**Outline for simulations manuscript**

**Introduction:**

* *Paragraph 1:* What are clines and what are their purpose
  + Topic: Clines are a change in the frequency of alleles, phenotpyes, or genotypes over some spatial extent (Haldane 1948; Endler 1977).
  + Clines are of continued interest to evolutionary biologists as they can help us disentangle adaptive (i.e. deterministic) from non-adaptive (i.e. stochastic) evolutionary mechanisms and inform speciation processes (Takahashi 2015).
  + The relative roles of deterministic vs. stochastic processes continues to be of central importance in evolutionary biology (Losos 1998; Simões 2008; Travisiano 1995).
* *Paragraph 2*: Parallel clines are the hallmark of adaptation: contrast parallel clines vs. single (i.e. unreplicated) clines
  + Topic: Parallel clines are considered strong evidence for the role of adaptive evolution (Stock et al. 2015; Samis et al. 2012; Huey 2007; Gilchrist 2001) especially clines in quantitative traits (Samis et al. 2012).
  + However, non-adaptive processes such as drift and spatially restricted gene flow (Vasemägi 2006), in addition to founder events and unequal sampling from ancestral populations (Keller and Taylor 2009) can generate phenotypic clines.
  + Disentangling the relative importance of stochastic and deterministic forces is essential prior to invoking the role of selection in generating adaptive phenotypic clines (Colautti and Lau 2015).
* *Paragraph* 3: Clines in complex traits (e.g. epistasis) may more readily evolve via neutral processes. Other ecological processes may also generate clines (e.g. inbreeding at range margins resulting in loss of heterozygosity)
  + Topic: Non-adaptive processes may more likely generate phenotypic clines in traits with complex genetic architectures (e.g. epistasis).
  + For two-locus traits with epistasis, stochastic processes (e.g. drift, founder events) can lead to predictable changes in the frequency of a phenotype within populations (a theoretical pop gen reference here would be nice).
  + Cite loss of S morph in *Eichhornia* as example (e.g. Husband and Barrett 1992a; 1992b; Barrett et al. 1989; Barrett et al. 2009).
  + Therefore, stochastic processes can lead to deterministic outcomes.
* *Paragraph 4*: Where do we study clines
  + Topic: Clines are studied latitudinally, altitudinally and across habitat types (e.g. urban vs. rural).
  + Clines often occur with demographic correlates – population size, inbreeding, founder events, stability, connectivity. DRIFT
  + In some cases we can study parallel clines; cities are a great place to study clines — parallelism
* *Paragraph 5*: Clover and urban-rural clines in cyanogenesis. Brief history of clover as a model system testing for clines. Urban-rural clines in cyanogenesis.
  + Topic: The purpose of this paper is to examine under what conditions non-adaptive processes such drift and gene flow can lead to phenotypic clines in traits with complex genetic architectures. Clover is a classic system for studying clines.
  + Thompson et al. (2016) recently detected parallel clines in the frequency of hydrogen cyanide (HCN) — an antiherbivore defense — across independent urbanization events; three out of four sampled cities show decreased HCN with increasing urbanization.
  + HCN is the result of two, independently assorting Mendelian genes and plants require a functional (i.e. dominant) allele at each locus to produce HCN.
  + Given that the loss of the dominant gene at either locus results in the loss of HCN, the frequency of HCN in populations in more likely to decrease than increase when stochastic forces are the only ones operating (Fig. 1: 3D allele freq. plot).
* *Paragraph 6*: Research questions
  + In this paper, we seek to address the following specific questions:
    1. How do genetic drift, migration, and selection interact to influence the formation of urban-rural 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)?
  + We then parametrize the simulations using estimates of contemporary migration rates and Ne obtained from macrosats to infer the likelihood that observed clines in HCN are the result of stochastic vs. deterministic processes.

**Methods**

* Section 1: Structure of simulations
  + *Paragraph 1*: Overview
    - Spatially-explicit, individual-based simulations coded in Python.
    - Allow independent control over stochastic and deterministic parameters across the landscape (Table 1: Parameters with meaning and ranges)
    - Life cycle of a generation in simulations (Fig. 2: Life cycle)
    - Model where we sample alleles from infinite pool (*What is this called?*)
  + *Paragraph 2*: Within population dynamics and drift
    - Logistic population growth
    - Population created in adjacent cells with some probability that depends on the population’s size (linear relationship, Equation)
    - Drift primarily controlled by the strength of the bottleneck during founder events. Stronger bottlenecks = stronger drift.
    - Drift can also be controlled either by varying *K* across the landscape (Alleaume-Benharira 2006).
    - Show how varying *K* or bottleneck proportion influence *Ne*, calculated as the harmonic mean of census population sizes over time (Wright 1938, supplementary figure)
  + *Paragraph 3*: Simulating migration
    - Migration declines with distance based on exponential distribution (Kimura and Weiss 1964). Makes migration like stepping-stone model since most migration occurs among populations close in space. (Equation)
    - For each population, allele frequencies are calculated based on Wright’s (1931) continent-island model (Equation), where *m* is the average proportion of immigrants arriving into the resident population across all existing populations (Equation), weighted by their population sizes (Equation).
  + *Paragraph 4*: Selection
    - Two-locus selection model (Likely multiple equations here)
    - Each cell in the matrix is given a selection coefficient (*s*). Selection declines linearly with increasing distance from the cell with the strongest selection, simulating an environmental gradient.
    - Selection acts independently on the dominant alleles of both loci
* Section 2: Cases examined in the paper
  + *Paragraph 1*: Overview
    - Ran 1000 simulations for all cases below
    - Number of generations for case 1 = 250, all others = 1000
  + *Case 1*: Matrix filled entirely
    - All cells initialized with populations
    - No creation or bottlenecks.
    - No population growth.
    - *K* varies across the landscape resulting in lower *Ne* inside cities.
    - Models a simple scenario where clover was always present but has reduced effective population sizes in urban environments
    - Vary migration as well.
  + *Case 2*: Drift through founder events and migration
    - Independently vary the strength of population bottlenecks and migration.
    - From results above, select interesting values from each and vary them interactively.
  + *Case 3*: Allele frequency variation
    - Vary frequency of dominant alleles at both loci in the first population interactively
  + *Case 4*: Selection
    - Under most interesting scenarios from above, add selection of varying maximum strengths.
    - Under panmixia, how strong does selection have to be to generate phenotypic clines? (*Should this be a question of the paper?*)
  + *Case 5*: Colonization history
    - How do above results change if initial population starts rural and colonize urban, start urban and colonizes rural, or was always present across the matrix (like case 1 but with selection).
* Section 3: Analyses
  + Response variables:
    - Mean strength of clines across all simulations
    - Proportion of significant positive or negative clines (Figure 3: Mock cline figure with description of positive vs. negative clines).
    - Look for differences based on confidence interval estimation.
    - Something about matrix filling and how data was subset for analyses (last generation only)?

**Results:**

* Present key results that mirror the cases identified in the methods section.

**Discussion**

*Paragraph 1:*Summary of key results and scope of discussion

* Results show that gradients in drift, either through changes in the maximum size of populations or serial founder events, consistently lead to parallel clines in cyanogenesis.
* Selection generates clines that are stronger in magnitude than those caused by drift and overwhelms opposing drift gradients.
* We begin by discussing the importance of our results for the evolution of clines in cyanogenesis. We then discuss the roles of adaptive and non-adaptive evolutionary processes in urban environments followed by a broader discussion on the role of genetic drift in a cline.

*Evolution of cyanogenesis clines in Trifolium*

*Paragraph 2:**Clover clines traditionally studied through the lens of selection*

* Clover has a long history of study among evolutionary biologists, in particular work by Hunor Daday showing continental-scale latitudinal and altitudinal clines in cyanogenesis.
* Longitudinal clines in cyanogenesis have also been documented and on a more local scale, urban-rural clines in HCN across multiple north American cities.
* In all cases, selection is the only mechanism invokes to explain HCN clines. Putative mechanisms often cited include: Minimum winter temperature (i.e. frost), herbivore pressure, and aridity.
* No previous work has explore the potential for drift to explain spatial variation in HCN frequencies, which is likely given the genetics of cyanogenesis.

*Paragraph 3: Results show that the genetic architecture of HCN lends itself to the preferential formation of parallel HCN clines*

* Independent of colonization scenario, drift induced the formation of parallel clines in HCN, with less HCN in urban populations.
* However, clines formed by drift alone were consistently weaker than clines that involved selection.
* This occurs because on its own, drift results in decreases in HCN frequencies in urban populations with low *Ne* but rural populations experience weak drift resulting in phenotype frequencies that oscillate mostly around the initial frequency. In contrast, selection in actively favouring HCN+ genotypes in rural populations and HCN– in urban populations, resulting in a steeper gradient in HCN frequencies across space and thus stronger clines.

*Paragraph 4: Drift alone CAN/CANNOT explain the presence of urban-rural cyanogenesis clines observed across cities*

* To understand whether drift can geneate observed urban-rural cyanogenesis clines, we compared the slopes of observed clines to the distribution of slopes obtained from simulations involving a strong gradient in drift only.
* This approach allowed us to address the following question: What is the probability of observing clines as strong or stronger than those actually observed across cities, if drift were the only mechanism acting?
* From this approach, we found that the proabability of observing a cline as strong or stronger than the weakest observed cline is X.
* This suggests that drift alone is SUFFICIENT/INSUFFICIENT to generate phenotypic clines in HCN.

*Urban environments as replicated systems to study clines*

*Paragraph 5*: Where do we study clines

* Clines are studied latitudinally, altitudinally and across habitat types (e.g. urban vs. rural).
* Clines often occur with demographic correlates – population size, inbreeding, founder events, stability, connectivity. DRIFT
* In some cases we can study parallel clines; cities are a great place to study clines — parallelism

*Paragraph 6*: Cities display gradients in the strength of drift and selection

* Numerous studies have identified changes in the strength of genetic drift with increasing urbanization
* Similarly, convergent evolutionary responses to selection in urban environments has been documented in a diverse range of taxa.
* Convergence in evolutionary mechanisms in urban environments combined with their global replication makes urban systems ideally suited to understanding how drift and selection interact to structure genetic and phenotypic diversity.

*Genetic drift and the formation of clines*

*Paragraph 7:* Parallel clines are the hallmark of adaptation, but drift can lead to parallelism with the right genetic architecture

* The presence of parallel clines has traditionally been considered strong evidence for the role of natural selection
* However, neutral processes can generate clines, which may be axacerbated in the presence of non-additivity.
* Crucially, drift should lead to equal frequencies of positive and negative clines at single loci underlying non-additive traits.
* Our results support this prediction as the mean slope of clines at *CYP* and *Li* was 0, while the mean slope of HCN clines was overwhelmingly positive. This strongly suggests that only the epistatically-determined phenotype is subject to directional change via drift.

*Paragraph 8*: Maintenance of clines under drift, selection, and gene flow

* While drift alone may generate clines in non-additive traits, it is not a sufficient mechanism for maintaining clines as local drift within populations will gradually weaken clines. Our results support this, as the mean slope of clines across simulations became gradually weaker over time in the absence of migration or selection (Figure SX).
* Maintenance of clines requires migration or selection. While too much migration weakens clines by homogenizing alleles across populations, minimal migration can prevent the loss or fixation of alleles from drifting populations, allowing for the maintenance of clines. In our simulations, migration rates on the order of X to X were sufficient to prevent the weakening of clines over time (Figure SX).
* Finally, selection can maintain clines indefinitely but clines become increasingly non-linear as selection is allowed to continue. Linear clines can only be maintained when small amounts of migration prevent the fixation of alleles due to selection.

*Paragraph 9*: Drift-selection balance

* An alternative way for clines in non-additive traits to be maintained is when selection is opporating in opposition to the prevailing drift gradient.
* The presence of opposing drift gradients leads to the preferential formation of clines in the direction opposite to that favoured by selection.
* In our simulations, selection in the presence of opposing drift needed to be X times stronger to generate clines as strong as those generated by selection alone.
* Nonetheless, when decreases in phenotype frequencies due to drift equal the increase in phenotype frequency due to selection, clines in non-additive traits can be maintained.

*Paragraph 10*: How common is this phenomenon?

* Directional change in phenotype frequencies due to drift in natural populations has been observed in other systems (e.g. *Eichhornia*).
* May be quite common and is expected to occur whenever traits are controlled by two or more loci in a non-additive manner.
* We would expect the extent to which phenotypes are susceptible to the formation of clines via drift to increase with the number of epistatically-interacting loci.
* While we focus on epistasis, the same pattern may occur when non-additivity results from dominance, although perhaps not as strongly.

*Conclusion*

* Drift can lead to the formation of parallel clines in non-additive traits despite equal frequencies of positive and negative clines at underlying loci.
* Given this, we suggest that drift should be considered a null model to be rejected prior to invoking selection in the formation of parallel clines in non-additive traits.
* Evidence in favour of selection could come in two ways: (1) Showing that more clines are observed than would be expected under drift alone. (2) Showing that single-loci underlying the focal trait consistently cline in the same direction, as only the phenotype is subject to directional change via drift.
* Addressing point (1) above requires large scale replication. Urban environments and the convergence in gradients in drift and selection they generate can provide the replication necessary to address this prediction.