

Exploring Mechanisms of Molecular Evolution and Their Representations in PCA

Matthew Aleck¹, Ronald J. Nowling¹

¹ Electrical Engineering and Computer Science, Milwaukee School of Engineering

Abstract

Principal Component Analysis (PCA) is widely used in population genetics for visualizing the relationships between samples (Neafsey et al. (2010)), correcting for stratification in genome-wide association studies (Price et al. (2006)), and with clustering to determine population structure (Lee et al. (2009); Patterson et al. (2006)). We wanted to explore how different mechanisms (e.g., random mutations, mating, and recombination) of molecular evolution lead to the patterns observed in PCA.

Using simulations, we evaluated the impact of these mechanisms on allele frequencies and the patterns observed in PCA. For this work, we implemented a population genetics simulator in Java. Our simulator allows us to enable or disable particular mechanisms so that we can study them individually as well in combination. We will present preliminary results and conclusions from this.

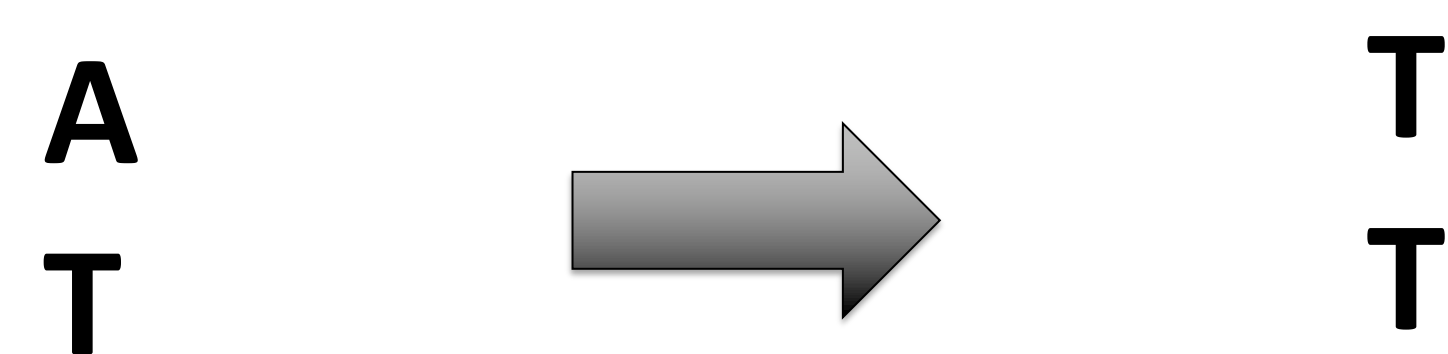
Hypothesis

Why can we detect inversions? We can detect inversion because inversions cause association between mutations.

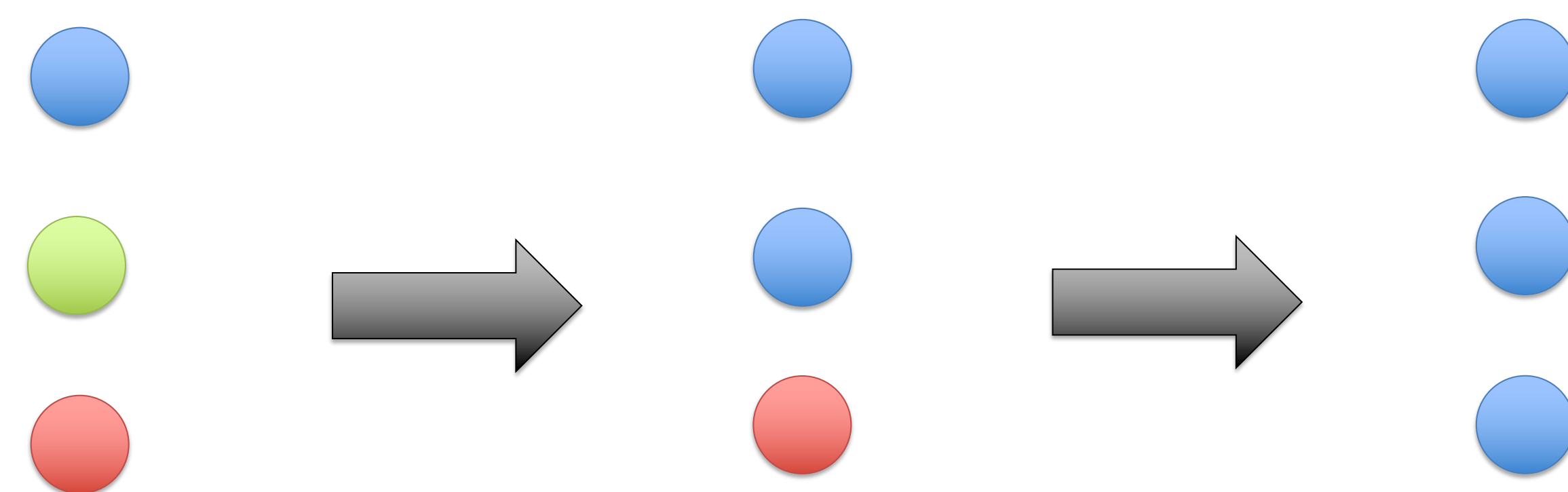
Inversions repress recombination. We simulated repressed recombination in this experiment. If recombination is partially suppressed, associated mutations can persist.

Methods

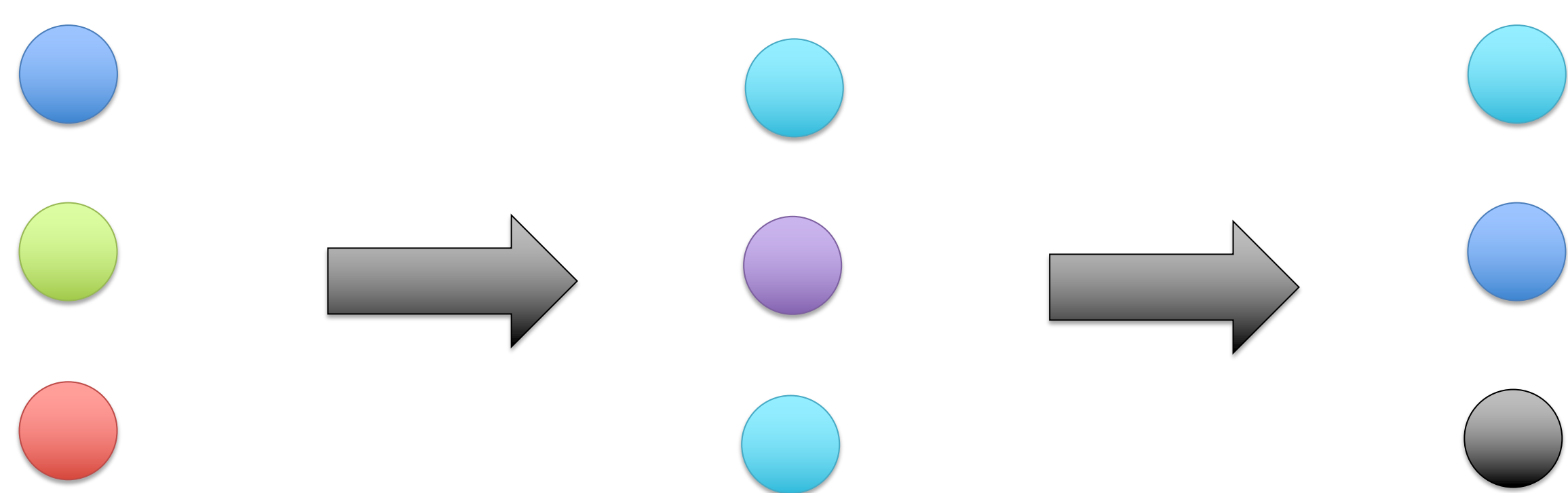
Mutation



Cloning



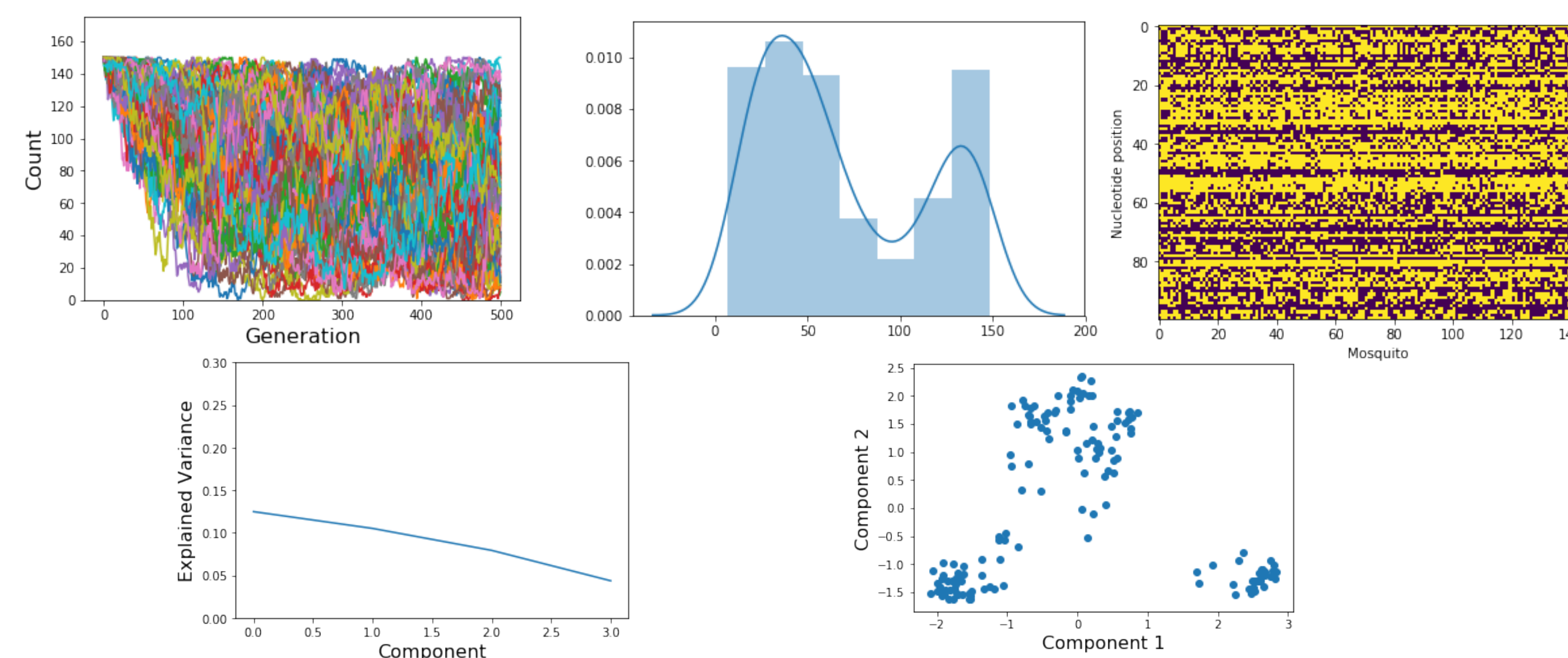
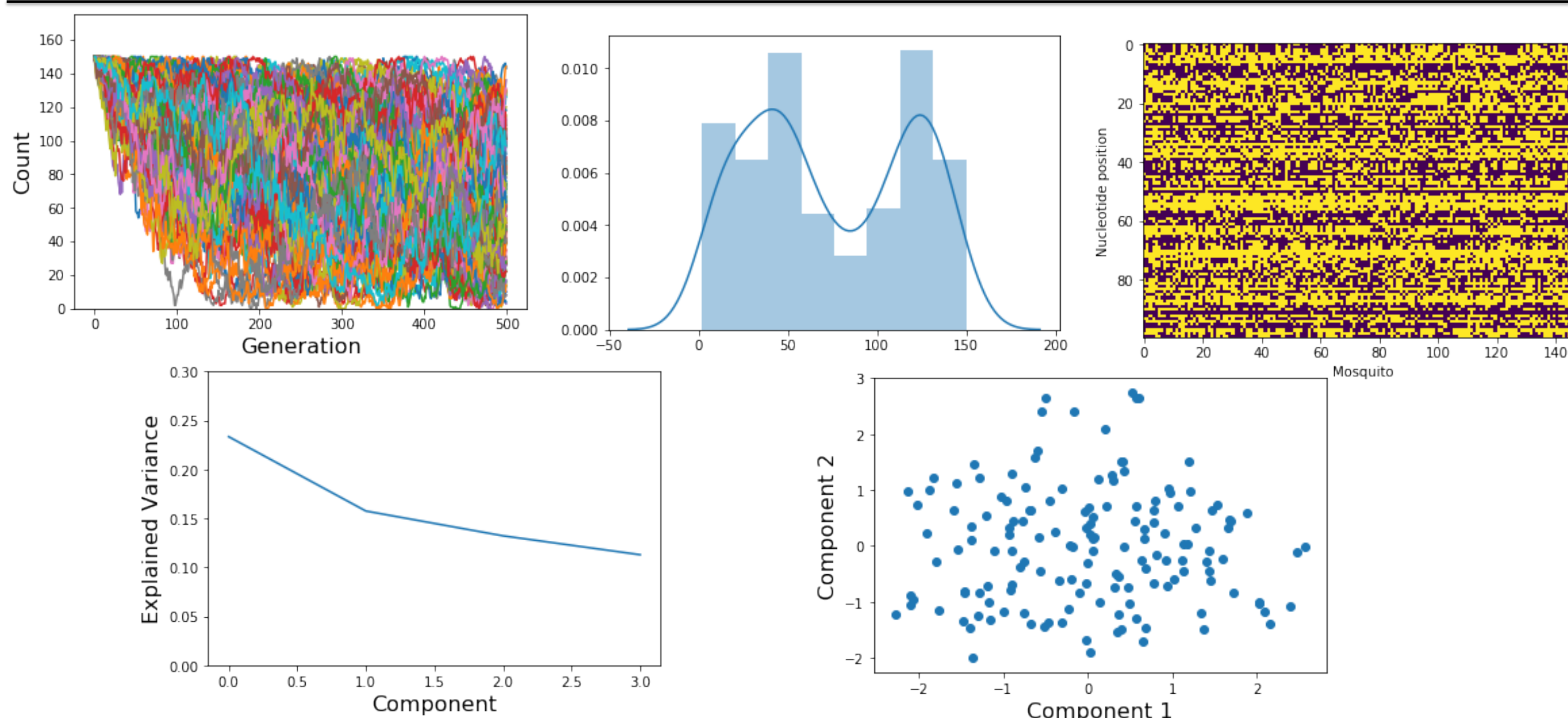
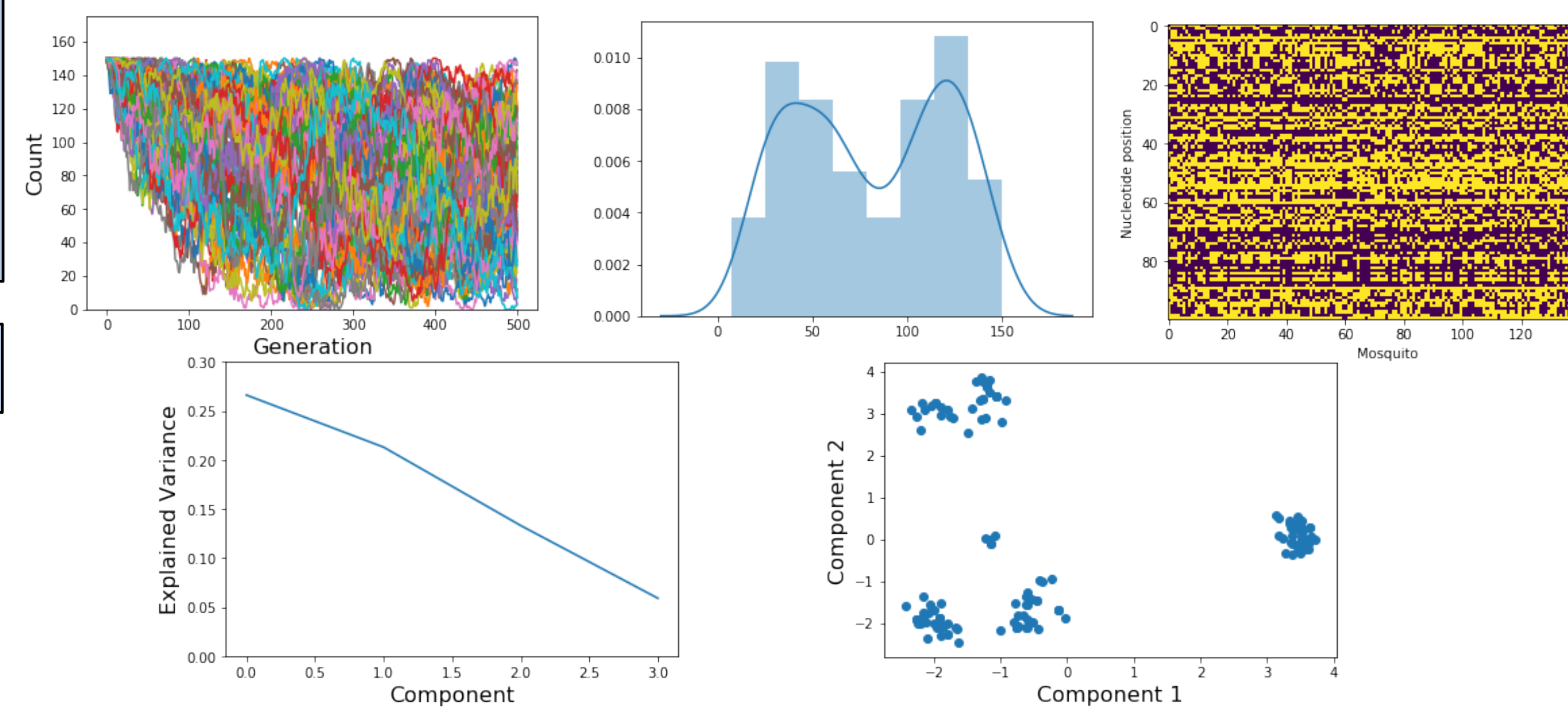
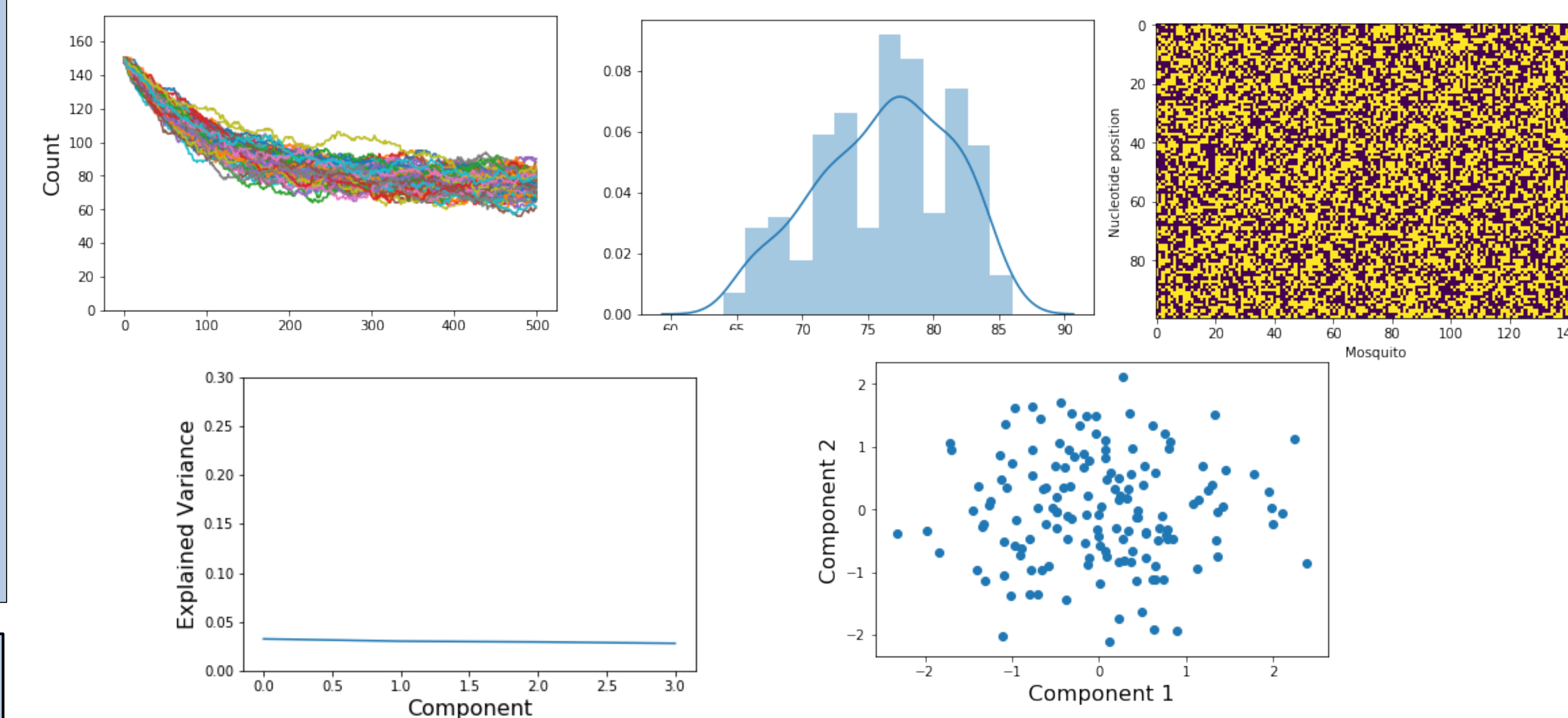
Mating/Recombination



Conclusions

Mating causes groups of mutations to be inherited together. Recombination disrupts the correlation. Inversions can be detected because they repress the recombination process, leading to the cluster structure in PCA.

Results



Discussion

Mutations Only

In experiment #1, we randomly mutated nucleotides. Figure 1 shows that the allele counts for the first nucleotide position clearly trend toward an average range of 60-100. Figure 2 shows the allele counts have a normal distribution. Figure 3 shows the genomes of the last generation of mosquitoes and few to no patterns can be discerned. Figures 4 and 5 shows that PCA can not find much association between the variants and that samples do not form into clusters.

Mutations and Cloning

In experiment #2, we combined random mutation with cloning (sampling with replacement). Figures 1 and 2 shows the allele counts form a bi-modal distribution. Figure 3 shows the genomes of the last generation of mosquitoes and repeating similar genomes can clearly be observed. Figures 4 and 5 shows that PCA found association between the variants and that the samples form distinct clusters. Cloning (random mating) causes correlation (linkage disequilibrium) among mutated nucleotides.

Recombination

In experiment #3, we added recombination. We randomly pick the two parents (sample with replacement) and a crossover point. Figures 1 and 2 show that the allele frequencies are not affected. Figure 3 show the same pattern as experiment #3. Figures 4 and 5 suggest that PCA does not find correlation among mutated nucleotides. Our experiment suggests that recombination counteracts the correlation caused by cloning.

Repressed Recombination

In experiment #4, we repressed recombination along half of the chromosome but enabled recombination in the other half. The repressed region simulates the effect of an inversion. Figures 4 and 5 show the cluster structure is restored because nucleotides are now inherited together. This pattern reflects what we see in real SNP data with inversions.