Speed and diversity of evolution in subdivided populations

by: Alec Kennedy, Diana Su, and Peter Ralph.

Introduction:

Between 1975 and1995, researchers on Vlieland, an island in the Netherlands, found an interesting genetic story. It was found that the subpopulation of a species of birds, the great tits (*Parus major)*, that were born in the Eastern part of the island had a much higher female survival rate than the females born in the Western part. This finding was interesting because it was hard to believe that two populations of a single species that lived in such close proximity would differ so greatly in survival on the same island. The study went on to find that this difference had to do with the number of immigrants coming into the island. The Western part of the island had almost four times as many immigrants come in than the East. The immigrants brought with them a trait that was not well suited for the local environment in Vlieland. The trait then took over in the Western part of the island because of the huge inflow of immigrants from other areas (Postma *et al.*, 2005). This study brings up an interesting point: migration increases the gene flow into a population while barriers to migration allow adaptation to local environments to occur.

How do these barriers to migration affect the time it takes a newly introduced advantageous trait to take over? Under different spatial conditions, how do the extents in which the trait takes over differ? The theory of evolution attempts to explain how genetic changes occur from one generation to the next. While it seems that often this theory covers the temporal aspect of abstract, spatially unstructured biological processes because of the relative ease of analysis of unstructured models, a number of articles have investigated the effects of spatial structure on evolution. Physical boundaries that affect the migration rates in and out of populations often influence the way genes become mixed together within a population (Avise, 1994).

How will local adaptation be affected by different migration and selection rates? When an organism adapts to its environment, the theory of natural selection says that the traits that make it more likely for the organism to survive and reproduce will tend to become more common over successive generations. Local adaptation, as we will define in this paper, will occur when the mean fitness (the capability of an individual of a certain genotype to reproduce) of a population is higher in its own habitat than in another (Gandon *et al.* 2002). There is no reason to believe that all locations a population occupies are homogenous. Some local environments may select against certain traits that are advantageous in others. When there are routes for migration of individuals between areas, genes can get mixed, and the intensity of this mixing generally increases with the rate of migration. Given a large network of subpopulations of a species it is interesting to look at the different effects of varying the number of connections between subpopulations as well as the heterogeneity of localized selection rates or the degree to which some areas cater more to a certain type compared to others.

The motivation for this project was to look at the role spatial structure plays in evolution, or more specifically the spread of advantageous genes in a population broken up into a number of subpopulations. In order to study the impact that physical boundaries play in gene flow we have created a simplified model in order to simulate a large number of generations of a haploid, semelparous (they die after they reproduce) species with a newly introduced advantageous mutation. With these simulations we have collected data from similar starting populations put into different situations with different spatial structures and selection heterogeneity. The differing selection heterogeneity was used to look at the effect of having different areas having different responses to the mutation. <u>The idea is to measure the speed of the adoption of new traits, how much the advantage takes over, how much it varies in the process, and how "local" the adaptations are</u>. We hope that these measures will provide some useful ideas of how spatial structures as well as heterogeneity in selection rates play roles in evolution.

Methods:

In building our model, we first assumed that individuals in the population are semelparous organisms, that is, each new generation of individuals (the offspring) replaces the old generation of individuals (the parents) upon birth. We further assumed that individuals in the population are haploid and reproduce asexually. Our model only considers two phenotypes: mutated or nonmutated (wild type). The mutation is in all cases advantageous for survival, although to varying degrees.

We are interested in the effect of spatial structure on the speed of evolution (computed as the number of generations before the majority of the individuals in the population adopts the mutation rather than the nonmutation), so we developed three separate migratory pattern models. We represent the migratory patterns by a graph, where equally-sized subpopulations exist at each node and migration occurs along the edges. These are: the *linear* graph; the partially linear or *partial* graph; and the *complete* graph. The linear graph is a single one-dimensional string of a finite number of locales. The partial graph is constructed from the linear graph by adding independently, with probability 0.5, to each locale an additional migratory path that connects the locale to a randomly chosen other locale. The complete graph is the free system, in which every pair of locales in the entire habitat space is connected by a migratory path. We often refer to the "amount" of spatial structure, defined as follows: the complete graph has the least spatial structure; the partial graph has the intermediate spatial structure; and the linear graph has the most spatial structure.

We also developed three separate selection scenarios: the *localized* selection scheme; the *incomplete* selection scheme; and the *homogeneous* selection scheme. We defined the localized selection scheme akin to a lottery system, in which a single randomly chosen locale's environment is extremely advantageous to the mutation. In the incomplete selection scheme, exactly half of the locales, chosen uniformly at random, possess environmental traits that are advantageous to the mutation. In both the localized and incomplete selection schemes, the remaining locales possess environmental traits that are neither advantageous nor disadvantageous

to the mutation. We defined the homogenous selection scheme so that all of the locales possess environmental traits that are advantageous to the mutation. The fitness values for each locale for the mutation are computed such that the wild type has fitness value one, and the average fitness value of the mutated type across the total habitat space is constant regardless of the selection scheme.

Based on these models, we examine the effect of the different spatial structures and selection schemes on the evolution of a starting population of two thousand individuals. We also considered three different numbers of distinct locales, the number of which also contributes to the spatial structure: the two groups case, where the total habitat space is separated into two separate locales; the ten groups case, where the total habitat space contains ten locales; and finally, the twenty groups case, where twenty locales exist in the total habitat space. In total, we consider twenty separate combinations of spatial structures and selection schemes, as listed in Figure 1. Note that when there are only two locales, the different spatial structures are all equivalent.

Figure 1. List of spatial structures and selection scenarios, separated by the number of locales in the total habitat space.

Locales	Spatial Structure; Selection Scenario
2	• Linear/Partial/Complete; Incomplete/ Localized • Linear/Partial/Complete; Homogeneous
10	• Linear; Localized Linear; Incomplete Linear; Homogeneous Partial; Localized

	Partial; Incomplete Partial; Homogeneous Complete; Localized Complete; Incomplete • Complete; Homogeneous
20	• Linear; Localized Linear; Incomplete Linear; Homogeneous Partial; Localized Partial; Incomplete Partial; Homogeneous Complete; Localized Complete; Incomplete • Complete; Homogeneous

We always begin each simulation with the same number of individuals in each locale, and this number is chosen so that the total starting number of individuals in the population is two thousand. For each generation of new individuals, we first *migrate* each individual independently of all other individuals. The probabilities of staying in the current locale or migrating from the current locale to the new locale are determined by the spatial structure. If an

individual is in a locale with k neighboring locales, then that individual will migrate with probability $k/20$, and stay put with probability $1-k/20$. If an individual migrated, it would choose a neighboring locale to migrate to uniformly at random. In none of our simulations did any locale have more than 19 neighbors. Based on these specifications, we constructed a migration probabilities matrix based on the type of spatial structure we are examining. Examples of spatial structures and their corresponding graphs and migration probability matrices are given in Figure 2.

Figure 2. Spatial structures and their corresponding graphs and migration probability matrices for a habitat space with 4 locales.

A. Linear

$$
\begin{array}{c c c c c}
 & 1 & 2 & 3 & 4 \\
1 & 0.95 & 0.05 & 0.00 & 0.00 \\
2 & 0.05 & 0.90 & 0.05 & 0.00 \\
3 & 0.00 & 0.05 & 0.90 & 0.05 \\
4 & 0.00 & 0.00 & 0.05 & 0.95
\end{array}
$$

B. Partial

$$
\begin{array}{c c c c c}
 & 1 & 2 & 3 & 4 \\
1 & 0.90 & 0.05 & 0.00 & 0.05 \\
2 & 0.05 & 0.90 & 0.05 & 0.00 \\
3 & 0.00 & 0.05 & 0.90 & 0.05 \\
4 & 0.05 & 0.00 & 0.05 & 0.90
\end{array}
$$
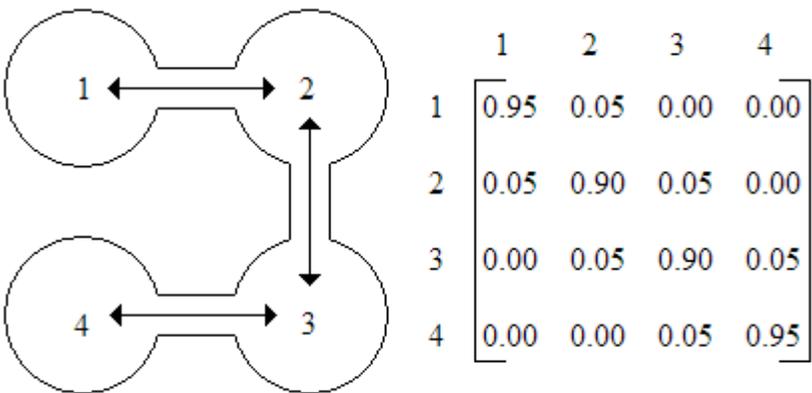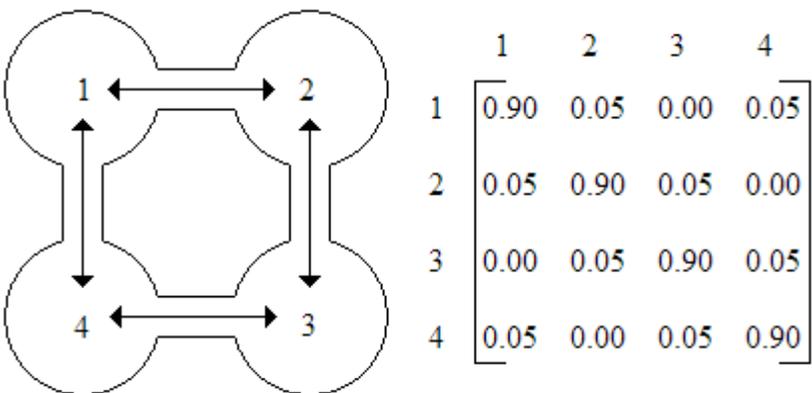
C. Complete
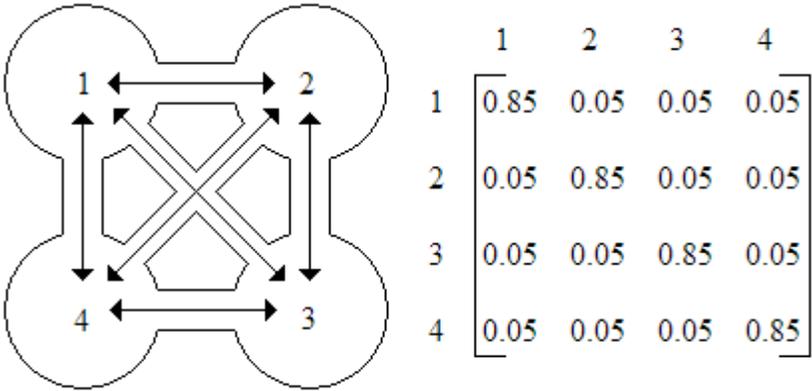
After migration, we then assigned fitness values to each individual in the population based on their current locale and their phenotype, where the fitness values are determined by the selection scheme. No individuals in the starting population are mutated. Because the wild type is not advantageous nor disadvantageous for survival, we assign a fitness value of 1.00 for individuals with the wild type, regardless of their current locale. We similarly assign a fitness value of 1.00 for individuals with the mutation currently in locales where the mutation offers no advantages. We compute the fitness value for individuals with the mutation currently in locales where the mutation is beneficial as follows:

$$f_l = 1 + nf$$

$$f_i = 1 + nf/2$$

$$f_h = 1 + f$$

where f_l is the fitness value for the localized selection scheme, f_i is the fitness value for the incomplete selection scheme, f_h is the fitness value for the homogeneous selection scheme, n is the number of locales present in the habitat space, and f is the advantage the mutation offers compared to the nonmutation. We fixed f to be 0.025. Examples of selection schemes and their corresponding diagrams and fitness value matrices are given in Figure 3.

Figure 3. Selection schemes and their corresponding diagrams and fitness value matrices for a habitat space with 4 locales.

A. Localized
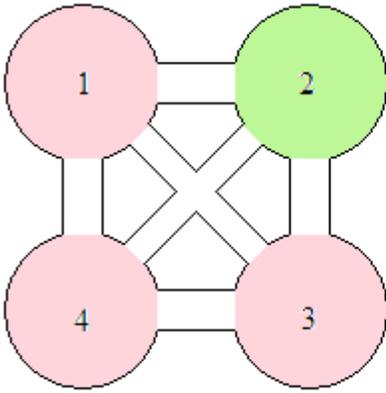
$$\begin{array}{cccc} & 1 & 2 & 3 & 4 \\ N & [1.00 & 1.00 & 1.00 & 1.00 \\ M & 1.00 & 1.10 & 1.00 & 1.00] \end{array}$$

B. Incomplete
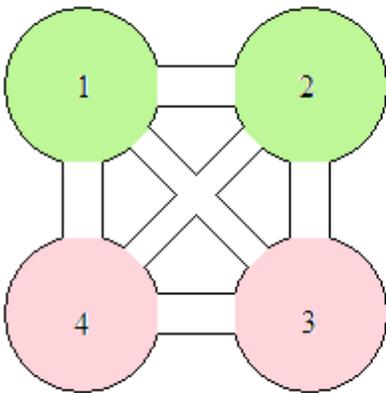
$$\begin{array}{cccc} & 1 & 2 & 3 & 4 \\ N & [1.00 & 1.00 & 1.00 & 1.00 \\ M & 1.05 & 1.05 & 1.00 & 1.00] \end{array}$$

C. Homogeneous
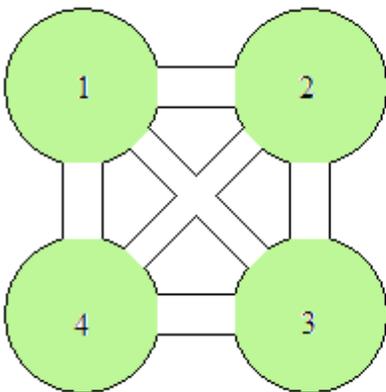
$$\begin{array}{cccc} & 1 & 2 & 3 & 4 \\ N & [1.000 & 1.000 & 1.000 & 1.000 \\ M & 1.025 & 1.025 & 1.025 & 1.025] \end{array}$$

We then allow individuals in the population to produce offspring. Each individual will be assigned a *birth rate* based on the current number of individuals in her locale and her fitness value, and given the birth rate, the number of offspring that each individual has is Poisson with mean b, independently of all other individuals. The birth rates are chosen so that the number of individuals in each locale exhibits density-dependent fluctuations about a carrying capacity (which is equal to the initial number of individuals in the locale). If the carrying capacity is denoted by C, f_{ik} is the fitness value for the individual based on its current locale i and its phenotype k, and f_i is the sum of the fitness values for all individuals in the same locale i, then we can compute the birth rate b as

$$b = C\, f_{ik} / f_i .$$

After reproduction, we rearrange the population to incorporate the new individuals. Since all individuals in the population are semelparous, we ignore previously existing individuals (the parents) and examine subsequently existing individuals (the offspring). We assign each offspring the same phenotype as the parent. We similarly assign each offspring the same locale as the parent.

We finally mutate the individuals to determine their new phenotypes. Each new individual is assigned the type different to that of their parent with probability 1/20, independently of all other individuals. Census data is then collected for each generation of individuals to calculate the number of mutated individuals and the number of nonmutated individuals for each of the locales.

Results:

We completed 50 simulations with 500 generations per run for each combination of spatial structure and selection scheme, as listed in Figure 1. We define a generation as the life cycle of an individual in the population, characterized by birth, migration, reproduction, and death. For each of the simulations, we collected the following census data each generation: the number of mutated individuals and the number of nonmutated individuals in each locale.

We first examined the counts of the nonmutated individuals in the population per generation compared to the counts of the mutated individuals in the population per generation. As seen in the sample plot in Figure 4, the colored solid lines are the counts of the nonmutated individuals per locale over the generations and the colored dotted lines are the counts of the mutated individuals per locale over the generations. The black solid lines are the averages of the counts of the mutated and the nonmutated individuals for the overall population over the generations. The intersection between the black solid lines occurs when half of the population adopts the mutation rather than the nonmutated phenotype.

Figure 4. Sample plot of a simulation run as counts.

complete structure; homogeneous advantage 20 groups
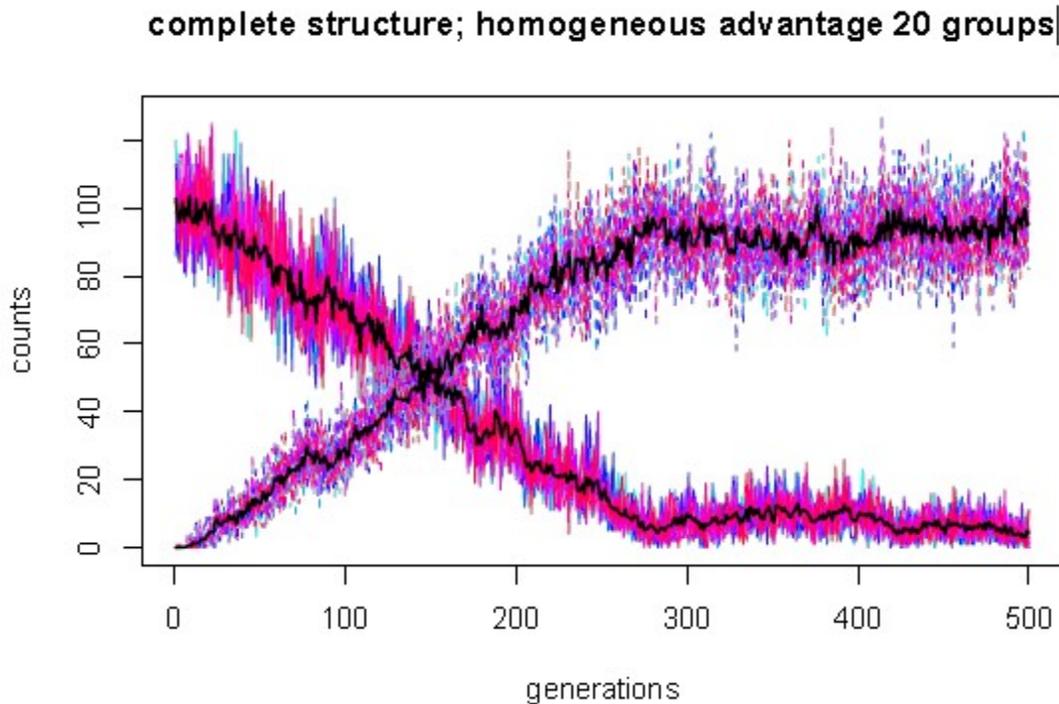
Note the density-dependent birth rates keep the population size more-or-less constant, although the degree of variation varied in different spatial structure and selection heterogeneity. We converted the counts into proportions to simplify the analysis of the plots. As seen in the sample plots in Figures 5 and 6, the colored solid lines now are the proportions of the nonmutated individuals (denoted as Type 1) compared to the mutated individuals (denoted as Type 2) per locale over the generations. The black solid line is the proportions for the overall population. About the black solid line are dotted lines at plus/minus two standard deviations of the by-locale proportion, summarizing the amount by which the proportion in each locale differs from the total population proportion, a quantity which we refer to as *spatially structured diversity*.The vertical dotted line is the first generation when the population proportion drops below 0.5, corresponding to the number of generations elapsed before half of the population adopts the mutated rather than the nonmutated phenotype. The horizontal dotted line is the population proportion at equilibrium, namely the frequency about which the population proportion fluctuates. The population proportion at equilibrium measures the degree to which both phenotypes are still around significantly, which we refer to as *unstructured diversity*.   The dotted lines surrounding the population proportion at equilibrium are at plus/minus two standard deviations of this same quantity, measured for each locale.

Figure 5. Sample plot of the simulation run in Figure 4 as proportions.

complete structure; homogeneous advantage; 20 groups
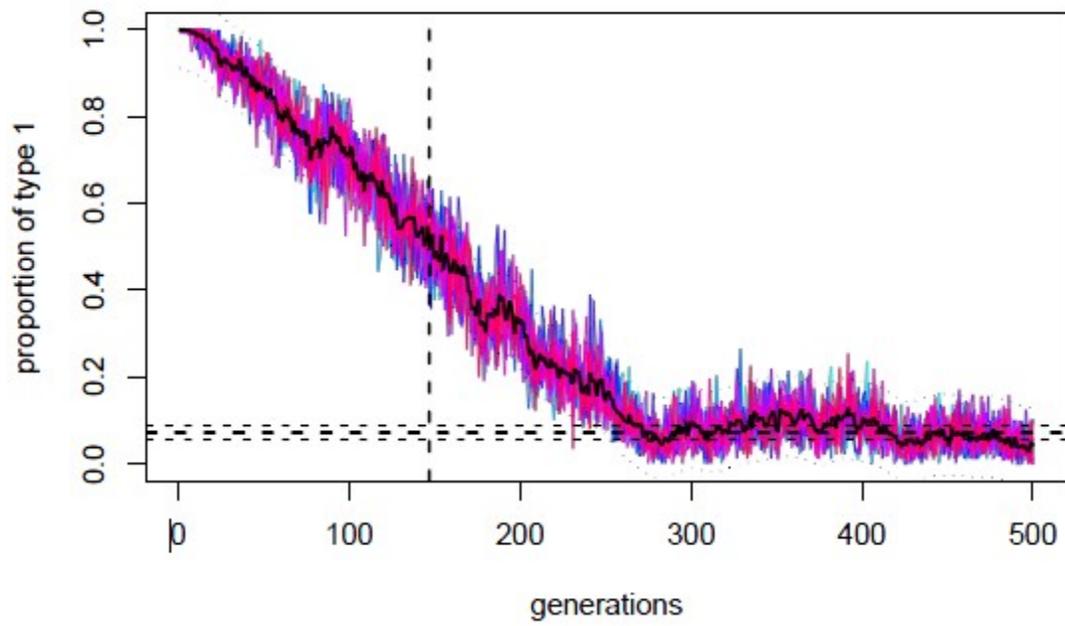
Figure 6. Sample plot of a simulation run as proportions, for comparison to Figure 5.

partial structure; localized advantage; 20 groups
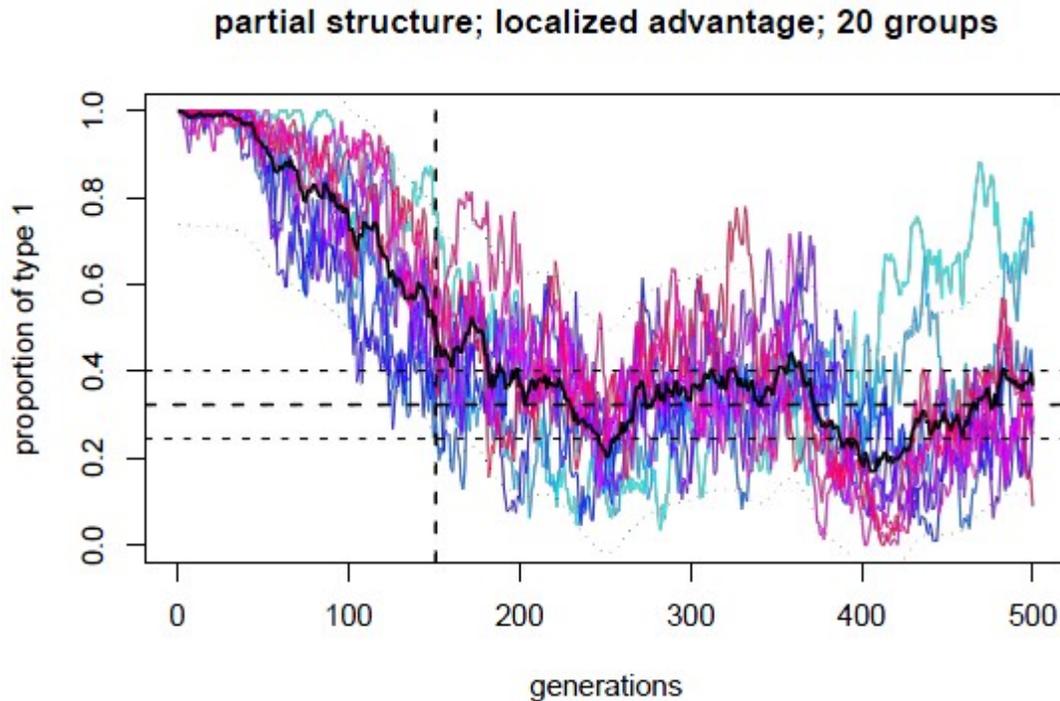
Since we are interested in the effects of spatial structure and selection heterogeneity on the speed of evolution, we examined the relationship between the population proportion at equilibrium (denoted *peq*) and the number of generations elapsed before half of the population adopts the mutation rather than the nonmutation phenotype (denoted *tz*). From Figure 7, we notice that *tz* increases with more spatial structure and more selection heterogeneity, particularly when the number of locales in the habitat space increases. We notice also that *tz* increases with more locales present in the habit space, with the exception of the combination of localized selection scheme and partial graph. For this particular combination, the 20-locales case has lower *tz* compared to the 10-locales case. We also observe that *peq* increases with more spatial structure and more selection heterogeneity. However, the effects of spatial structure and selection heterogeneity work in combination on both *tz* and *peq*. With complete spatial structure, selection heterogeneity has little effect. With homogeneous selection scheme, spatial structure has little effect. Hence, spatial structure and selection heterogeneity cannot be treated as wholly separate factors influencing the adoption of the mutation phenotype over the nonmutation phenotype.

Figure 7. Plots of the number of generations elapsed before half of the population adopts the mutation rather than the nonmutation phenotype (*tz)* against the population proportion at equilibrium (*peq*). Each circle represents a single simulation run, characterized by the combination of spatial structure and selection scheme. The circles have differing sizes that represents the number of locales present in the habitat space: the 2-locales case has the smallest circles; the 10-locales case has the intermediate circles; and the 20-locales case has the largest circles. Points are colored according to variance in the equilibrium frequency between locales.
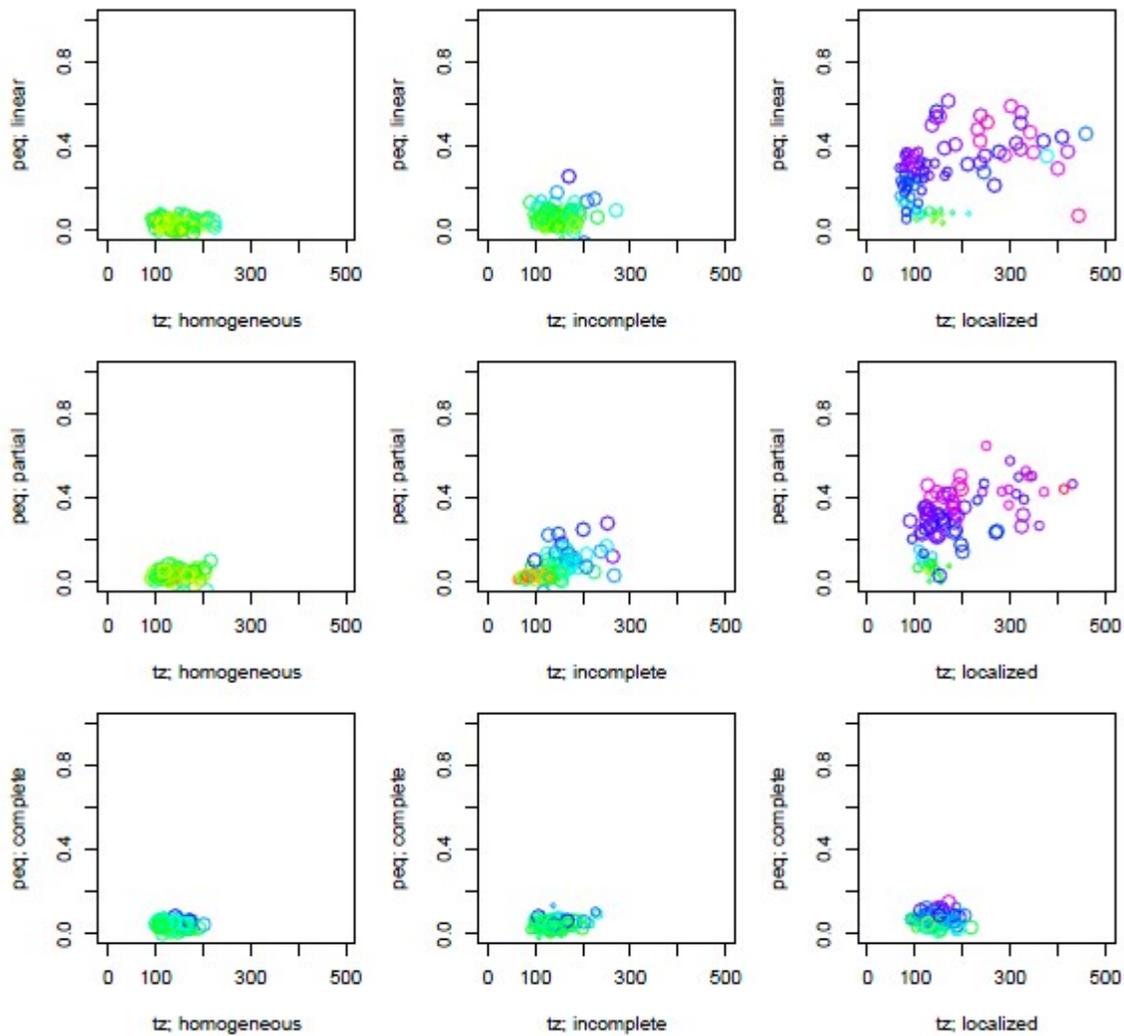
From Figures 5, 6, and 7, we clearly observe that the standard deviations of the population proportion at equilibrium varies across the different combinations of spatial structures and selection schemes. From Figure 8, we conclude that these differences are significant. The standard deviations for the homogeneous selection scheme are the lowest out of all the selection

schemes, followed by the incomplete selection scheme and finally the localized selection scheme. The standard deviations for the complete graph are the lowest out of all the spatial structures, followed by the partial graph and finally the linear graph. Again, the standard deviations increases in combination of the spatial structure and the selection heterogeneity.

Figure 8. Box plot of the standard deviations of the proportion at equilibrium by locale, as separated by the different combinations of spatial structures and selection schemes listed in Figure 1. In the labels on the *x*-axis, the spatial structure is listed first (complete/partial/linear) and the selection heterogeneity is listed second (local/incomplete/homogeneous), abbreviated by the first three letters.

equilibrium frequency variation
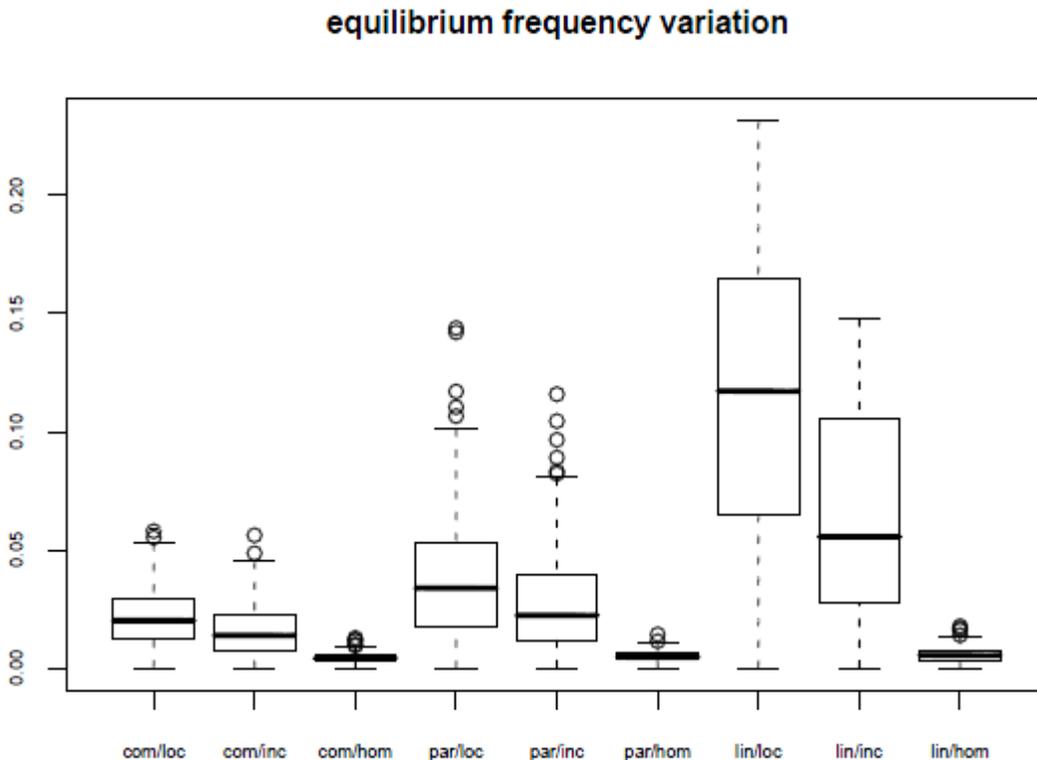
From Figures 5 and 6, we also clearly observe that the standard deviations of the subpopulation proportions per locale differs significantly. Since we are interested in the effects of spatial structure on local adaptation of the mutation phenotype, we examined the between-subpopulation variations based on the different spatial structures. As seen in Figure 9, the complete graph has the lowest variation values out of all the spatial structures. While the partial graph has lower variation values compared to the linear graph, the variation values for the partial and linear graphs differ less significantly to each other in comparison to the variation values for the complete graph.

Figure 9. Box plot of the standard deviations of the subpopulation proportions per locale, as separated by the different spatial structures.

Between-supopulation variation
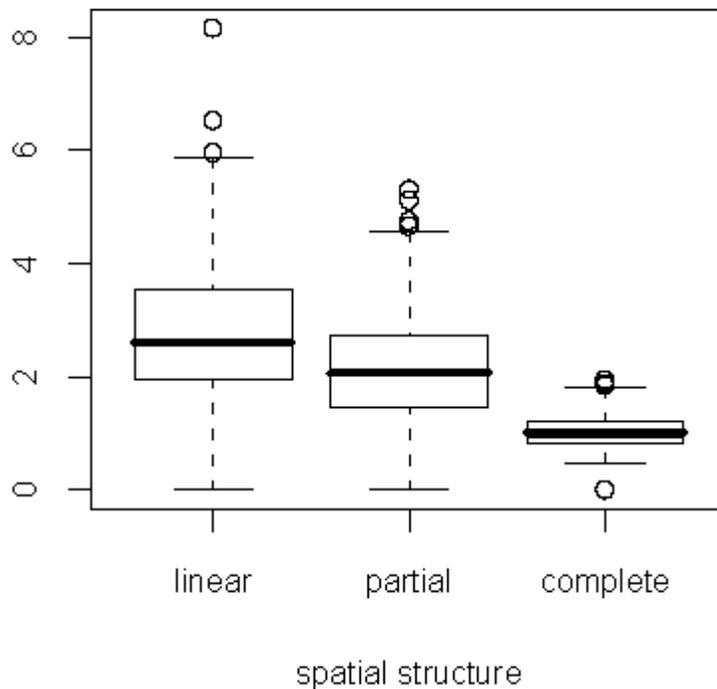
Conclusions:

       We are interested in how different spatial structures and selection schemes affect the speed of adoption of a newly introduced advantageous trait and local adaptation of this mutation phenotype. We are also interested in the differences in how much the population adopts the mutation and how much this adaptation fluctuates in the process.

       We expected that the more spatial structure the habitat space has, the longer it takes before the population adopts the mutation. We also expected that the more heterogeneous the selection scheme is, the longer it takes before the population adopts the mutation.

       Based on our analysis, increasingly spread out spatial structure, at least in the partial and linear graphs compared to the complete graph, increases the number of generations required before half the population adopts the advantageous trait. More spatial structure also prevents the population from adopting the advantageous trait as much as compared to less spatial structure; more spatial structure increases unstructured diversity. Similarly, selection schemes that favor local adaptation compared to overall adaptation increases the number of generation required before half the population adopts the advantageous trait and prevents the population from adopting the advantageous trait as much. The effects appear only in combination: with complete

spatial structure, selection heterogeneity has little effect; and with homogeneous selection scheme, spatial structure has little effect.

We observe that increasing the number of locales present in the habitat space also increases the number of generations required before half the population adopts the advantageous trait and prevents the population from adopting the advantageous trait as much. Adding more locales to the habitat space is like to adding more spatial structure. However, we had noted that adding more locales to the habitat space from 10 locales to 20 locales for the partial graph and localized selection combination actually decreased the number of generations required before half the population adopts the advantageous trait and facilitates the population in adopting the advantageous trait. We surmise that the anomaly arises because of the differing number of added migratory paths: the 10-locale case had 5 additional migratory paths while the 20-locale case had 10 additional migratory paths.

Finally, we notice that increasingly spread out spatial structure leads to more local adaptation and more structured diversity. We also noted that increasingly heterogeneous selection scheme leads to more structured diversity. However, as stated before, the effects work in combination. Thus, selection heterogeneity contributes little to structured diversity when the spatial structure is the complete graph.

Future Directions:

First of all, we did not look at all possible spatial structures, just the two extremes and one in-between. The study described in this paper just gave a general idea of spatial structures part in evolution. Future simulation studies can take into account more structures if interested.

An assumption made was that the mutation was either advantageous or the same in all areas. We did not consider the situation in which the mutation was worse off in some areas. It would be interesting to see if having the mutation be deleterious in some places would change the results.

The model also only applied to haploid, semelparous organisms. We only looked at two types of genes (mutated or non-mutated) and assumed that populations died after giving birth to offspring. The model did leave open the possibility to add more than two types and could be easily manipulated to include diploid individuals. It would be interesting to look at the effects of spatial structure and heterogeneity in selection rates on individuals with dominant and recessive traits. Also, the model does not necessarily have to describe semelparous organisms. It could easily be assumed that some of the offspring from the next generation are still the same as some of the individuals from the previous generation who continue to survive.

Furthermore, it has been shown that a wide class of similar models in population genetics satisfy a "central limit theorem", in that the trajectories of the allele frequencies converge to the

Wright-Fisher diffusion, or related island models, as the total population size tends to infinity. See Ewens (2004) or Shiga (1988). Our model certainly falls in this category, and so our results can be expected to apply to populations of sufficient size that do not satisfy the above assumptions.

References

Avise, John C. *Molecular Markers, Natural History and Evolution*. Chapman & Hall, London. (1994)

Ewens, Warren. *Mathematical Population Genetics*. Springer . (2004)

Gandon, Sylvain and Yannis Michalakis. "Local adaptation, evolutionary potential and host-parasite coevolution: interactions between migration, mutation, population size, and generation time." *Journal of Evolutionary Biology*. Vol. 15 Issue 3, 451-462. (2002).

Postma, Erik and Arie J. van Noordwijk. "Gene flow maintains a large genetic difference in clutch size at a small spatial scale." *Nature* 433, 65-68 (2005).

Shiga, Tokuzuo. "Stepping stone models in population genetics and population dynamics."
    In: *Stochastic processes in physics and engineering (Bielefeld, 1986)*, Math. Appl.
    Vol. 42, 345--355, Reidel, Dordrecht (1988).