

Spread of an Advantageous Allele through Space and Time

Kyle Shaw and Peter Beerli, Department of Scientific Computing, Florida State University, Tallahassee FL USA

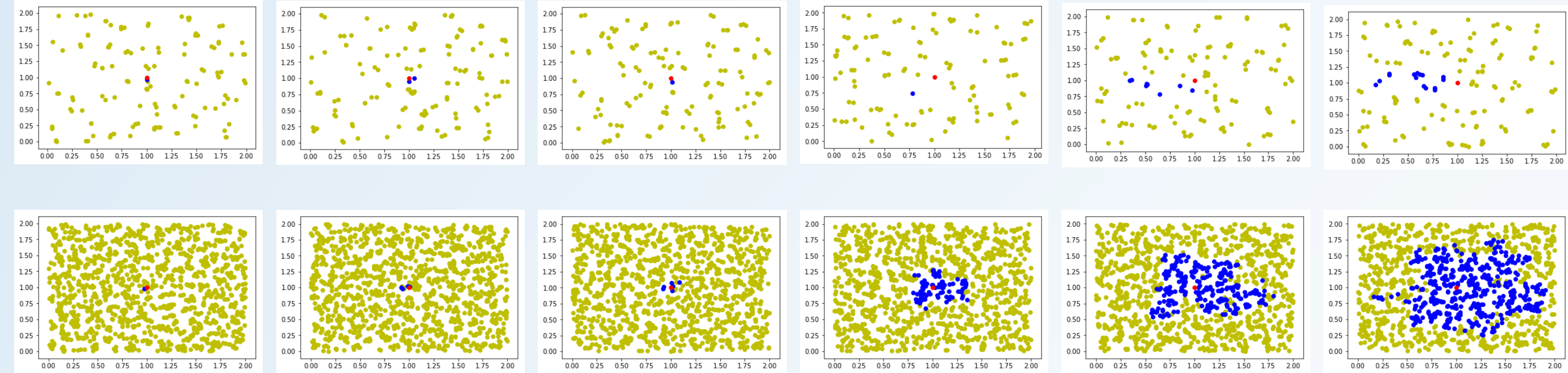
Genetic data is intricately tied to the geographic point in which the sample was collected. Each member of a species can only travel so far, thus genetic data can only travel so far in each generation. Thus, population structure and geographic features can have stark effects on the patterns found in genetic data. Yet few modern methods include geographic data, and of those few none do it very well. Many researchers are attempting to change this. I have several simulation experiments that will help the field understand how geography effects the genetic patterns. A simulated advantageous allele experiment shows that new mutations often spread through the population like a wave in a pond. A structured population simulation helps reveal how difficult it can be for genetic data to travel through sparse population patches. These methods will in turn help with the creation of new more accurate methods.

Introduction

Every species lives, moves, reproduces, and dies in an ever-changing environment. This environment affects how individuals interact. Harsh environments might increase the competition between members of a species and lower the population density. This in turn decreases the number of mates an individual has available to choose from. Low population sizes are known to decrease genetic diversity. On the other hand, Bountiful environments might allow a population to grow to plentiful sizes, thus increasing genetic diversity. The number of mates available to choose from is intrinsically tied to genetic diversity. However, population density can not be simply defined as number of individuals per square mile. Some populations, such as wolves, deer, and even humans can travel quite a ways to find a mate, while other species such as frogs, snails, and turtles may not travel so far. Thus, it is more realistic to define density in terms of number of individuals within a given dispersal distance.

It is also worth noting that every single new mutation can only spread from the point of origin 1 dispersal distance every generation. Thus, all relatively new mutations will be highly correlated with geographic location, while old mutations that have saturated the population will have lost that correlation. The same is true for genetic patterns which can be created/modified through a process known as recombination. This correlation between genetic patterns and geographic location is often ignored in panmictic models. Non-panmictic models, or continuous population models are emerging as a hot topic to explore how geography affects evolution over time.

Travelling Wave Affect



Here are several snapshots of a population simulated through time and space. Each dot represents an individual; with blue dots representing individuals who contain a new advantageous allele. Yellow dots represent individuals without the new allele. The first graph on the left is first in time, with only one individual containing the new mutation. Every snapshot to the right is taken after several generations have passed. The red dot marks the point of origin for the mutation. The top slide show is a sparsely populated, while the second row is very densely populated. Notice how the spread is much more symmetric in the dense population. This is because randomness has less of an effect. The mutation also tends to spread faster and farther in dense populations.



This graph shows the velocity vs Density. Each simulation was run over several geographic ranges as shown in the key. Here we can see two things. First, as density increases, velocity increases as well, with diminishing returns. Second, beyond a certain density, velocity is independent of time.

Using random walk theory and taking the expected Maximum distance from the origin of k random walks. We can calculate a probabilistic lower bound for the speed of a travelling wave.

$$v = \frac{\sigma k \sqrt{\frac{t\pi}{2}} \sum_{i=0}^{k-1} \binom{k-1}{i} \frac{-1^i}{(i+1)^{1.5}}}{t}$$

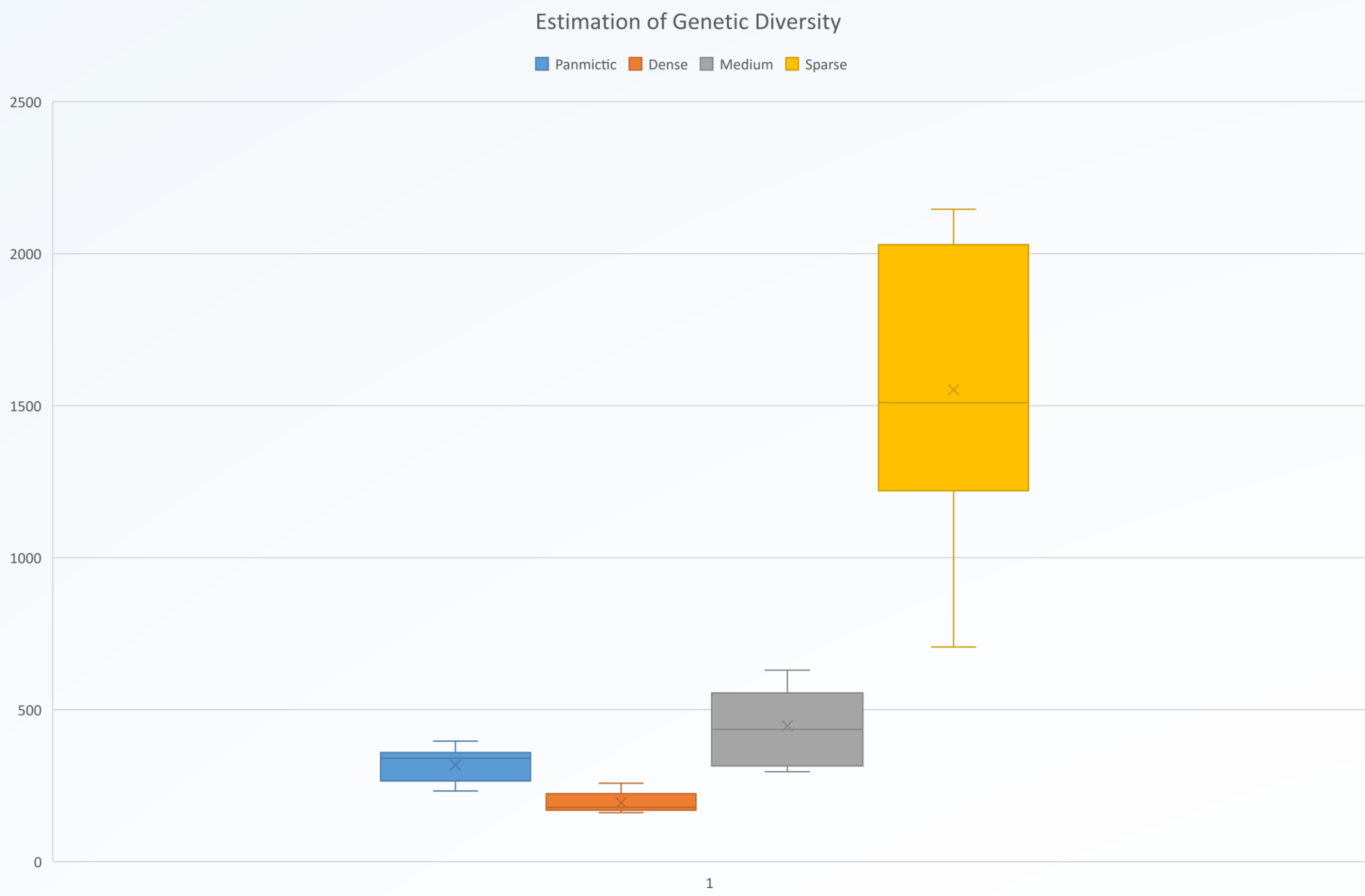
$$= \sigma k \sqrt{\frac{\pi}{2t}} \sum_{i=0}^{k-1} \binom{k-1}{i} \frac{-1^i}{(i+1)^{1.5}}$$

This formula is not very useful. First, it is very unstable to calculate. And second, k changes every generation as new individuals are born. I know from experimentation that the velocity is independent of time, thus somehow the summation must cancel the root of t.

Affect of Density on Genetic Diversity

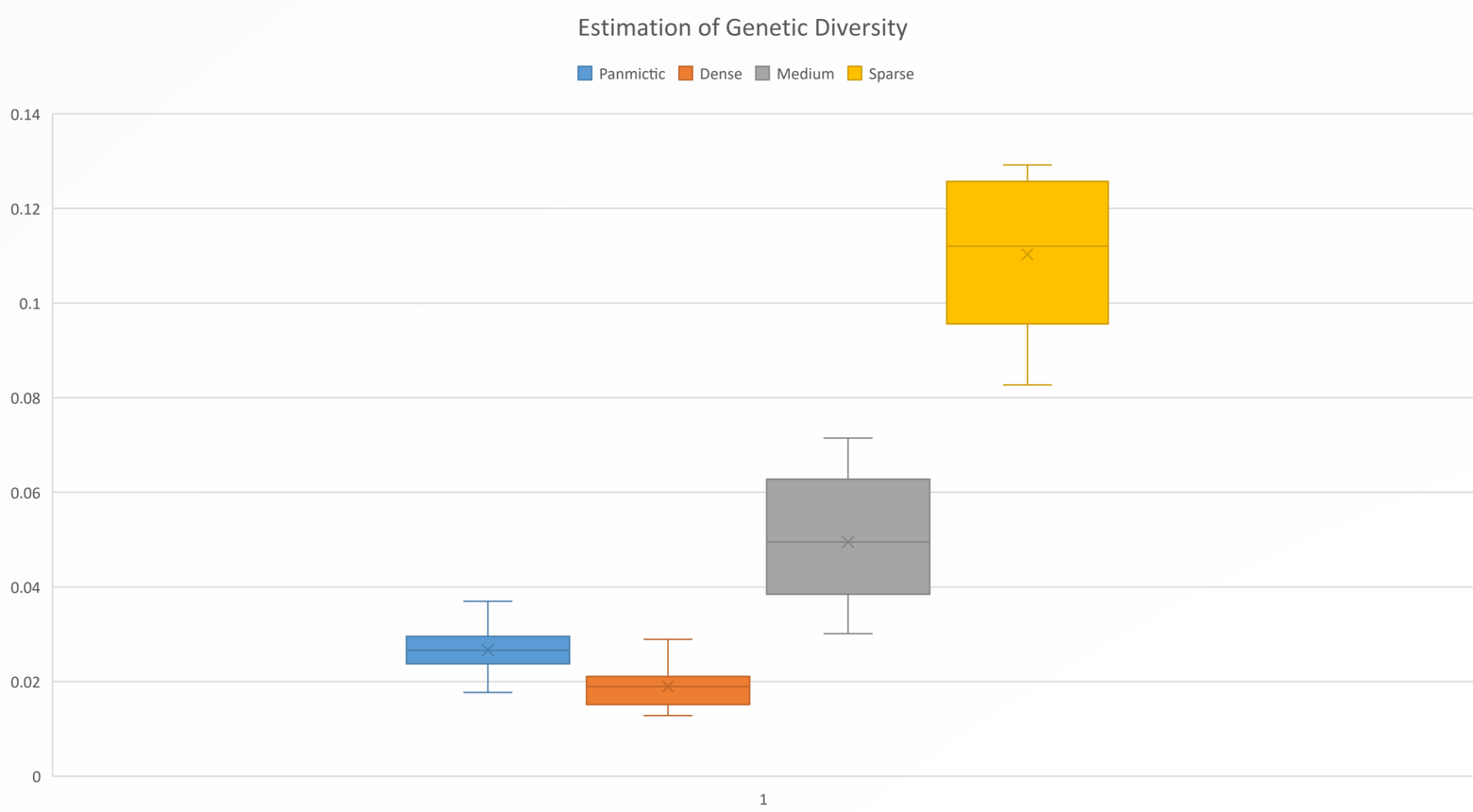
A recent paper shows how dispersal distance can affect genetic diversity. They found that there is a significant affect densities are in the range of 100 or less. Populations with densities of greater than 1000 seem to be very similar to panmictic populations. However, in their simulations they only vary the dispersal distance. Which under their model will increase competition and lower population size, thus, I believe, confounding their results slightly. As a reminder, increasing dispersal distance will increase local density and thus increase genetic diversity. However lowering population will decrease genetic diversity. To truly understand the effect of population density on a populations genetic diversity we must hold all parameters constant except for those being tested.

I built a similar model, but I hold population size and dispersal distance constant. I vary density through the size of the landscape. Larger landscapes allow individuals to spread out more and facilitate a sparser population. Population size is held constant through the competition parameter. I simulated 3 populations, a Sparse, a dense and a medium, as well as a panmictic population. I then measured genetic variability and show the results below in some box and whisker plots.



Watterson Theta

Watterson theta is a commonly used summary statistic used to measure genetic diversity in a sample of a population. Calculated as $K/\ln n$ where K is the number of SNP's and n is the nth harmonic Number with n as the number of samples.



Migrate

Migrate is a program that uses the Monte Carlo Markov Chain algorithm to estimate several population parameters. I used Migrate to estimate Theta, a representation for genetic diversity.

While my results are very similar to that discovered by Battey et al. It is interesting to not that the dense population shows less genetic diversity than that of the panmictic population. More exploration into this is needed.

Future work

Further work on the analytical solution to the traveling wave. A computational model that could predict the speed at which new mutations travel across a population could be very useful. At the moment, diffusion approximations look promising.

Exploring the effects of population structure on genetic diversity, and even on genetic patterns will continue. The next step is to determine the relative strengths of the effects population density and population size have on diversity. It would also be interesting to see how more complicated population structures (i.e. pockets of sparsity) affect genetic variation.

Another promising vein of research is the incorporation of machine learning. Rather than personally exploring the data and discovering genetical patterns, then developing summary statistics to describe the data and make predictions we let a neural network discover those patterns and make a prediction. Currently this is being done by Dan Schnieder. He recently presented at our colloquium about his use of neural networks with predicting selective sweeps. I have started to work with him to include geographic coordinates as input into the neural networks to determine if the extra data could help increase the accuracy.