Thesis Summary Report

Brian Davis

Introduction

This report is a brief summary of the work carried out thus far as part of my master's thesis. The intention of this report is to provide Dr. Guillaume Beslon with an overview of my research goals and my progress towards reaching these goals.

The overall goal of this thesis is to examine some of the factors which might lead to reductive evolution in simulated organisms. Reductive evolution is the process of an organism's genome being reduced in size, often leading to a loss of genes and thereby gene functions. Since it is very difficult to isolate the various causes of this phenomenon in real-world organisms, we are instead focused on using the *in silico* experimental evolution tool **aevol** to perform experiments, isolating one factor at a time to try to determine which are the most influential in reducing the genome of our simulated organisms.

Methods

As stated above, we are using **aevol** to create artificial organisms and then allowing them to evolve over several thousand generations. We began the experiments with five clonal populations, where each of the 1,024 organisms in each population has an identical genome. These genomes were initially completely random but they had been allowed to evolve for 10 million generations. The genome for these organisms was provided by Dr. Beslon. These five clonal populations form the *control* condition in our experiments, as they were allowed to evolve for 500,000 generations without any changes in their mutation rates, selection rates, or population sizes from their original generation. That is to say, the following aevol parameters have these values in the *control* condition for all five populations:

INIT_POP_SIZE - 1024

POINT_MUTATION_RATE - 1e-7
SMALL_INSERTION_RATE - 1e-7
SMALL_DELETION_RATE - 1e-7
MAX_INDEL_SIZE - 6

SELECTION_SCHEME - fitness_proportionate 1000

The only differences between the five clonal populations are that they each use a different random seed, which is to say that their pseudorandom number generators are seeded with different numbers, allowing random events to occur differently each time (e.g. mutations, insertions, deletions, etc.).

Starting with the same genome as the control conditions, we furthermore created additional clonal populations, also allowing them to evolve over 500,000 generations, but this time with different conditions. We had six conditions: +/- mutation rate, +/- selection rate, and +/- population size. For each of the conditions, the respective parameter value was changed, and the population was then also allowed to evolve for 500,000 generations. For each condition, the parameter values are shown in the table below:

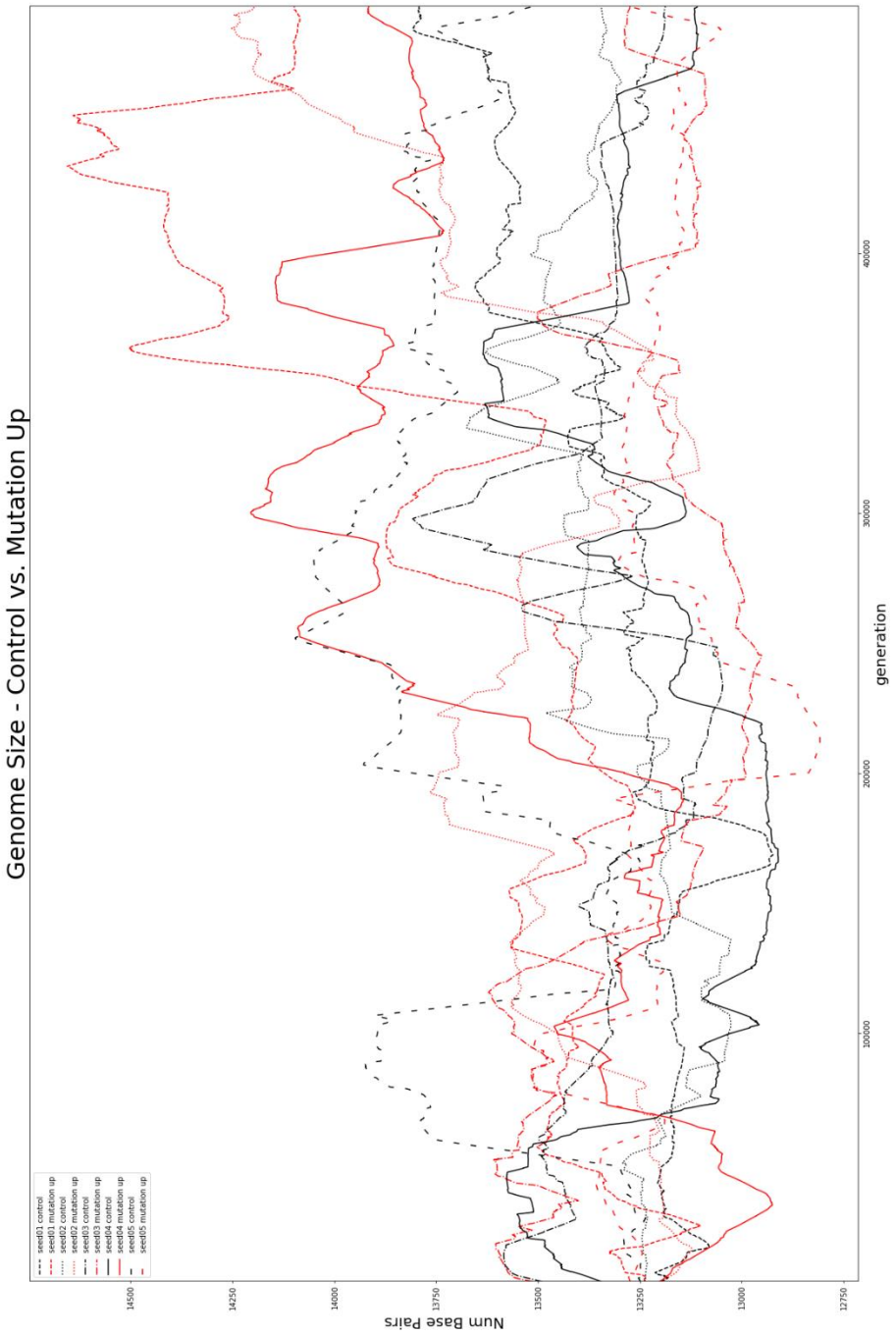
Parameter	Condition						
	C	M+	M-	S+	S-	P+	P-
INIT_POPULATION_SIZE	1024	1024	1024	1024	1024	**4096**	**256**
POINT_MUTATION_RATE	1.00E-07	**4.00E-07**	**2.50E-08**	1.00E-07	1.00E-07	1.00E-07	1.00E-07
SMALL_INSERTION_RATE	1.00E-07	**4.00E-07**	**2.50E-08**	1.00E-07	1.00E-07	1.00E-07	1.00E-07
SMALL_DELETION_RATE	1.00E-07	**4.00E-07**	**2.50E-08**	1.00E-07	1.00E-07	1.00E-07	1.00E-07
SELECTION_SCHEME=fitness_proportionate	1000	1000	1000	**4000**	**250**	1000	1000

In this table, C=control, M+=mutation up, M-=mutation down, S+=selection up, S-=selection down, P+=population up, P-=population down. Values which are **bolded** have been changed from the control condition.

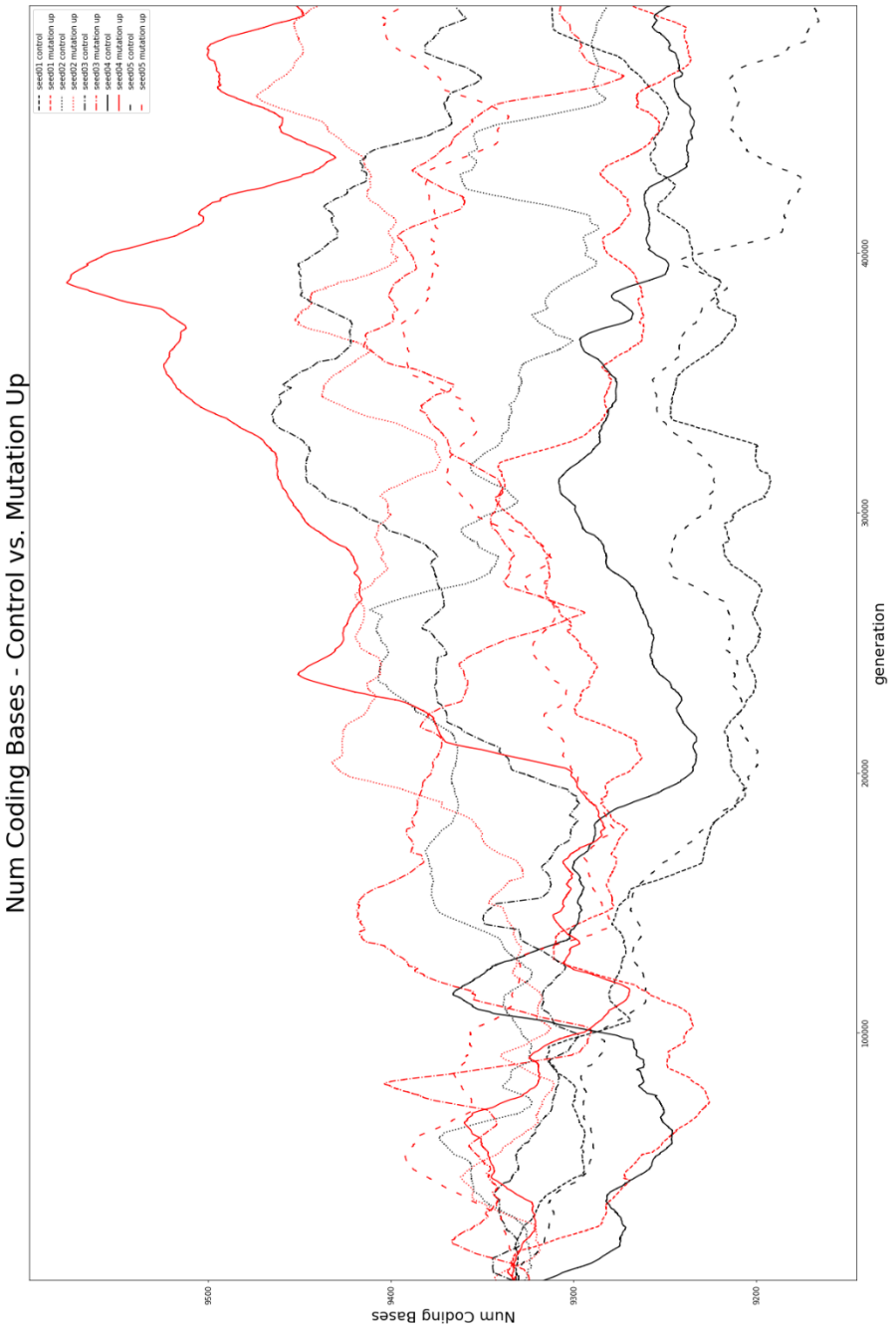
Results

Below we present our results of these experiments. So far, not all runs for all conditions have been completed: seed01 and seed02 are fully complete but seed03, seed04, and seed05 have only been completed for the *control* and *mutation up* conditions. All other conditions for seed03, seed04, and seed05 have so far only been completed to 100,000 generations, but the experiments are ongoing. Below, we present our findings for one condition, *mutation up*, for all seeds. In the graphs below, the conditions from the same seed (i.e. *control* and *mutation up*) have the same line pattern, and the *control* condition is always black, whereas the *mutation up* condition is red.

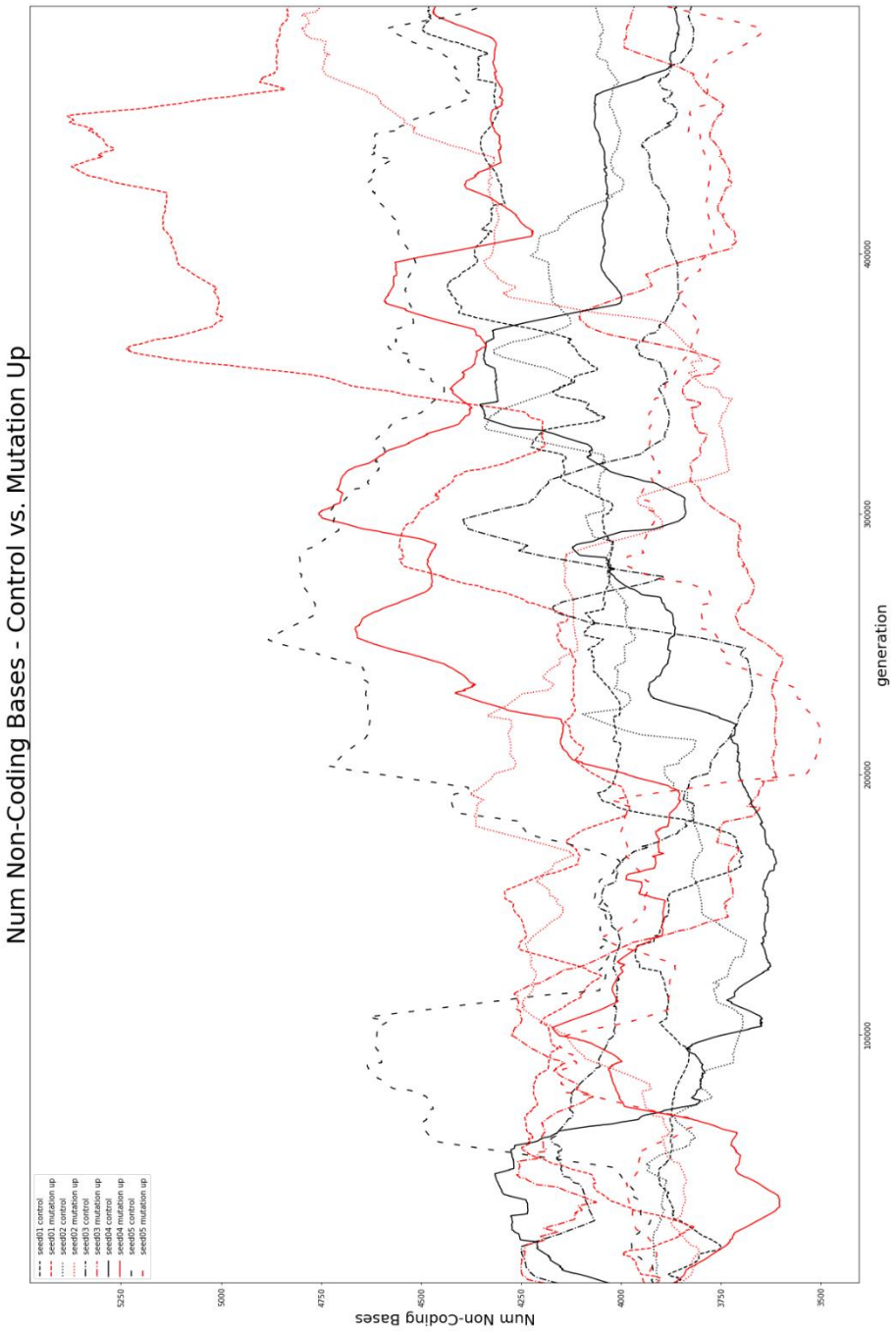
Genome Size


Genome Size - Control vs. Mutation Up
Num Base Pairs
generation
seed01 control
seed01 mutation up
seed02 control
seed02 mutation up
seed03 control
seed03 mutation up
seed04 control
seed04 mutation up
seed05 control
seed05 mutation up
14500
14250
14000
13750
13500
13250
13000
12750
100000
200000
300000
400000

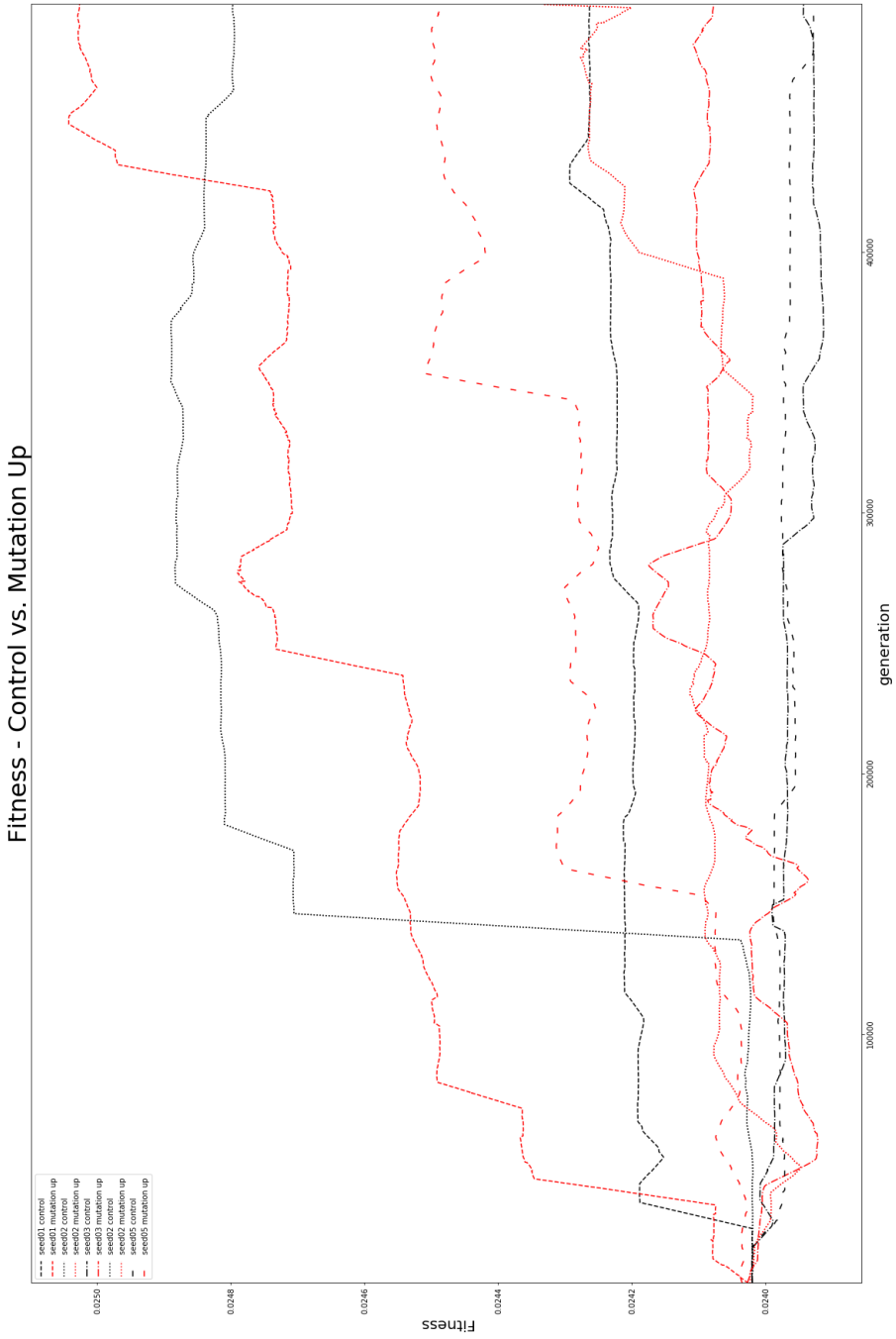
Number of Coding Bases



Number of Non-Coding Bases



Fitness


Fitness - Control vs. Mutation Up
seed01 control
seed01 mutation up
seed02 control
seed02 mutation up
seed03 control
seed03 mutation up
seed04 control
seed04 mutation up
seed05 control
seed05 mutation up
Fitness
0.0250
0.0248
0.0246
0.0244
0.0242
0.0240
100000
200000
300000
400000
generation

Discussion

TODO – Fill out discussion section. **Discuss relevant papers and expectations vs. our findings.**