

The effects of linkage disequilibrium on maintaining balanced polymorphisms in seasonal adaptation

Thomas O’Leary, Csenge Petak, Alison Hall

13 December 2019

1 Abstract

Understanding the mechanisms of preserving genetic diversity is paramount to predicting the capacity of organisms to adapt to changing environmental conditions. In environments characterized by seasonal fluctuations, certain alleles may be beneficial under certain conditions and harmful in other environments. This fluctuation of additive benefits results in a strong directional selection that varies according to the season. Bergland et al. (2014) experimentally observed that seasonal variation is able to preserve polymorphisms that maintain genetic diversity. Wittmann et al. (2017) modeled how allelic dominance changes according to the season in an organism to account for different seasonal life-history strategies. They present a term called “segregation lift” which describes instances of seasonal allele-fluctuations which serve to maintain genetic diversity over time. In this paper, we build upon the Wittmann et al. (2017) model, using the same parameters to study how genetic linkage and seasonal balance may affect segregation lift. The Wittmann et al. (2017) model found that in an unlinked genome the dominance parameter needed to be greater than 0.5 for segregation lift to occur. Our model suggests that in diploid organisms exhibiting genetic linkage, segregation lift occurred when dominance=0.5, suggesting that multilocus polymorphisms are maintained if, at a certain locus, one allele over the other is beneficial during one season.

2 Introduction

Genetic polymorphisms can provide a cache of diversity that contributes to adaptation in highly variable ecosystems [Gulisija2016]. Fluctuating selection can maintain polymorphisms in seasonal environments if strong selective pressure towards both homozygous genotypes varies seasonally (Novak and Barton 2017, @Bergland2014). Seasonality can drive powerful selection fluctuations, especially for organisms which have short a generation time of less than a month (Messer, Ellner, and Hairston 2016). Understanding the role of seasonality plays in maintaining genetic polymorphisms can allow more accurate predictions to be made about the capacity for organisms to adapt to climate change conditions (Fournier-Level et al. 2016).

There has been some debate regarding whether balancing selection is capable of maintaining genetic polymorphisms in highly variable environments, and it is often thought that this mechanism alone is not enough to maintain balanced polymorphism (Asthana, Schmidt, and Sunyaev (2005), Philip W. Hedrick, Michael E. Ginevan (1972)). A study by Bergland et al. (2014) observing *Drosophila melanogaster* in a temperate Pennsylvanian orchard revealed that strong seasonal selection can drive balancing selection favoring different life history traits over a short time span. These results suggest that seasonal variation contributes to the maintenance of two alleles to be present at the same loci long-term. If there is no fitness cost associated with polymorphic variation and there is a benefit gained through heterozygosity, then it is suggested that that balancing selection could preserve genetic variation (Asthana, Schmidt, and Sunyaev 2005, @Wittmann2017).

Wittmann et al. (2017) created a model simulating the seasonal selection observed by Bergland et al. (2014) to clarify the role this allelic fluctuation plays in maintaining multilocus polymorphisms. This model introduces the term “segregation lift” which considers the importance of dominance in maintaining polymorphism across many loci in the genome and proposes that segregation lift plays an important role in preserving genetic variation.

Our model recreates and builds upon the work of (???) to understand how linkage disequilibrium contributes to maintaining polymorphisms in a natural population. We study how the rate of crossover, epistasis, number

of generation per season, population size, allelic dominance, and seasonal balance contribute to segregation lift. Climate change models predict a change in seasonal timing, with a lengthened and earlier summer and a milder shorter winter @[Bradshaw2006]. The seasonal balance metric allows our model to include effects of changing length of seasons to understand how a longer summer may influence segregation lift. Finally, we test how allelic interactions affect the dynamics of seasonal adaptation.

3 Methods

3.1 Basic Model

The method for calculating fitness is based on the work of Wittmann et al. (2017) who originally proposed *segregation lift* as a general mechanism which may maintain balanced polymorphisms in a population under certain conditions.

The model first determines an individual’s seasonal score (z) according to the equation:

$$z = n_{hom} + d * n_{het}$$

where n_{hom} is the number loci homozygous for the allele particular to that certain season and d is a dominance parameter (which ranges between 0 and 1) and n_{het} is the number of heterozygous loci. A dominance parameter of greater than one ($d > 1$) would correspond to a traditional heterozygote advantage, and therefore is not of interest in this model.

The fitness is calculated according to the equation:

$$w(z) = (1 + z)^y$$

where $w(z)$ is the fitness of a particular z score. The exponent here is introduced to allow for non-linear interaction between genotype and fitness, an interaction that approximates the affect of epistasis.

The model we propose differs from the *segregation lift* model in a few important ways. For one, the loci in our model are located on diploid linked chromosomes, so the homologous pairs can recombine during a simulated meiosis before mating. Expand on why this may be interesting or important... Linked genes not require a $d > 0.5$. Some genes that may be important for seasonal traits (*i.e.* heat shock proteins) are located close to each other on chromosomes and have a high amount of linkage disequilibrium...

Our model also does not allow for “selfing” where an individual can asexually reproduce with itself to create an offspring... Other stuff maybe too

Table 1: Default values are a list the values that were used generally for each simulation run. The alternative values are values that were used to determine the affect of each variable individually. Some alternatives were tested in combination – maybe explain which ones were tested in combination.

Parameters	Default	Alternatives values
Crossover probability	0.05	0, 0.001, 0.01, 0.1, 0.5
Generations per year	20	10, 20, 50
Seasonal balance	2	1.25, 1.5
Population size	500	100 1000
Exponent of the fitness function	1	0.5, 2
Dominance	0.5	0.2, 0.8
Mutation probability	10^{-4}	
Number of loci	100	
Years	300	

3.2 Simulated seasonal selection

The general outline for the seasonal selection simulation:

1. A population of N individuals is created with a single diploid chromosome containing L loci, all of which contribute to the individual's seasonal fitness.
2. Seasonal score is calculated according to Eqn 1 and the fitness is determined. Pairs of parents – of size N – are sampled stochastically with a probability proportional to their seasonal fitness.
3. Each parent in the pair undergoes recombination with a per locus recombination rate of P_c .
4. A chromosome from each parent is selected at random to create the diploid offsprings.
5. The new generation replaces the old generation in the population.
6. The new individuals undergo mutation with a per locus mutation probability P_m which was fixed at the relatively high rate of 1×10^{-14} (Haag-Liautard et al. 2007).
7. A new generation begins and the current season is determined by the current generation and seasonal balance parameter S_b .
8. A new generation of parents is selected according to their seasonal fitness and the process continues to cycle for a number of years.
9. Loci specific allele frequencies are tracked throughout the simulated evolutionary run.

3.3 Interaction between loci: an alternative scenario

We decided to test the scenario when loci have a certain interaction between them. In real genomes mutations often have different fitness effects depending on the genetic background i.e. the presence of other mutations. Thus, linkage can be advantageous if specific combinations of mutations are better kept together because of their reinforcing effect on each other. In this simulation we kept the same fitness function as before but also rewarded sequences of repeated values of the beneficial allele. For example, in the summer an individual with chromosome 01010101 would have a fitness of 4 for the number of 1's + 1 for the longest sequence of 1's, while an individual with chromosome 01110000 would have a fitness of 3 + 3 which would then outcompete the former individual even though the total number of 1's is less.

4 Results

We compared the effect of different variables by looking at the allele frequencies at the last generation of the simulation and calculated the standard deviation (the higher this value the less balancing polymorphism was present in the population) and proportion of fixed loci. Since we had 5 repetitions for each parameter value we used an Analysis of Variance (ANOVA) to look for significant differences.

We first tested whether we were successful in reimplementing the model made by Wittmann et al. (2017). We too found that with no linkage (i.e. cross-over happens at every loci) both the proportion of fixed loci ($p = 2.52 \times 10^{-7}$) and standard deviation of loci frequencies ($p = 4.63 \times 10^{-7}$) are lower when dominance is 0.8 compared to when dominance is 0.5. See Figures 1 and 2.

Then, we started playing with different parameters.

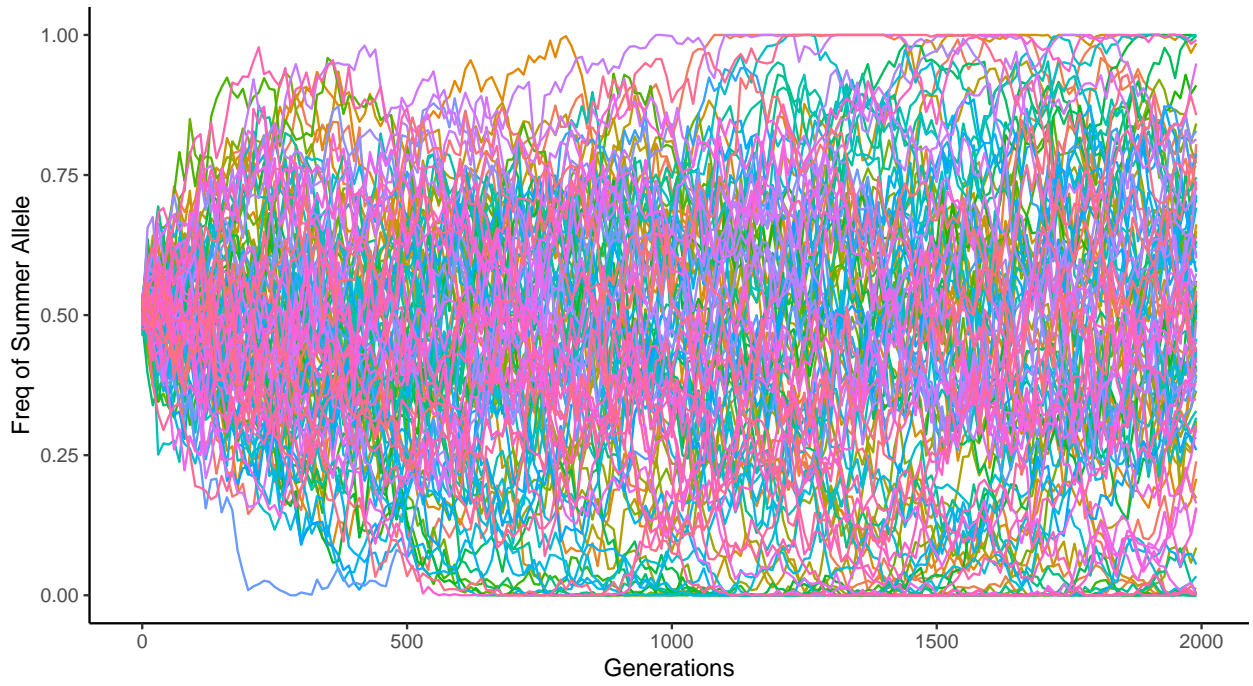
4.1 Crossover rate

Crossover rate didn't seem to have an effect at all in terms of standard deviation ($p = 0.123$) and proportion of fixed loci ($p = 0.462$). The greatest variation in both measures was when there was no crossover introduced at all and changes in the chromosomes were only due to mutations. See Figure 3.

4.2 Dominance

Since the crossover rate didn't have an effect on whether or not balanced polymorphism is achieved, our tests with different dominance values resulted in the same dynamics as when we tested with crossover rate 0.5. See Figure 4.

A Dominance = 0.8, no linkage



B Dominance = 0.5, no linkage

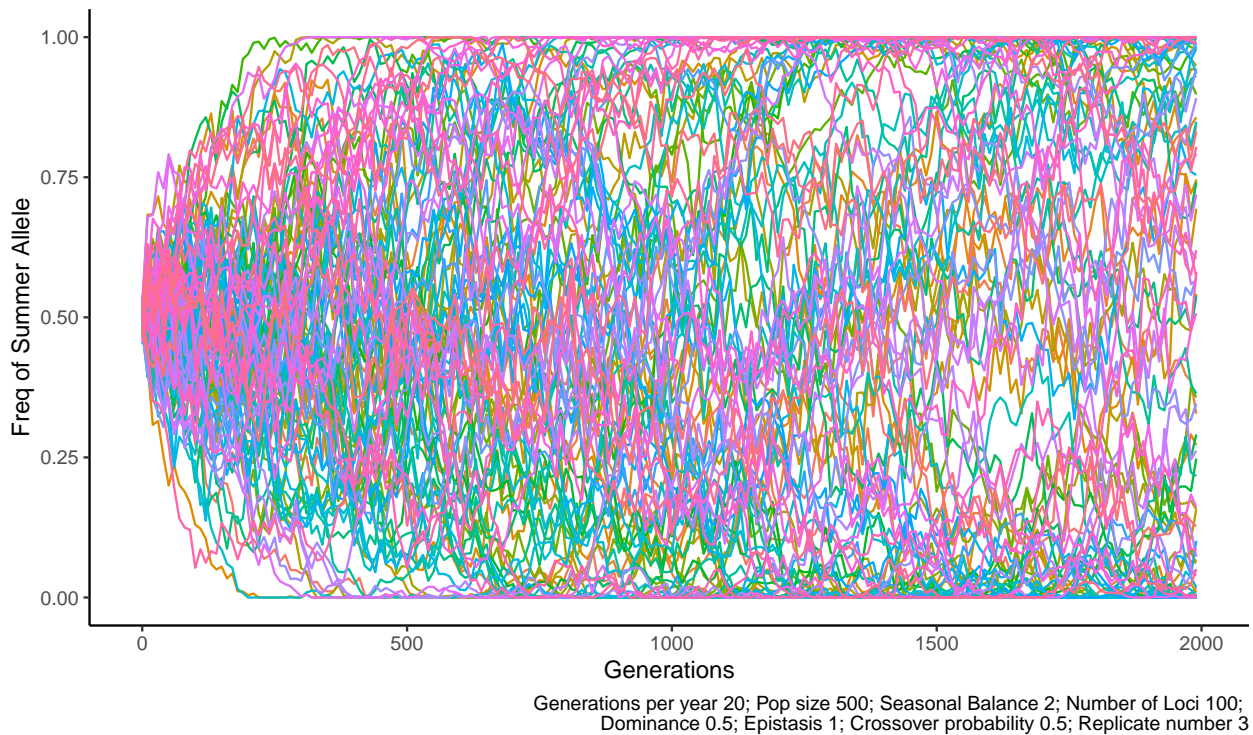


Figure 1: Summer allele frequency at each loci over time. (A) When $d = 0.8$ fixation of a few alleles start only after about 700 generations and most alleles finish with frequencies closer to 0.5. (B) On the other hand, when $d = 0.5$ fixation starts as soon as 200 generations and most alleles fixate or get very close to fixation by the end of the simulation.

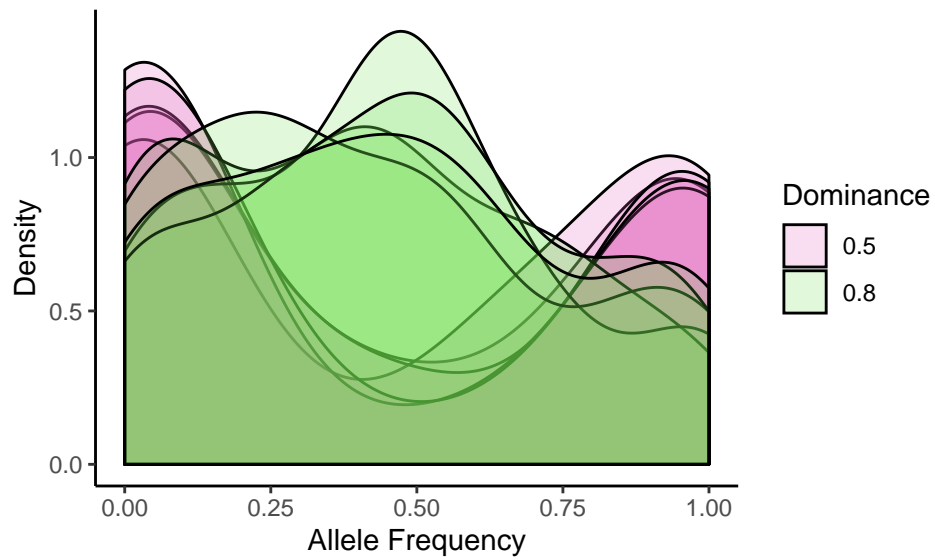


Figure 2: A density plot of the allele frequencies at the final generation for five replicates of each value of the dominance parameter.

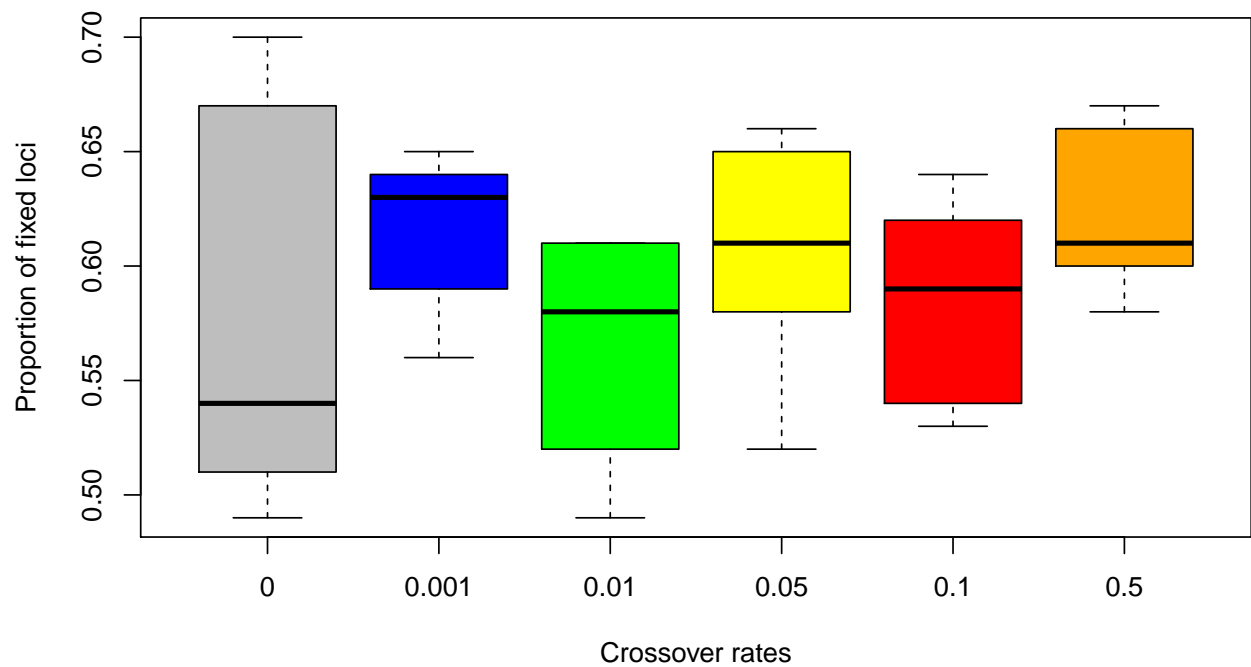


Figure 3: Different crossover rates did not effect the allele frequency distribution at the end of the simulations.

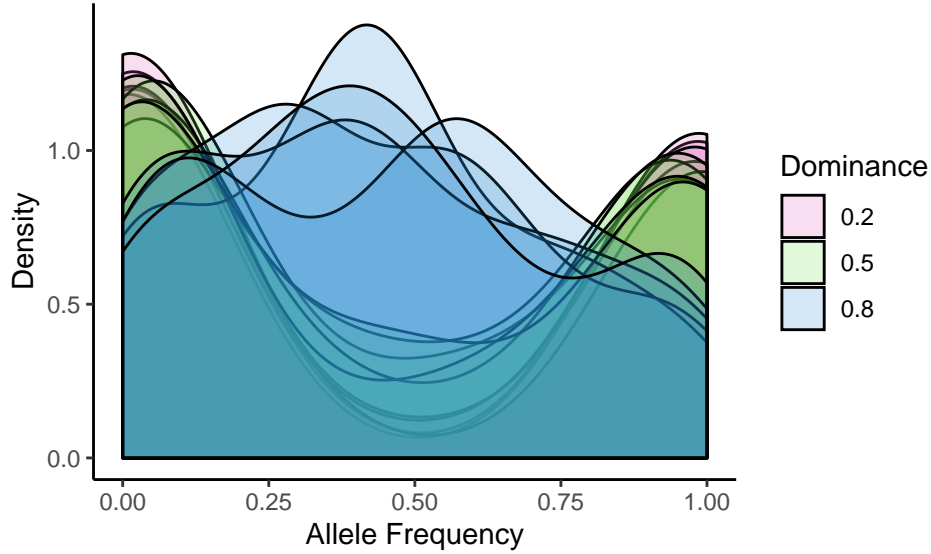


Figure 4: A density plot of the allele frequencies at the final generation for five replicates of each value of dominance with a crossover rate of 0.05.

4.3 Epistasis

Under default crossover rate (0.05) epistasis didn't have an effect (standard deviation $p = 0.986$, proportion of fixed alleles $p = 0.613$). See Figure 5.

4.4 Generations per season

When the generations per season was 10, the standard deviation was significantly lower than for longer generation times per season ($p = 0.00161$). See Figure 6.

4.5 Population size

Population size had the biggest effect by far (standard deviation $p = 1.51e-10$, proportion of fixed loci $p = 1.63e-10$). Purely by changing population size from 100 to 1000 average proportion of fixed loci decreased from 0.9 to 0.3. See Figure 7.

4.6 Seasonal balance

Having a season that has more generations than the other didn't result in a higher proportion of fixed alleles ($p = 0.289$). Rather, while when both seasons had equal number of generations (seasonal balance = 2) alleles fixed by reaching frequencies of either 0 or 1, when summer had more generations more alleles had frequencies of 1 than 0, but the overall proportion stayed the same. See Figure 8.

4.7 Mutational interaction

Introducing interaction between loci made seasonal balance to be reached more difficult. Every variable kept the same the new simulation produced a higher average standard deviation of allele frequencies ($p = 3.58e-05$) and proportion of fixed loci ($p = 7.76e-06$) regardless of crossover rate. See Figure 9.

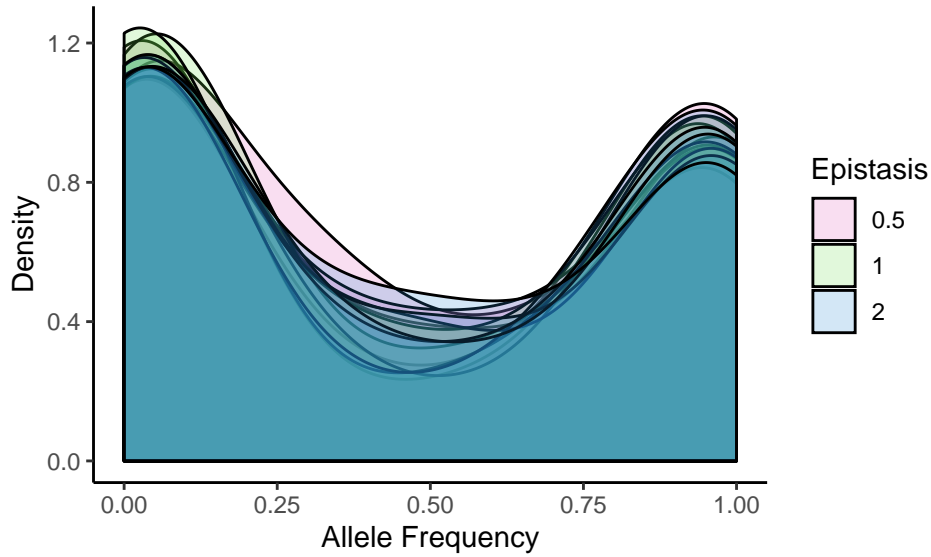


Figure 5: A density plot of the allele frequencies at the final generation for five replicates of each value of the epistasis parameter.

5 Discussion

5.1 Crossover rate

Crossover rate didn't have an effect on seasonal balance. Our hypothesis is that two opposing forces created by it canceled out. First of all, higher crossover rates create more genetic variation in the population "helping" selection but on the other hand it also creates stochasticity since a selected high fitness individual can pass on an unfit chromosome after crossover by chance.

When there was no crossover and so linkage was maximized there was a much bigger variation in the proportion of fixed loci in the last generation. This could be because since in this case the only way for an individual to change was through mutations, which is pretty infrequent, and if the starting population contained a lot of individuals with more of one allele than the other many loci went to fixation while in a starting population with individuals whose genomes contained 50%-50% of the two alleles balancing polymorphism could be maintained more successfully.

5.2 Epistasis

We'll think more about an explanation on why epistasis didn't have an effect.

5.3 Generation

We hypothesize that when the generation time is shorter the allele frequencies oscillate less over time as they experience selection for a shorter period of time. Thus, alleles are less likely to go to fixation.

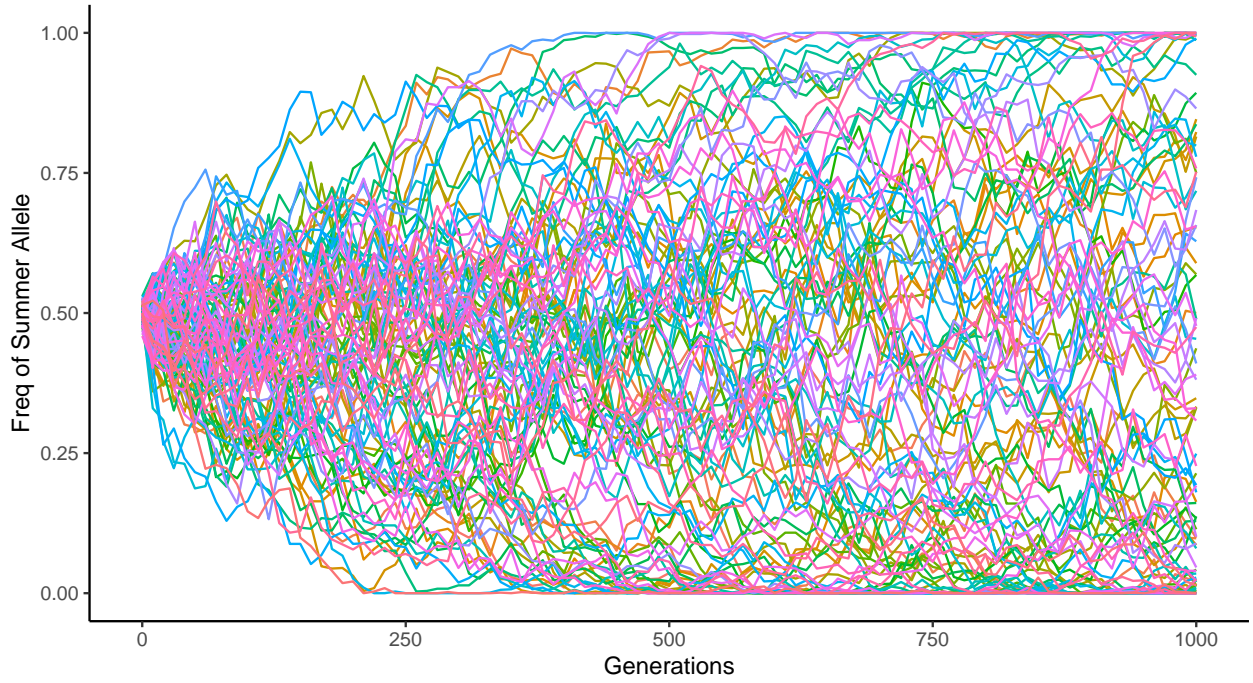
5.4 Population size

Population size had the greatest effect of all the variables we tested. This is probably due to the effect of random genetic drift; when the population is small alleles go to fixation due to the large effect of drift.

5.5 Mutational interaction

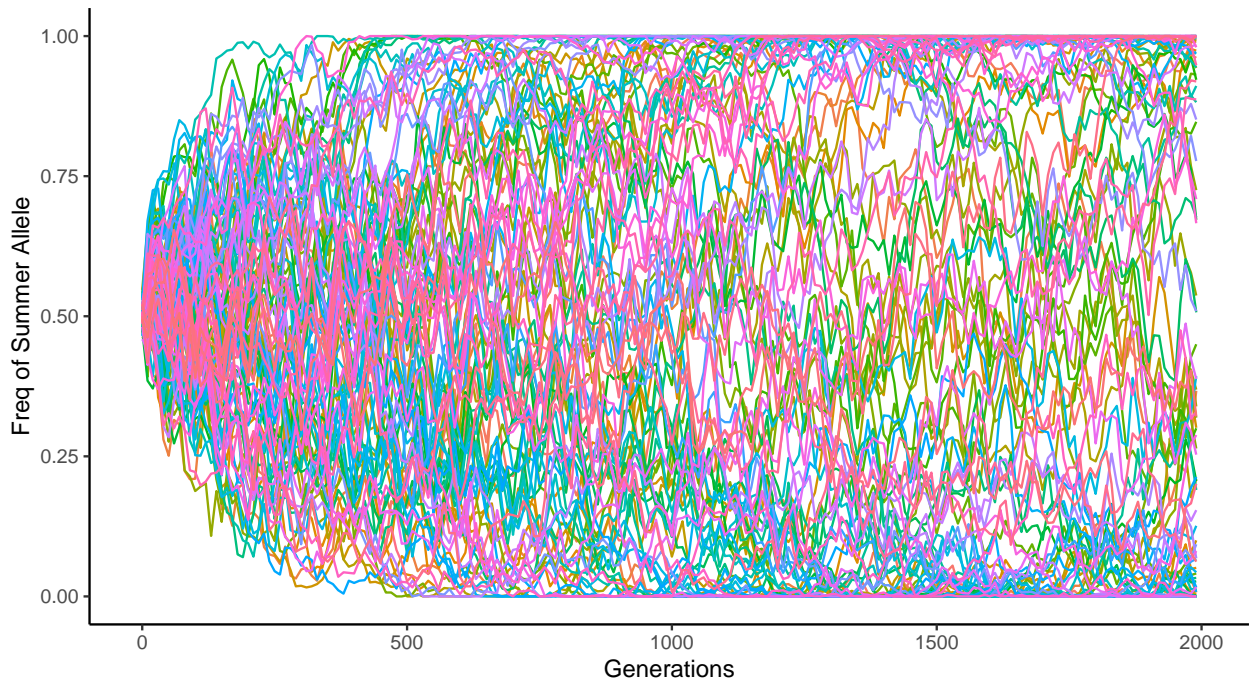
We'll think more about an explanation on why even here crossover didn't seem to have an effect.

A Generations per year = 10



Generations per year 10; Pop size 500; Seasonal Balance 2; Number of Loci 100;
Dominance 0.5; Epistasis 1; Crossover probability 0.05; Replicate number 2

B Generations per year = 50



Generations per year 50; Pop size 500; Seasonal Balance 2; Number of Loci 100;
Dominance 0.5; Epistasis 1; Crossover probability 0.05; Replicate number 1

Figure 6: Summer allele frequency at each loci over time. When the generation time is shorter the allele frequencies oscillate less over time as they experience selection for a shorter period of time. Thus, alleles are less likely to go to fixation.

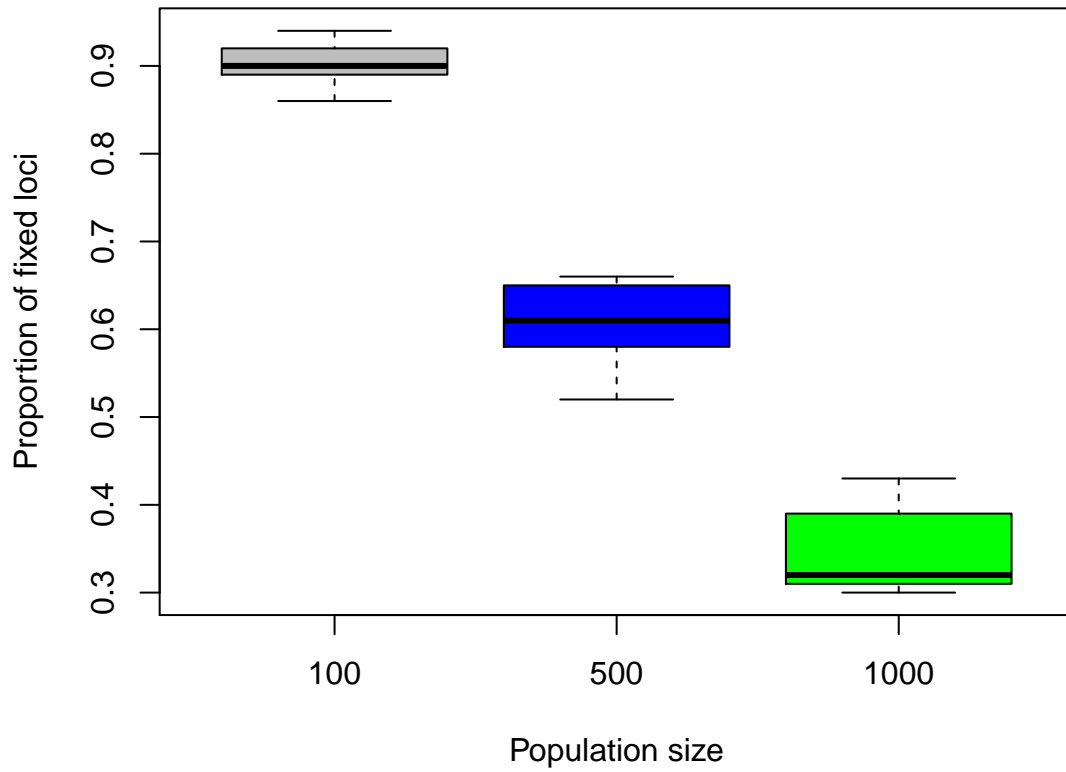


Figure 7: Population size had a dramatic effect on both standard deviation and proportion of fixed loci.

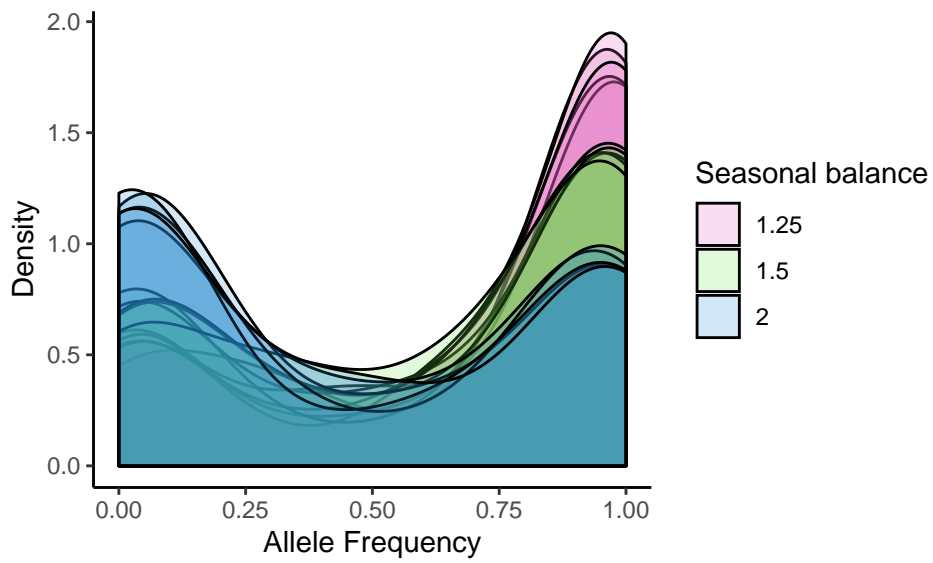


Figure 8: A density plot of the allele frequencies at the final generation for five replicates of each value of the seasonal balance parameter.

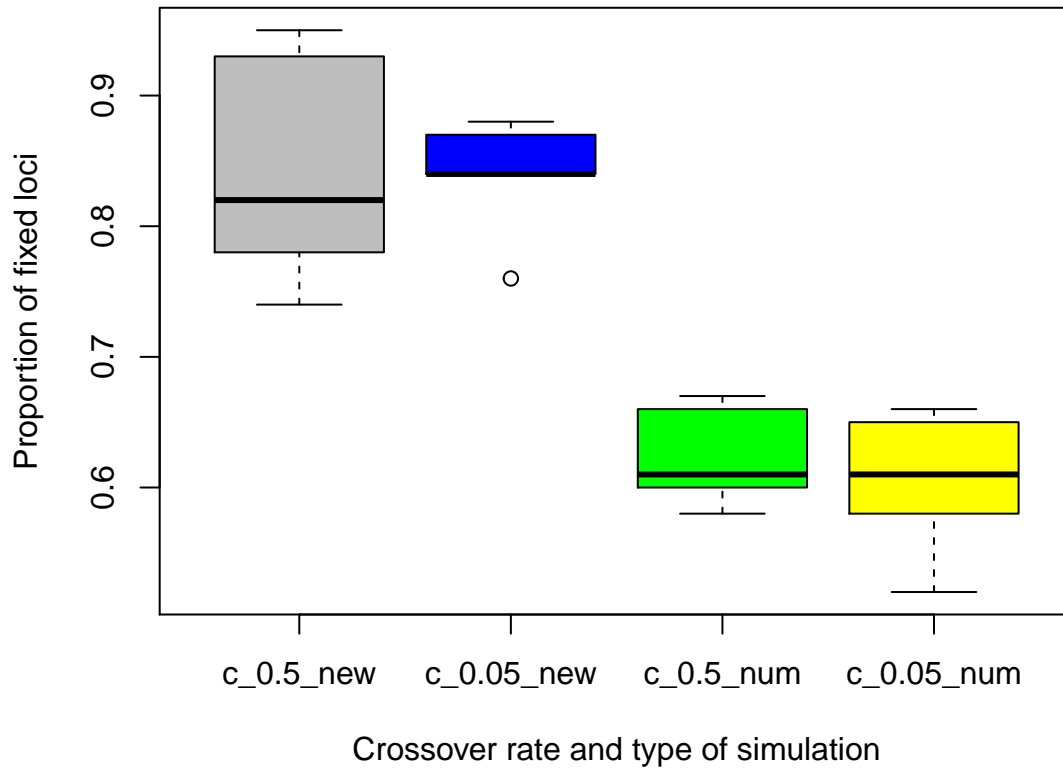


Figure 9: Interaction between loci increased proportion of fixed loci regardless of crossover rate.

6 Easton comments

- I was surprised that you opened the introduction with text on phenotypic plasticity. Your work really gets at genetic change from season to season, as opposed to only phenotypic changes.
- Good second introduction paragraph to explain the two most important papers that motivate your work
- In your introduction, you need an additional paragraph explaining what gaps remain after the two papers you describe. This then sets up the reasons for why your study will be important
- Good job describing the model
- Eventually section 3.2 could be made into a nice conceptual figure
- Overall, your figures came out very nice. They are simple and aesthetically appealing. In the real manuscript, you won't want the figure interpretations within each caption
- As I mentioned in class, I think there are some interesting parts of your analyses in the context of climate change predictions. With climate change, we expect many places to experience longer (and earlier) summers. In addition, generation times for some species are expected to shorten with warmer temperatures. To me, your results would then say that with climate change, more loci are expected to become fixed because of summer changes, but that shorter generations might present an opposing force.
- I think your work is certainly possible. I am not an expert on this literature, but it does seem that you address several interesting points by focusing on epistasis, seasonal balance, and generation length.

Here is what I need from you for the final project:

Nothing. Your first draft is good enough. If you make edits and changes, I am happy to read additional drafts, this semester or next. Otherwise, you are done for the semester! Great work!

References

- Asthana, Saurabh., Steffen. Schmidt, and Shamil. Sunyaev. 2005. "A limited role for balancing selection." *Trends in Genetics* 21 (1): 25–30. <https://doi.org/10.1016/j.tig.2004.11.007>.
- Bergland, Alan O., Emily L. Behrman, Katherine R. O'Brien, Paul S. Schmidt, and Dmitri A. Petrov. 2014. "Genomic Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in *Drosophila*." *PLoS Genetics* 10 (11). <https://doi.org/10.1371/journal.pgen.1004775>.
- Fournier-Level, Alexandre, Emily O. Perry, Jonathan A. Wang, Peter T. Braun, Andrew Migneault, Martha D. Cooper, C. Jessica E. Metcalf, and Johanna Schmitt. 2016. "Predicting the evolutionary dynamics of seasonal adaptation to novel climates in *Arabidopsis thaliana*." *Proceedings of the National Academy of Sciences of the United States of America* 113 (20): E2812–E2821. <https://doi.org/10.1073/pnas.1517456113>.
- Haag-Liautard, Cathy, Mark Dorris, Xulio Maside, Steven Macaskill, Daniel L. Halligan, Brian Charlesworth, and Peter D. Keightley. 2007. "Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*." *Nature* 445 (7123): 82–85. <https://doi.org/10.1038/nature05388>.
- Messer, Philipp W., Stephen P. Ellner, and Nelson G. Hairston. 2016. "Can Population Genetics Adapt to Rapid Evolution?" *Trends in Genetics* 32 (7). Elsevier Ltd: 408–18. <https://doi.org/10.1016/j.tig.2016.04.005>.
- Novak, Sebastian, and Nicholas H. Barton. 2017. "When does frequency-independent selection maintain genetic variation?" *Genetics* 207 (2): 653–68. <https://doi.org/10.1534/genetics.117.300129>.
- Philip W. Hedrick, Michael E. Ginevan, Evelyn P Ewing. 1972. "Genetic polymorphism in heterogeneous environments." *The Scientific Basis of Medicine Annual Reviews*, 174–93.
- Wittmann, Meike J., Alan O. Bergland, Marcus W. Feldman, Paul S. Schmidt, and Dmitri A. Petrov. 2017. "Seasonally fluctuating selection can maintain polymorphism at many loci via segregation lift." *Proceedings of the National Academy of Sciences of the United States of America* 114 (46): E9932–E9941. <https://doi.org/10.1073/pnas.1702994114>.