Simulating and Modelling Adaptive Walks with the NK Model

USRI Final Research Output

Abigail Kushnir

Western University

Summer 2022

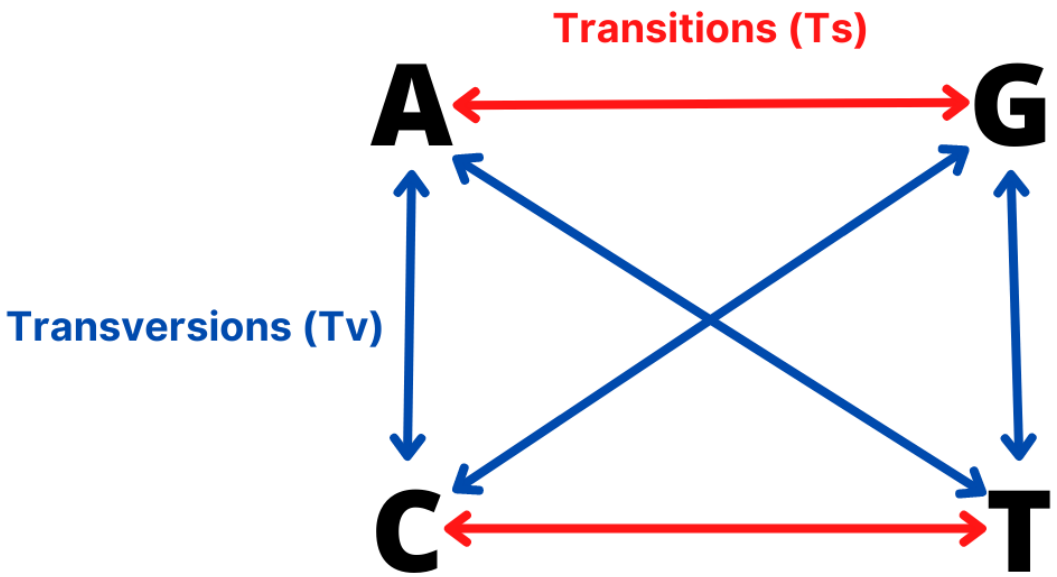
Supervised by Dr. Lindi Wahl

Biological Background

- Mutations to an organism's genome can either be beneficial or detrimental to it.
  - This can be determined by whether the mutation increases or decreases the genome's **fitness**, which can be interpreted as the number of viable offspring an organism would produce.
  - A higher fitness is more desirable from an evolutionary and biological standpoint.
- There are many different types of mutations and organisms will often undergo certain types more often than others; we call this a **mutation bias**.
  - For this project, we are most interested in mutations at a base pair (Adenine, Thymine, Cytosine, Guanine) level – specifically **transitions** and **transversions**.

Biological Background

- Each base has 3 possible mutations: 1 transition and 2 transversions.
- Transitions and transversions have their own bias: the Tv/Ts bias. When shifted, the fraction of transversions that occur changes.



Biological Background

- Each position in the DNA sequence is referred to as a **locus**.
  - For example, the second locus in the sequence A C A A G would be C.
- Each locus contributes to the genome's overall fitness. When we mutate a locus, that contribution changes.
  - For simplicity, we'll assume that changing the contribution of one locus does not affect the contributions of the other loci, i.e. no epistasis.
  - The genome's total fitness is then given by:

$$W = \sum_{i=1}^{N} \frac{w_i}{N}$$

where N is the number of loci and w_i is the fitness contribution of each of them.

Overview

- Over the summer, I worked under the supervision of Dr. Lindi Wahl and with M.Sc. candidate Tenoch Morales to explore how changes to an organism's mutation bias affect its **distribution of fitness effects** (DFE).

- We simulated adaptive walks (i.e. allowing a genome to undergo a set number of mutations) using Python to gain insight towards theoretical equations to model the behaviour.

Simulating an Adaptive Walk

To simulate adaptive walks artificially, we use an algorithm that essentially:

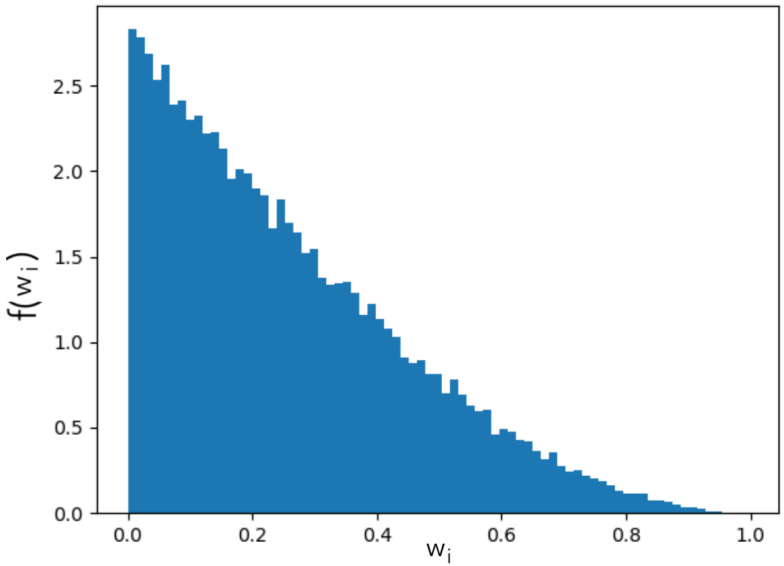
1. Randomly generates a genome of a set length and assigns fitness contributions, which are pulled from a uniform distribution ($U(0,1)$), to each locus.
2. Attempts to make mutations to the genome; not all mutations will fix. The probability that a mutation is accepted is dependent on $2s$, where s is the effect that the mutation has on fitness.
3. After a set number of mutations, determines how many of the next possible mutations would be beneficial or detrimental. This is what produces the **DFE**.

A significant portion of my time was spent updating this algorithm to make it more efficient, accurate, and finding ways to save insightful data (e.x. how many tries before a mutation fixes, the fitnesses of loci before and after they mutate, the distrubtion of fitnesses for loci that didn't mutate, etc.).

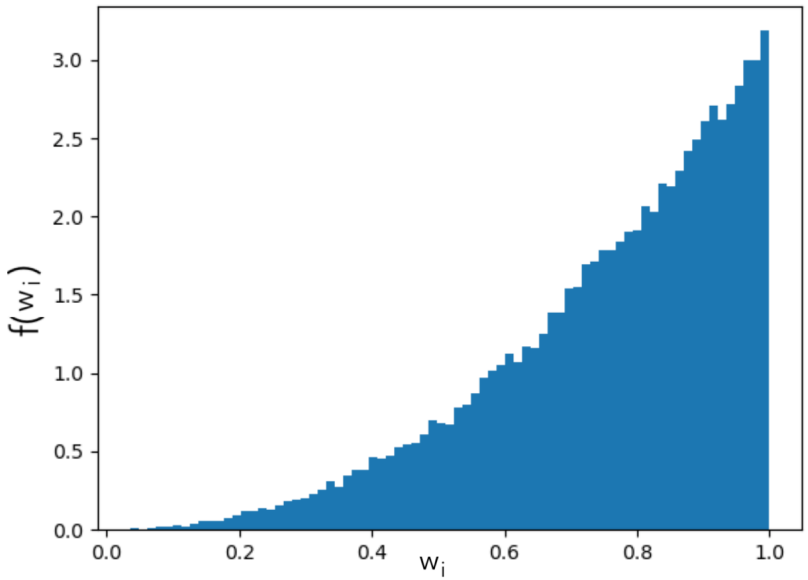
- Since the fitness contributions of the loci are distributed uniformly, the mean at the beginning of the walk is 0.5. When we make the first mutation, the average fitness of the loci where the mutation occurs is around 0.25; a mutation is more likely to fix on a locus with a lower fitness. After the mutation fixes, the average fitness of the mutated loci jumps to around 0.75.

- As the walk progresses and we perform more mutations, these means increase slightly with each step.

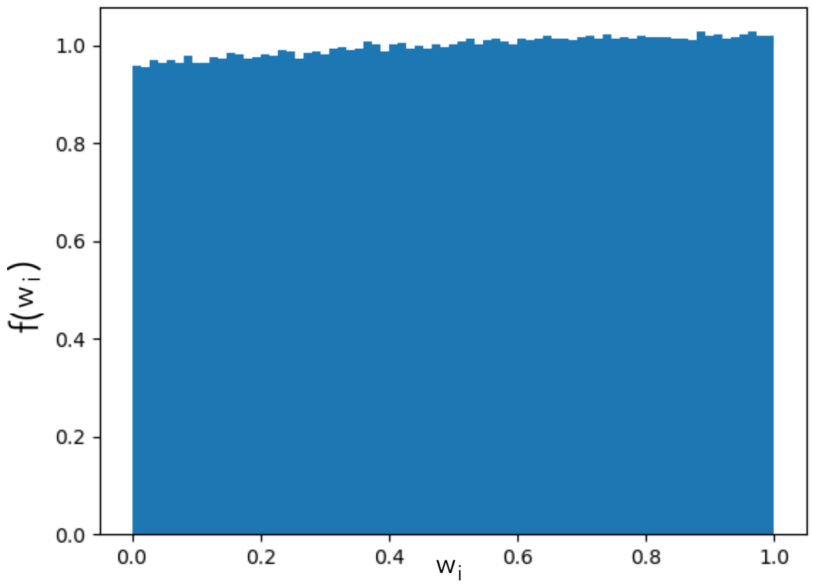
Fitness of Selected Loci Before Mutation 1 (PDF)



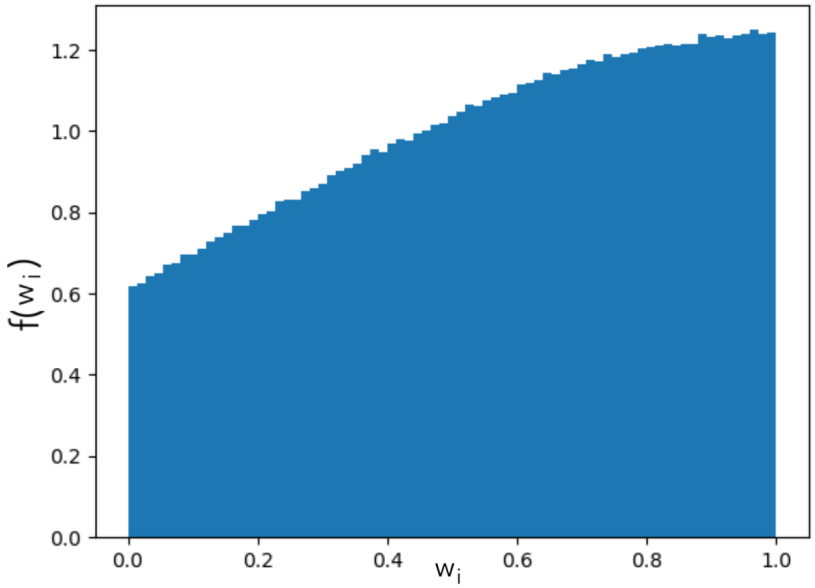
Fitness of Selected Loci After Mutation 1 (PDF)



Fitness of Unmutated Loci After Mutation 1 (PDF)

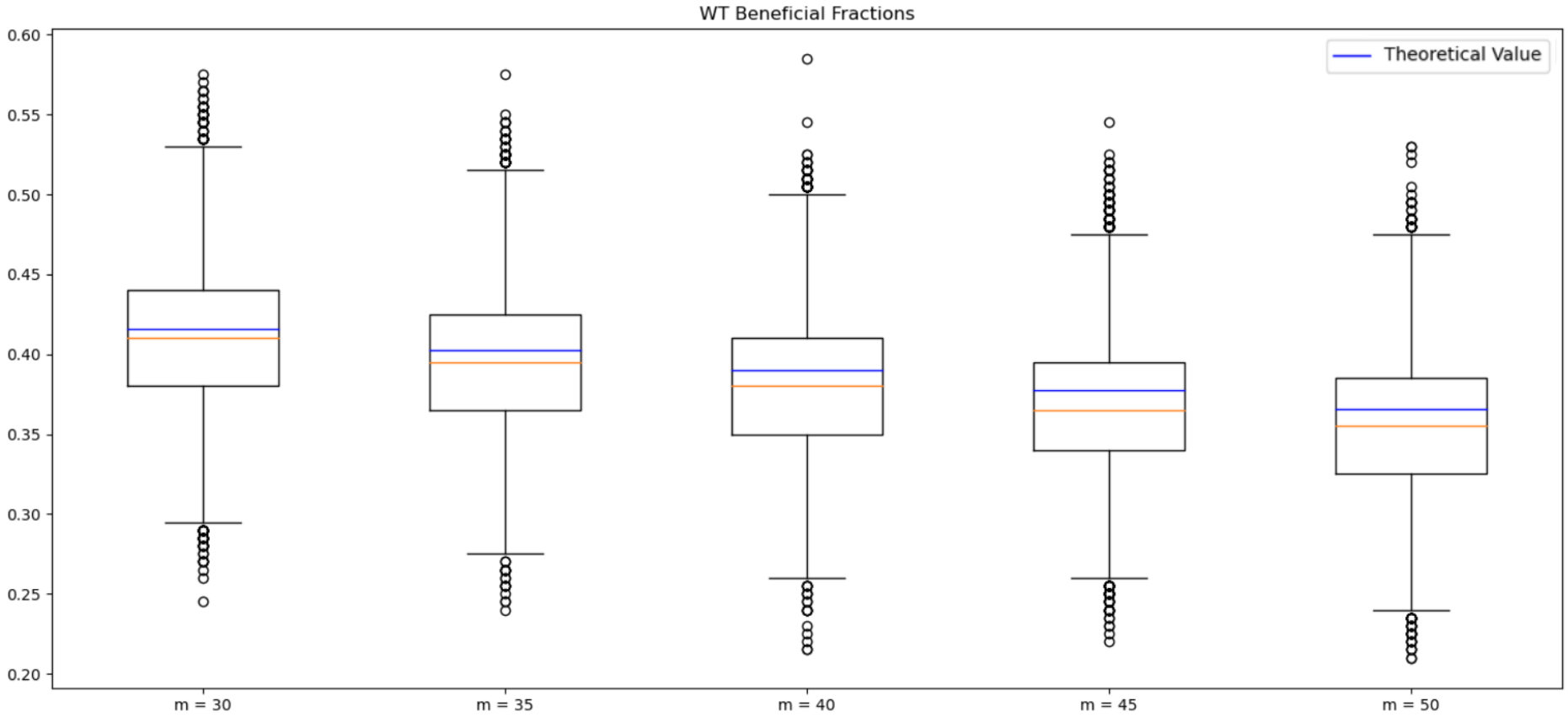


Fitness of Unmutated Loci After Mutation 10 (PDF)

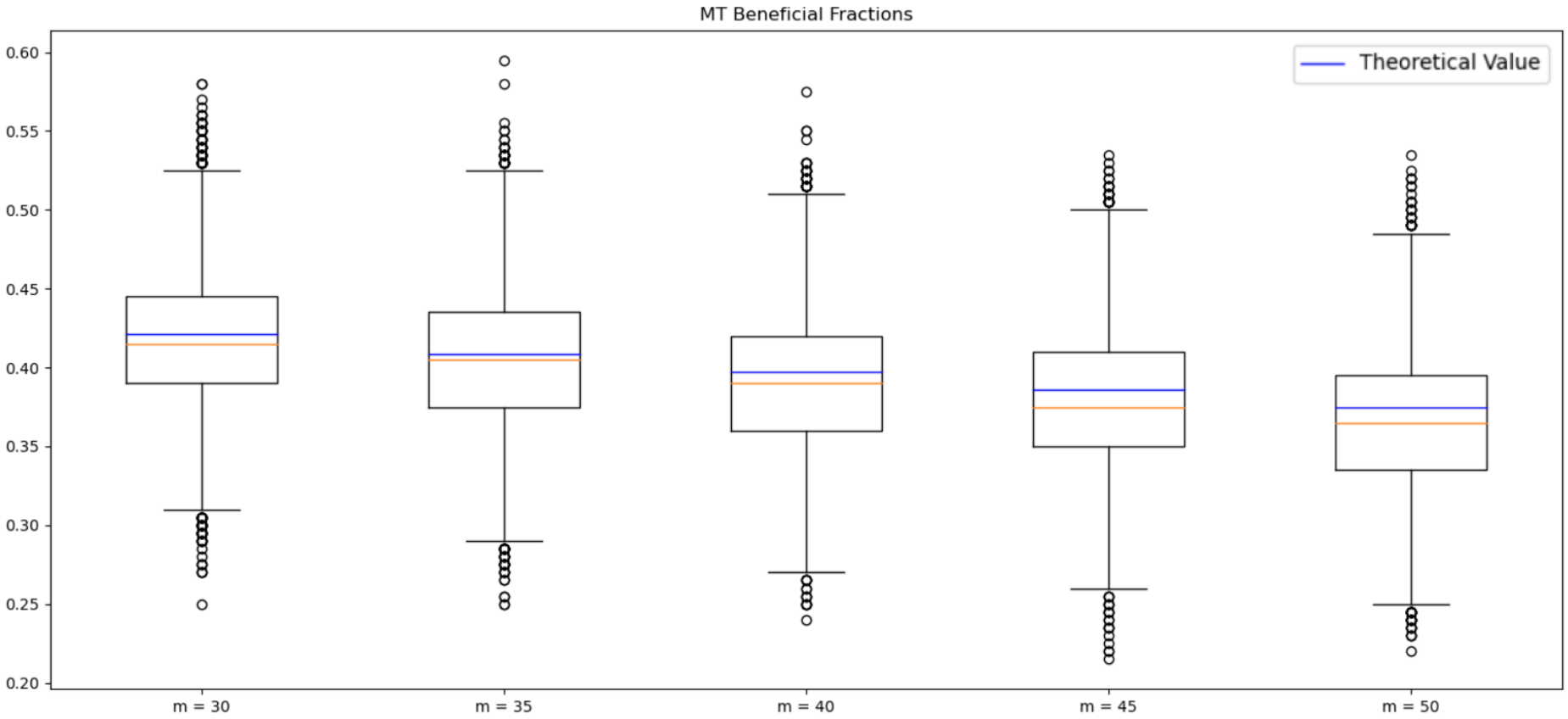


After only 1 mutation, the distribution of the unmutated loci looks very close to uniform, but as the walk progresses we see the depletion of lower-fitness loci. The distributions above come from data using genomes with 50 loci and so after only 10 mutations, the difference is significant.

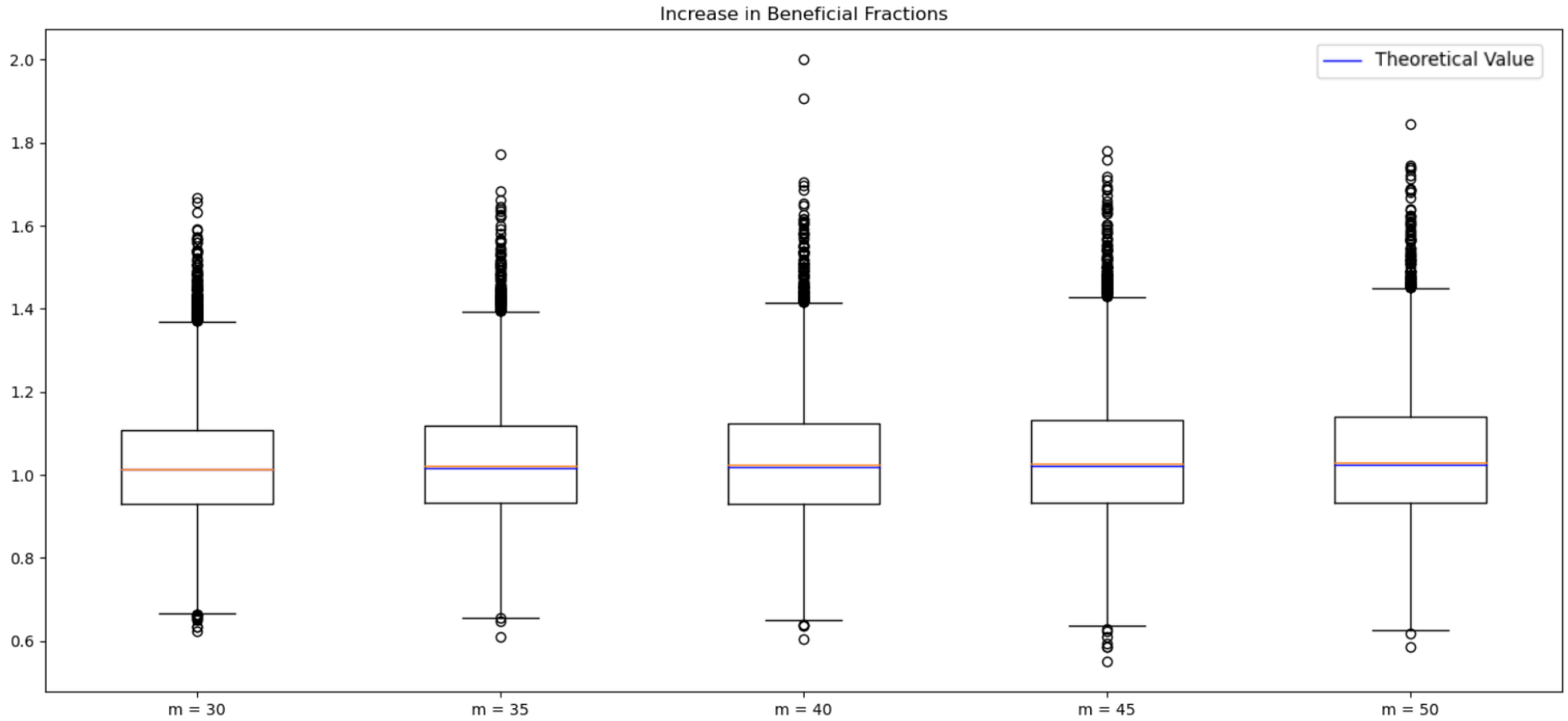
To test the accuracy of the theoretical equations, we ran simulations with $N = 200$ and went up to 50 mutations. Using a wild type Tv/Ts bias of 0.45, the fraction of possible mutations that are beneficial after 30 mutations is around 0.411 and decreases to about 0.354 after 50 mutations. The theoretical results are quite close, but we noticed that they get less accurate as mutations increase.



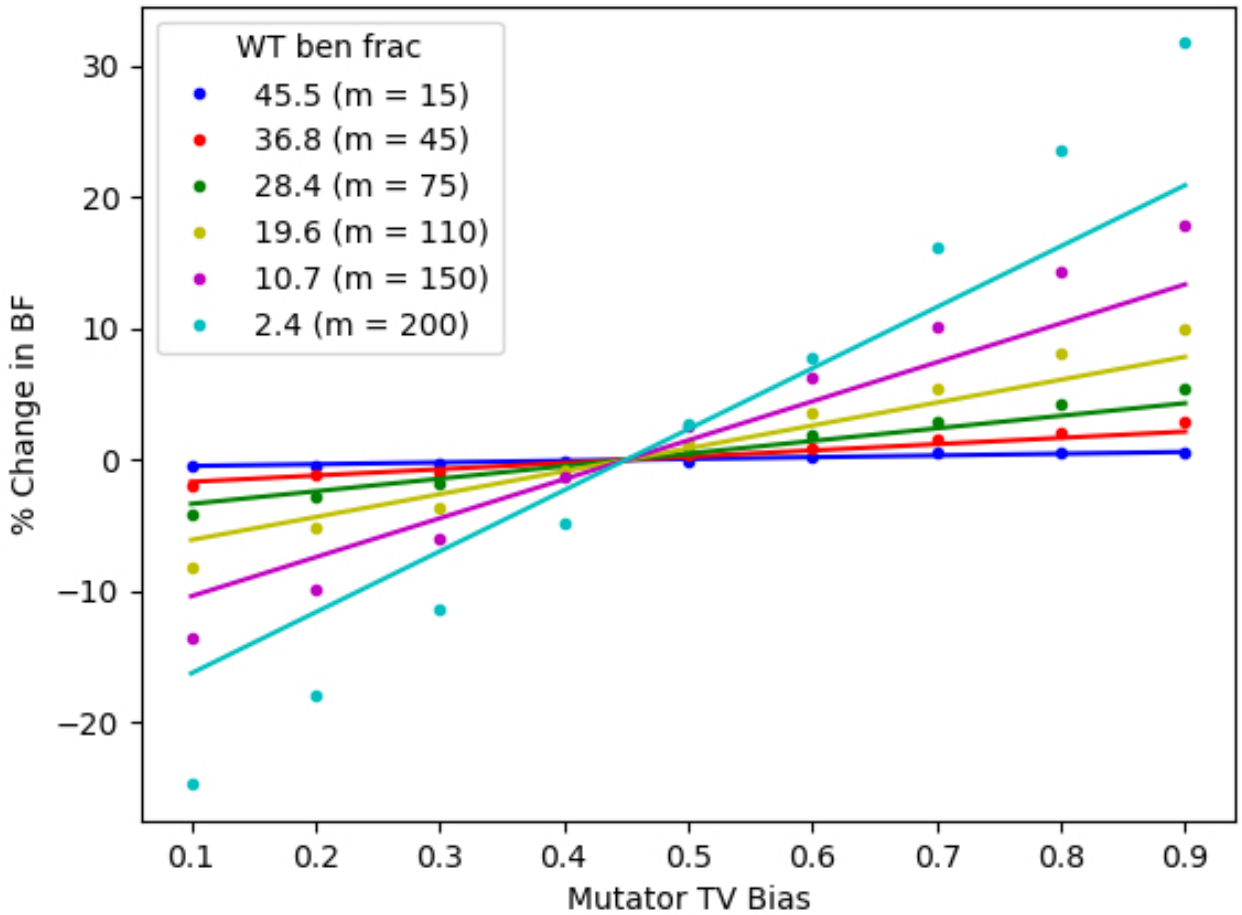
For the mutator strain, we shifted the Tv/Ts bias to 0.9. The beneficial fraction after 30 mutations was now about 0.417 and decreased to 0.365 after 50 mutations. We noticed similar amounts of error with the theoretical values as we did for the wild type.



Looking at the increase in the beneficial fractions from the wild type strain to the mutator, we noticed that the theoretical results were very accurate.



We went on to significantly decrease the wild type beneficial fraction by allowing up to N mutations and explored how using a different mutator bias affects the increase. As expected, the theoretical results became less accurate as we allowed more mutations. This could be due to biological assumptions being taken in the theoretical equations.



Solid lines represent theoretical values, points represent simulation results.

Next Steps

- This project has made considerable progress over the past few months and we've explored behaviours of adaptive walks in more depth than originally planned. The effect of the mutation bias on the DFE has not been widely researched and so there are still many areas to be explored.

- Going forward, we aim to further examine characteristics of adaptive walks and run more simulations with different numbers of loci, mutations, and different biases to try and determine why the theoretical results differ increasingly from the simulations as more mutations occur. Hopefully, they're able to be improved for longer adaptive walks.