confirming early suggestions (e.g. (Bateman 1950)) that biological dispersal often differed from that of a Brownian particle, with thicker tails than Normal (leptokurtotic). Despite discussion of a diversity of approaches (e.g. (Diekmann 1978)), the parent-offspring displacement distribution, keystone of spatial population genetics, was almost always described only in terms of a variance (a proxy for scale, explored later) because it was implicitly assumed to only ever have one shape: the Brownian Normal (Bateman 1950, Nathan, Klein et al. 2012). But Normal is just one dispersal kernel – other kernel shapes are possible, and kernel shape makes a difference.

Neutral clines

The effect of dispersal kernels on cline shape can be illustrated in the neutral case by contrasting two stable distributions for parent-offspring displacement: Normal vs Cauchy (with thicker tails), and the clines they generate (Figure 1). As with the displacement of a Brownian particle, the width of a Brownian cline increases with the square root of time (Figure 1a,b). In contrast, the width of clines for the thick-tailed Cauchy kernel increases faster – linearly with time (Figure 1c,d). If we were going to infer time since neutral contact between populations from the widths of clines in traits, we could reduce our error by co-estimating the shape of the clines. If we were going to infer

the existence of a polygenic species barrier from stepped clines, we should be aware that a thick-tailed dispersal kernel can give stepped (thick tailed) clines even in the neutral case (Figure 1d). But we are getting ahead of ourselves: Figure 1 shows neutral clines which are CDFs of stable location dispersal kernels, and the only selection mentioned so far is that of Fisher's wave of advance of an advantageous gene. While that work gave rise to its own entire field of endeavour (Invasion Biology, (Skellam 1951, Kot, Lewis et al. 1996, Lindström, Håkansson et al. 2011, Phillips 2015)), it is selection against admixture which has been key to understanding the body of cline theory most entangled with speciation: hybrid zone clines.

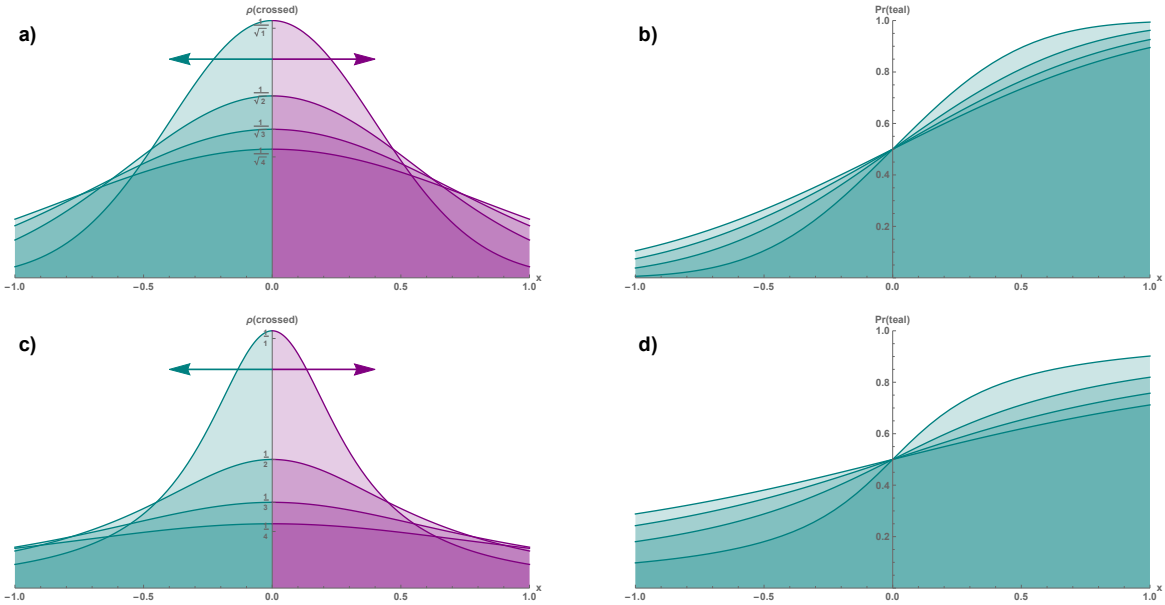


Figure 1: The neutral contact case for stable dispersal kernels over four discrete generations: Two populations (purple arriving from x -ve, teal from x +ve) meet centrally and (left panes) their alleles spread across the contact following a dispersal location kernel. Top panes: When the kernel is (a) the Normal distribution PDF, the expected neutral cline is (b) the 'sigmoid' Normal Distribution CDF. The Normal PDF (a) can be convoluted over generations to give the characteristic widening bell-shaped curves of the diffusion of Brownian particles. The Brownian cline width (b) increases with the square root of time, as with the displacement of a Brownian particle (Einstein 1905). Bottom panes: When the kernel is (c) the thick-tailed Cauchy distribution PDF, the expected neutral cline is (d) the 'stepped' Cauchy Distribution CDF. The Cauchy PDF (c) can also be convoluted over generations to give widening leptokurtotic curves. The neutral Cauchy kernel cline width (d) increases linearly with time, faster than Normal. All panes: initial PDFs are scaled (see Table 1) such that CDF clines have unit width at unit time. (Right panes) Cline widths are the inverse of the central gradients of the CDFs, and so by definition equal to the denominators of the central values of the PDFs (left panes). Only rising (teal) clines are shown; the complementary (purple) falling clines are redundant, and omitted for clarity.

Glossary (Concepts in approximate order of first occurrences, as underlined in the text)

PDF: Probability density function

CDF: Cumulative distribution function. In 1D: convolution of a step function with a PDF.

Dispersal kernel: The contribution of a specific core (source point, parent, parental copy) to the re-organisation of certain units (e.g. offspring) in a larger entity (population). See (Nathan, Klein et al. 2012) for history and proposed usage in ecology and evolution. **Herein 'dispersal kernel', unless otherwise stated, will mean dispersal location kernel.**
Dispersal location kernel: The PDF for the end location of a dispersal vector. In an *n*-dimensional field area this is a *n*-dimensional PDF (See Box 1, first pane). The Normal dispersal location kernel has a zero-centred bell shape in any dimension.
Dispersal distance kernel: The PDF for distance covered by a dispersal vector. (See Box 1, second pane). No information is lost when summarising a radially symmetrical dispersal *location* kernel in any dimension as a dispersal *distance* kernel of one dimension. The Normal dispersal distance kernel is a half bell shape in 1D, but this shape changes as the mode shifts away from zero in higher dimensions.
Effective dispersal kernel: the dispersal location kernel as observed after the effects of selection.

Shape of a distribution: All moments of a distribution other than the first two, location and scale. This leaves skewness (asymmetry), kurtosis (thick-tailed-ness), and further moments with increasingly subtle descriptions. Kurtosis exceeding that of the Normal distribution is leptokurtosis.

Cline width: The inverse of the maximum gradient of a smooth change in trait.
Cline centre: The turning point of maximum gradient of a smooth change in trait.
Shape of a cline: By analogy with the shape of a distribution, all aspects of a cline other than centre and width. In particular asymmetry and stepped-ness (leptokurtotic kernel distributions give rise to stepped clines, figure 1c,d).
Unit cline: A zero-centred cline of unit width (unit central gradient). Distinct unit clines differ only in shape. Any unit cline $U(x)$ can be {recentered,rescaled} to any {centre,width} $\{c, w\}$ as $U\left(\frac{x-c}{w}\right)$. See unit time clines in Figure 1, black and grey zone clines in Box 1.
MAD: Maximum absolute difference: a comparison of two functions across a set of points. Here, unless otherwise stated, unit clines are compared for points $-4 \leq x \leq +4;\ \Delta x = 0.01$, i.e. over 8 cline widths.

Stable distributions: A distribution is stable if a linear combination of two independent random variables with this distribution has a distribution of the same shape, i.e. differing at most in location and scale parameters.
Convolution of a distribution: A linear combination of *n* independent random variables with this distribution has a distribution described by its *n*-fold convolution. The 2-fold convolution of 1D continuous dispersal kernel $k(\cdot)$ is $k_2(x) = \int k(z).k(x-z)\,dz$ (note that for all values of z the sum of the arguments of k is x). Only stable distributions do not change shape under convolution. Only a small proportion of distributions have analytic solution under convolution, however *convolution is simple to approximate to arbitrary accuracy by simulating (large) arrays of random variates and adding them*.

Exogenous/endogeneous (extrinsic/intrinsic) selection: The cause of exogeneous selection is tied to the environment (e.g. taxa or genes are adapted to different regions or niches). Endogeneous selection is caused by genome interactions independent of environment. See (Kruuk, Baird et al. 1999) on expectations for clines maintained by endogeneous vs exogeneous selection.
Tension zone: A hybrid zone maintained by endogeneous selection, and thus free to move across the environment. Tension zones move down density gradients to become trapped in density troughs or at physical barriers to geneflow (Barton 1979).

Indirect selection: Change in allele frequency at one locus due to selection acting at another locus (or loci) in statistical association (linkage disequilibrium). As admixture generates linkage disequilibrium genome-wide, *indirect selection will affect unlinked loci*. Calling indirect selection 'linked selection' is therefore unnecessarily confusing (Stankowski, Chase et al. 2019).

Context	PDF Scale	UnitCline CDF; $\{c, w\} = \{0, 1\}$ $p = U(x)$	Inverse UnitCline $x = U^{-1}(p)$	Cline names $T(\cdot)$
Dispersal kernel	Normal $\sqrt{2\pi}$	$Erfc[-\pi x]/2$	$-Erfc^{-1}[2p]/\sqrt{\pi}$	Brownian $T(\infty)$ $\approx T(116)$ *MAD* 0.001
Thick-tailed Dispersal kernel	Cauchy π	$\frac{1}{2} + Tan^{-1}[\pi x]/\pi$	$Tan[\pi(p - 1/2)]/\pi$	Cauchy $T(1)$
Selection against admixture of a single non-recombining genome region	Logistic 4	$\frac{1}{1 + e^{4x}}$ $\frac{1}{2} + Tanh[2x]/2$	$Log\left(\frac{p}{1-p}\right)/4$ $Tanh^{-1}[1 - 2p]/2$	Bazykin $\approx T(6.67)$ *MAD* 0.001
Viability selection against admixture of a single non-recombining genome region	Student-*t* (shape 2) $2\sqrt{2}$	$\frac{1}{2} + \frac{x}{\sqrt{1 + 4x^2}}$	$Sgn[p - ½]\frac{\sqrt{4p - 1 - 4p^2}}{4\sqrt{p^2 - p}}$	Gavrilets (stepped extreme) $T(2)$
model of symmetric clines	Student-*t* (shape v) $\beta\left[\frac{v}{2}, \frac{1}{2}\right]\sqrt{v}$	$\frac{1}{2}I_{z(x)}\left[\frac{v}{2}, \frac{1}{2}\right]; \quad x < 0;$ $1 - \frac{1}{2}I_{z(x)}\left[\frac{v}{2}, \frac{1}{2}\right]; \quad x \geq 0;$ $z(x) = \frac{1}{1 + x^2\beta\left[\frac{v}{2}, \frac{1}{2}\right]}$	$z^{-1}\left(I^{-1}\left[2p, \frac{v}{2}, \frac{1}{2}\right]\right); \quad p < \frac{1}{2};$ $z^{-1}\left(I^{-1}\left[2(1-p), \frac{v}{2}, \frac{1}{2}\right]\right); \quad p \geq \frac{1}{2};$ $z^{-1}(\omega) = \frac{\sqrt{1-\omega}}{\beta\left[\frac{v}{2}, \frac{1}{2}\right]\sqrt{\omega}}$	*T*-cline $T(v)$

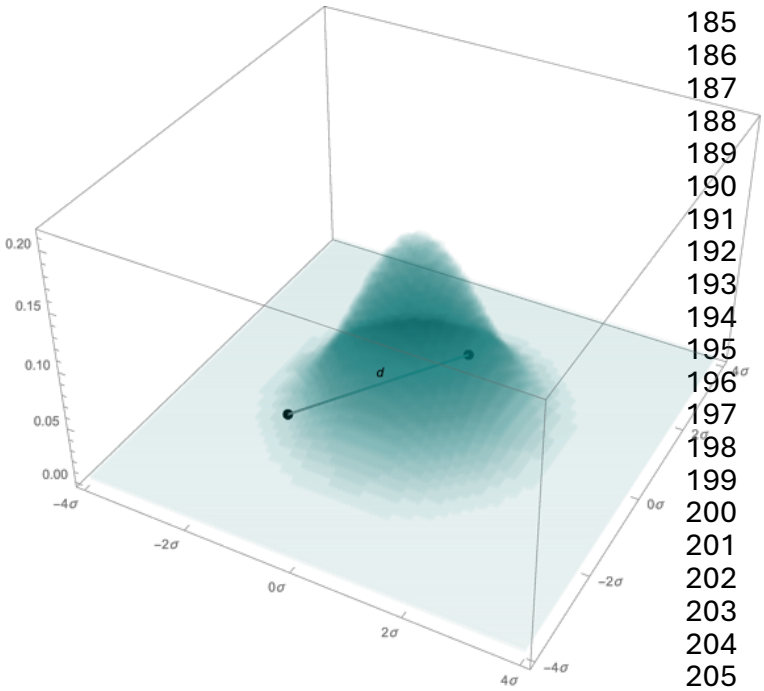
Table 1: Symmetric $UnitClines$ and their inverses. 'Scale': For distribution φ, $CDF[\varphi(Scale), x] = UnitCline_{\varphi}[x]$. Blue: alternate forms. *MAD* Maximum absolute difference (between a $UnitCline$ and the closest *T*-cline). Key to functions: $\{Erfc, Erfc^{-1}\}$ the complementary error function and its inverse; Tan tangent; Log natural logarithm; $Tanh$ Hyperbolic tangent; $Sgn[x]$ the sign function; $\{I_z, I^{-1}\}$ the incomplete beta function and its inverse$\left(I^{-1}[\ I_z[a, b]\ , a, b]\ = z, \quad z^{-1}(z(x)) = x\right)$; β the beta function.

Selection

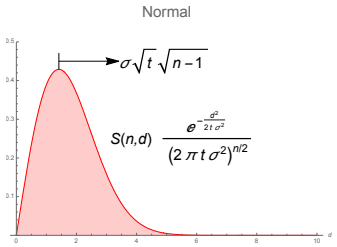
Can selection against genome admixture arrest the ever-widening neutral clines of Figure 1(b,d) to produce a stable 'gradation' between populations of the sort Darwin envisaged in the opening quote? For the Normal dispersal kernel, diploid genomes, no recombination, and selection against F1s, Bazykin showed the answer is yes (Bazykin 1969). Bazykin's selection-stabilised clines can also be described in terms of free recombination with selection against heterozygotes at one single locus. This might seem a simpler description, but it must be caveated that no indirect selection is acting. Because admixture generates genome-wide associations across loci (Baird 2015), potential sources of indirect selection could lie anywhere in the genome, and so the Bazykin cline, when framed in terms of heterozygotes, is an expectation for the causal locus of a monogenic species barrier. In both these framing of the Bazykin result, only one non-recombining genome region is under selection, and only three genome types are distinguished: two pure and one eqi-admixed (a heterozygous diploid locus or an entire F1 individual's genome). Selection against the admixed type arrests cline spread and distorts the equilibrium cline shape from that of the neutral Normal CDF to that of the CDF of the logistic distribution (See Table 1). Despite this change in shape both Normal and logistic clines are called sigmoid. The Bazykin result applies at two recombination extremes, and it is useful to bear both in mind, for while a species barrier maintained by selection against heterozygotes at a single locus may sound infeasible, one maintained by selection against F1 individuals, genome wide, may not. At the F1 extreme the potential for indirect selection is maximal because *cis* associations between all genes in diverged genomes are maintained, while every site of the genome is heterozygous by source. Then, even in the presence of recombination, strong selection against F1-like individuals will reduce the effective recombination rate, slowing the decay of pure *cis*-genome associations, and increasing the potential for indirect selection at multiple loci to further distort cline shape. The paths of these initial changes in multilocus cline shape, dependent on both hybrid indices and degree of heterozygosity by source, are spatially explicit cousins of the paths taken through Fisher's geometric fitness space during admixture (Simon, Bierne et al. 2018). Returning to the no-recombination case, Bazykin not only demonstrated a spatial dispersal-selection equilibrium existed, but also calculated, to a weak selection approximation, the expected width of his clines as $\frac{2\sigma}{\sqrt{s}}$, s being the selection acting against heterozygotes and σ, the scale of per generation dispersal (see Box 1). With hindsight, and on considering the effects of kernel shape, we see the expected width of a cline maintained by Bazykin selection will depend not only on the selection acting, but also on the dispersal kernel shape.

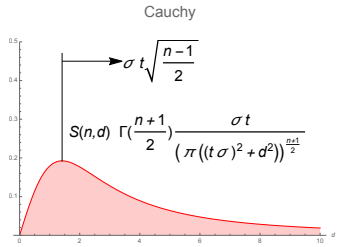
While (Nagylaki 1975) showed that Haldane's (Haldane 1948) exogeneous cline could be recast as a special case of a Fisher wavefront, it was Barton ((Barton 1979)a, section 3: (i)) who pointed out that there is a family of solutions between Bazykin's endogeneous symmetric cline and Fisher's (endogeneous) asymmetric travelling wave, depending on whether two pure genome types are equally fitter than their admixed type. Equality gives the symmetric Bazykin cline, inequality, traveling clines moving toward the less fit pure type. Barton (Barton 1979)b also showed that selection on multiple loci (a polygenic barrier) distorts cline shape through indirect selection, steepening the cline centre relative to its tails to produce stepped clines. In fact, in those two seminal 1979 papers

Box 1: Scale, speed and neighbourhood size. **Scale** What are the natural units of cline width? Most biologists would suspect a 1 km wide hybrid zone (HZ) is narrow for birds, but wide for snails, i.e. a cline is not narrow or wide based only on its width in SI units. Instead we use units specific to the study organism: the per generation scale of dispersal σ (a length, smaller for snails, larger for birds). Setting the origin of a frame of reference at a parent, a zero-centered radially symmetrical multivariate PDF can be used to define a dispersal location kernel K with scale σ for offspring: Here, the field area has $n = 2$ dimensions (e.g. {x,y}) measured in σ, and two offspring of the same parent are joined by a line length d. The 1 km HZ is narrow for birds if their σ is >>1 km, and conversely wide for snails if their σ is << 1km. If we measured bird and snail HZs in 'natural' σ units, and they had similar histories, governed by similar processes, then they would have similar width. This is how cline theory 'scales' over diverse study systems. The distance d between two offspring of the same parent follows the dispersal distance kernel found by sampling two location vectors from K and adding them to derive K_2, the 2-fold convolution of K. **Speed** For the stable distributions in Figure 1, the t-fold convolution can be expressed for any dimension n of field area:



Here the PDFs are $S(n,d)K_t(n,d)$ for the respective kernels, and are plotted for $n = t = 2$ where $S(n,d)$ is the surface area of the n-sphere radius d. The position of the modal distance (black verticals) shows us how the spatial scale of a chain of inheritance increases over time: linearly for the Cauchy kernel, but with the root of time for the Normal kernel, generalizing the (Figure 1) observation that different kernel shapes give different speeds of process, from the widening of clines in 1D, to spatial inheritance in any dimension of field area. This has consequences for the rate of coalescence. **Neighbourhood size** As the probability of pairwise coalescence in the previous generation, in an idealised population without spatial context, is $\frac{1}{2N}$, so in an idealized spatial population, for the two offspring above, this coalescence probability is $\frac{1}{\rho} K_2(n,d)$. N is population size, ρ population density (per spatial extent), and $K_2(n,d)$ expresses how pairwise parent sharing probability falls off with distance d in n dimensions. We can remove the dependance on distance by letting d tend to zero (offspring found in contact). Then, for a 2D field area and Normal dispersal, this coalescence probability is $\frac{1}{4\pi\rho\sigma^2}$, the inverse of Wright's neighbourhood size. For a 2D field area and Cauchy dispersal, the inverse of the neighbourhood size is $\frac{1}{8\pi\rho\sigma^2}$; the rate of coalescence is halved by changing the shape. Neighbourhood size is affected by both the dimension and shape of the dispersal kernels as $2^n\pi^{n/2}\rho\,\sigma^n$ and $2^n\pi^{(n+1)/2}\rho\,\sigma^n\,\Gamma\left(\frac{1+n}{2}\right)$ respectively. Note 1: The scale σ is always raised to the dimension of spatial extent, whereas the density is always per spatial extent, leaving neighbourhood size a non-spatial quantity commensurate with population size N. Note 2: The Normal distribution is special in that its variance is equal to its scale squared: σ^2. Where other dispersal kernels are considered, the variance is no longer a useful proxy for scale.





Barton laid out the properties of spatial genome admixture as we understand it today, and went on to show, with Hewitt (Barton and Hewitt 1985), that the majority of hybrid zones had cline widths too narrow to be consistent with maintenance at the scale of environmental change (as explored by Haldane), and so instead were likely to be tension zones maintained by endogeneous barriers. Here we see how deeply species and cline concepts can become entangled, as it is tempting to define species as only those taxa kept distinct by intrinsic barriers (all other barriers being context dependent, and therefore potentially ephemeral). Arguments for a more operational taxonomy (e.g. (Mallet 1995)) have largely fallen on deaf ears, with one exception: It appears acceptable to describe multilocus endogeneous barriers as species barriers irrespective of how taxonomists rank the organisms on either side (see (Krieber and Rose 1986) e.g. *Mus* 'subspecies' (Albrechtová, Albrecht et al. 2012)).

With (Barton 1979)a came the possibility of inferring the existence of a multilocus species barrier (polygenic selection) by deciding whether or not an observed cline had a steepened central portion. (Barton and Bengtsson 1986) developed a continuous explicit model of stepped cline shape, and (Kruuk, Baird et al. 1999) went on to show this stepped shape was similar whether multilocus selection was endogeneous or exogeneous. The Kruuk result is not for one cline shape, but rather a continuum of multilocus stepped shapes with the Bazykin shape at the single locus limit. In the meantime, (Gavrilets 1997) had shown how viability selection, even acting on a single locus, can also result in stepped cline shapes. As with Kruuk, the Gavrilets result is for a continuum of stepped shapes, Bazykin-shaped at one extreme of the selection model, but with the CDF of the Student's-*t* distribution (with 2 degrees of freedom, or shape parameter 2) at the other extreme ((Gavrilets 1997) Equation 14a). I will use $T(2)$ to refer to the unit cline with the shape of this CDF (see Table 1).

All of these selection cline results assume the Normal dispersal kernel implicit in the diffusion method. Together with the previous section we now see that stepped cline shapes can be expected with or without selection (given variation in dispersal kernels), and with or without multiple loci (for Normal dispersal kernels, given a diversity of selection regimes). What then can we hope to infer if we observe a stepped cline shape? Further: Neither the Kruuk nor the Gavrilets stepped result is even expressed as a cline function that could be fitted to data – this is why they have no entries in Table 1. The Kruuk result ((Kruuk, Baird et al. 1999) eq. 14) is in the form of an ordinary differential equation parameterised by a coupling coefficient $\phi = (L-1)\frac{s}{r}$ summed over the joins between L loci. Instances of the equation can be numerically solved using e.g. the NDSolve tool in Mathematica (Wolfram Research 2019). The Gavrilets result ((Gavrilets 1997), Eq 14b) takes the form of the inverse function of a cline $g^{-1}(C,p) = x$ which, unlike the functions in Table 1, itself has no obvious inverse. Instead Gavrilets clines can be numerically approximated by tabulating $\{g^{-1}(C,p),p\}$ over values of p and interpolating. When we talk of stepped clines then, what is our model?

Cline models and inference

While Haldane himself said he lacked sufficient data to support estimates of cline parameters for deer mice, he dedicated a large section of his discussion on how data hungry such estimates are (Haldane 1948). An exemplary field sampling effort and allozyme allele counting allowed estimates of the parameters of a stepped cline between *Bombina* subspecies (Szymura and Barton 1986, Szymura and Barton 1991), a model-based analysis that set the paradigm for hybrid zone inference software for decades (Analyse, (Baird and Barton 1995), ClineFit (Porter, Wenger et al. 1997), Cfit (Gay, Crochet et al. 2008), HZAR (Derryberry, Derryberry et al. 2014)). That original stepped cline model is a tri-partite composite (see Figure 2) of a 'sigmoid' (logistic) central portion joined to exponential tails (Szymura and Barton 1991). There are four shape parameters corresponding to a barrier strength and a tail decay rate in either direction. This allows for both cline asymmetry and central steepening using parameters that have direct interpretation for evolutionary process. Further, likelihood comparison with simpler nested models (where parts of the parameter vector are fixed) allows powerful likelihood ratio tests for asymmetry and stepped-ness (e.g.(Macholán, Munclinger et al. 2007)). The inference framework built around the tripartite model allowed stepped clines to be identified in several further field systems, *Podisma* (Barton and Gale 1993),*Pontia* (Porter, Wenger et al. 1997), *Mus* (Macholán, Munclinger et al. 2007)), but perhaps not as many as expected under null (multilocus, polygenic) models of speciation (Barton and Charlesworth 1984). Thirty years after demonstrating the *Bombina* zone was stepped, Barton commented on the paucity of further examples "This may be because dense spatial sampling is needed to identify a step, but more likely is because the genetic map is typically long enough that selection does not often maintain a strong barrier." (Barton 2020). The first potential explanation is the data hunger noted by Haldane, the second refers to the balance between selection against admixture, and recombination, which admixes genomes (hence the coupling coefficient of the Kruuk result). Recombination breaks down the linkage disequilibrium generated by admixture and/or epistasis, weakening indirect selection and opposing epistatic selection. A given amount of selection against admixture might then be overwhelmed by a long genetic map (high recombination). While it is clear how this applies when selection and recombination are each described by a single parameter (Barton 1983, Baird 1995, Kruuk, Baird et al. 1999), it is less clear when both selection and recombination densities vary along the genome (Martin, Davey et al. 2019, Stankowski, Chase et al. 2019). Here instead we explore a third potential explanation for the paucity of stepped cline observations. Sampling data is not the only thing needed to identify a step: one also needs the model of what a stepped cline is, and how that differs from a non-stepped 'sigmoid' cline. The more free parameters this model has, and the greater the distance between the model and the process generating the observations, the less power available to infer a step. The flexibility of the tripartite stepped model is in natural trade-off against both high parameter numbers and distance from 'reality'. It has four free parameters for asymmetric stepped shape whereas the analogous distribution-shaping moments are only two: skewness and kurtosis; it also has two discontinuities where its parts join, corners that are not a feature of the underlying expectations (Fitzpatrick 2013) (Contrast the 2-parameter *T*-cline fit to a tripartite fit in Figure 2); further, as the 'sigmoid' central portion of the tripartite cline is the logistic CDF for selection against admixture, the

neutral shape of the 'sigmoid' Brownian cline is not strictly nested within the tripartite model. This might sound merely a technical issue, but modelling the neutral case using the Bazykin cline, but no selection, returns a flat line of $p = 1/2$ everywhere. This is because the Bazykin result is for equilibrium, and the equilibrium for no selection against admixture is (eventual) infinite spread.

Stepped clines with fewer parameters?

We have shown a relationship between dispersal kernels (PDFs) and clines (CDFs). If probability distribution leptokurtosis and skew can be expressed as two moments, perhaps cline stepped-ness and asymmetry can be expressed with just two parameters?

T-clines Each cline result touched on to this point has been linked to the CDF of a probability distribution modelling either a dispersal kernel, in the neutral case, or for the Bazykin case, an equilibrium post-selection effective dispersal kernel (The exception being the asymmetric Fisher wave). Three of these CDFs are unified within the Students-*t* distribution: The neutral stepped Cauchy cline of Figure (1) is shape $T(1)$, the Gavrilets single locus stepped extreme is shape $T(2)$ and the sigmoid Brownian cline shape is shape $T(\infty)$. This suggests the continuous shape parameter υ of the Student's-*t* distribution (whose whole numbers correspond to degrees of freedom), a parameter which smoothly alters PDF kurtosis, could be used as a stepped-ness parameter for a continuum of CDF cline shapes $T(\upsilon)$ from Brownian neutral 'sigmoid' to extreme stepped $T(0 < \upsilon < 1)$. Further, the non-central Student's-*t* distribution can be re-expressed (For this and other mathematical details, see the Supplementary Material) with an asymmetry parameter α, giving a plane of cline shapes $T(\upsilon, \alpha)$ from Brownian neutral to stepped and from left biased ($\alpha < 0$) through symmetric ($\alpha = 0$) to right biased ($\alpha > 0$).

This T-cline model reduces the four tripartite cline shape parameters to two, has no discontinuities and includes both the neutral case and one extreme of the Gavrilets continuum. $T(\upsilon, \alpha)$ as a model of the Fisher wave is shown in Figure 2: the maximum absolute difference (MAD) between $T(8.365, 1.367)$, for any value in (Fisher 1937) table IV is <0.001. This is in contrast to the best fit tripartite model at MAD 0.018. All these features of the T-cline as a shape model are encouraging, however the tripartite model's potential to estimate distinct barrier strengths in each direction has been sacrificed, the biological interpretation of the $T(\upsilon, \alpha)$ parameters (away from matches to existing cline theory) is approximate or unclear, and the relationships between the symmetric $T(\upsilon)$ model and the Kruuk and Gavrilets step continua remain unexplored.

The Students-*t* is not the only candidate distribution for expressing stepped-ness and asymmetry of clines; the generalised Logistic with parameters (1,1) matches logistic shape, where the Students-*t* has no exact match. The Kruuk, Gavrilets shape continua results start with that shape at their sigmoid extremes; perhaps one or the other follows the shape of the generalised logistic as they become more stepped? To explore such possibilities and decide between shape models it seems best to construct a shape space within which all the clines discussed thus far can be compared. Previous cline shape comparisons (Barton and Gale 1993, Gavrilets 1997) have focussed on zero-centred width-rescaled clines as here ($UnitClines$, though Gavrilets chose to rescale to half

width), then Logit transformed. The Logit function is the inverse of the Logistic CDF, and so the transformed Logistic (Bazykin) cline is linear, other shapes deviating from linearity (Figure 2.2, (Barton and Gale 1993)). However, these deviations remain difficult to interpret, and the justification for comparing all other cline shapes to Bazykin is weakened when (i) we see other clines also have simple inverses (Table 1) and (ii) we remember the special (Normal kernel) nature of the Bazykin result. Instead, here we seek a shape comparison framework 'outside' of all the cline shapes we wish to compare.

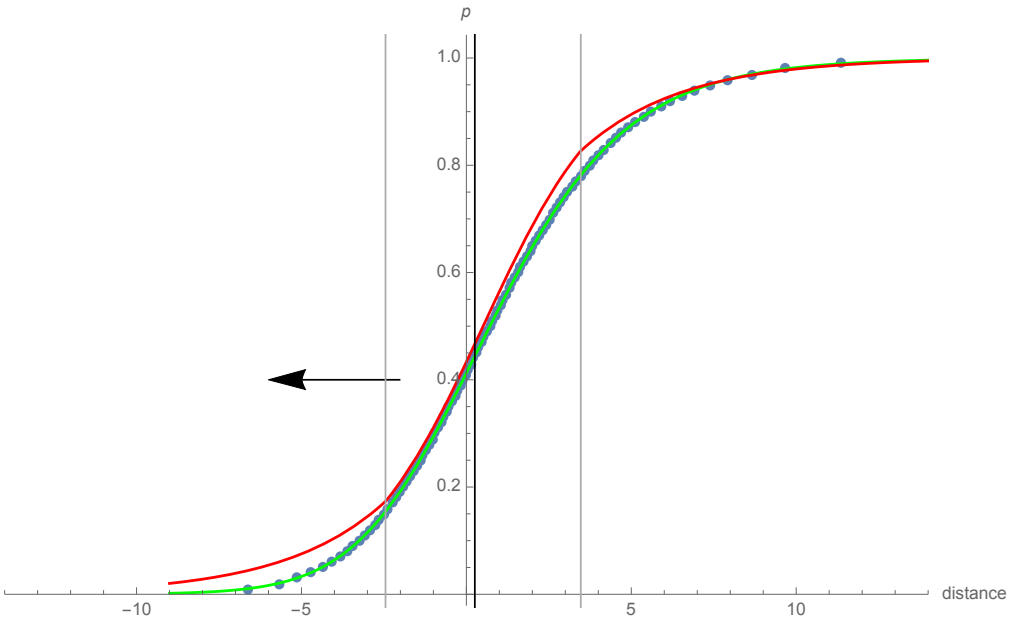


Figure 2: Fitting models to the shape of Fisher's wave of advance. Blue points: Fisher tabulated values for the wave of advance in Table IV, (Fisher 1937). This is an asymmetric cline of {centre,width} $\{c, w\} = \{0.256, 8.110\}$ as estimated from Fisher's table. The centre is marked with a black vertical line. Green line: Degrees of freedom 8.365 and non-centrality 1.367, parameters of the Student-*t* distribution PDF, can be used respectively to shape stepped-ness and asymmetry of the cline $T\left(\frac{x-c}{w}, 8.365, 1.367\right)$ that differs nowhere by more than 0.001 from the Fisher values (MAD 0.001). Red line: A tripartite cline fit to the Fisher data. Vertical grey lines mark joins between the central logistic part and exponential tails. The right hand join is shown for the best fit, the best fit left hand join falls further to the left than shown (out of frame), but least squares fitting with $\{c, w\}$ and all four shape parameters (two for each tail) free to vary only achieves MAD 0.018, due to a poor match at the best fit 'corner' marked by the right hand grey vertical.

A continuous model of barrier effects: The T-cline model is a re-expression of existing probability distribution results. In this section we take an approach closer to the development of the tripartite cline model: We construct a cline model from simpler building blocks. This differs from the tripartite cline in that it is continuous, i.e. without joins or 'corners'. It differs from the T-cline in that it requires two variables, not one, to parameterise stepped-ness.

Barriers to geneflow have units of distance, may be found at cline centres, and are expected to change the shape of clines (Barton and Bengtsson 1986). Suppose the effect of a barrier to gene flow is, from the gene perspective, to increase the 'subjective' distance experienced when crossing the cline centre. Box 2 shows how such a model can be developed, and Figure 3 places all the symmetric cline shapes, and continua of cline shapes, discussed thus far in the context of this *lamdal* barrier effect model.

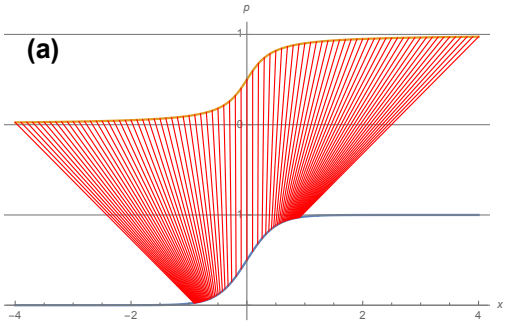
Box 2: '*lamdal*' model of continuous cline shape

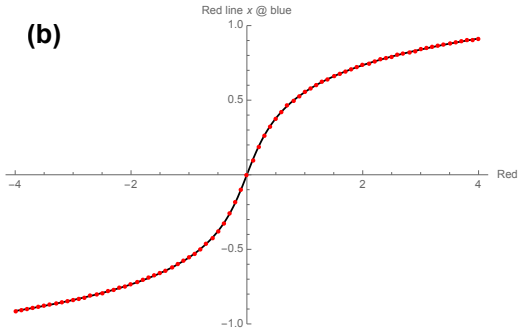
We seek a continuous approximation to a stepped unit cline, assuming a barrier distorts the gene's eye view of distance travelled during geneflow. First, exploring the exact nature of such a distortion, suppose unit cline U_N is 'sigmoid' and U_C is stepped. By the nature of inverse functions

$$U_N\left(U_N^{-1}(U_C(x))\right) = U_C(x)$$

$$U_N(\,f(x)\,) = U_C(x); \quad f(x) = U_N^{-1}(U_C(x))$$

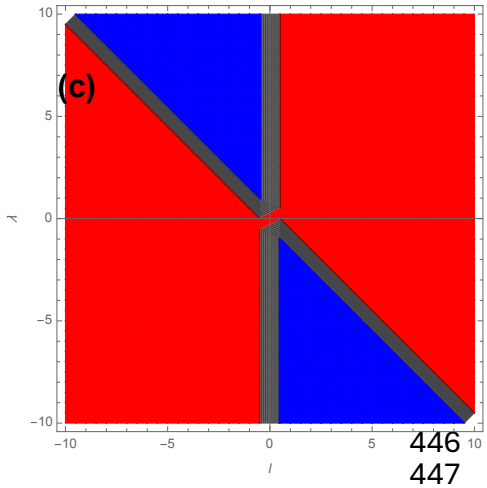
Here $f(x)$ is an exact distortion of distance x such that sigmoid U_N becomes stepped U_C. (a) Red lines connect points of equal p on two unit clines, above $UnitNormal$, below $UnitCauchy$. (b) Red points are the x-coordinates of the ends of each red line in (a), the black line is $f(x)$ for these two cline shapes. Under this distance distortion $UnitNormal$ becomes $UnitCauchy$.





The form of $f(x)$ suggest a two parameter approximation to the distance distortion for stepped cline shapes in general:

$$f(x) \approx f_{\lambda,l}(x) = X + \lambda(\,U_B(X/l) - U_B(0));$$
$$X = \frac{x}{1-\frac{\lambda}{l}}$$



The '*lamdal*' (λ, l) distorted distance increases linearly other than with the central effects of $U_B(\cdot)$, a barrier $UnitCline$ p-recentered by $U_B(0)$, rescaled by l and reweighted by λ. Distance is rescaled $x \to X$ such that any $UnitCline$ will retain unit width under *lamdal* distortion. (c) shows the (red) space in which *lamdal* distortion of $UnitClines$ results in (monotonic rising) $UnitClines$. The apparently simple form of the *lamdal* approximation suggests simple interpretation of $\{\lambda, l\}$, but the valid (red) space in (c), which includes negative values of both arguments (due to the $x \to X$ rescaling) belies this suggestion; the *lamdal* approximation is instead best interpreted in the context of how it fits to known cline shapes, as in Figure 3.

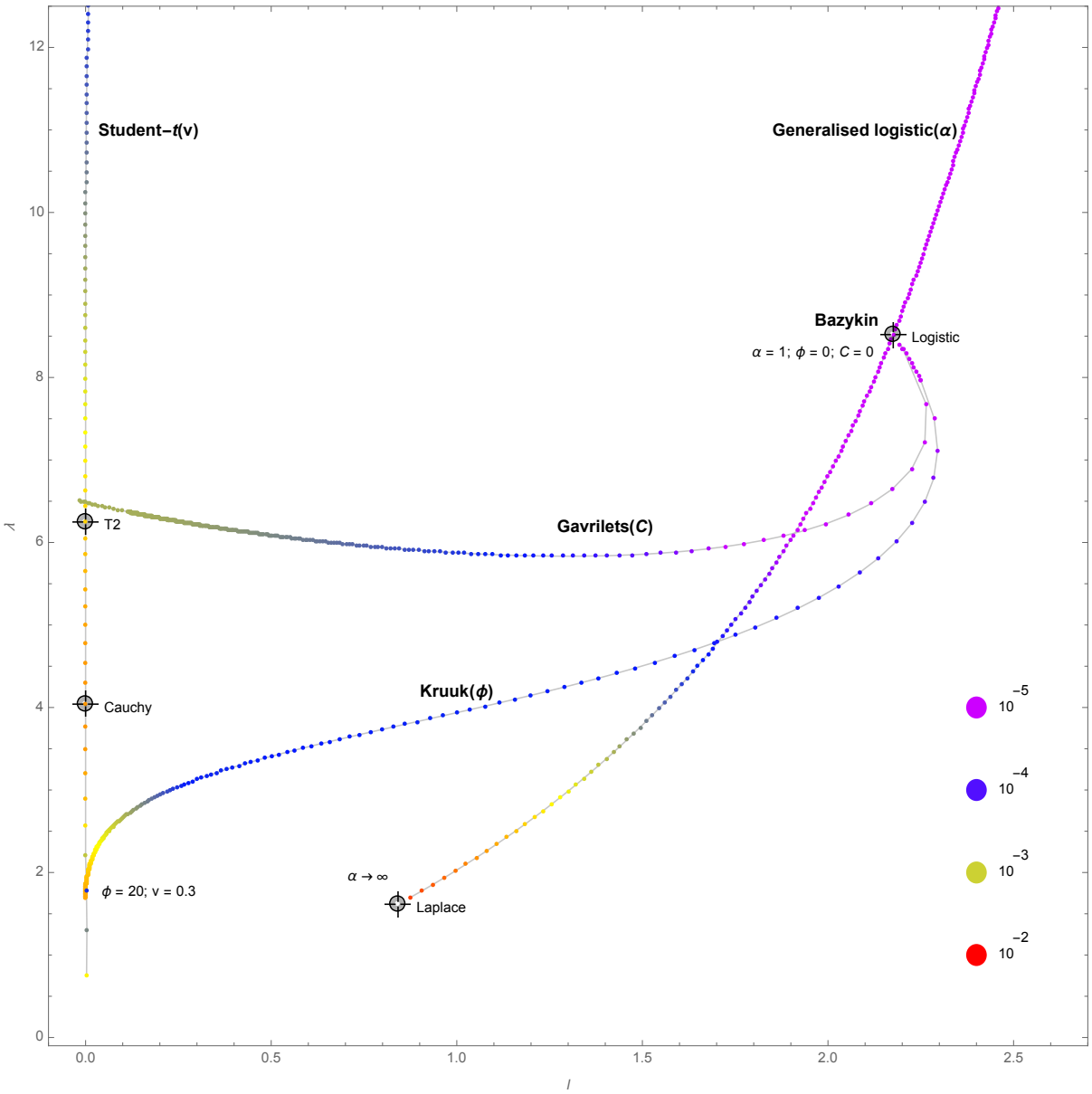


Figure 3: A comparison of cline shapes. Fits to cline shapes are plotted on the plane formed by the $\{l, \lambda\}$ parameters of the *lamdal* barrier effects model $UnitNormal\left(f_{\lambda,l}(x)\right)$, with $U_B = T(0.15)$, (See Box 2). Goodness of fit is shown in colour, with a key to levels of maximum absolute difference (MAD). Sigmoid shapes are found toward the top (single locus selection against admixture, Bazykin; high v Student-t and high α generalised logistic). Stepped shapes are to the bottom and left (single locus viability selection Gavrilets with $C \to \infty$, the neutral Cauchy kernel, multilocus admixture selection Kruuk with high ϕ). The Gavrillets and Kruuk continua are shown for a series of numerical integrations of their relevant parameters. The *T*-cline and generalized logistic results are analytic.

The *lamdal* model captures sigmoid and barrier cline shapes well where colours are cool in Figure 3. It becomes a poor approximation on the generalized logistic continuum as cline shape tends to that of the Laplace (double exponential) distribution. It becomes poor on the Kruuk continuum when the summed coupling exceeds ~10. It is a reasonable approximation throughout the Gavrilets continuum. It is a poor approximation to *T*-cline shapes of intermediate v, e.g. Cauchy and *T2*. While the spatial population genetics continua of cline shapes cross paths with the shapes of CDFs of known probability distributions, they do not align. In the *lamdal* space of figure 3 the former run horizontally, the latter more vertically. It therefore seems unlikely that there is an oven ready CDF waiting in the literature for us to discover and use as a model that will better embody the notion of smooth stepped clines for inference. Of the two CDF continua in Figure 3 the *T*-cline family appears preferable for inference, as it allows for both neutral cline shape (at high v) and stepped shape, whereas the generalized logistic distribution terminates at Laplace shape. It also seems that an analytical solution unifying the Kruuk and Gavrilets notions of stepped clines, and generalizing this over diverse kernel shapes is not on the horizon. We can however fit the *lamdal* model to data and see where the shape parameter estimates fall in relation to these existing results (See Supplementary materials).

The *lamdal* model naturally extends to the asymmetric case when we allow its central barrier $UnitCline$ to be asymmetric, for example defined by an asymmetric *T*-cline such as in Figure 2. In fact, the *lamdal* model should not be used for inference *without* this possibility of asymmetry. In Figure 3 we use the *lamdal* model to compare the shapes of clines known to be symmetric, our first visualization of how the various models of cline shape relate to each other. The situation for inference is qualitatively different. If we make an assumption of symmetry during inference, and we are wrong, then even estimates of cline centres and widths can be mis-inferred ((Baird and Macholan 2012), Box 14.3). This issue of mis-shapen mis-inference, like the data hungry nature of cline fitting, will not go away no matter what perspective we take, and this can be illustrated in the context of genome clines.

Genome clines

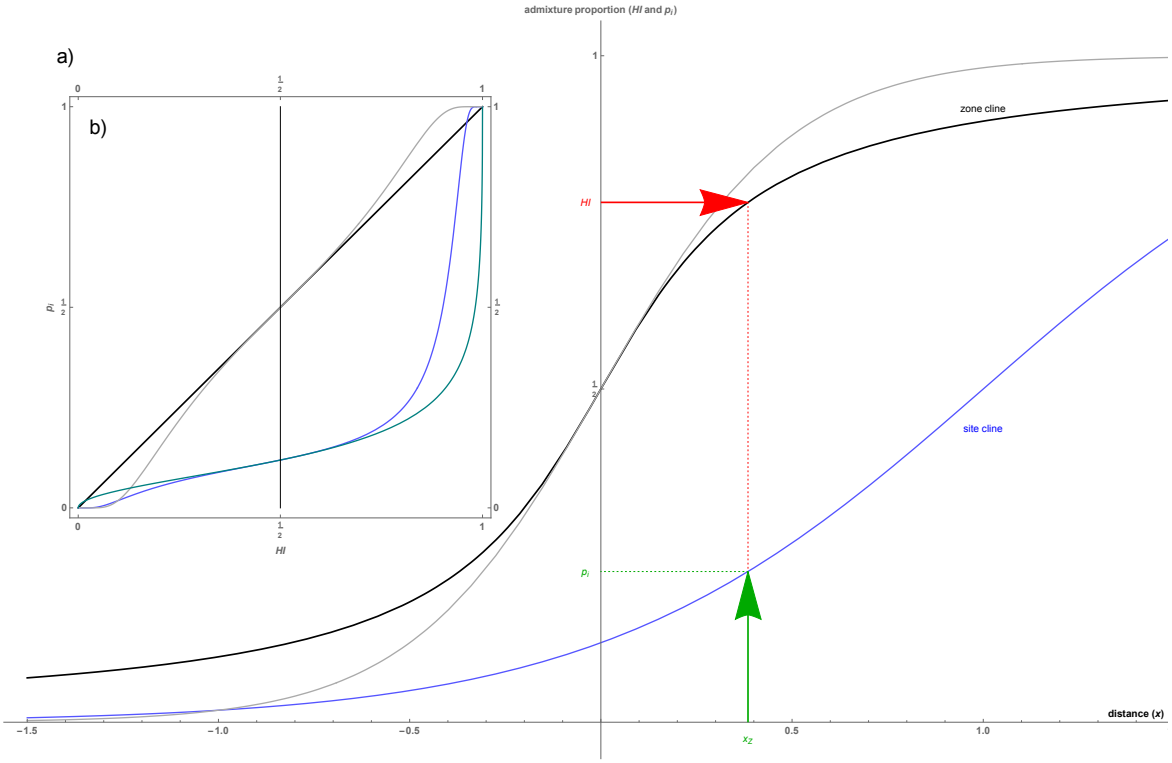
Suppose we infer a sigmoid symmetric cline 1 km wide as in Box 1, but the field data is insufficient to distinguish between a Normal sigmoid shape and a Bazykin (logistic) sigmoid. The data are then consistent with two very different scenarios: either a 1 km cline is being maintained by e.g. Bazykin selection (and we do not for how long) or a neutral cline has expanded in width to 1km (and we do not know how long this took). If we knew the per generation dispersal scale of the organism, then we would be able to estimate the selection acting, or the time since contact. (If we also knew the real time since contact, we might find the estimated time implausible, and finally be able to decide between the scenarios: selection must be acting). Unfortunately, the key to unlocking such evolutionary puzzles σ, the natural measure of spatial processes such as hybrid zones, is notoriously difficult to measure. Just as with cline studies, capture-mark-recapture studies are notoriously data hungry, and do not actually report on per generation σ, but rather within-generation movements. These two can obviously be very different even within the same organism, for example migratory birds with high nest-site

fidelity (Ruegg 2008), leaving direct estimates of σ difficult, and the scenarios we can resolve through cline fitting reduced.

Where there is poor prior knowledge of the per generation scale of dispersal, an alternative is to measure trait cline widths relative to a global estimate, a kind of outlier scan analogous to Fst scans (but using a statistic that does not confound dispersal with diversity). It is perhaps this idea that has driven exploration of Barton's 'concordance' transform (Szymura and Barton 1991), where the distance axis of a cline is substituted by a global hybrid index (HI) axis. This was suggested as a convenient relative {centre, width} comparison of clines where geographic sampling coordinates were difficult to interpret, and has been used, for example, to compare hybrid indices of parasites and their hosts (Goüy de Bellocq, Ribas et al. 2018). The distance→HI axis replacement was further developed as the genome clines approach ((Gompert and Buerkle 2009, Gompert and Buerkle 2012), see (Macholan, Baird et al. 2011) for comparison with concordance). It has been suggested the convenience of the axis replacement extends to the case where "the geographic model [...] takes a more complicated form than the simple logistic function, for example, when clines are asymmetric or stepped" (Fitzpatrick 2013), presaging a surge in genome cline fitting software (Bailey 2024, Gompert, DeRaad et al. 2024), but this suggestion is over optimistic: no {centre, width} cline model can capture variation in cline shape, and using a model with the wrong shape leads to genome landscape mis-inference (Box 3). The logit-logistic genome cline model proposed by (Fitzpatrick 2013) is not freed of assumptions by a change in x-axis: in fact it implicitly assumes the shapes of *two* clines: one in (genome-wide) hybrid index, and one in trait frequency. A generalised genome cline (Box 3) relaxes this assumption to cases where the genome-wide cline function is inversible, and Table 1 details invertible cline functions of different stepped-ness, allowing a generalised genome cline approach to be applied assuming a variety of symmetric hybrid index cline shapes, and even stepped and asymmetric T-clines.

What would be the knock-on effects of extending the genome cline model to one that allows for a better match to a globally asymmetric stepped hybrid zone such as the house mouse hybrid zone (Macholán, Munclinger et al. 2007)? First and for most, any 'free lunch' impression that genome clines allow confident estimates with little spatial sampling of a hybrid zone will likely be reduced, because fitting more parameters requires more data, and may reveal confidence to be overconfidence. Second, we might expect the variance in cline centre and width estimates to be reduced, because shape mis-match effects such as those in Box 3 should be minimised. From this perspective the shape parameters of hybrid zones are now acting as nuisance parameters, uninteresting for the comparison of trait clines with each other, but necessary if the widths and centres of those trait clines are to be comparable, and estimated without overconfidence. This is not to say that the overall shape of a hybrid zone or genomic wavefront is uninteresting, but rather to recognise that intensive field sampling has now often become the most expensive, and potentially controversial, part of such analyses.

BOX 3: Genome clines



a) Black curve: a multilocus (genome wide) hybrid index (HI) changes across a hybrid zone. The spatial units are scaled and offset so that this is a unit cline $U_Z(x)$. Blue curve: allele frequency at site i changes as cline $U_i\left(\frac{x-c_i}{w_i}\right)$. The site cline is w_i-fold wider than the zone cline and offset by distance c_i . Grey curve: illustrating a zone cline of an alternate shape; the black curve is Cauchy, the grey logistic.
b) (inset) The curves of pane (a) when the distance axis is replaced with zone hybrid index. One further (teal) curve is shown: the blue site cline of (a) when the alternate (grey) zone cline shape is assumed.

Switching x-axes. (a) The distance x coordinate of the site cline can be replaced by the HI coordinate of the zone cline. Red arrow: applying the inverse of the zone cline to a hybrid index gives a zone x coordinate $x_Z = U_Z^{-1}(HI)$. Green arrow: applying the site cline at this x coordinate gives a site allele frequency $p_i = U_i\left(\frac{x_Z-c_i}{w_i}\right)$. Putting these two steps together, distance x is replace by HI in a generalised genome cline

$$p_i = U_i\left(\frac{U_Z^{-1}(HI) - c_i}{w_i}\right)$$

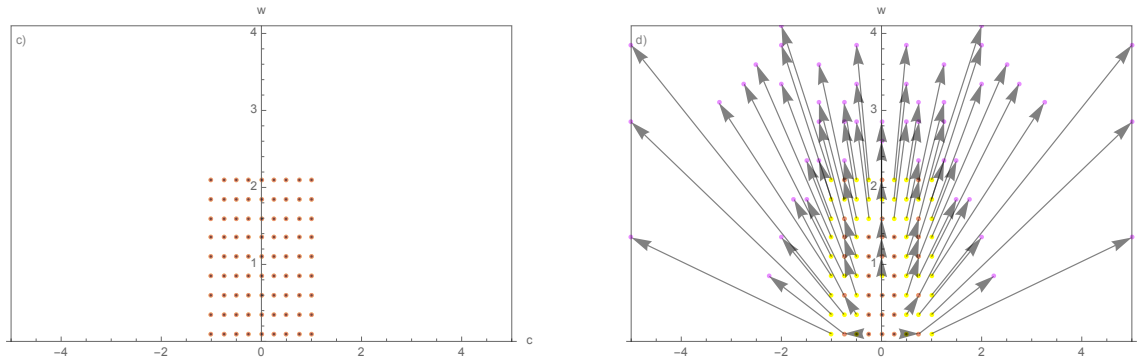
If both cline shapes $\{U_Z, U_i\}$ are logistic, the inverse U_Z^{-1} is a rescaled logit function and this is the "logit-logistic" special case genome cline, equation (4b) of (Fitzpatrick 2013), used by (Goüy de Bellocq, Ribas et al. 2018), and implemented in softwares gghybrid (Bailey 2024) and bgchm (Gompert, DeRaad et al. 2024):

$$p_i = \frac{S^{v_i}}{S^{v_i} + (1-S)^{v_i} e^{u_i}}$$

$$S = HI;\ v_i = \frac{1}{w_i};\ u_i = 4\frac{c_i}{w_i};\ U_Z = U_i = UnitLogistic$$

BOX 3: Genome clines (continued)

Site cline shape expectations can vary along the genome, and the zone hybrid index shape may be stepped and/or asymmetric. The two parameters $\{u_i, v_i\}$ of a logit-logistic genome cline are insufficient to capture this richness of cline shapes variation.



Misinference using the logit-logistic function: (c,d) True site cline parameter points $\{c, w\}$ are plotted in yellow, estimates of them are plotted in purple; when an estimate and a truth coincide, that point becomes brown. When estimate and truth differ, an arrow joins truth to estimate. (c) When the zone cline shape is logistic and site cline shapes are also logistic the logit-logistic function is appropriate and estimation returns the truth. (d) When the zone cline shape is instead Cauchy (as in pane (a)) but the site cline shapes remain logistic, the logit-logistic function is not appropriate for inference: estimates are only accurate when the truth is close to the origin (site cline centre, widths resemble those of the zone cline). For other site clines, errors are large and there is no simple error correction: centre and width error effects are not independent. Estimation procedure: For each true point a grid of parameter values was searched exhaustively for minimum difference between the logit-logistic function and genome cline shapes for $U_i = UnitLogistic$ and (c) $U_Z = UnitLogistic$, (d) $U_Z = UnitCauchy$. Eight estimations lie on the boundary of this parameter search grid, and so may represent even more extreme errors.

555

Conclusions and future directions

Cline shapes are intimately related to the shapes of location dispersal kernels, and this allows us to draw on more than a century of work on probability distributions as a source of, and perspective on, the shapes of cline and wavefront models.

Most people find it easier to distinguish the bell shaped curves on the left of Figure 1, than to distinguish the s-shaped curves on the right, generated from those bells. A cognitive bias against distinguishing cline shape suggests decisions regarding whether clines shapes matter should be based on objective measures, such as whether shape differences affect the results of inference ((Baird and Macholan 2012), Box 14.3). Here, we have seen evidence that cline shape matters not only for spatial and genomic cline inference (Figure 2, Box 2), but we are also reminded more generally, that the shape of dispersal kernels matters for the speed of spatial process, and that part of the gathering phase of the coalescent during which geographically distant coalescence is improbable (Box 1).

Cline expectations are more diverse than can be usefully captured with two parameters {centre, width}; in particular, stepped and asymmetric shapes have biological importance as they can result from multilocus or non-standard forms of selection, and movement or asymmetric geneflow respectively. Further, here we have only considered the case where population density is sufficiently high that the effects of drift on cline expectations are negligible. Where drift acts it is expected to steepen the centre of site clines, while widening the overall cline in hybrid index (Polechová and Barton 2011), suggesting again (as with Box 3) that assuming these shapes are the same is unwise. It seems a wide variety of (stepped) symmetric shapes can be smoothly captured at high fidelity with just two shape parameters (Figure 3), and perhaps reasonably approximated with just one shape parameter – the shape parameter of the Student's-*t* distribution allows cline stepped-ness to be adjusted. The non-central Student's-*t* allows variation in both stepped-ness and asymmetry, and captures Fisher's wave of advance at high fidelity (Figure 2). It appears these continuous cline models allow better fits with fewer parameters than tripartite cline models, and computationally efficient Python tools for *T*-cline implementations are made available here: (Baird and Daley 2023). There is no free lunch: these potential advantages are at the cost of reduced interpretability of the fitting parameters, though cline fits can be projected onto the space of Figure 3, allowing comparison with classical models. While these developments may allow for improvements over the tripartite cline model commonly used for spatial clines, the inverse *T*-cline (Table 1) may also allow cline shape to be accounted for in genome cline approaches.

Regarding our opening question: Why, after 30 years of searching, have so few stepped hybrid zones been identified? For spatial clines the answer remains unclear, but perhaps, asking simpler questions of better models with fewer parameters will allow us to decide whether stepped clines are actually rare, or just rarely proven. For genome clines, there is a simple answer: because we were not looking; there was no genome cline model of a stepped hybrid zone. We might hope that generalisation of genome clines such as proposed here would allow a similar resolution of the opening question, but this is by no

means obvious because the way data is sampled has changed profoundly since Haldane first fit a cline to genetic data and estimated selection (Haldane 1948). It is now extremely rare that hundreds of genetic samples are gathered from the field. Instead relatively few genomes are gathered, and from them very many nucleotide variants (SNVs) are sampled. In spatial genetics the latter does not compensate for the former, because few genomes means few spatial sampling locations, irrespective of how many SNVs are then inspected. This is equally true for genome clines, because few genomes means few sampling locations on the global HI x-axis. In these circumstances cline shape may be reduced to a necessary, but nuisance, parameter. Further, we cannot simply scale up cline fitting to cover (and compare) data at every one of potentially millions of SNVs – aside issues of from computation tractability, there are only one or two recombination events per chromosome per generation, so neighbouring sites in admixture systems are clearly not independent witnesses of the evolutionary process (Baird 2015). Assuming they are independent will lead to overconfident inference. This highlights that, to make sound admixture inference over modern genomic data, the blockwise nature of admixture tracts must be recognised ((Shipilina, Pal et al. 2023, Ebdon, Laetsch et al. 2024)). To estimate the boundaries of such tracts we should leverage every single SNV that forms a cline – their non-independence under admixture is now a positive, not a negative, and it turns out that for genomic data, introgressing blocks become obvious when SNVs are co-polarised by their association, such that all their clines are rising (or all falling) (Baird, Petružela et al. 2023, Ebdon, Laetsch et al. 2024). The polarisation operation scales linearly with genome size and reports a 'diagnostic index' matching-statistic between each SNV and a global estimate over individuals. This global estimate is a *superset* of the information necessary to plot a global cline in hybrid index, as it also contains the analogous global central bump in heterozygosity caused by admixture (cf p_{12} in (Simon, Bierne et al. 2018)). Downstream inference can be targeted on regions that differ from this global estimate. If in future it becomes commonplace to genome sequence individuals sampled at very many different field co-ordinates, then genome cline shapes may stop being a necessary nuisance, and start again to be of active interest. In the meantime it seems the data appetite of the questions we would wish to ask of genomic clines may best be fed blockwise, and relative to change in both hybrid index and heterozygosity (cf (Simon, Bierne et al. 2018)).

Not all modern field sampling is directed toward high throughput sequencing of genomes. Advancing technology has also increased the potential throughput of geographic locations for measures of quantitative traits and genetic markers, in particular SNV assays such as KASP (He, Holme et al. 2014) allow for individuals from very many locations to be cost effectively assayed for scores of genetic markers (Touchard, Cerqueira et al. 2024). Because clines in genetic and quantitative traits are governed by similar dispersal and selection processes, cline models apply equally to both (Barton and Gale 1993), and so, likewise, any developments in shaped cline models. True high throughput field sampling to take advantage of these developments risks however perturbing that which we wish to observe – slowing a wave of advance or narrowing a hybrid zone by reducing population density through destructive sampling, so if there is to be a renaissance in cline inference field studies, these new models and technologies should be carefully coupled with the parallel advances that have been made in non-destructive sampling (Lefort, Boyer et al. 2015).

Acknowledgements

1. SJEB was supported by Czech Science foundation grant number 22-32394S. SJEB would like to thank Jitka Polechova for help with the numerical solution to the Kruuk continuum. Metacentrum computational resources: This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic through the e-INFRA CZ (ID:90140)."

References

Abbott, R., D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, N. Bierne, J. Boughman, A. Brelsford, C. A. Buerkle, R. Buggs, R. K. Butlin, U. Dieckmann, F. Eroukhmanoff, A. Grill, S. H. Cahan, J. S. Hermansen, G. Hewitt, A. G. Hudson, C. Jiggins, J. Jones, B. Keller, T. Marczewski, J. Mallet, P. Martinez-Rodriguez, M. Möst, S. Mullen, R. Nichols, A. W. Nolte, C. Parisod, K. Pfennig, A. M. Rice, M. G. Ritchie, B. Seifert, C. M. Smadja, R. Stelkens, J. M. Szymura, R. Väinölä, J. B. W. Wolf and D. Zinner (2013). "Hybridization and speciation." Journal of Evolutionary Biology **26**(2): 246.

Albrechtová, J., T. Albrecht, S. J. E. Baird, M. Macholán, G. Rudolfsen, P. Munclinger, P. K. Tucker and J. Piálek (2012). "Sperm-related phenotypes implicated in both maintenance and breakdown of a natural species barrier in the house mouse." Proceedings of the Royal Society B: Biological Sciences **279**(1748): 4803-4810.

Bailey, R. I. (2024). "Bayesian hybrid index and genomic cline estimation with the R package gghybrid." Molecular Ecology Resources **24**(2): e13910.

Baird, S. J. (2015). "Exploring linkage disequilibrium." Molecular ecology resources **15**(5): 1017-1019.

Baird, S. J. and M. Macholan (2012). What can the Mus Musculus/M. m. domesticus hybrid zone tell us about speciation? Evolution of the house mouse. Oxford, Oxford University Press. **3:** 334-372.

Baird, S. J., J. Petružela, I. Jaroň, P. Škrabánek and N. Martínková (2023). "Genome polarisation for detecting barriers to geneflow." Methods in Ecology and Evolution **14**(2): 512-528.

Baird, S. J. E. (1995). "A simulation study of multilocus clines." Evolution **49**(6): 1038-1045.

Baird, S. J. E. and N. H. Barton (1995). Analyse–An application for analyzing hybrid zones.

Baird, S. J. E. and N. Daley (2023). Clinalyse.

Barton, N. and B. O. Bengtsson (1986). "THE BARRIER TO GENETIC EXCHANGE BETWEEN HYBRIDIZING POPULATIONS." Heredity **57**: 357-376.

Barton, N. and K. Gale (1993). Genetic analysis of hybrid zones. Hybrid Zones and the Evolutionary Process,. R. Harrison. London/New York/ Oxford, Oxford University Press**:** 13–45.

Barton, N. H. (1979). "DYNAMICS OF HYBRID ZONES." Heredity **43**(DEC): 341-359.

Barton, N. H. (1979). "GENE FLOW PAST A CLINE." Heredity **43**(DEC): 333-339.

Barton, N. H. (1983). "MULTILOCUS CLINES." Evolution **37**(3): 454-471.

Barton, N. H. (2020). "On the completion of speciation." Philosophical Transactions of the Royal Society B **375**(1806): 20190530.
Barton, N. H. and B. Charlesworth (1984). "GENETIC REVOLUTIONS, FOUNDER EFFECTS, AND SPECIATION." Annual Review of Ecology and Systematics **15**: 133-164.
Barton, N. H. and G. M. Hewitt (1985). "Analysis of hybrid zones." Annual Review of Ecology and Systematics **16**(1): 113-148.
Bateman, A. (1950). "Is gene dispersion normal?".
Bazykin, A. (1969). "Hypothetical mechanism of speciaton." Evolution **23**(4): 685-687.
Darwin, C. (1859). On the origin of species, Routledge London.
Derryberry, E. P., G. E. Derryberry, J. M. Maley and R. T. Brumfield (2014). "HZAR: hybrid zone analysis using an R software package." Molecular ecology resources **14**(3): 652-663.
Diekmann, O. (1978). "Thresholds and travelling waves for the geographical spread of infection." Journal of Mathematical Biology **6**(2): 109-130.
Ebdon, S., D. R. Laetsch, R. Vila, S. J. E. Baird and K. Lohse (2024). "Genomic regions of current low hybridisation mark long-term barriers to gene flow in scarce swallowtail butterflies." bioRxiv: 2024.2006.2003.597101.
Einstein, A. (1905). "On the motion of small particles suspended in liquids at rest required by the molecular-kinetic theory of heat." Annalen der physik **17**(549-560): 208.
Endler, J. A. (1977). Geographic variation, speciation, and clines, Princeton University Press.
Fisher, R. A. (1937). "The wave of advance of advantageous genes." Annals of eugenics **7**(4): 355-369.
Fitzpatrick, B. M. (2013). "Alternative forms for genomic clines." Ecology and evolution **3**(7): 1951-1966.
Gavrilets, S. (1997). "Single locus clines." Evolution: 979-983.
Gay, L., P.-A. Crochet, D. A. Bell and T. Lenormand (2008). "COMPARING CLINES ON MOLECULAR AND PHENOTYPIC TRAITS IN HYBRID ZONES: A WINDOW ON TENSION ZONE MODELS." Evolution **62**(11): 2789-2806.
Gompert, Z. and C. Buerkle (2012). "bgc: Software for Bayesian estimation of genomic clines." Molecular Ecology Resources **12**(6): 1168-1176.
Gompert, Z. and C. A. Buerkle (2009). "A powerful regression-based method for admixture mapping of isolation across the genome of hybrids." Molecular Ecology **18**(6): 1207-1224.
Gompert, Z., D. A. DeRaad and C. A. Buerkle (2024). "A Next Generation of Hierarchical Bayesian Analyses of Hybrid Zones Enables Model-Based Quantification of Variation in Introgression in R." Ecology and Evolution **14**(11): e70548.
Goüy de Bellocq, J., A. Ribas, J. Bryja, J. Piálek and S. J. Baird (2018). "Holobiont suture zones: Parasite evidence across the European house mouse hybrid zone." Molecular Ecology **27**(24): 5214-5227.
Haldane, J. (1948). "The theory of a cline." Journal of genetics **48**(3): 277-284.
He, C., J. Holme and J. Anthony (2014). "SNP genotyping: the KASP assay." Crop breeding: methods and protocols: 75-86.
Huxley, J. (1938). "Clines: an auxiliary taxonomic principle." Nature **142**(3587): 219-220.
Kot, M., M. A. Lewis and P. van den Driessche (1996). "Dispersal Data and the Spread of Invading Organisms." Ecology **77**(7): 2027-2042.

Krieber, M. and M. R. Rose (1986). "Molecular aspects of the species barrier." Annual review of ecology and systematics: 465-485.
Kruuk, L. E. B., S. J. E. Baird, K. S. Gale and N. H. Barton (1999). "A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids." Genetics **153**(4): 1959-1971.
Lefort, M.-C., S. Boyer, A. Barun, A. E. Khoyi, J. Ridden, V. R. Smith, R. Sprague, B. R. Waterhouse and R. H. Cruickshank (2015). Blood, sweat and tears: non-invasive vs. non-disruptive DNA sampling for experimental biology, PeerJ PrePrints.
Lindström, T., N. Håkansson and U. Wennergren (2011). "The shape of the spatial kernel and its implications for biological invasions in patchy environments." Proceedings of the Royal Society B: Biological Sciences **278**(1711): 1564-1571.
Macholan, M., S. J. E. Baird, P. Dufkova, P. Munclinger, B. V. Bimova and J. Pialek (2011). "ASSESSING MULTILOCUS INTROGRESSION PATTERNS: A CASE STUDY ON THE MOUSE X CHROMOSOME IN CENTRAL EUROPE." Evolution **65**(5): 1428-1446.
Macholán, M., P. Munclinger, M. Šugerková, P. Dufková, B. Bímová, E. Božíková, J. Zima and J. Piálek (2007). "Genetic analysis of autosomal and X-linked markers across a mouse hybrid zone." Evolution: International Journal of Organic Evolution **61**(4): 746-771.
Mallet, J. (1995). "A species definition for the modern synthesis." Trends in Ecology & Evolution **10**(7): 294-299.
Martin, S. H., J. W. Davey, C. Salazar and C. D. Jiggins (2019). "Recombination rate variation shapes barriers to introgression across butterfly genomes." PLoS biology **17**(2): e2006288.
Nagylaki, T. (1975). "Conditions for the existence of clines." Genetics **80**(3): 595-615.
Nathan, R., E. Klein, J. J. Robledo-Arnuncio and E. Revilla (2012). Dispersal kernels. Dispersal Ecology and Evolution. J. Clobert, M. Baguette, T. G. Benton and J. M. Bullock. Oxford, UK, Oxford University Press Oxford, UK.
Phillips, B. L. (2015). "Evolutionary processes make invasion speed difficult to predict." Biological Invasions **17**: 1949-1960.
Polechová, J. and N. Barton (2011). "Genetic drift widens the expected cline but narrows the expected cline width." Genetics **189**(1): 227-235.
Porter, A. H., R. Wenger, H. Geiger, A. Scholl and A. M. Shapiro (1997). "The Pontia daplidice-ed usa hybrid zone in northwestern Italy." Evolution **51**(5): 1561-1573.
Pybus, O. G., M. A. Suchard, P. Lemey, F. J. Bernardin, A. Rambaut, F. W. Crawford, R. R. Gray, N. Arinaminpathy, S. L. Stramer and M. P. Busch (2012). "Unifying the spatial epidemiology and molecular evolution of emerging epidemics." Proceedings of the national academy of sciences **109**(37): 15066-15071.
Ruegg, K. (2008). "Genetic, morphological, and ecological characterization of a hybrid zone that spans a migratory divide." Evolution **62**(2): 452-466.
Shipilina, D., A. Pal, S. Stankowski, Y. F. Chan and N. H. Barton (2023). "On the origin and structure of haplotype blocks." Mol Ecol **32**(6): 1441-1457.
Simon, A., N. Bierne and J. J. Welch (2018). "Coadapted genomes and selection on hybrids: Fisher's geometric model explains a variety of empirical patterns." Evolution Letters **2**(5): 472-498.
Skellam, J. G. (1951). "Random dispersal in theoretical populations." Biometrika **38**(1/2): 196-218.

Stankowski, S., M. A. Chase, A. M. Fuiten, M. F. Rodrigues, P. L. Ralph and M. A. Streisfeld (2019). "Widespread selection and gene flow shape the genomic landscape during a radiation of monkeyflowers." PLoS biology **17**(7): e3000391.
Szymura, J. M. and N. H. Barton (1986). "GENETIC-ANALYSIS OF A HYBRID ZONE BETWEEN THE FIRE-BELLIED TOADS, BOMBINA-BOMBINA AND BOMBINA-VARIEGATA, NEAR CRACOW IN SOUTHERN POLAND." Evolution **40**(6): 1141-1159.
Szymura, J. M. and N. H. Barton (1991). "THE GENETIC-STRUCTURE OF THE HYBRID ZONE BETWEEN THE FIRE-BELLIED TOADS BOMBINA-BOMBINA AND B-VARIEGATA - COMPARISONS BETWEEN TRANSECTS AND BETWEEN LOCI." Evolution **45**(2): 237-261.
Terborgh, J. (2020). "At 50, Janzen–Connell Has Come of Age." BioScience **70**(12): 1082-1092.
Touchard, F., F. Cerqueira, N. Bierne and F. Viard (2024). "Adaptive alien genes are maintained amid a vanishing introgression footprint in a sea squirt." Evolution Letters: qrae016.
Wolfram Research, I. (2019). Mathematica. Champaign, Illinois, Wolfram Research, Inc.