

# Modelling Evolution Through Simple Genetic Algorithms in Java

Jackson Shen, Spencer Socoloski

CSSE 220, Spring 2021

## 1 Introduction

The goal of this project was to demonstrate and clearly visualize simple genetic evolution. In order to do so, we created a model and visualization of randomly mutable chromosomes, whose populations were determined through various genetic algorithms in Java.

## 2 Structure

The model consists of a population, a list of chromosomes. Each chromosome within the population has a genome, a bitstring of 0s and 1s. Key characteristics of the overall fitness of the population are graphed, showing the average fitness of the population, in addition to fitness of the fittest chromosome, least fit chromosome, and the average Hamming distance between the chromosomes. If the chromosome is within the strongest  $n$  chromosomes, as defined by the user, the chromosome is immune to mutations and is automatically chosen for the next generation.

## 3 Functionality

This project allows for the visualization of various fitness functions and evolutionary selection functions, in addition to showing the diversity of a population.

### 3.1 Fitness Functions

The default basic fitness function used in this project was the number of 1s in the chromosome's bitstring genome. The more 1s there were, the stronger the chromosome, and vice versa.

In addition to the default fitness function, we also utilized other fitness functions, such as similarity to an "ideal" chromosome, or the longest consecutive string of 1s or 0s. In the case of similarity to an ideal chromosome, we can expect similar results to the default basic function, as the "ideal" in that case

would be a chromosome with a genome of all 1s. However, the expected reaction to evolution for the longest consecutive string of 1s and 0s is different, as a mutation within the genome could have far stronger effects than the other fitness functions.

### 3.2 Selection Functions

The default evolution selection function is truncation: taking the top 50 percent of the population relative to the fitness function, and repopulating the next generation with the chosen 50 percent. Repopulation occurred through two methods: with and without crossover. With crossover, chromosomes were combined using a one-point crossover function, then mutated to repopulate the next population. Without crossover, chromosomes were directly mutated, then placed into the next generation.

There are two other selection functions, both of which follow the same repopulation rules: roulette selection and ranking selection. Roulette selection chooses the chromosomes who repopulate the next generation through random choice, based on the chromosome's fitness relative to the rest. Rank selection chooses chromosomes in a similar manner, with the probability weights of choosing a chromosome tied to the chromosomes rank, such as first or second, instead of the chromosome's relative fitness.

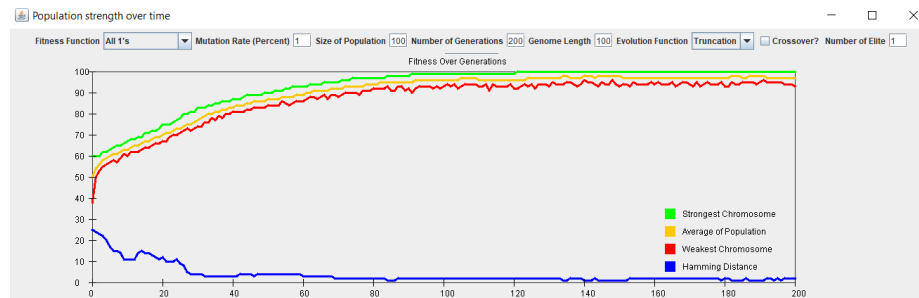
### 3.3 Diversity

The measure of the diversity of the population, the Hamming distance, can be calculated from a given population. To put it simply, the Hamming distance between two chromosomes is the number of bits in the genome that differ from each other. The average of all the Hamming distances between chromosomes is what we will use to measure the diversity of the population. As the population evolves, we will expect the average Hamming distance to decrease.

## 4 Experiments

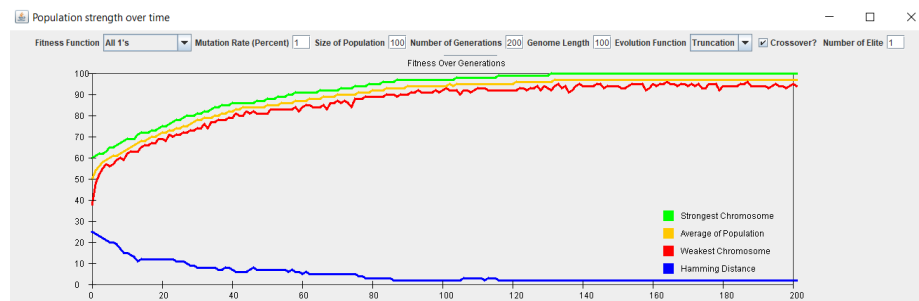
This simulator was used within various experiments to gain further insight into how different functions, events, and properties would affect the evolution of a population. All of the experiments were conducted using a genome length of 100, a population size of 100, a mutation rate of one percent, and one elite chromosome.

## 4.1 Evolution Without Crossover



Without crossover, the evolution took approximately 120 generations for the first chromosome to reach the maximum fitness. As the population evolved, the fitness of the fittest chromosome within the population did not decrease, as expected from the elitism function. The average fitness and the fitness of the weakest chromosome in the population increased over time as well, showing that the entire population is undergoing improvement. In addition, the average Hamming distance decreased, as the population became closer and closer to maximum fitness, since chromosomes are approaching the same ideal genome.

## 4.2 Evolution With Crossover



With crossover, the evolution took approximately 130 generations for the first chromosome to reach the maximum fitness. This is surprising, as we expected crossover to decrease the number of generations it took to have a chromosome reach maximum fitness. From these tests, we found that crossover actually increases the time it took for a population's fittest individual to reach an ideal genome.

Other than the number of generations it took to reach maximum fitness, the evolution with crossover showed results that were consistent with evolution without crossover.

## 4.3 Evolution Without Mutation

When evolving a population without mutation, the fittest chromosome in the population never reached the ideal fitness. This is likely due to the patterns

within the chromosomes, making it impossible to evolve beyond a certain point. After the top 50 percent of the population stabilizes, the crossover function no longer has any effect on the population, instead just copying the population generation after generation.

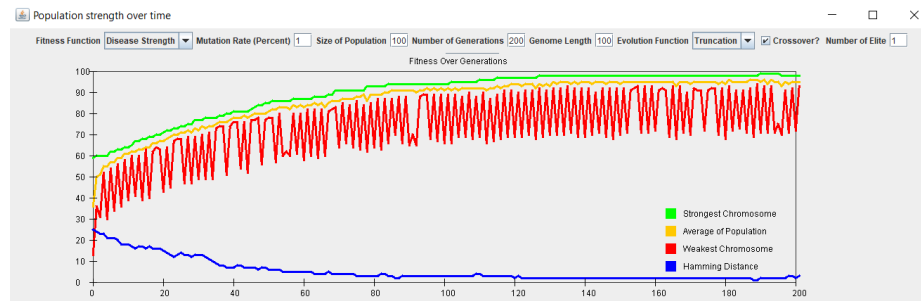
## 4.4 Recreating a Classic Study

[insert smth]



[insert smth about nolan paper]

## 4.5 Expanding Results



In order to understand what would happen in the case of a genetic mutation that causes a disease, we created a fitness function that was similar to the basic fitness function, but with one bit in the chromosome that when turned on, decreased the fitness by 25. In doing so, we found extreme spikes within the least fit chromosome within the population. This up and down spiking is due to the removal of weak chromosomes every generation, but as the generation mutates and evolves, the "diseased" gene turns on, making a chromosome with a low fitness. The diseased chromosomes are also reflected in the average strength of the population, the curve of which is significantly bumpier than what was seen within the standard fitness functions. Through this model, we have demonstrated the real-life situation of genetic diseases, where a faulty genome can cause a significant decrease in an organism's so-called fitness.