**Introduction**

Evolutionary clines—changes in the frequency of a measurable character (e.g. allele, genotype, phenotype) over a geographical area (Huxley 1938, 1939)—have long served as model systems in evolutionary biology (Endler 1977). Clines arise and are maintained via the interplay of natural selection generating spatial differentiation in morph frequencies, gene flow homogenizing morph frequencies between populations and local genetic drift acting within populations (Haldane 1948; Felsenstein 1975; Slatkin and Maruyama 1975; Endler 1977; Saccheri et al. 2008; Volis and Zhang 2010; Takahashi 2015). The multiple evolutionary mechanisms structuring clines has prompted their continued use by evolutionary biologists seeking to explore the relative contributions of non-adaptive and adaptive evolutionary processes in structuring patterns of genetic and phenotypic diversity and differentiation within natural populations.

While phenotypic or allelic clines are often interpreted as strong evidence of adaptive evolution, non-adaptive processes (e.g. genetic drift and gene flow) may also generate covariance between allele frequencies and environmental gradients. For example, local genetic drift in combination with spatially restricted gene flow (i.e. isolation by distance, Wright 1943) can generate single-locus clines (Vasemägi 2006). Similarly, serial founder events can generate clines in additive quantitative traits (Colautti and Lau 2015) and phenotypic clines can arise via multiple introductions from a species’ native range during invasion (Keller et al. 2009). Disentangling the relative importance of stochastic and deterministic forces in the formation of clines is thus essential prior to invoking the role of selection in generating adaptive differentiation between populations.

A corollary of the formation of single-locus or additive trait clines via neutral processes is that clines in all directions should occur with equal frequency; assuming the same initial allele frequencies, alleles should be lost or fixed with equal probability under only the effects of genetic drift. Thus, the presence of multiple independent clines in the same direction (i.e. parallel clines) is strong evidence for the role of natural selection, as putative adaptations are unlikely to evolve repeatedly via stochastic forces (Samis et al. 2012). However, when traits have a non-additive genetic basis (e.g. epistasis), clines may occur more frequently in a particular direction because stochastic changes in allele frequencies at one locus may have a disproportionate effect on phenotype frequencies. In other words, stochastic processes may lead to deterministic outcomes. For example, stochastic forces have caused the repeated loss of the Mendelian inherited short-style (S) morph from tristylous populations of *Eichhornia paniculata* in North-eastern Brazil, Jamaica, and Cuba (Barrett et al. 1989, 2009, Husband and Barrett 1992a,b). This occurs because only the S-morph produces gametes containing the dominant allele necessary to produce the S-morph; crosses between the medium-style (M) and long-style (L) morphs cannot produce plants with short styles, making the S-morph particularly susceptible to loss via drift (Barrett 1993). The fact that drift can lead to directional changes in non-additive traits across multiple, independent populations suggests that the presence of parallel clines in such traits is insufficient evidence for the role of selection in generating adaptive differentiation. In such cases, replicating phenotypic clines across many independent spatial locations can help us better understand the roles that drift, selection and gene flow play in the formation and maintenance of parallel clines.

Clines are typically studied along latitudinal (e.g. Huey et al. 2000; Caicedo et al. 2004; Campitelli and Stinchcombe 2013), altitudinal (Clausen et al. 1948; Luo et al. 2015; Zhao and Wang 2015) or longitudinal (Samis et al. 2008, 2012) transects, since such transects are often associated with gradients in putative selective agents (e.g. temperature, precipitation, biotic interaction strength). Importantly, geographical transects can also be associated with variation in demographic factors important for determining the extent to which drift will act in populations. This is especially true during invasions or rapid range expansions when founder events result in smaller marginal populations exchanging fewer migrants, leading to reduced genetic diversity and increased differentiation relative to more central populations (Alleaume-Benharira et al. 2006; Eckert et al. 2008; Excoffier and Ray 2008). Such spatial variation in demography may provide the foundation upon which parallel clines in non-additive traits can form. Studies of non-additive trait differentiation across replicate geographical gradients in selection and demography would provide the strongest test of the relative contribution of drift, selection, and gene flow in the formation of parallel clines. However, logistical challenges associated with replicating continental-scale latitudinal, altitudinal or longitudinal transects preclude such replication for many systems.

Urbanization is one of the most widespread human disturbances on earth and it provides an excellent large-scale replicated system to study how adaptive and non-adaptive evolutionary processes contribute to the formation of parallel clines. Urban environments continue to expand across the globe and are a leading cause of species extinction (McKinney 2006; Seto et al. 2010), stimulating research into how they alter the evolutionary processes that enable species to persist under these drastically altered environmental conditions. The widespread fragmentation associated with urbanization has resulted in gradients in the strength of genetic drift and gene flow. For example, small, more isolated urban populations of white-footed mice (*Peromyscus leucopus*) in New-York City have reduced genetic diversity due to increased drift and reduced gene flow relative to non-urban populations (Munshi-South and Kharchenko 2010; Munshi-South 2012; Munshi-South et al. 2016). Similarly, contemporary urbanization has reduced the effective population size of urban fire salamanders (*Salamandra salamandra*) in Oviedo, Spain (Lourenço et al. 2017). While the influence of urbanization on non-adaptive evolutionary processes is well-studied (Johnson and Munshi-South 2017), evidence is emerging that urban environments alter natural selection and species adaptation as well.

Urban environments are associated with variation in biotic and abiotic factors (McKinney 2006; Johnson et al. 2015), many of which can be potent agents of selection for many taxa. For example, industrial production around urban centres in eastern North America has increased concentrations of PCBs—a persistent environmental pollutant— in coastal estuaries, leading to parallel evolution of PCB resistance in four populations of Atlantic killifish (*Fundulus heteroclitus*) (Reid et al. 2016). Similarly, increased urban temperatures have driven the evolution of greater thermal tolerance in urban Acorn ant (*Temnothorax curvispinosus*) populations across multiple cities in the United States (Diamond et al. 2017, Diamond et al., 2018 *this issue*). In addition, Thompson *et al.* (2016) recently identified parallel urban-rural clines in the frequency of plants producing hydrogen cyanide (HCN) (i.e. cyanogenesis)—a potent antiherbivore defence—in populations of white clover (*Trifolium repens*) across multiple cities. They found that HCN defended genotypes were less frequent in urban populations in 3 of the 4 cities examined(Thompson et al. 2016). While the authors identified lower winter surface temperatures in urban populations as a putative selective agent structuring these clines, they did not consider the alternative hypothesis that these clines could be caused by genetic drift.

The purpose of this paper is to examine under what conditions non-adaptive (e.g. genetic drift, gene flow) and adaptive processes (e.g. selection) can generate clines in phenotypes with a non-additive genetic basis, using the cyanogenesis polymorphism in white clover as a model to this end. White clover is native to Eurasia but has been intentionally introduced into temperate regions worldwide due to its importance in agriculture (Burdon 1983; Kjærgaard 2003). White clover and the cyanogenic polymorphism has a long-history of study among evolutionary biologists as both latitudinal and altitudinal clines in cyanogenesis have been reported in clover’s native (Daday 1954a,b; de Araújo 1976) and non-native (Daday 1958; Ganders 1990; Kooyers and Olsen 2012) ranges with higher frequencies of HCN reported in populations from warmer (e.g. lower elevation and further south) (Daday 1965) and drier environments (Kooyers et al. 2014). HCN is controlled by two independently segregating Mendelian loci (*CYP79D15* and *Li*) and plants require a dominant allele at each of two loci to produce HCN. As such, this trait exhibits duplicate recessive epistasis (i.e. complimentary epistasis, Sackton and Hartl 2016) where recessivity at either locus (or both) results in individuals lacking HCN (i.e. acyanogenic, HCN–). A consequence of this genetic architecture is that the frequency of HCN can only increase via drift if the frequency of the dominant alleles at both underlying loci drift upward (Table 1). Any other combination of frequency changes in the dominant alleles result in decreases in the frequency of HCN, making population particularly susceptible to loss of HCN via drift.

In this paper, we develop a series of spatially explicit, agent-based simulations to address the following specific questions: (1) How do drift, gene flow and selection interact in the formation and maintenance of spatial clines in HCN? (2) Is the formation of urban-rural clines in HCN contingent upon the colonization history of white clover in cities (i.e. urban to rural, rural to urban, always colonized)? While we focus on cyanogenesis in clover as a model for answering these questions, our results will have implications for the formation of clines in traits with a non-additive genetic basis in a range of different systems.

**Table 1:** Direction of change in the frequency of HCN as the frequency of dominant alleles at underlying loci either increases or decreases. In three out of four cases, HCN frequencies decrease. This makes populations particularly susceptible to loss of HCN via drift.

|  |  |  |
| --- | --- | --- |
| ***CYP79D15*** | ***Li*** | **HCN** |
| Increase | Increase | Increase |
| Increase | Decrease | Decrease |
| Decrease | Increase | Decrease |
| Decrease | Decrease | Decrease |