UNIVERSITY OF CALIFORNIA SAN DIEGO

Simulating the Evolutionary Effects of Environmental and Genetic Variation on Life History in *Caenorhabditis*

A thesis submitted in partial satisfaction of the requirements
for the degree Master of Science

in

Biology

by

Rachel Erin Ross Goodridge

Committee in Charge:

Professor Scott Rifkin, Chair
Professor Jonathan Shurin, Co-Chair
Professor Shannon Ellis

2021

The thesis of Rachel Erin Ross Goodridge is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

ACKNOWLEDGMENTS

I would first like to acknowledge Professor Rifkin for his support as the chair of my committee. He has provided endless guidance throughout every step of this process and dedicated much of his time to assisting me in the lab, with creating and using this model, and with organizing and editing this thesis. His support was crucial for me to continue exploring, problem solving, and figuring out the next steps in my work. Without him, I would not have discovered my great interest in coding and this project certainly would not have been possible.

I would also like to acknowledge Professor Shurin and Professor Ellis for their support as members of my committee. Professor Shurin always gave me a clear picture of what exactly my project was about, why it is interesting, and where it is headed. Meeting with him helped me to look at the big picture and determine which things were both reasonable and feasible. Professor Ellis was my Python coding instructor and taking her class greatly improved my understanding of and skills in this language. Not only was the class a huge turning point in my project, but also she gave some excellent pointers on how to better display some of my data.

In addition, many thanks to the current and former members of the Rifkin lab for their support and advice, specifically Antonia Darragh, Joanna Bundus, Bing Yang, Jessica Bloom, Shea Summers, Michael Cradeur, Alexis Cugini, Matthew Campos, Rohan Kanchana, and Alexander Bevier. I especially could not have navigated through the lab without Jessica, Antonia, and Joanna helping me learn many procedures and techniques. Finally, a big thank you to my parents, family, and friends who have provided endless support and encouragement throughout my six years at UCSD, and especially through a global pandemic. I could not have done it without all of these people supporting and assisting me. Thank you.

ABSTRACT OF THE THESIS

Simulating the Evolutionary Effects of Environmental and Genetic Variation on Life History in *Caenorhabditis*

by

Rachel Erin Ross Goodridge

Master of Science in Biology

University of California San Diego, 2021

Professor Scott Rifkin, Chair
Professor Jonathan Shurin, Co-Chair

Nematodes such as *Caenorhabditis elegans*, *C. briggsae*, *C. remanei*, and *C. nigoni* are model organisms that primarily reside in rotting fruit and plant matter, feeding on the bacteria that inhabit these degrading vegetation (Frézal & Félix, 2015). When conditions are poor, worms must decide whether to go into a larval stage called dauer (Avery, 2014). Entering dauer would allow the worms to survive for months; however, the risk of dying in dauer is high and this results in a tradeoff. I created an extensive model in Python to simulate the population dynamics and decision-making strategies of worms and their responses to various environmental

conditions. This model includes genes related to both dauer and travel direction decision making strategies. Experiments showed significant evolution of both genes when there is stronger selection against worms in dauer, both genes when there is higher frequency of food availability, the dauer gene only when the dauer genotype to phenotype mapping is altered, and neither gene when seasonality in terms of environmental productivity is introduced. Stronger selection against worms in dauer led them to evolve a lower likelihood of dauer and a preference for traveling away from neighbors. A higher frequency of food availability also led them to evolve a lower likelihood of dauer but a preference for traveling towards food. A higher genotype to phenotype mapping value led them to evolve a lower likelihood of dauer as well. However, there may be some underlying patterns present in many of these experiments that require further study.

CHAPTER I : INTRODUCTION

The genus of nematodes, *Caenorhabditis*, including *C. elegans*, *C. briggsae*, *C. remanei*, and *C. nigoni* are model organisms that primarily reside in rotting fruit and plant matter. They feed on the bacteria that inhabit these degrading vegetation and, when the food supply depletes, they are forced to disperse in search of the next source (Frézal & Félix, 2015). This promotes a vigorous boom/bust life cycle. During periods of high bacterial abundance, these worms thrive and become populous. However, when the food has been consumed, the worms generally die before they are able to find better conditions (Félix & Duveau, 2012).

When the food begins to run out or when conditions are poor for other reasons (e.g. overcrowding), worms must decide whether to go into a larval stage called dauer. In general, worms go through a life cycle after hatching that consists of four larval stages (L1, L2, L3, and L4) and an adult stage, during which they are capable of reproducing. However, there is a secondary cycle that includes a stage called dauer as well as a pre-dauer stage (L2d) which bypasses the L2 and L3 stages. In fact, there are two dauer decisions - one whether or not to enter L2d and another whether to continue into dauer or exit back into the normal reproductive cycle (Avery, 2014). Choosing dauer would allow the worms to survive for months while they migrate via a larger invertebrate, travel small distances on their own, or simply wait for more favorable conditions. However, dauer do not eat so there is a tradeoff - being able to eat and develop immediately upon conditions improving but possibly dying before this occurs or dying in dauer while looking for better conditions. While a worm that did not choose dauer may have much faster progeny production, a worm that returned to the reproductive cycle after coming out of dauer may save its genetic lineage from extinction, assuming conditions improve enough for it to survive that long (Félix & Duveau, 2012).

These alternate phenotypes depend on more than just the environment. In addition to lack of food and high pheromone levels signaling overcrowding, other factors include abnormal temperatures, high environmental uncertainty, past events, genetics, and some element of randomness (Ailion & Thomas, 2000; Avery, 2014). For example, dauer worms form at a more moderate rate under temperatures ranging from 15 to 25 degrees, but at a much faster rate when 27 degrees is reached (Ailion & Thomas, 2000). Other factors, such as environmental uncertainty and past events, play into a worm's dauer decision as well. It is beneficial to be able to predict the future environment with some certainty. If conditions remain constant relative to the past, the future can easily be predicted. Continuously poor circumstances will likely lead to more of the same, and vice versa. However, if the environmental condition has some sort of temporal fluctuations or spatial variations, the worm must also take that uncertainty into account. The more uncertain the environment, the more a worm will favor the pre-dauer L2d (and potentially dauer) stage (Avery, 2014). Based on their own past or on the past of their ancestors, worms also have some experiential and epigenetic influence on their dauer decision. The likelihood of choosing dauer can be partially determined by the environmental condition and volatility experienced both previously and by the mother. Worms that go into dauer will produce offspring that are more inclined to make the same choices (Avery, 2014). Lastly, there is an element of randomness to this process. Just as each worm experiences slightly different circumstances, each worm may also make an individual decision that cannot be perfectly predicted. In reality, dauer is the more dangerous option, so the worms have to rely on a sort of bet hedging system to survive (Avery, 2014).

Given this information, I created an extensive model in Python to simulate the population dynamics and decision-making strategies of worms and their responses to various environmental

conditions. This program takes into account many basic known facts, but through the process of its creation, unknowns have been discovered and assumptions made in their place. All claims made in the model can be divided into one of the following three categories: things that have been simplified, things that are unknown, and things that are known facts. These lists can then be individually dissected and items either supported or rejected based on experimental evidence. The first category (things that have been simplified) is the result of avoiding complications that would make the model too unwieldy. The second category reflects assumptions that were important to have in the model but for which supporting data is sparse or non-existent. These point directly to new experiments that can improve the realism of the model. The third category are parts of the model that are based directly on the known biology of these worms.

CHAPTER II : MATERIALS AND METHODS

2.1 DEVELOPMENT OF THE MODEL

2.1.1 Grid Setup, Food, and Pheromones

The model operates on a multidimensional grid system which contains information about the worms, food, and pheromones in each location. It is set up as a 3-dimensional numpy array with 22 layers that span a 100 by 100 unit region. It wraps around into a torus shape so as to keep worms from falling off or running into the edge. This creates a space that maximizes roaming ability, yet limits physical distance. Each layer of the grid contains various different information. The first two layers in combination provide a 2-dimensional description of every possible position on the surface of the torus. The next two layers describe the amount of food and pheromones existing in each of those locations. The remaining layers describe the number of worms in each of those locations, divided into subcategories based on gender and developmental stage (Figure 1).

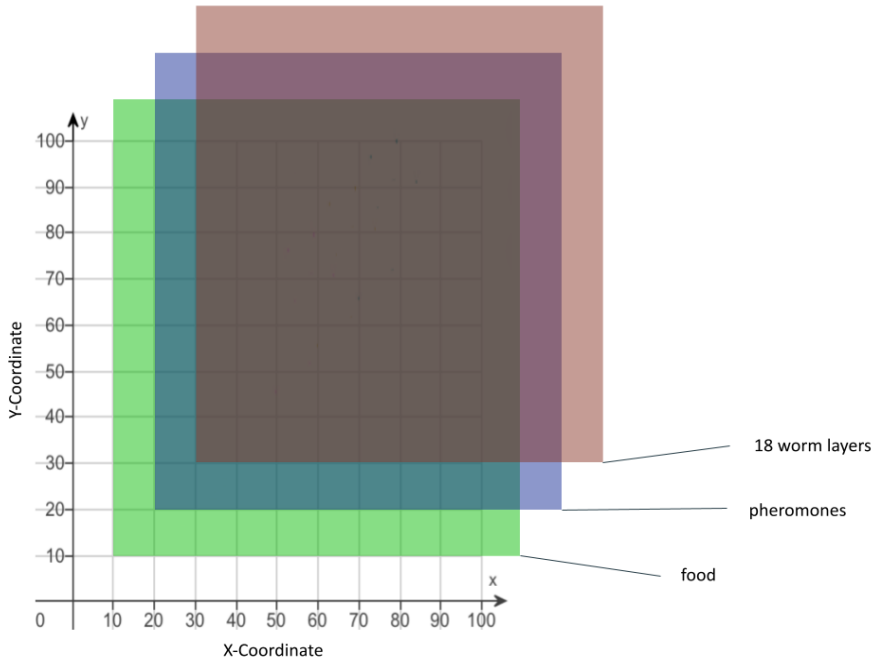

Figure 1: Flattened visual representation of the grid. There is a grid system setup in the program that contains information about location, food, pheromones, and worms. The first two layers describe the x and y locations two-dimensionally (pictured in white). The next two layers describe the amount of food (pictured in green) and pheromones (pictured in blue) at each location. The next 18 layers describe the amount and types of worms, broken down by gender and stage, at each location (pictured in red).

The food on the grid exists only within specified patches. The default patch size and orientation covers a 10 by 10 square and is 10 units away from other patches. Assuming the default grid size, the maximum number of patches this would allow is 25 (Figure 2). A simulation will begin by populating the grid with a specified number of food patches in locations randomly chosen without replacement from the 25 options. Throughout each simulation, food growth occurs logistically based upon a set rate and carrying capacity. There is also the option for this rate to fluctuate. Intermittently, new patches will appear, adding onto any existing patches, at some oscillating rate. This is meant to simulate rotten fruit dropping from a tree or bush and, thus, making more *E. coli* available for consumption. The option for oscillation can introduce some element of seasonality, during which there are periods of rapid patch repopulation and periods of low environmental productivity.

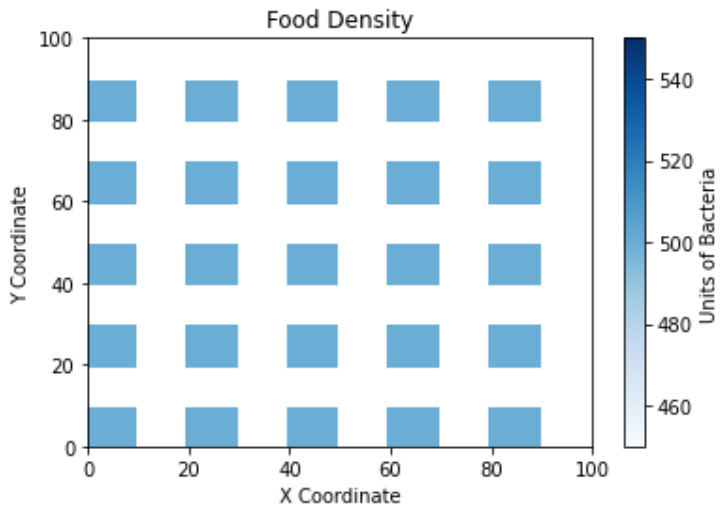

Figure 2: Possible food patch locations. The third layer of the grid which corresponds to food values in each location is pictured here. Based on the x and y coordinates in a two-dimensional view, food density is shown in units of bacteria based on the default initial amount and occupying all possible locations.

One advantage to setting up the grid with this many layers of information is the locations of specific types of worms can be easily identified. Any worm can determine exactly who and where each of its neighbors are. To do this, four additional, almost identical versions of the original grid are created with slight modifications to account for each of the four cardinal directions. The north, south, west, and east grid give the same information, but shifted one unit north, south, west, or east respectively. For instance, looking at the same ordered pair (x, y) and its corresponding data in all four of these grids will give information about what food, pheromones, and neighbors are surrounding the point (x, y) in the original grid. This will become useful for various worm decision-making processes later on.

2.1.2 Worm Properties

At the beginning of every simulation another numpy array is created to contain information about each individual. Every worm has the following 44 properties: (1) an identifier, (2) gender, (3) tally of food consumed, (4) life stage, (5 - 7) location, (8) energy, (9 - 12 & 25 -

44) various genes, (13 & 14) record of parents, (15 - 22) tally of time spent per stage, (23) alive or dead status, and (24) decision whether or not to travel. They are arranged in the order indicated in a 1,000,000 by 44 array, allowing for the maximum number of worms alive at any given time to be one million by default. Identifier, gender, genes, living status, and parents are constant throughout the lifetime of an individual worm. Other properties are continuously updated, including the amount of food consumed, life stage, location, energy, time spent, and decision to travel. These properties factor into the decisions made by the worms and also allow the program user to keep a record of the life history of every worm.

There are basic properties that can describe each worm (the following numbering system is indicated above). (1) Their identifier, for example, makes it possible to uniquely recognize every individual and examine their specific information. (2) Their gender will play a role in decision-making as well as determining their reproductive method. Different species of *Caenorhabditis* have different mating systems - adrodiecy vs. dioecy. Species including *C. elegans* and *C. briggsae* have mostly self-fertilizing hermaphrodites and some males, while other species like *C. remanei* have outcrossing males and females (Chasnov et al., 2007). (3) The amount of food they have consumed will directly relate to when they progress to the next life stage. Once they have consumed some predetermined amount of *E. coli*, they may then molt into the next life stage (Lee et al., 2006). (4) Worms also keep track of which stage they are in and (15 - 22) how long they have spent in each stage. Their stage will determine how much food they can eat in one time step, the amount of pheromones they release, and their amount of energy expenditure. Depending on what stage they are in, they can molt only to specific other stages. They also only reproduce if they are adults. Stages include egg, L1, L2, L2d, L3, dauer, L4, adult, and old (the point when adult worms no longer reproduce). The lengths of time spent in L1

and L2d have an effect on their dauer decisions. (5 - 7) A worm's location is important for determining food availability and neighbors. Location is described in 2D in terms of x and y coordinate position and also separately in 1D using a singular position value. (13 & 14) Worms record who their parents are. If a hermaphrodite self-fertilizes, then the identifier of the worm will be listed twice. This allows for lineage tracing and counting the number of generations that have passed. (23) Whether or not a worm is alive is the first thing determined in each function in Python. All dead worms will not be considered in calculations. Lastly, (24) worms decide whether or not to travel based on a multitude of factors (detailed below). All worms are continuously moving, but not all worms will always travel to a new location. Once they have decided to travel, then they can calculate which direction is desired.

In addition to these basic properties, worms have other more complex properties (the following numbering system is a continuation from above). (8) The amount of energy a worm has determines whether, or for how long, that worm will survive its particular conditions (Van Raamsdonk et al., 2010). Energy can be gained and lost in several ways. Energy must be gained through the consumption of food and can be thought of as a proxy for the increase in ATP count (Shi et al., 2015). Energy expenditure can happen due to metabolic processes and reproduction. It is continuously costly to stay alive, and an additional cost is added for laying an egg (Zanni et al., 2015). If worms can not find sufficient amounts of food quickly enough to replenish their energy supply, they will perish from starvation (Dilberger et al., 2019). Additionally, all energy-related events occur in proportion to the life stage of each worm. Smaller worms will gain less and spend less energy than larger worms (Zanni et al., 2015).

Another more complicated property is their genetics. (9 & 10) Different strains of *C. elegans* show different tendencies towards going into dauer under the same conditions,

suggesting some genetic influence (Viney et al., 2003). In reality there are many genes that combine to influence this, but in the model I have simplified them down to one single gene that affects the likelihood that the worm will choose to go into dauer. My focus is not on the genetics *per se* but on the effects of ecology, demographics, and selection on this choice. Since these worms are diploid, they have two copies of each gene which can each take on a different value from a range of options, in agreement with the "infinite alleles" model (Kimura & Crow, 1964). Each new version of the gene (a new mutation) can be considered a new allele (Ewens, 2016) and these are represented quantitatively in the program by an expression level. The expressed value of this trait is the average of their two alleles and represents the effect of the gene (e.g. how much mRNA/protein gets produced) which translates to a genetic bias towards or against choosing dauer.

Similarly, (11 & 12) there are genetic influences on social behavior. Some strains of *C. elegans* (like N2) will clump while feeding, whereas other strains (or when a mutation is introduced) may prefer solitary behavior (Dorado-Morales et al., 2014; Serena Ding et al., 2019). This can again be simplified to a single gene that affects which direction a worm prefers to travel. Worms generally tend to move away from neighbors and towards food, a decision that is dependent on their surroundings but skewed by genetics. Depending on the strain and possible mutations, they will have some predisposition for favoring food more or favoring solitude more. Finally, (25 - 44) a neutral set of genes can be used to distinguish lineages from the original members of the population. All of the worms from the beginning of a simulation have ten genes each with two alleles unique to them that pass on through the generations. By studying these, worms can be matched to their ancestors.

2.1.3 Concept of Time

The passing of time is tracked throughout a simulation. Each time step in the program is one iteration of the simulation (all functions run one time) and represents approximately one hour. During this iteration, all worms are simultaneously performing many tasks, such as moving, eating, decision making, and reproducing. Other non-worm functions that run one time per iteration include updating the grids, food growth, and pheromone decay. All of these tasks put together make up one singular time step and it would be considered equivalent to the events that occur during an hour. The number of hours that have passed is counted so events can be assigned to particular locations in time. A typical simulation would run for 30,000 iterations which is equal to about 3.4 years and can reach upwards of 200 generations of worms.

During every iteration, there are four tasks that are not performed by worms themselves. First, the list of neighboring grids must be updated to reflect the north, south, west, and east directions on the flattened torus. Worms can then make decisions based on this refreshed information. Later, logistical food growth and exponential pheromone decay can proceed according to Equations 1 and 2, respectively. The carrying capacity (K) and rate of growth/decay (r) are parameters determined in the setup of the simulation, while N_0 and N represent the current and new values, respectively. As new food patches fall on top of the old ones (like a new rotten fruit falling to the ground), the carrying capacity will increase accordingly. Shown in Figure 3, the food patches will fall with some probability (p, blue line) that can oscillate over time based on the amplitude (a), frequency (f), and center value (c, gray line) of a sine curve function (Equation 3). Lastly, the grid must be updated at the end of every iteration to reflect what changes have occurred throughout that time step. This can properly show the new locations and

stages of all worms who have traveled and molted during that hour, in addition to the new food and pheromone amounts.

$$N = \frac{K \cdot N_0 \cdot e^r}{K + N_0 \cdot (e^r - 1)}$$

Equation 1: Logistical growth. This equation is used to simulate the growth of the bacteria on the grid. N_0 is the current amount of *E. coli* and N is the new amount. The carrying capacity (K) and growth rate (r) are parameters in the setup of the model.

$$N = N_0 \cdot e^r$$

Equation 2: Exponential decay. This equation is used to simulate the decay of pheromones on the grid. N_0 is the current amount of pheromones and N is the new amount. The decay rate (r) is a parameter in the setup of the model.

$$p = a \cdot sin(ft) + c$$

Equation 3: Seasonality. This equation is used to vary the rate at which food patches appear (p) on the grid over time (t). The sine curve can be altered by changing the amplitude (a), frequency (f), and center value (c).

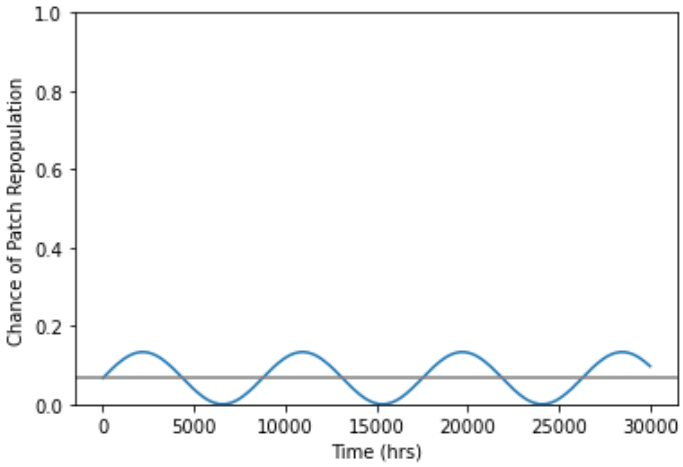

Figure 3: Seasonality of food patch repopulation rate. Every iteration, there is some chance that a new patch of food will appear. This can be altered over time, based on a sine curve with set center, amplitude, and frequency values. The center line (the default patch repopulation rate) is shown in gray and the oscillations that may occur throughout time are shown in blue. The default frequency allows for a period of one year.

2.1.4 The Dauer Stage

When worms go into dauer, they act a bit differently from other worms. The major benefits are due to the fact that they do not need to eat food and are able to withstand enough stress to survive for months (Hu, 2017). In the model, worms in dauer will not eat, but they will

count the food they come across. The purpose of this is to determine when conditions are becoming more suitable. Once they reach another food patch, they can begin making their decision to come out of dauer. Along with not eating, worms in dauer do not gain or spend any energy, so they can live for a very long time. After 4 months (or 2880 iterations), worms in dauer will finally die from having wandered too long with no success (Altun & Hall, 2009). There are also significant costs associated with choosing the dauer stage. Most dauer do not survive (Félix & Duveau, 2012) and, for those that do find food, there is a slim chance that they will choose to come out of dauer when there is not quite enough food for reproduction. To mimic this survival rate, dauer worms in the simulation will be automatically killed off at some high rate when they first enter the stage. This introduces a tradeoff which can allow worms to evolve some compromised strategy between high and low genetic likelihood of dauer.

2.1.5 Movement and Travel Decision Making

At every time step, worms decide whether or not they should move and then, if they decide to move, which direction. The decision to travel is primarily based on how much food and surrounding pheromones are sensed in the location of each individual (Ben Arous et al., 2009; McGrath & Ruvinsky, 2019), but the strength of these factors depends on genetics (Dorado-Morales et al., 2014). If the food in their location is gone or less than one unit remains, then they will be forced to travel. In all other cases, their probability of choosing to travel is based on Equation 4 and pictured in Figure 4. Their genotype (g) acts as a weighted average between the fraction of food and pheromones and must be between the values of zero and one, but is typically closer to one half. Fractions are calculated based on the food (f) at the current location divided by the maximum amount of food (f_m) a worm can sense and the pheromones (w)

in all directions surrounding the worm divided by the maximum amount of pheromones (w_m) a worm can sense. Figure 4 assumes that the factors are weighted equally, the maximum food is 2000, and the maximum pheromones is 500. This displays the probability of any individual deciding to travel based on a range of conditions.

$$p = g \cdot (1 - \frac{f}{f_m}) + (1 - g) \cdot \frac{w}{w_m}$$

Equation 4: Probability of traveling. Each worm calculates the probability (p) of whether or not it chooses to travel at every time step. This is based on the value of their travel direction gene (g), the amount of food (f), perceived maximum food (f_m), surrounding pheromones (w), and perceived maximum pheromones (w_m).

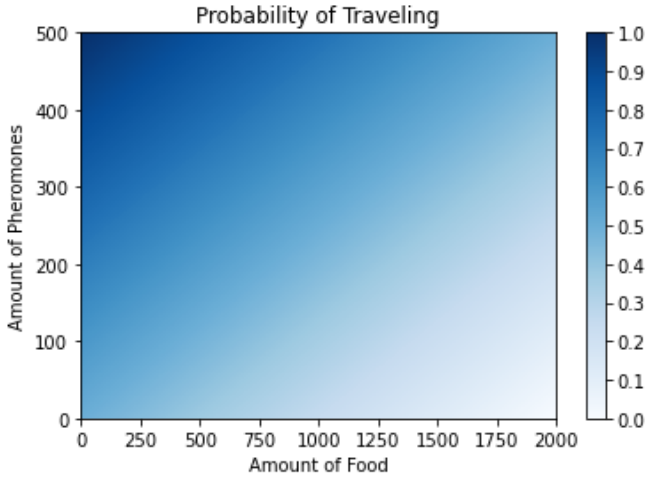

Figure 4: Probability of traveling. The probability that a worm will choose to travel is based on food and surrounding pheromones but is weighted depending on worm preference. In this figure, the assumed weight is 0.5, meaning that these factors are averaged equally. Then some probability of travel between 0 and 1 can be calculated.

Once the worm has decided to travel, it must then decide on the direction. The direction will differ based on the gender of the worm. Adult males tend to move towards both food and neighbors (White et al., 2007), while all other worms will choose to go towards food, but prefer a less crowded environment (Hu, 2017). Again based on the weighted average of the same gene (g) as before, the food (f) and pheromones (w) in each direction are calculated as a fraction of the total surrounding food and pheromones and the subscripts on these variables indicate the direction (Equations 5 and 6). The probability of all four directions (north, south, west, and east)

will add up to one, and this constitutes the random chance that a worm will pick each of those directions. Since they are simply probabilities, the direction of travel chosen is not necessarily going to be the best as far as achieving the worm's goal of finding food and avoiding or approaching neighbors. However, the best direction will be the most likely to get picked.

$$p = g \cdot \left(\frac{f_{direction}}{f_n + f_s + f_w + f_e + 4}\right) + (1 - g) \cdot \left(\frac{w_{direction}}{w_n + w_s + w_w + w_e + 4}\right)$$

Equation 5: Adult males' travel directions. This equation is calculated four separate times, one for each cardinal direction, to determine the probability (*p*) of moving in that particular direction. The sum of all four probabilities is one and the worms must choose between them. This decision is based on the value of the travel direction gene (*g*), fraction of food (*f*) in that direction (indicated by subscripts), and the fraction of pheromones (*w*) in that direction (indicated by subscripts). To avoid dividing by zero, one unit of bacteria and one unit of pheromones was added to each direction (hence + 4 in the denominators).

$$p = g \cdot \left(\frac{f_{direction}}{f_n + f_s + f_w + f_e + 4}\right) + (1 - g) \cdot \left(\frac{\frac{1}{1+w_{direction}}}{\frac{1}{1+w_n} + \frac{1}{1+w_s} + \frac{1}{1+w_w} + \frac{1}{1+w_e}}\right)$$

Equation 6: All other worms' travel directions. This equation is calculated four separate times, one for each cardinal direction, to determine the probability (*p*) of moving in that particular direction. The sum of all four probabilities is one and the worms must choose between them. This decision is based on the value of the travel direction gene (*g*), fraction of food (*f*) in that direction (indicated by subscripts), and the fraction of pheromones (*w*) in that direction (indicated by subscripts). To avoid dividing by zero, one unit of bacteria and one unit of pheromones was added to each direction.

2.1.6 Eating

All living worms will choose to eat at every time step if there is food in their location. The amount of *Escherichia coli* available in each location on the grid will be divided up among all the worms in that same location with portions weighted based on the stage of the worm. Each life stage corresponds to a different portion of food. In general, larger/older worms will eat more than the smaller/younger ones and will take a larger cut of the available supply (Rodríguez-Palero et al., 2018). However, worms will not eat more than their specific portion nor more than what is available. After worms consume the food, it gets subtracted from the grid. They also gain a proportional amount of energy and add to their tally of food consumed, similar to assessing nutritional value (Ben Arous et al., 2009; MacNeil et al., 2013). The only exceptions

are eggs and worms in dauer which do not eat or gain energy, but they still add to their food consumed tally. This allows them to molt to the next stage at the proper time. Eating is the key to keeping worms alive and growing.

2.1.7 Molting and Dauer Decision Making

A worm must eat a specific quantity of bacteria before it can molt to the next stage (Klass, 1977; Lee et al., 2006). This can be achieved by counting the total units of food consumed. As worms progress through their life cycle, they will be required to eat increasing amounts of food before they can molt. For example, in the model, a worm in L2 will require roughly twice as much food as a worm in L1. However, no matter how much food is available to them, they will not be allowed to eat so much that they are immediately able to advance. In fact, the portions were created in such a way that if they eat one full portion of food at every time step, they will be ready to molt after the typical length of time spent in their stage (assuming one time step is roughly equivalent to one hour). Once they have consumed enough food for that particular stage, then they will automatically molt into the subsequent stage, with the exception of worms in L1, L2d, and dauer. The only difference is that when these worms are ready to molt, they have to make decisions about whether or not to choose (or come out of) dauer. After a worm has molted, its energy level will be diminished to a specified amount to account for the cost of the transition.

After consuming the required amount of food, worms in L1 and L2d calculate the probability of going into pre-dauer or dauer (Equation 7) based on how long they have spent in that stage (t) which reflects the availability of *E. coli* for consumption, the total surrounding pheromones (w) divided by the perceived max (w_m), and their genetic susceptibility for picking

dauer (*d*). There is also another variable that determines how strongly their genotype will relate (or map) to their phenotype (*m*). In Figure 5A, the value of this mapping component is 1 and gene values to the left of the average "time spent" line will mostly make the worm favor going into dauer while gene values to the right will make the worm favor returning to the reproductive cycle. In Figure 5B, the value of the mapping component is 3 and this smooths out the probabilities more on the diagonal so worms with any gene values could potentially select either option given the proper conditions. There is also a "cutoff" in dark red in the top left corner. This occurs when a worm has spent too much time in L2d without consuming the required amount of food (Equation 8). In this case, the assumption is made that there is not enough food for the worm to reach molting and it will die unless given the option to molt into dauer sooner. Based on circumstances like those shown in Figure 5 and when some minimum cutoff value (*c*) is reached, these worms will be forced into dauer. However, the threshold for this can be adjusted.

$$p = 0.5(\frac{1}{1 + e^{\frac{d-t}{m}}}) + 0.5\frac{w}{w_m}$$

Equation 7: Probability of going into pre-dauer (L2d) or dauer. This probability (*p*) is calculated when a worm in L1 or L2d consumes the required amount of food. It is based on the value of the dauer gene (*d*), the amount of time spent in L1 or L2d (*t*), a genotype to phenotype mapping value (*m*), and the fraction of surrounding pheromones (*w*) out of the perceived maximum (w_m).

$$\frac{1}{1 + e^{\frac{d-t}{m}}} > c$$

Equation 8: L2d cutoff into dauer. If this statement is true, a worm will get sent from L2d into dauer without needing to consume the required amount of food. The cutoff value (*c*) and genotype to phenotype mapping value (*m*) are parameters in the setup of the program. This calculation is also based on the value of the dauer gene (*d*) and the amount of time spent in L2d (*t*).

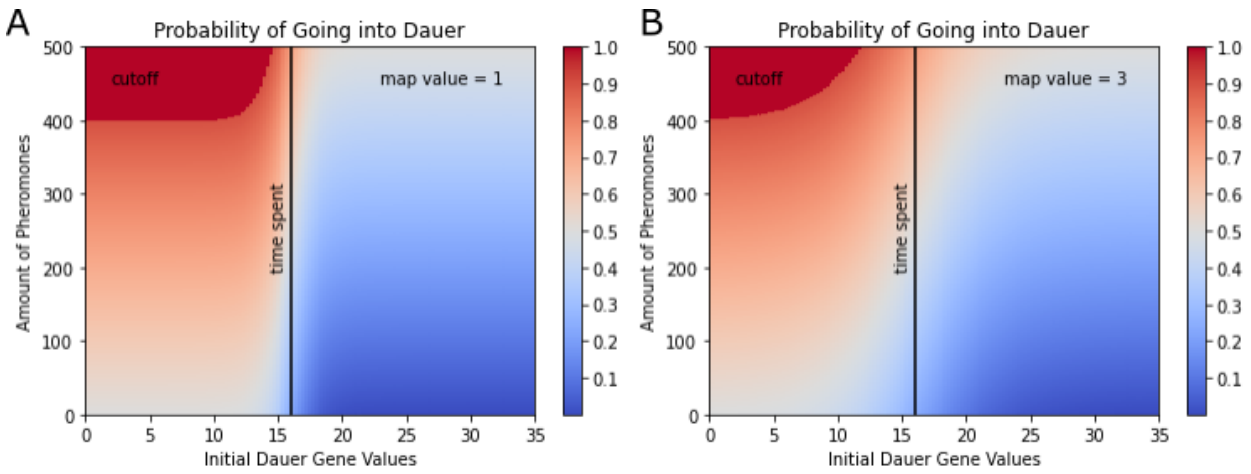

Figure 5: Probability of going into pre-dauer (L2d) or dauer. There is a complex probability that a worm will go into pre-dauer or dauer based on many factors, including surrounding pheromones, time spent in L1 or L2d (as a proxy for food availability), and some genetic influence. The more pheromones exist and the lower the quantitative value of the expressed dauer gene, the more likely the worm is to go into dauer, and vice versa. The coloring in this figure will shift horizontally depending on how much time a worm has spent in the previous stage. In this case, a vertical line is drawn at 16 to show how a worm who spent 16 time steps in L1 or L2d would decide. Generally, there is a clear difference between the left half of the graph being more likely to go into dauer (red) and the right half of the graph being less likely to go into dauer (blue). To soften this effect, another variable that maps the genotype to phenotype is introduced. In Figure 5A, the mapping value is 1 which produces a harsher divide, while in Figure 5B the mapping value is 3 and this blurs the line. There is also a dark red area in the upper left corner that represents when worms are forced into dauer (from L2d only), even though they did not consume the amount of food required for molting.

Once a worm has gone into dauer, it then has to decide whether or not to come out of dauer. Recovery from dauer is dependent on both population density and food availability (Hu, 2017). After finding enough food, the probability that a worm will re-enter the normal reproductive cycle depends on the amount of food in its current location (f) divided by the perceived food max (f_m) and the total surrounding pheromones (w) divided by the perceived pheromone max (w_m). This is averaged to determine if current conditions are suitable and the result is the probability that it will choose to stay in dauer (Equation 9). Once it is immediately able to return, it still may or may not choose to do so.

$$p = 0.5(1 - \frac{f}{f_m}) + 0.5\frac{w}{w_m}$$

Equation 9: Probability of staying in dauer. This probability of staying in dauer (p) is calculated when a worm has found the required amount of food. It is based on the fraction of food (f) out of the perceived maximum (f_m) and the fraction of surrounding pheromones (w) out of the perceived maximum (w_m).

2.1.8 Reproduction

Caenorhabditis reproduction can occur via two systems, either androdioecy or dioecy (Chasnov et al., 2007). The default parameter assumes androidecy, because *C. elegans* are the primary focus. However, the code is set up to accommodate for either system, depending on the species being studied. First the females or hermaphrodites will collect sperm from whatever male crosses their path. At every time step, if an adult male and an adult female or hermaphrodite are occupying the same space on the grid, their names will be added to a list of those who mated. Then later, this information can be accessed when the female or hermaphrodite is choosing to lay an egg. Eggs will be created in the same location as the worm who laid them and every egg costs the female or hermaphrodite some energy to produce, but there are a couple methods.

One reproductive system involves self-fertilization, which has a few requirements. This will only occur when the reproductive method is set to androdioecy and when the hermaphrodites either do not have any sperm or choose not to use it. At every time step, hermaphrodites decide whether or not to lay an egg and then whether or not to use sperm (if they have any). Once self-fertilization has been determined, the resulting offspring have a 99% chance of being hermaphrodite and a 1% chance of being male (Yeh et al., 2018). Each of the offspring will select two copies of every gene from their mother, since they are diploid organisms (Brenner, 1974), with some very low set rate at which mutation may occur. This method tends to produce more homozygous offspring over time.

Both reproductive systems involve outcrossing, which will only occur when females or hermaphrodites have collected sperm (and the hermaphrodites have chosen to use it). At every time step, females or hermaphrodites have some probability of reproducing. Once sperm is obtained and used, the resulting offspring have a 50% chance of being hermaphrodite/female and a 50% chance of being male (Yin & Haag, 2019). Each egg will be assigned to a father so it can more easily collect a copy of each gene from the proper parent. Again, when the genes are passed down, there is some small chance of mutation. This method may produce more heterozygous offspring and could potentially help maintain a bit of diversity in the population.

2.1.9 Logistics of Running Simulations

This model includes many separate functions (such as eating, moving, and reproducing) that must all come together to run the simulation. A master function incorporates all these modular elements in a logical order (Figure 6). For example, before a worm can reproduce, it should know if it can mate with any of the worms in its location. In Figure 6, the superscript on each of the functions indicates the order in which they run. In addition to calling all the functions in this specific order, the master function must also keep track of variables. The parameters initially passed in will become incorporated into a dictionary and used to create the other variables, only three of which are shown in Figure 6. These are then passed into whichever functions will use them and can be altered inside. When new variables are created, they will be returned from that particular function and the master function must return all the variables at the end as well.

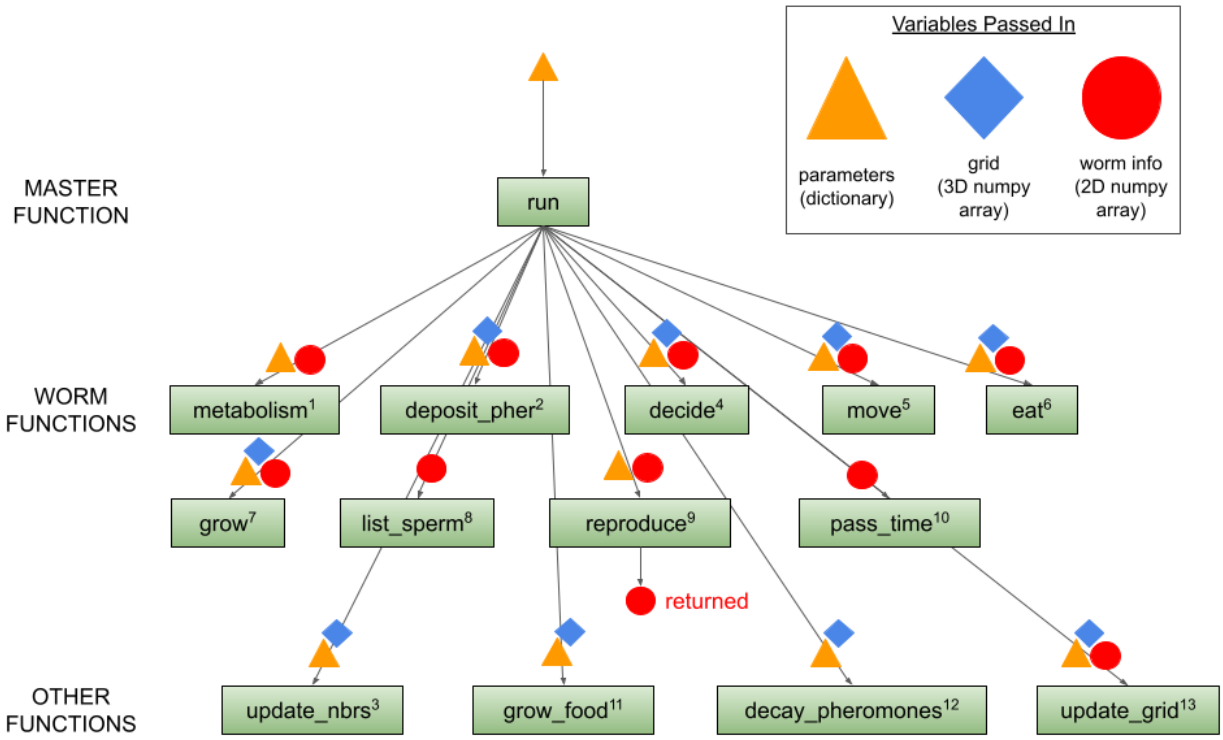

Figure 6: Integration and order of functions. This diagram shows all functions that are called within the master function "run." The superscript in the corner of each secondary function indicates the order in which it is called. The orange triangles (representing parameters), blue diamonds (representing the grid), and red circles (representing worm information) are shown above each function that requires them. Most of the parameters are defined in the initial call of the master function, but all variables are ultimately created within the master function and then passed into the other functions for use. Worm information is also returned from the "reproduce" function. Note that this only shows three of the many variables used during a simulation.

A crucial component to the creation of this model was dealing with storage space. One variable is an array that can hold up to one million worms and their properties (array size can be chosen by the user), but throughout a simulation, many more than one million worms may exist. To improve the storage efficiency of the program, rather than just adding onto the existing array, it can be overwritten with new information. When all the one million slots are full or there is not enough room for the new progeny, all dead worms get exported into a file which makes space for more eggs. This is a relatively clean and simple solution, but it assumes that the number of worms alive at one time will not exceed one million.

Another key to running successful simulations is to save snapshots of the program along the way. At specific intervals, all current parameters can be written out to a file in a binary format. This is fairly straightforward to save and then access later and is useful for understanding what events may have occurred throughout the simulation. Additionally, at a more frequent rate, the size of current genetic lineages and number of alleles for the dauer gene can be written out to a text file. This allows for a finer resolution understanding of the change in abundance of worm lineages and dauer alleles over time. Throughout a simulation, many files are created to save various information at different time points and these will all be used later in analysis functions for a clearer picture of that experimental trial.

2.2 RANDOMNESS AND ASSUMPTIONS IN THE MODEL

2.2.1 Randomness in the Model

At the beginning of and throughout every simulation, there are a multitude of random choices made by the program which ultimately have an effect on the outcome. The more randomness introduced, the more unpredictable the result will be. Each of the random events, while seemingly minute, can accumulate quickly and will build on each other over time to create very different circumstances given the same set of parameters. This can be controlled by selecting the seed in each simulation. The seed dictates the choices made in each random occurrence and thus allows for duplication. Using the same seed will produce all the same choices and consequently the same outcome. Therefore, each repetition of the same experiment with the same parameters should make use of a different seed than the ones before it. Only then can a more full range of possibilities be discovered.

There are four random events that occur in the setup of each simulation. First, the number of initial food patches is a parameter input into the simulation, but the location of those patches is not predetermined. The location of each patch is randomly chosen without replacement from a list of specified possible locations. Second, the gender of each of the original worms is randomly chosen, but weighted based on which reproductive system has been selected. If androdioecy is the method, then each worm will have a 99% chance of being hermaphrodite and a 1% chance of being male (Yeh et al., 2018). If dioecy is the method, then each worm will have a 50% chance of being female and a 50% chance of being male (Yin & Haag, 2019). Third, the placement of the initial population is random. Each worm will choose a food patch as well as a location within the patch to begin as an egg. Fourth, the values of the genes that relate to dauer and travel direction are randomly chosen in the initial population. The dauer gene selects any random real number between two fixed values. The travel direction gene selects any random real number based on a normal distribution with a fixed center and standard deviation. Since each worm is diploid and these random choices are independent of one another, it is likely that all worms in the initial population will begin as heterozygotes at these loci.

There are also sixteen random events and/or choices the worms make that occur during every iteration throughout the simulation. (1) The growth rate of the food is randomly chosen based on a normal distribution, assuming the fluctuation parameter is not set to zero. (2) There is some probability that a new food patch will fall at every time step and (3) where that patch falls is also randomly chosen. (4) The direction that a worm chooses to travel is random but is weighted based on its surroundings. (5) When a worm decides whether or not to molt into dauer, there is some randomness in that decision, either from L1 to L2/L2d or (6) from L2d to L3/dauer. (7) There is a very high chance that a worm will die in dauer, and when worms first enter dauer,

the ones who die are randomly selected based on this probability. Additionally, (8) worms must choose when to come out of dauer and this is based on surroundings, but somewhat random as well. (9) During reproduction, a hermaphrodite or female must choose when to lay an egg and similarly, (10) when to use available sperm. The new progeny then has a multitude of initial choices to make. They must choose (11) their gender based on the reproductive strategy, (12) whether a genetic mutation will occur in each of the genes received, (13) what genetic mutation occurs if it does mutate, and (14) which of the mother's (and the father's) genes will be passed down to them. (15) For those that have reproduced using outcrossing methods, the eggs must each randomly choose which one is their father from the list collected by their mother. Lastly, (16) every worm must decide whether or not to travel. All of these events put together are what create the simulation and they make each outcome unique.

2.2.2 Assumptions in the Model

All claims made in the model can be divided into three categories: things that have been simplified, things that are unknown, and things that are known facts. Some choices in the model were made in an effort to avoid overcomplication of the simulations. For example, the food patches have been simplified so they are all the same size, have a set distance from each other, and get replaced randomly in space. These items are all declared out of ease. Other simplifications made are related to the mating and reproduction process. It is assumed that all adult hermaphrodites or females and adult males in the same location will successfully mate at every time step and that once a male has died, his sperm will not persist. The latter point can prevent the need to dig male information out of a file once dead worms have been removed from the memory of the program. Additionally, hermaphrodites in the model that have sperm will tend

not to use it. In reality, male sperm will displace hermaphrodite sperm and be used almost exclusively in succeeding mating (Stewart & Phillips, 2002). This introduced bias against using sperm accounts for the ineffective mating of males and lower fitness of outcrossed offspring which leads to selection against these progeny (Anderson et al., 2010). Furthermore, hermaphrodites (or females if they have sperm) can decide whether or not to lay an egg at every time step. In reality, hermaphrodite *C. elegans* produce 250-350 progeny each (Hodgkin & Barnes, 1991). However, it is not feasible to have such extreme exponential growth in the simulation, so this number was significantly reduced to control the exploding population. Some other simplifications are related to hatching and molting. Eggs will all hatch after the same amount of time has passed and the amount of food a worm requires to molt to the next stage is the same for each worm. While this may not realistically be the case, the simulation still provides some variation of length of each developmental stage nevertheless. As far as the energy requirement for molting, this has no direct cost in the model, but instead energy levels are limited preceding the molt to a specific value for that stage. Lastly, and arguably most noteworthy, temperature has not been factored into this model, but could be incorporated in the future.

The creation of this model has also brought up a multitude of things that are unknown about these worms. Many unknowns are related to the movement direction of the worms. For example, if there is no food, the model assumes that a worm must immediately travel to a new area to seek more. The simulation also operates under the assumption that worms can not only smell their surroundings (Ben Arous et al., 2009) but also that there is some maximum amount of food and pheromone sensory input that can be experienced. This point was crucial for a few of the equations. Another food-related unknown is whether worms are constantly eating whatever *E. coli* is available to their maximum capability and how much each worm consumes relative to

its developmental stage. The portions of food and timing of consumption rate for each stage was just assumed to be double after nearly every molt. However, it is possible to quantify the growth of a worm and tie this to the development of each life stage, because there are size thresholds for each molt (Uppaluri & Brangwynne, 2015). Yet it is nonetheless difficult to directly measure food consumption or assume how this translates into the model. Lastly, this simulation does not account for differences in the likelihood of going into dauer by gender. This could be easily studied in the laboratory in the future, but currently remains unclear. For the most part, this model has been created based on known facts, which have been substantially previously stated. However, information from future experiments can be incorporated to further improve the realism of the model.

CHAPTER III : RESULTS AND DISCUSSION

3.1 QUALITY CHECK

Each simulation was checked to ensure that it ran properly and contained reasonable information. This can be done by looking through a collection of additional functions written alongside the program that display graphs and calculate statistics. These data include the final generation number, the density of worms across the grid, the fraction of worms by gender and stage, the fraction that die in each stage, and the amount of time spent per stage. Excluding the generation number, these results remain relatively consistent across every experimental variation and simulation conducted. In addition, all functions can be run at each of the saved time points in order to view changes that occurred throughout time.

The goal of running simulations for 30,000 iterations was to reach at least 100 generations so that sufficient time for evolution would be allowed. Typically these simulations reach somewhere between 175 and 215 generations, which is well above the goal. However, each one is different and can also be food-dependent. In environments with less resource availability, the simulations may only reach 150 generations due to the reduced environmental productivity which is capable of supporting fewer worms. All simulations were long enough to show some sort of evolution (i.e. there were clear trends and stable results).

The next thing to check is the worm density throughout the simulation. There are a variety of ways this plot can be organized, namely by the stage or by the gender of the worms. Density tends to be highest within food patches, which confirms the clumpy feeding behavior observed in *C. elegans* (Dorado-Morales et al., 2014; Serena Ding et al., 2019). Once a food patch becomes depleted, the worms will spread out in search of the next patch. In Figure 7, tight clumps that are more dense can easily be distinguished from groups that are dispersing.

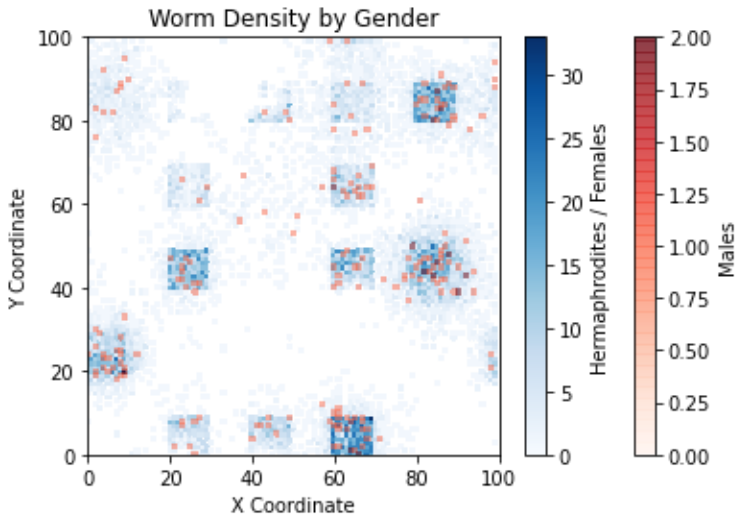

Figure 7: Density of worms plotted by gender. This is an example of what the worm density may look like at any given time during a simulation. Here, a heatmap of males (red) is overlaid on a heatmap of hermaphrodites (blue) spread across the flattened torus. While the male data is translucent, it still tends to cover some of the female data in those locations.

It is also pertinent to check the number of existing worms of each particular gender and stage throughout time. In androdioecious systems, such as *C. elegans*, few to no males should be maintained within the population. To verify that the parameters have been set properly, a ratio of roughly 99:1 hermaphrodites:males is to be expected (Figure 8A). By stage, there tend to be primarily eggs and L1 larvae in these simulations (Figure 8B). This is likely due to the rapid rate of egg production. In addition, simulations always have an abundance of adult worms, which is realistically the stage in which they spend the longest amount of time and should be very plentiful. As for the worms in dauer, on the other hand, there are very few displayed in Figure 8B, but this is likely because a high proportion die instantly in the program instead of lingering and so may not be an accurate representation of the number of worms in dauer in a wild population. Besides that, the other amounts seem relatively similar and all are almost always consistent across simulations as well.

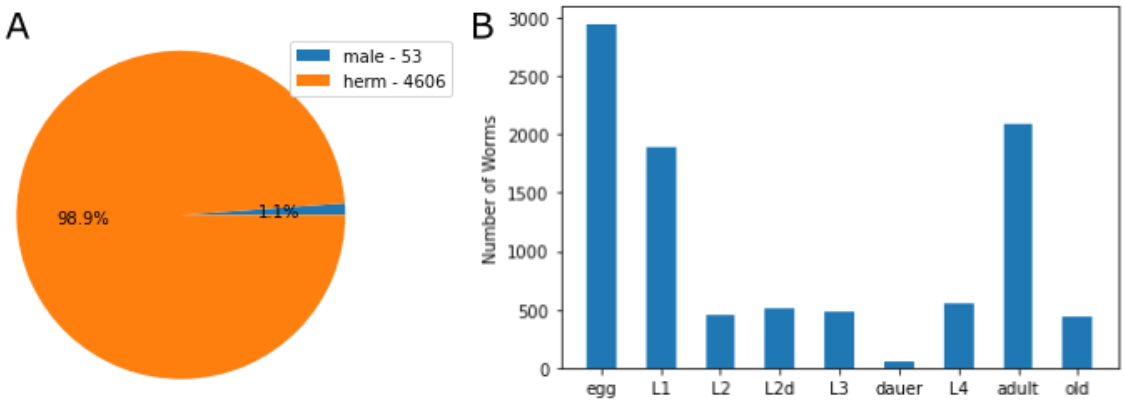

Figure 8: Number of worms by gender and stage. Figure 8A is an example of what the ratio of hermaphrodites:males looks like throughout any given simulation. The percentages are listed within the pie chart and the abundances are listed in the legend. Figure 8B is an example of what the abundance of worms per stage looks like throughout any given simulation. In general, worms mostly exist as eggs, L1 larvae, and adults. On occasion, populations can exist exclusively in dauer, but this figure shows the most typical distribution.

Another thing to check would be the number of worms that die in each stage (Figure 9). Most worms die in L1, likely because they did not hatch in a food patch or they ran out of food and could not get to the next patch in time. The chance of a worm dying in L1 is roughly 65-69% in every simulation. Furthermore, there is an even higher chance of dying as an adult, usually around 85-89%, rather than by natural causes during the old stage, it is unclear why this is the case. It could simply be due to chance, since worms spend much of their time in this stage, there is more opportunity for death. In any case, all other death rates are remarkably lower, aside from the preset rate for dauer worms.

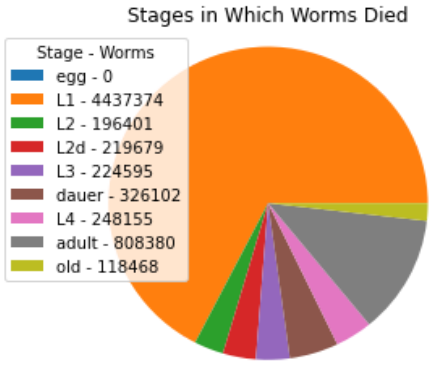

Figure 9: Number of worms dead by stage. This is an example of how many worms have died in each stage at the end of any given simulation. The quantities are listed in the figure legend beside their respective stages. While the exact number of worms may vary, the size of the pie chart sectors remain roughly consistent. Note that eggs don't die in this program. Something else of note is that L1 arrest is not included in this simulation. If an egg hatches into an environment without food, this L1 worm can arrest and survive an additional two weeks (Lee et al., 2012). The lack of this feature may be a reason why so many L1 worms die in the simulations. However, it is unclear whether surviving for an extra two weeks in L1 would affect the worm's chances of finding the next food patch.

The last item that should be checked after every simulation is the amount of time the worms spent in each stage on average. The typical amount of hours *C. elegans* spends in each stage is known (Avery, 2014) and practically worked into the parameters in the model. Assuming each worm eats the maximum amount of food possible at every time step, the shortest amount of time it could take to reach the next stage is predetermined. However, the length of time each individual spends is variable and may be shorter or longer than the preset amount. The only cases where the lengths of time spent are shorter are when L2d worms are cut off and forced into dauer or when a worm dies. When the length of time spent is longer, it is due to lack of food availability. Thus, it is important to check these averages to verify that the lengths of each stage are sensible. In Figure 10, the averages indicate that worms have generally been spending the proper amount of time and none are living for longer than they should.

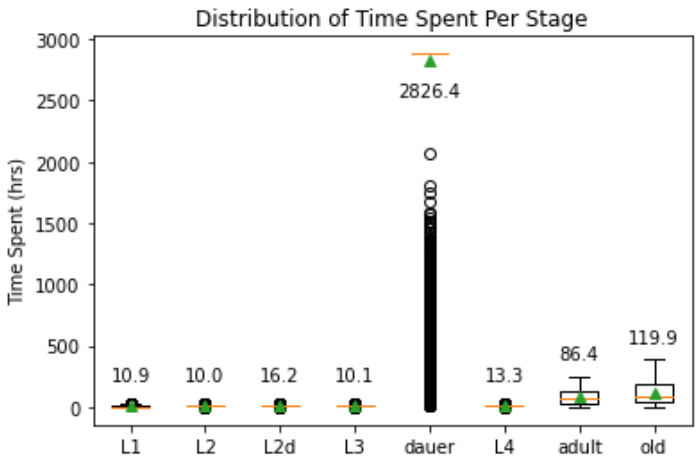

Figure 10: Distribution of time spent per stage. This is an example of the lengths of time spent in each stage, and is typical of every simulation. The yellow bars represent the median and the green triangles represent the mean. The mean is also printed above each box plot. There are few outliers for each except in the case of the dauer worms. Most will be immediately culled to represent the ones that die in search of food (shown in this plot at the maximum limit of 2880 as if they've spent 4 months searching), and all those who survive are plotted as outliers below the average.

3.2 PARAMETERS MEASURED

3.2.1 Clumpiness and Diversity

Throughout a simulation, the diversity of the population can be assessed by observing the total number of genetic lineages, calculating the diversity within each individual patch of food on the grid, and measuring the diversity across food patches throughout the grid. During every simulation, the total number of lineages will always decrease over time as unsuited or unlucky ones do not survive. Since new lineages cannot be introduced, this trend tends to be the natural progression. However, it can happen at different rates. Generally, only one or two lineages will remain by the end of each simulation, but on occasion, three or more persist. These unusually diverse situations could potentially be resolved with additional iterations. It may be the case that only one lineage will ultimately persist in most simulations given ample time. However, no simulations were run past 30,000 iterations, so this remains unknown.

In general, the diversity within each patch of food is relatively high early in the simulation and then declines as the overall number of lineages decreases. Every patch begins with a mixture of worms - each with different unique lineage tracing markers. As some lineages start to become dominant over other ones, patches will become more and more homogenous. The diversity of each location on the grid can be measured by calculating the Shannon diversity index. This is the proportion of each lineage (*p*, i.e. the number of worms in that lineage out of the total number in that location) times the natural log of that fraction, summed across every lineage (*n* in total) occupying that location (Equation 10). While diversity within food patches remains higher than in surrounding areas, there is an overall decrease over time (Figure 11).

$$H = -\sum_{i=1}^{n} p_i \ln p_i$$

Equation 10: Shannon diversity index (*H*). This diversity measure is calculated separately for each location on the grid. The measurement is based on the proportion of each lineage occupying that spot (*p*) times the natural log of that proportion for all *n* lineages in that location.

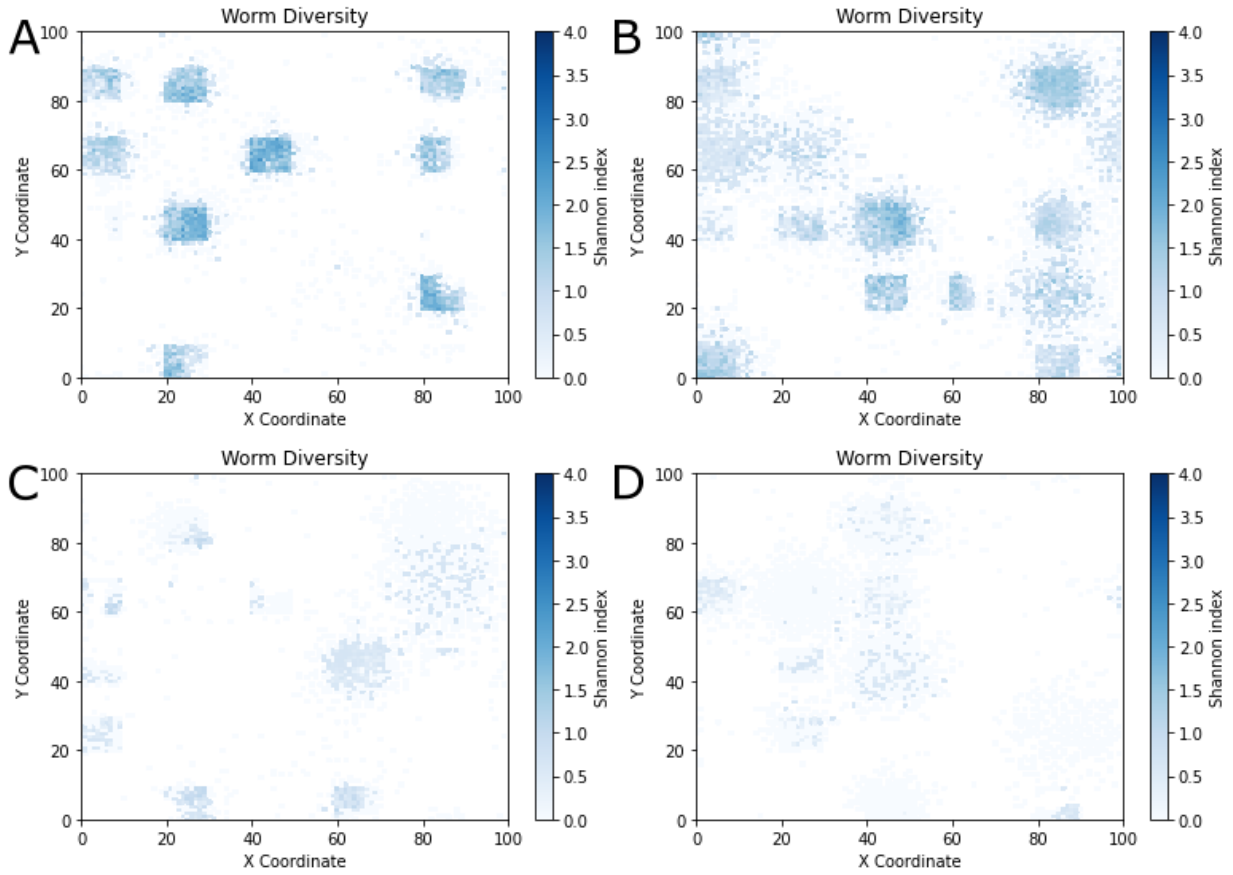

Figure 11: Worm diversity over time. These heatmaps show the change in diversity over time. Figure 11A represents time point 500, B is 2000, C is 15000, and D is 30000. Each location on the grid is calculated and colored according to its Shannon diversity index. Darker colors represent more diverse areas. For consistency between figures, the Shannon diversity index values are limited between zero and four.

In general, the diversity across food patches is also relatively high early in each simulation and then decreases as worms spread out across the grid. Throughout the initial parts of the simulation, lineages tend to be somewhat clumpy. The progeny of each worm will not wander too far from the remainder of their genetic line. They may occupy nearby patches, but remain a cohesive group. Then, as there are fewer and fewer lineages over time, the worms will become fully dispersed across the grid. The distance between worms can be calculated according to Equation 11, where (x_1, y_1) and (x_2, y_2) are the locations of two worms in two-dimensional space, w represents the width of the grid, and h is the height. This can then be computed pairwise across all worms in a lineage to find the average clumpiness of each genetic line. By the end of

the experiment, the few lineages remaining will not be clumpy at all compared to a random sample of the population (Figure 12).

$$d = \sqrt{min(|x_1 - x_2|, w - |x_1 - x_2|)^2 + min(|y_1 - y_2|, h - |y_1 - y_2|)^2}$$

Equation 11: Euclidean distance on the two-dimensional surface of a torus. This formula is a distance measurement (d) between two locations on the grid (x_1, y_1) and (x_2, y_2) using the height (h) and width (w) of the grid.

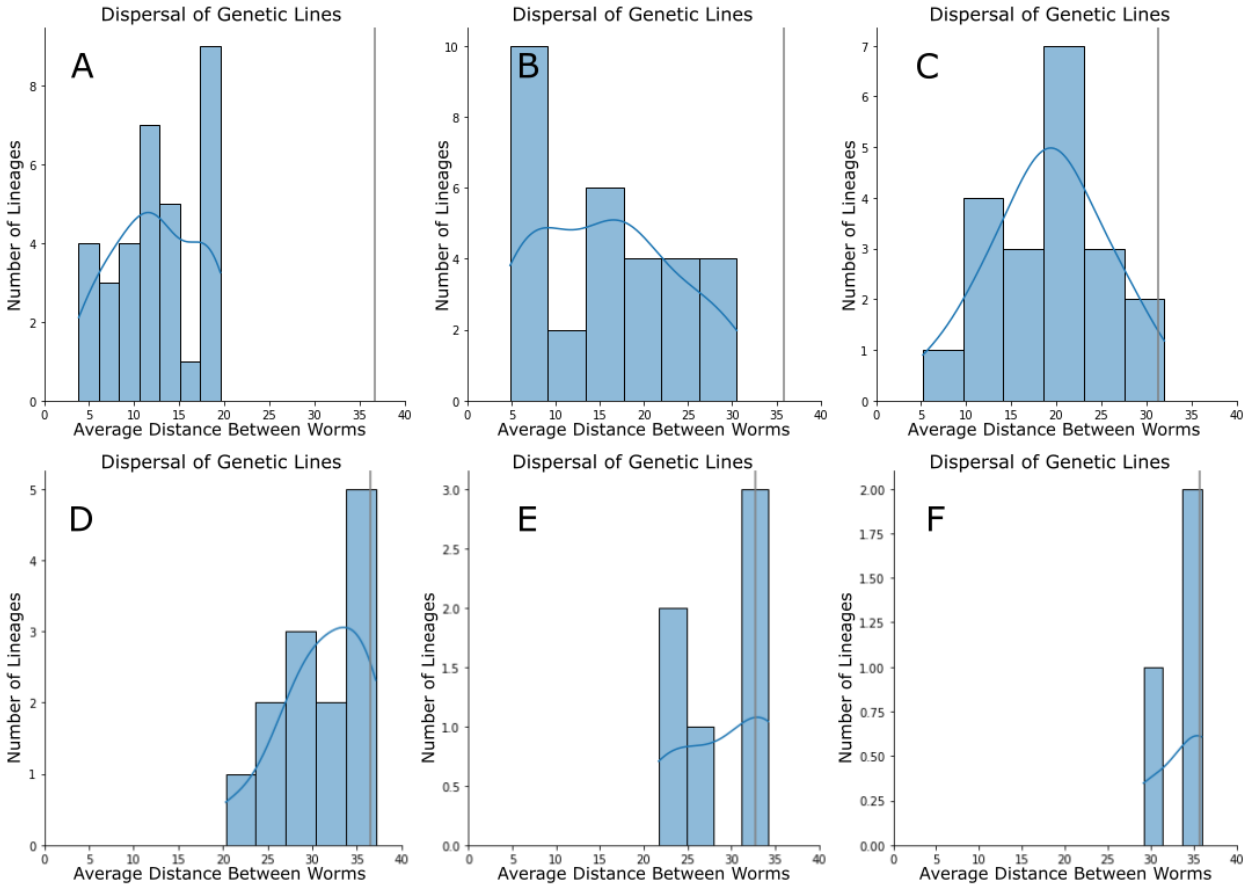

Figure 12: Dispersal of genetic lineages. A histogram of average distances between worms describes the dispersal of each lineage throughout time. The x axis represents the calculated units of distance across the torus repeated and then averaged for all worms within each lineage and the y axis represents the number of lineages with that particular average distance. The lineages included are limited to just those that make up more than 1% of the population. The gray vertical line on the right end of each graph represents the average distance between worms taken from a random sample of the entire population. Figure 12A is time point 500, B is 1000, C is 2000, D is 10000, E is 20000, and F is 30000. Throughout time, the total number of lineages decreases and each of them tends to be more and more dispersed in comparison to the random sample. This dispersal calculation can be used as a proxy to measure the diversity between patches. The more dispersed the lineages are, the less diversity there will be.

3.2.2 Mutation

Mutants can be determined by comparing the current allele pool to the original allele pool and searching for differences. Since alleles are real numbers, it is very unlikely for a mutation to take on the same value as one of the original alleles, and so this method is fairly reliable. The fraction of mutant genes can be plotted over time (Figure 13). This fraction tends to steadily increase throughout a simulation, but mutations can both pop in and out of the population. Thus, this trend does not always hold. Sometimes lineages with mutations become dominant and then their frequency will increase quickly over time. Other times the mutant worms do not survive. The result by the end of any particular simulation is seemingly random, but the gene related to travel direction often has a higher fraction of mutations surviving than the gene related to dauer. This suggests that there may be strong selection for specific values of the travel gene and lineages that do not initially possess a desirable allele may favor the mutant worms who have that allele. However, the general tendency for mutations to persist remains unclear.

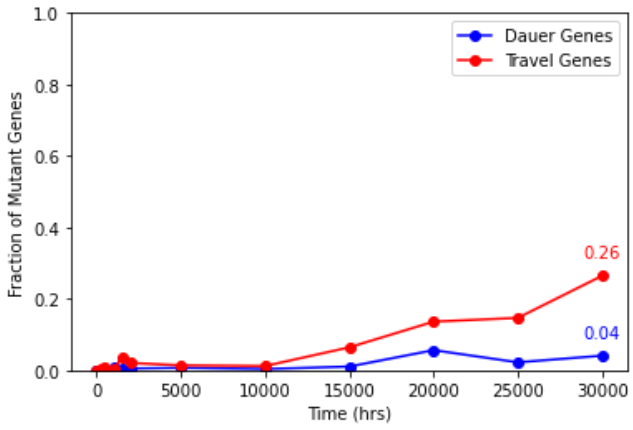

Figure 13: Fraction of mutants over time. The fraction of mutations (i.e. the number of alleles that are not present in the original population divided by the total number of alleles) can be plotted over time for the genes related to both travel direction (red line) and dauer (blue line). These plots only take into account the genes of worms who are alive at each of the time points. The final value of this fraction is printed above the last data point for each gene.

3.2.3 Dauer Tendency

Throughout each simulation, the number of worms in dauer can be measured as well as the evolution of their tendency to go into dauer (i.e. their changing genetics) depending on the circumstances. In general, each lineage of worms will have at least some members in dauer. This is ultimately how they persist when resources in their area of the grid become depleted. However, due to the high death rate of these worms, the percentages will usually remain relatively small. The only exception is when environmental conditions are so harsh that just about every worm on the grid is in dauer.

The fraction of worms that chose to go into dauer can be measured not only by which lineage they belong to but also by their genetics. A higher value associated with this dauer gene will tend to reduce the number of worms that go into dauer in a logistically curved pattern (Figure 14A). Studying this in combination with the amount of time a worm spent in L2d can give a more complete picture. As the value of the gene increases and the number of hours spent in L2d decreases, the fraction of worms that go into dauer tends to get smaller (Figure 14B). The converse is also true. Worms who spend more time in L2d and have lower dauer genetic values will be more likely to go into dauer. However, this only occurs up until a specific point. Once these combined factors reach the L2d cutoff limit, all worms will go into dauer, which appears as a diagonal red line in Figure 14B. These results are evidence that the simulations generally perform as expected, according to the dauer probability calculation (Equation 7).

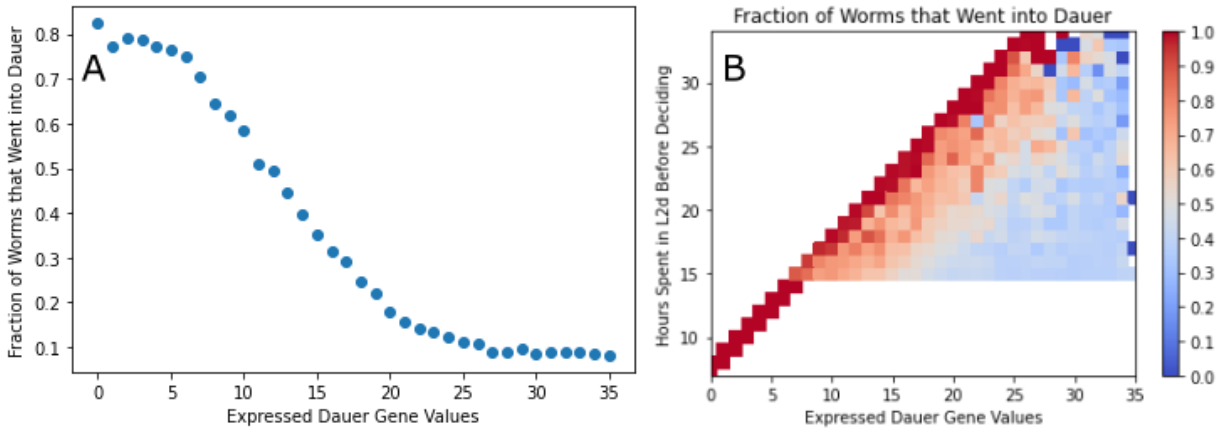

Figure 14: Empirical fraction of worms in dauer. In Figure 14A, to calculate the fraction of worms that went into dauer, the total number of worms that went into dauer throughout a simulation can be collected and divided by the total number that went into either L3 or dauer combined. These values, organized into categories by expressed dauer genotype (i.e. the average of their two alleles), are what make this plot. In Figure 14B, the fraction of worms that went into dauer in this figure is calculated similar to that of A, but only includes worms that completed the L2d stage (and not worms that molted from L2 to L3). Worms are sorted into categories based on both genetics and time spent in L2d. Colors ranging from blue to gray to red indicate the fraction of worms (on a scale of 0 to 0.5 to 1) that chose dauer. White squares indicate there were no worms with that specific combination of traits/circumstances.

In addition to calculating the fraction of worms in dauer, there is also a change in the distribution of alleles present over time (Figure 15). At the beginning of each simulation, the alleles present in the population are relatively uniformly distributed across the possible values (Figure 15A). Over time, these bounce around a bit as circumstances allow. Some lineages tend to be faster to reproduce, while other lineages are faster to die. This produces a bit of turbulence in the allele pool, and consequently in the expressed dauer gene value as well. Often, there will be some mixture of alleles, representing worms with different dauer strategies, maintained throughout most of a simulation. At the very end, this diversity may be eliminated as a single lineage of worms, all with roughly the same genes, becomes dominant on the grid.

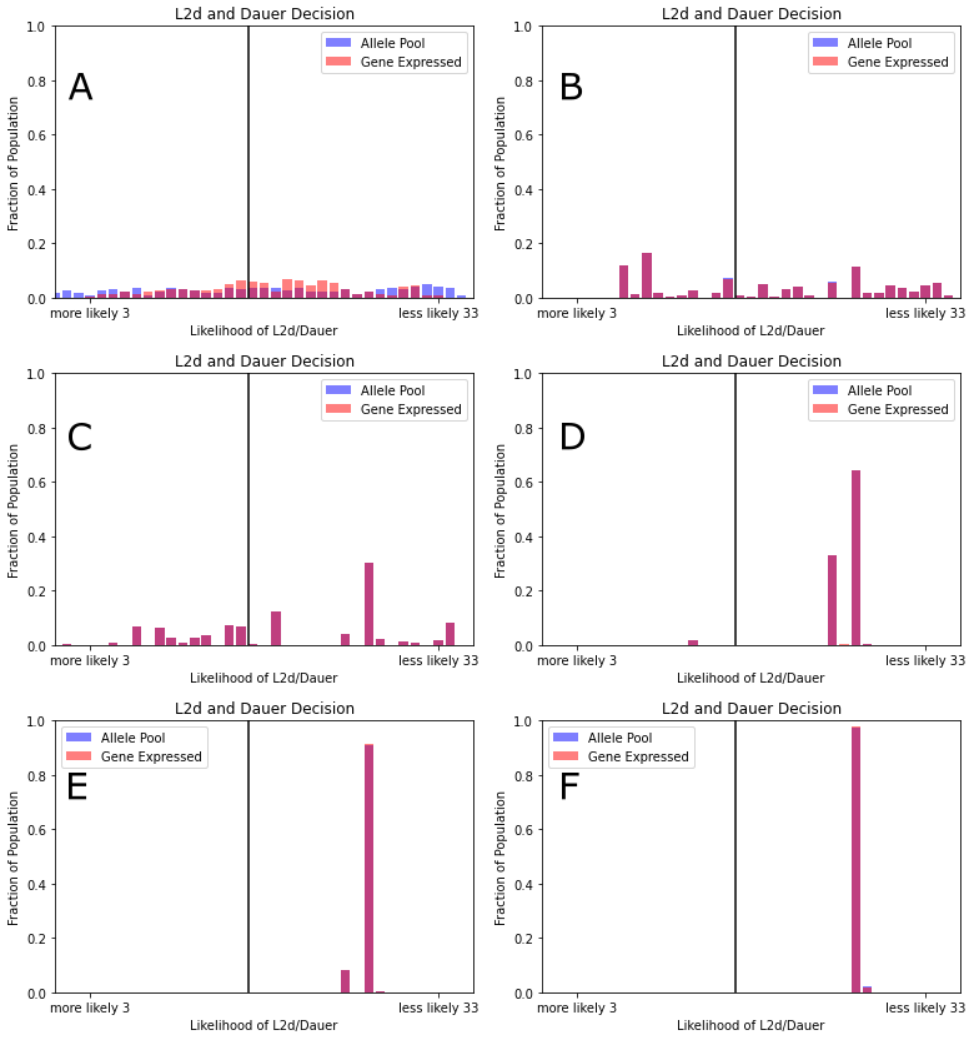

Figure 15: Distribution of dauer alleles over time. A distribution of both the allele values (red bars) and expressed values (blue bars) can be plotted over the course of time throughout a simulation. These overlaid colors are translucent, so a purple bar suggests that the alleles take on the same value as the expressed gene, meaning those individuals are homozygous. The x axis is arranged by the value of each allele, i.e. the effect it has on the individual in terms of dauer tendency. This value can be thought of as a proxy for the quantity of protein produced from transcribing this gene. Larger values (shown on the right end as “less likely 33”) will have the effect of smaller likelihood of going into dauer and smaller values (shown on the left end as “more likely 3”) will produce more worms in dauer (Equation 7). This graph is set up to match that of Figure 5, where the vertical bar is the average amount of time spent in L2d. Anything to the left of this bar is generally more likely to go into dauer and anything to the right of this bar is less likely. Figure 15A represents time point 1, B is 1000, C is 2000, D is 10000, E is 20000, and F is 30000. Over time, the alleles tend to settle on a particular value.

A simpler way to look at this data is to plot the average of all the dauer allele values over time for a clearer way to track the overall changes in the population. Depending on which parameters are input into the program, a simulation may have one of four common results shown

in Figure 16. In Figure 16A, the average line increases and then levels off, signifying that worms become much less likely to go into dauer throughout the simulation. This result shows a strong selection against lineages who have some tendency towards dauer. This may be the case in specific circumstances including both a high rate of food availability and a high probability of dying in dauer. Additionally, the upward shift may happen even faster when these conditions are more extreme. In Figures 16B and C, the average line will increase at first and then drop down, signifying that worms early in the simulation tend not to go into dauer and this tendency changes by the end. This result may occur when there is less food availability or when the probability of dying in dauer is lower. Figure 16B shows the most typical result - some value near the middle. However, when these parameters are set much lower, the shift in genetics may happen faster or be more extreme, such as in C. This graph shows a strong selection for worms in dauer. The fourth and most common result, shown in Figure 16D, can be the same as any of the above, but with some added fluctuations throughout the simulation. This variation in preferred dauer tendencies often occurs when environmental conditions are changing, whether localized or across the grid, and/or when the most common genetic lineage periodically changes. Even when overall conditions are constant, there may still be some variation throughout the experiment. To summarize, the change in dauer tendency over time can sometimes be closely tied to environmental conditions, but is not always predictable. One apparent trend is the initial decrease in dauer tendency at the beginning of every simulation. These worms are clearly the fastest to produce new progeny, but this strategy will not necessarily last.

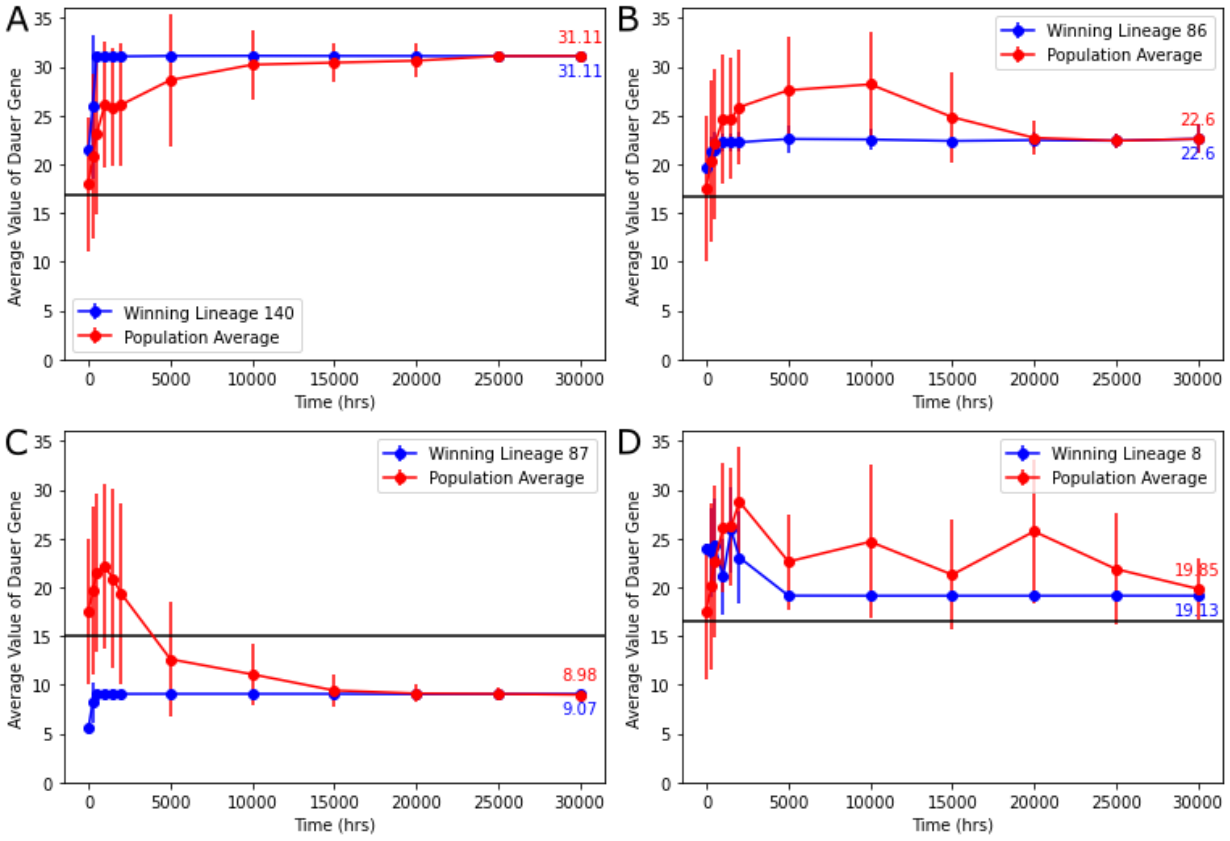

Figure 16: Average dauer tendency over time. Here are four different cases showing the change in average dauer allele values over time. Higher values represent worms that are less likely to go into dauer and lower values represent worms that are more likely to go into dauer. The horizontal line is the average length of time spent in L2d for reference as to which values are considered high or low (see Figures 5 and 15 for more explanation). Figure 16A shows a strong decrease in dauer tendency. B shows a gradual decrease and then subsequent increase in dauer tendency. C shows an initial decrease and then a strong increase in dauer tendency. D shows a fluctuating result. Simulations A-C were run under constant conditions, while simulation D included some seasonality as far as food availability. Each plot includes a line for the average dauer value across the entire population (red line) with the standard deviation plotted as error bars on each data point and a line for the average dauer value of the dominant "winning" lineage remaining at the end of the simulation (blue line) with the standard deviation plotted as error bars. The name of the original worm that the winning lineage is descended from is listed in the key for each figure. Additionally, the average values at the last time point are printed and colored to indicate the line with which they are associated.

3.2.4 Travel Direction Tendency

Throughout each simulation, there is a change in the distribution of alleles related to the travel direction the worm prefers (Figure 17). Based on the value of this gene, worms decide the importance of traveling towards food or away from neighbors. At the beginning of each

simulation, the alleles present in the population are normally distributed and both factors evenly influence the decision (Figure 17A). Over time, these alleles tend to shift towards the right end of the plot, which signifies that worms favor traveling towards food over avoiding their neighbors.

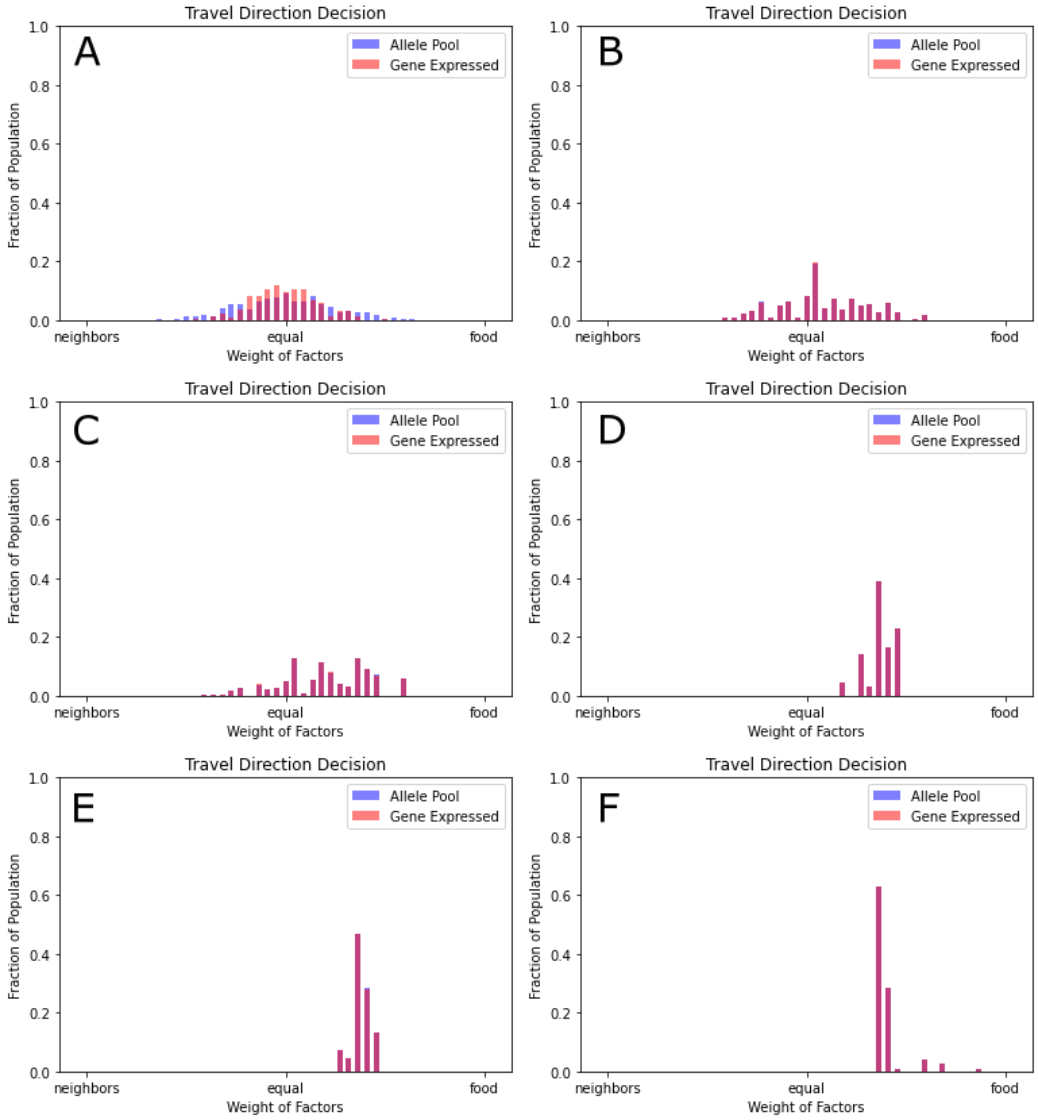

Figure 17: Distribution of travel direction alleles over time. A distribution of both the allele values (red bars) and expressed values (blue bars) can be plotted over the course of time throughout a simulation. These overlaid colors are translucent, so a purple bar suggests that the alleles take on the same value as the expressed gene, meaning those individuals are homozygous. The x axis is arranged by the value of each allele, i.e. the effect it has on the individual in terms of travel direction. Values shown on the right end mean that the worm prefers to travel towards food and values shown on the left end mean that the worm prefers to travel away from neighbors (or towards neighbors for adult male worms). The limits of this axis extend from 0.25 to 0.75 and 0.5 is labeled as “equal,” meaning equally considering the two factors. Figure 17A represents time point 1, B is 1000, C is 2000, D is 10000, E is 20000, and F is 30000. Over time, the alleles tend to settle on the right side of the graph.

Like before, a simpler way to look at this data is to plot the average of all the travel direction allele values over time for a clearer way of tracking the overall changes in the population (Figure 18). There seems to be an optimal range of values for this gene which are roughly between 0.50 and 0.65. Interestingly, these values always shift towards the preference for finding food over avoiding neighbors, sometimes more strongly depending on the circumstances in the simulation. This will produce a clumpy feeding behavior, such as naturally observed in some strains of *C. elegans* (Dorado-Morales et al., 2014; Serena Ding et al., 2019), and it clearly shows the evolution of this behavior.

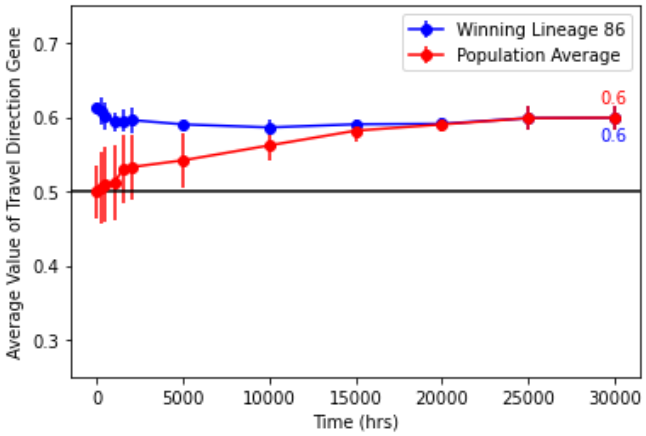

Figure 18: Average travel direction gene over time. Here is an example of the change in average travel direction allele values over time. Higher values represent worms that are more likely to travel towards food and lower values represent worms that are more likely to travel away from their neighbors (or towards their neighbors for adult male worms). The horizontal line marks an equal interest in both of these factors. In almost every simulation, worms will tend to be more interested in reaching their food. This plot includes a line for the average travel direction value across the entire population (red line) with the standard deviation plotted as error bars on each data point and a line for the average travel direction value of the dominant "winning" lineage remaining at the end of the simulation (blue line) with the standard deviation plotted as error bars. The name of the original worm that the winning lineage is descended from is listed in the key. Additionally, the average values at the last time point are printed and colored to indicate the line with which they are associated.

These travel direction gene values can be collected from all simulations conducted, averaged, and plotted in a distribution (Figure 19). Out of 128 simulations, each can boil down to two single statistics - the average value of all the travel direction alleles present in the population at the beginning (Figure 19A) and the same measurement at the end (Figure 19B). These two

separate distributions can be compared against each other for significant changes. The beginning average value across all simulations was 0.5007 with a standard deviation of 0.0021 and the end average was 0.5747 with a standard deviation of 0.0378. To compare these distributions, a paired t-test was performed which resulted in a t-statistic of -22.36 and a p value of 7.45×10^{-46}. This indicates that these two distributions are highly significantly different. It should also be noted that the same seed was used repeatedly in different experiments. This allows for identical starting conditions, but different resulting conditions and is the reason for the abnormally shaped distribution in Figure 19A. In summary, there is some advantage to those worms who tend to travel towards the food and, thus, this behavior will somewhat evolve.

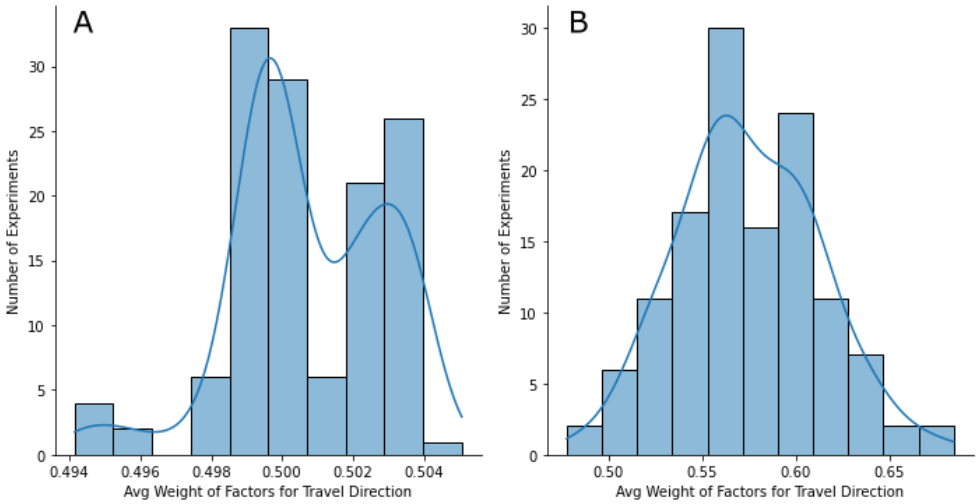

Figure 19: Distribution of average travel direction values across simulations. These histograms each include 128 data values which are the averages of the travel direction gene value from different simulations conducted in varying environments. Values closer to 1 suggest that worms prefer to advance towards the food, while values closer to 0 suggest that worms prefer to avoid their neighbors when choosing their direction of travel. A value of 0.5 indicates that worms are equally as likely to travel towards their food as they are to travel away from their neighbors. These distributions show the number of experiments conducted with that particular average value of the gene and there is a density plot drawn on top. Figure 19A is time point 1 and B is 30000.

3.2.5 Population Size

Throughout every simulation, there is some sort of fluctuation occurring in the overall population size. Whether there is a consistent rise and fall or more sporadic changes, the number

of worms alive is never constant. This could occur for a multitude of reasons including changes in the dominant/most common lineage of worms, localized poor environmental quality in a particular region, and/or globally poor productivity on a large scale. It seems fluctuations tend to happen even when environmental conditions remain constant. If conditions are poor, worms will have difficulty surviving and the population size will be very small. Ones that are better adapted to the environment may succeed and start to repopulate. On the other hand, if conditions are really good, the population size will explode and then overcrowding becomes problematic for the worms. Either way, this can create some oscillations in the overall size of the population, as predicted by the boom/bust model (Félix & Duveau, 2012).

These fluctuations become even more dramatic and less unpredictable when they are tied to the productivity of the environment (Figure 20). As the rate of food patch replacement on the grid becomes slower, the population size will crash. In extreme cases, it becomes almost zero and worms may exist exclusively in dauer. Then, when food patches become plentiful again, the size of the population will explode. The more food is available, the less worms will tend to rely on dauer, and the speed of reproduction can increase. When environmental conditions are very extreme, the population size will tend to closely track this (Figure 20B). Otherwise, the fluctuations tend to rely more on localized events, sometimes producing a more inconsistent boom/bust cycle (Figure 20A).

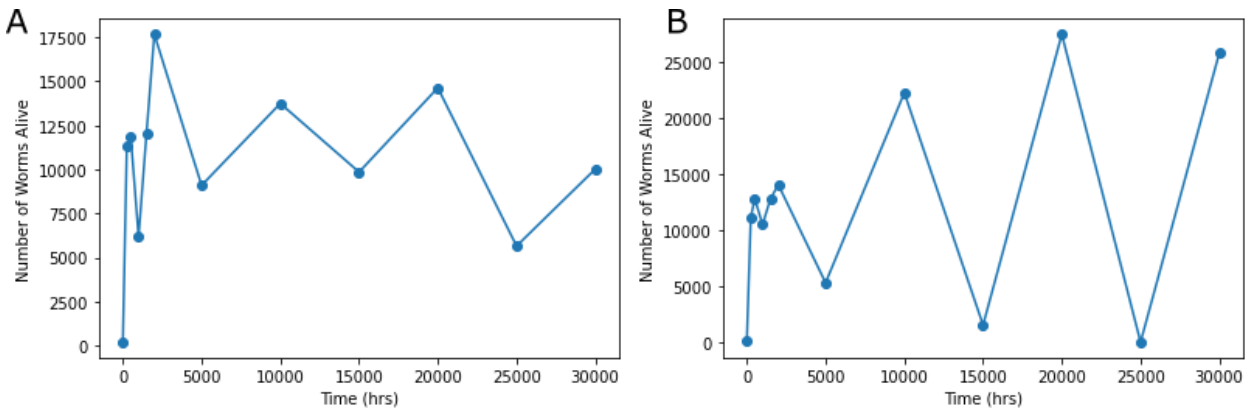

Figure 20: Population size over time. These plots show the number of living worms at a handful of points throughout a simulation. The experiment in Figure 20A features no cycling of food patch replacement rates, just constant conditions. Yet here, there is some relatively regular fluctuation in the overall population size, potentially produced by local areas of *E. coli* drying up. The experiment in Figure 20B, in contrast, was performed with heavy seasonality and this produces some extreme cycling of the population size. At times, the population crashes nearly to zero. These dips correspond to periods of almost zero chance of patch replacement as shown in Figure 3. Interestingly, it seems adding seasonality has just enhanced the natural population cycle (in some cases), but appears to occur on a longer/larger scale.

3.2.6 Lineage and Allele Frequencies

Each worm in the simulation has a control genome, or grouping of genes, which does not mutate and serves the purpose of tracing worms back to their original genetic ancestor from the beginning of the experiment. Each of the original worms can be grouped together to form a genetic lineage with their progeny. All of these simulations have 200 original worms, so there is a maximum of 200 possible genetic lineages. The worms in these lineages tend to have the same traits, since they mostly breed by selfing, and can be monitored over time in a variety of ways to see how long they survive. At each time point, the most prevalent lineage on the grid can be determined and, by the end of every simulation, one lineage can be declared as the winning lineage. Since this lineage occupies a majority of the space, the dauer strategy of its members must have been suitable enough for that particular environment.

The most straightforward way to observe how lineages spread across the grid over time is to plot a heatmap of them (Figure 21). Throughout a simulation, different lineages can be

observed occupying different patches at different times. There will often be clusters of lineages, and each travels around as more or less a cohesive group. On occasion, it is possible for a lineage to almost completely "disappear," or exist fully in dauer, and then make a complete comeback if they hit an unoccupied patch. As a result, it seems like the dynamics across the grid are almost constantly changing as different lineages take over, at least until the end of a simulation or whenever the diversity dies down. At that point, the lineages will be fairly dispersed across all the patches and one may begin to dominate the grid.

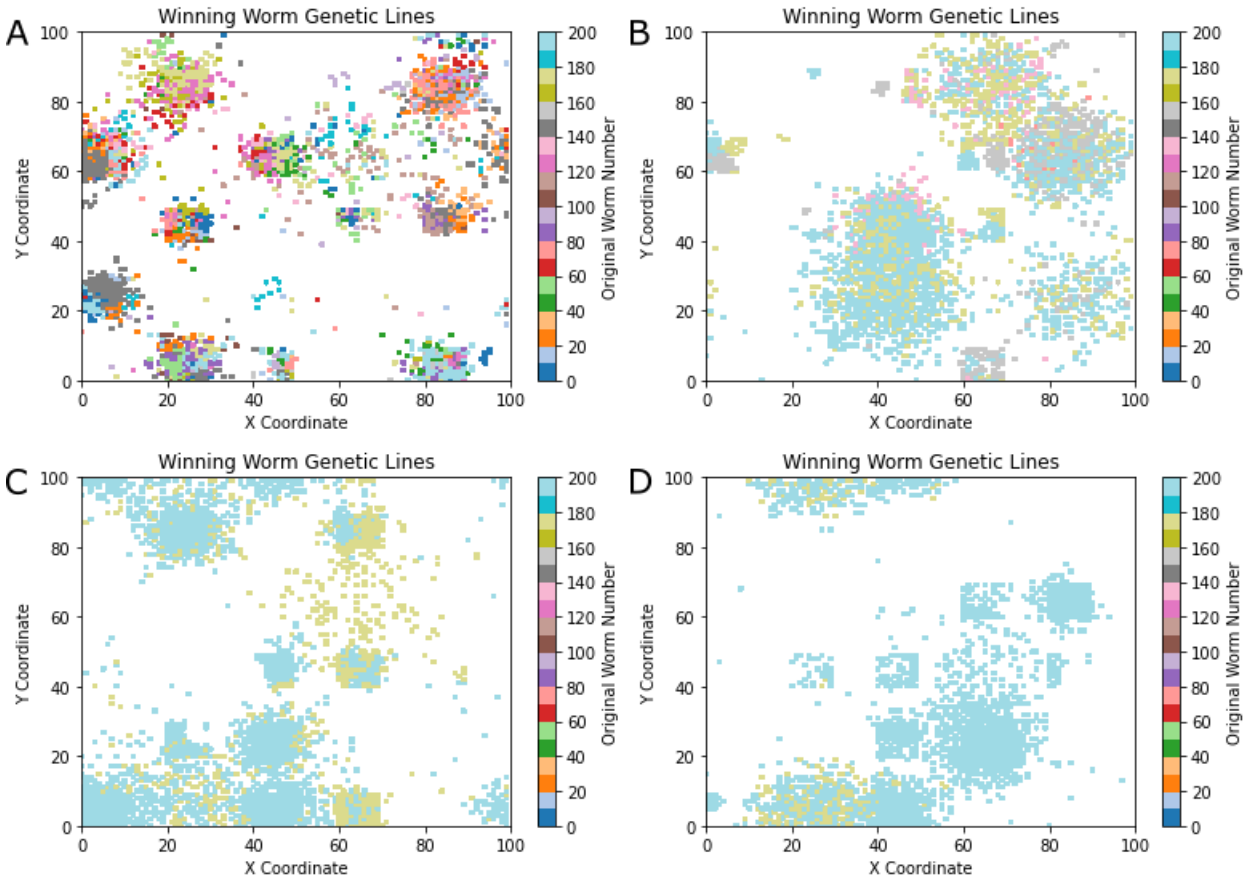

Figure 21: Heatmap of worm lineages over time. In each location of the grid, the winning lineage can be determined by finding the mode. These are then plotted on a two-dimensional surface and colored based on the name of the original worm in the lineage (shown in the legend to the right). Over time, the number of lineages decreases as they spread across the grid, until there are only one or two left. Figure 21A represents time point 500, B is 10000, C is 20000, and D is 30000.

The abundance of each lineage can be seen more clearly in Figure 22. Just like with gene values (Figures 15 and 17), lineages will also bounce around over time. Some worms may reproduce faster than others, while some may die faster than others. After the majority of this turbulence settles, it becomes clear which lineages are succeeding. A simulation may go through multiple phases of different winning lineages, but usually will settle on one or two by the end.

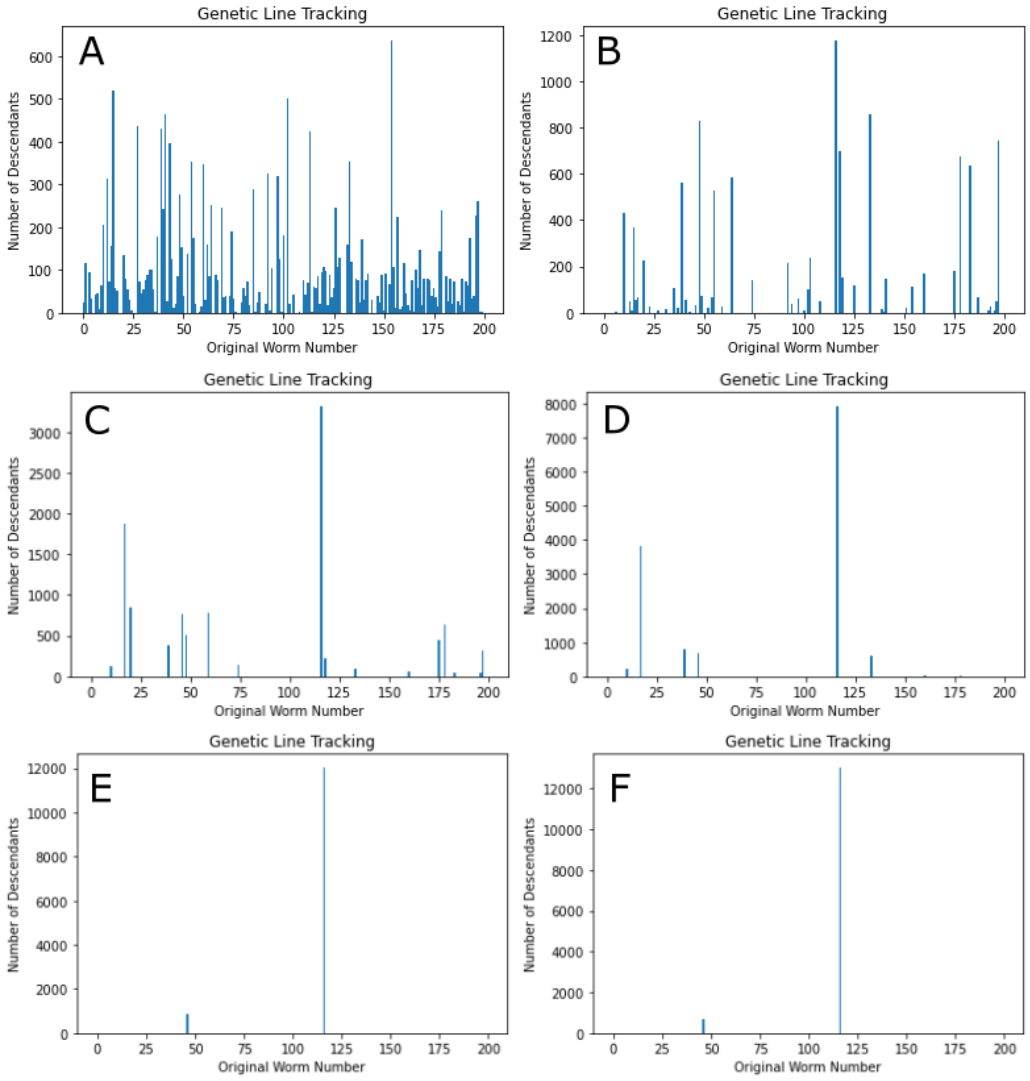

Figure 22: Lineage abundance over time. At each point in time, all worms can be sorted into lineages according to their original ancestor. Lineages are measured by the name of the original worm (x axis) and the number of living descendants at that point in time (y axis). Change in abundance over time can be seen here in this simulation, and by the end, there is a clear winning lineage. Figure 22A represents time point 500, B is 2000, C is 5000, D is 10000, E is 20000, and F is 30000.

It can be a challenge to determine exactly why each winning lineage has taken over at a specific moment in time. Changes in the dominant lineage often occur simultaneously with changes in average dauer tendency (Table 1). In this case, it seems that the new winning lineage has a more dominant dauer strategy, or at least possesses more optimal gene values than the previous lineage. Depending on the circumstances, this change could happen immediately (when selection against dauer is strong) or it could take some time. However, sometimes changes in the winning lineage do not relate to the average dauer tendency. In this case, there must be some other factor that created the change. Most likely, it was due to chance. If a worm happened to be in the right place at the right time, this could create a rippling effect. Many events in these simulations seem quite circumstantial and it is possible for a lineage to survive just by luck.

Table 1: Winning lineages and their dauer gene values. The column labeled "Time (hrs)" lists all the time points at which measurements were collected throughout this particular simulation. The "Winning Line" is the name of the original worm ancestor who's living descendants occupy a majority of the grid at that time point and the "Frac of Pop" is the fraction of living worms that belong to the winning lineage. Lineage 0 is always listed as the winning lineage at time point 1, because all lineages have just one individual (the original member) and the first one will automatically be determined as the mode. The columns "Avg Dauer Gene" and "Std Dev" measure the average dauer gene value and standard deviation of the worms in the winning line only.

Time (hrs)	Winning Line	Avg Dauer Gene	Std Dev	Frac of Pop
1	0	26.620	0.000	0.005
250	81	32.011	0.060	0.014
500	107	25.790	5.643	0.044
1000	172	32.273	5.011	0.100
1500	182	20.396	0.019	0.188
2000	55	28.130	0.544	0.130
5000	182	20.423	0.453	0.305
10000	3	17.756	0.256	0.599
15000	3	17.625	0.194	0.718
20000	3	17.723	0.109	0.746
25000	3	17.597	0.202	1.000
30000	3	17.690	0.360	1.000

Changes in lineage abundance seem to track pretty closely with changes in the frequency of the different dauer alleles (Figure 23). This would suggest that each of the lineages tends to have homozygous individuals as well as a unique dauer allele carried by the entire line of

descendants. Since all worms in the beginning of a simulation are heterozygous, this would mean that one of the alleles has been quickly selected. Additionally, there are times when different lineages have very similar dauer gene values. In this case, there may be some competition between these lineages for space and resources. However, the resulting average dauer strategy would be the same no matter which lineage, if any, succeeds in occupying the grid.

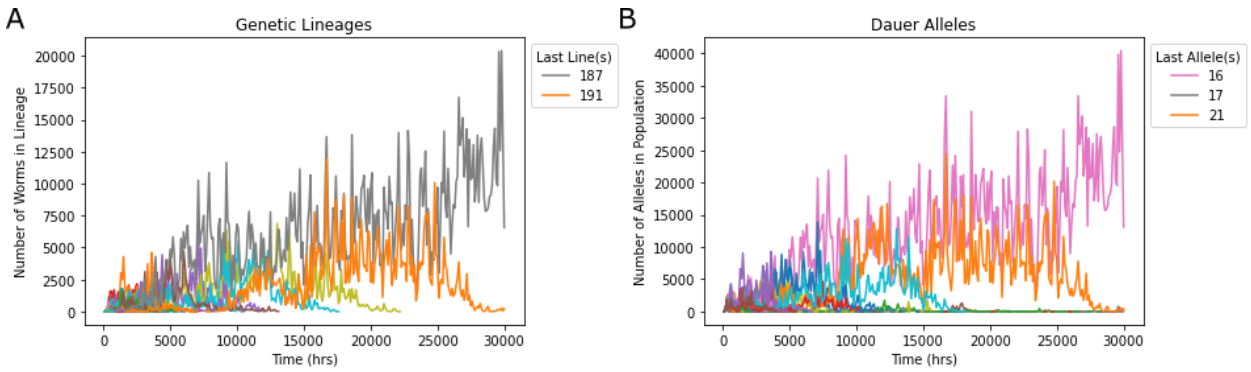

Figure 23: Abundance of genetic lineages and dauer alleles over time. These high resolution plots show detailed changes in the abundance of each genetic lineage (Figure 23A) and each dauer allele value (Figure 23B) over time. The genetic lineages include all descendants from the original population, labeled by the name of the ancestor. The dauer allele values are rounded to non-negative whole numbers for sufficient grouping. Each lineage and dauer allele is colored separately, but only the remaining ones at the final time point are shown in the key. Notice that these plots are almost identical and can be matched to each other. In this case, the descendants of worm 187 possess dauer allele 16 and the descendants of worm 191 possess dauer allele 21. These plots will not have the same shape when there are multiple lineages with similar genetics or when their genetics change over time. Additionally, the low frequency dauer alleles at the bottom of Figure 23B show that there is some slight variation from the more populous alleles. For example, at the last time point, the dauer allele value 17 is likely a mutation from allele 16.

Each simulation operates on a slightly different timeline. Sometimes the winning lineage will quickly occupy the grid, but other times it takes time for the winning lineage to settle. This lineage can take over as early as around 5,000 time steps, but also as late as 25,000 or 30,000 time steps. Simulations produce a variety of results (Figure 24) and this may also be aided by environmental conditions. In Figures 24A and B, there are constant environmental conditions and changes seem to occur at no particular time point. Whereas in Figures 24C and D, the rate of food patch repopulation is fluctuating and, during crashes in population size, bottlenecks seem to occur, especially during the first crash. This leaves the winning lineage to take over following

that period of decline. In summary, the changing of lineages seems to be due to a combination of being more genetically fit for the situation and surviving environmental obstacles by chance.

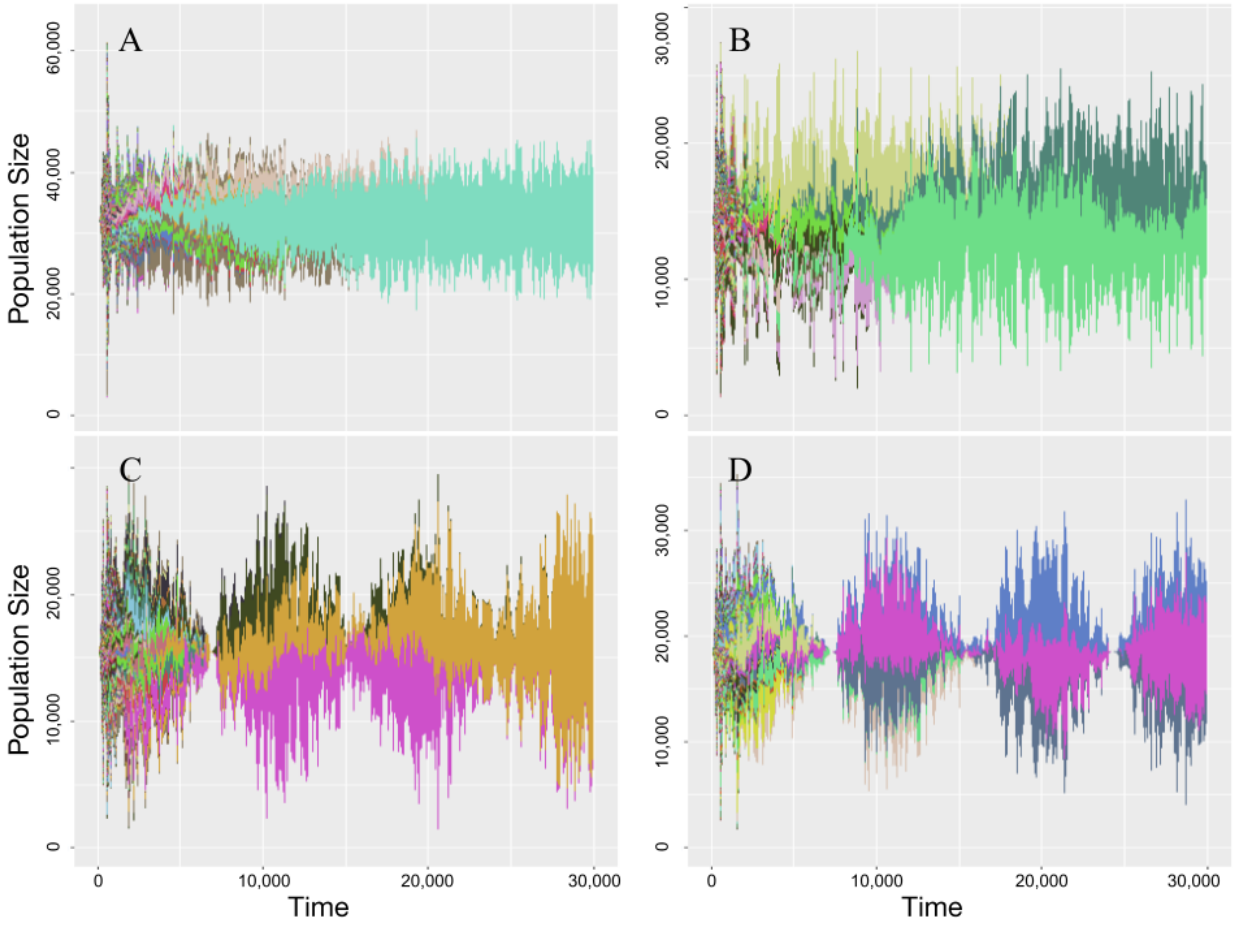

Figure 24: Muller plots of genetic lineages. These high resolution R plots are made from the ggmuller package. Using lineage abundance data, a muller plot can be produced that depicts decreasing population diversity. Over time, each lineage is plotted as a different color and the vertical space it takes up represents the number of descendants from that genetic line. Figures 24A and B are from simulations run with constant environmental conditions, while Figures 24C and D have fluctuations in food availability. It is clear at what points the food is scarce based on the large dips in population size.

3.3 EXPERIMENTAL SIMULATIONS

3.3.1 Preliminary Experiments

Some of the first experiments were meant to determine how the starting conditions of the simulation affect the results. Whether the entire initial population began in the same food patch or spread out across the grid did not seem to change the outcome in terms of the average

resulting dauer gene value or the number of lineages remaining. Additionally, the number of patches in the beginning of the simulation had no effect on the outcome whatsoever. Variations of this experiment include starting with 5, 10, 15, 20, and 25 food patches. It should also be noted that both of these preliminary experiments were run with no selection against dauer, which does have a large effect on the resulting dauer gene value. For more convincing and comparable results, these experiments should be performed again with a strong dauer death rate, like in following experiments. Other preliminary experiments that did have strong selection against dauer were run. From these early tests, it seems like a range of dauer gene values may be suitable for each condition. For example, there are some simulations with all the same parameters and very different outcomes in terms of the dauer gene value. There are also some simulations that result in a mixture of dauer gene values. These preliminary tests were useful in determining that the program works and what possible outcomes may look like.

3.3.2 Strength of Selection

Among the experiments performed in the model, one tested how altering the probability that a worm dies in dauer affects the evolution of the population. Over the course of each simulation, the average values of both dauer (Figure 25A) and travel direction genes changed and were affected by the selection imposed on dauer worms. Stronger selection against dauer led worms to evolve a lower likelihood of entering that stage - reflected in their dauer gene value (Figure 25B). The strength of selection strongly predicted the average likelihood of entering dauer with a positive slope of 163.11 (95% confidence interval of 89.439 to 223.147). Since the confidence interval does not include zero, this would suggest that the strength of selection is driving the adaptation of this gene.

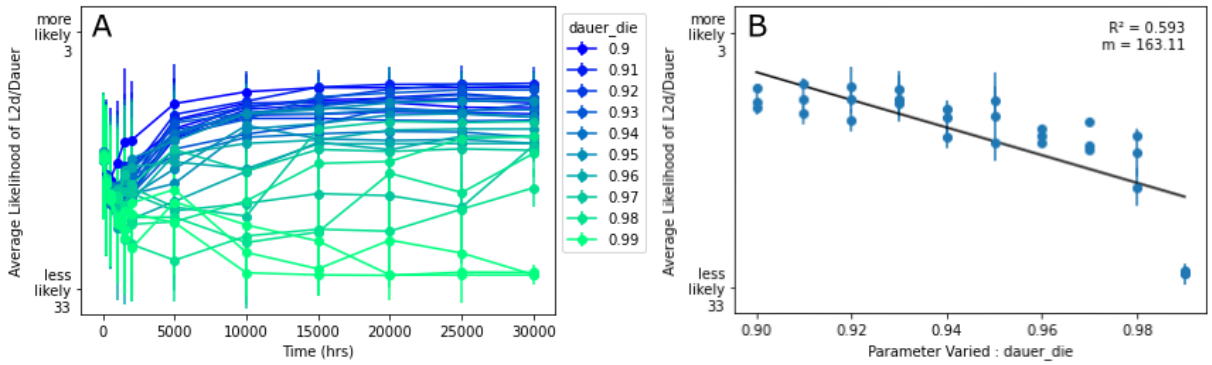

Figure 25: Average dauer value by strength of selection. These plots combine the results of every simulation run as part of this experiment. The y axis is the average value of all dauer alleles present in the population at a specific time point. Lower dauer values (plotted at the top of the y axis) indicate higher likelihood that a worm will choose dauer and higher dauer values (plotted at the bottom of the y axis) indicate lower likelihood that a worm will choose dauer. In Figure 25A, the average value of all dauer alleles present in the population is plotted as a data point at each time point of data collection. These lines (a separate one for each simulation) are colored based on the value of the input parameter: "dauer_die." The value of this parameter indicates the fraction of worms that will die in dauer throughout the simulation. In Figure 25B, only the average value of all dauer alleles present in the population at the last time point is plotted. Each data point in this plot represents an individual simulation from this experiment and the line of best fit is drawn through them. The x axis is the parameter varied: "dauer_die." Additionally, R^2 (the fit of the line) and m (the slope) are listed in the top right corner.

Similarly, a stronger selection against dauer led worms to evolve less attraction to food (Figure 26). This parameter was found to be negatively correlated with this variable and there was a slope of -0.482 (95% confidence interval of -0.810 to -0.104). Since the confidence interval does not include zero, this would suggest that the strength of selection is driving the adaptation of this gene. However, it is unclear exactly why this might be the case. One possible explanation is that since worms in dauer die more often, other worms will want to avoid going into dauer, and this can be accomplished in part by avoiding their neighbors. Other explanations may also be plausible, but all will require further exploration.

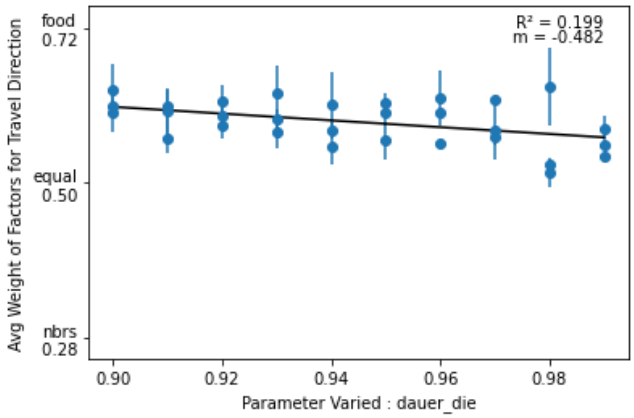

Figure 26: Average resulting travel direction value by strength of selection. This plot combines the results of every simulation run as part of this experiment. The x axis is the parameter varied: "dauer_die." The value of this parameter indicates the fraction of worms that will die in dauer throughout the simulation. The y axis is the average value of all travel direction alleles present in the population at the last time point. Lower travel direction values indicate that worms favor traveling away from their neighbors (or towards neighbors in the case of adult males) and higher travel direction values indicate that worms favor traveling towards their food. These are labeled as "nbrs" for neighbors, "equal" meaning they weigh their decision 50/50 on each factor, and "food." Each data point in this plot represents an individual simulation from this experiment and the line of best fit is drawn through them. Additionally, R^2 (the fit of the line) and m (the slope) are listed in the top right corner.

3.3.3 Environmental Productivity

Among the experiments performed in the model, one tested how altering the rate of food patch repopulation affects the evolution of the population. Over the course of each simulation, the average values of both dauer (Figure 27A) and travel direction genes changed and were affected by the productivity of the environment. The faster the rate of food production, the more the worms evolved a lower likelihood of entering dauer - reflected in their dauer gene value (Figure 27B). The rate of food patch repopulation predicted the average likelihood of entering dauer with a positive slope of 55.80 (95% confidence interval of 22.678 to 86.934). Since the confidence interval does not include zero, this would suggest that productivity of the environment is driving the adaptation of this gene.

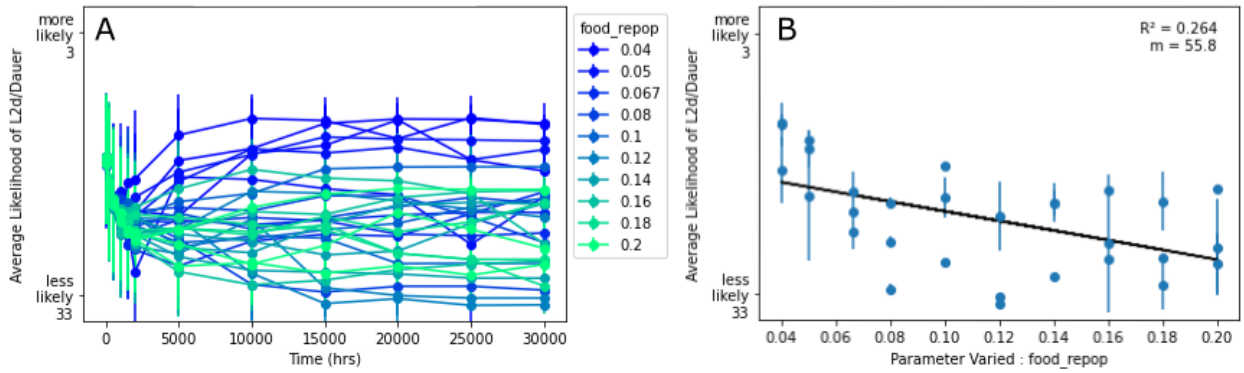

Figure 27: Average dauer value by food patch repopulation. These plots combine the results of every simulation run as part of this experiment. The y axis is the average value of all dauer alleles present in the population at a specific time point. Lower dauer values (plotted at the top of the y axis) indicate higher likelihood that a worm will choose dauer and higher dauer values (plotted at the bottom of the y axis) indicate lower likelihood that a worm will choose dauer. In Figure 27A, the average value of all dauer alleles present in the population is plotted as a data point at each time point of data collection. These lines (a separate one for each simulation) are colored based on the value of the input parameter: "food_repop." During each iteration, the value of this parameter indicates the likelihood that a new patch will appear on the grid. In Figure 27B, only the average value of all dauer alleles present in the population at the last time point is plotted. Each data point in this plot represents an individual simulation from this experiment and the line of best fit is drawn through them. The x axis is the parameter varied: "food_repop." Additionally, R^2 (the fit of the line) and m (the slope) are listed in the top right corner.

Similarly, a more productive environment led worms to evolve more attraction to food (Figure 28). This parameter was found to be positively correlated with this variable and there was a slope of 0.660 (95% confidence interval of 0.443 to 0.863). Since the confidence interval does not include zero, this would suggest that the productivity of the environment is driving the adaptation of this gene. However, it is unclear exactly why this might be the case. One possible explanation is that if there is more food available, worms will be able to smell it more often, hence more opportunity for selection to occur. Other explanations may also be plausible, but all will require further exploration.

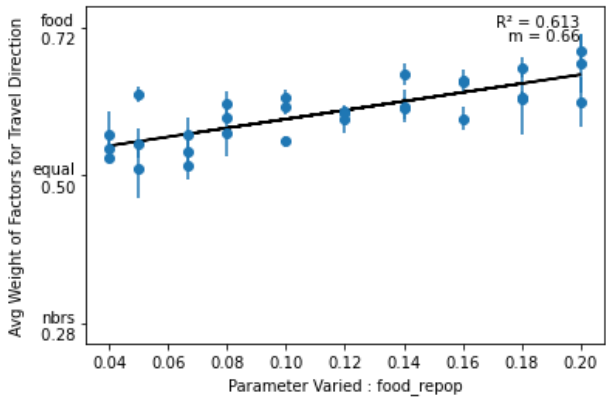

Figure 28: Average resulting travel direction value by food patch repopulation. This plot combines the results of every simulation run as part of this experiment. The x axis is the parameter varied: "food_repop." During each iteration, the value of this parameter indicates the likelihood that a new patch will appear on the grid. The y axis is the average value of all travel direction alleles present in the population at the last time point. Lower travel direction values indicate that worms favor traveling away from their neighbors (or towards neighbors in the case of adult males) and higher travel direction values indicate that worms favor traveling towards their food. These are labeled as "nbrs" for neighbors, "equal" meaning they weigh their decision 50/50 on each factor, and "food." Each data point in this plot represents an individual simulation from this experiment and the line of best fit is drawn through them. Additionally, R^2 (the fit of the line) and m (the slope) are listed in the top right corner.

3.3.4 Genotype-Phenotype Map

Among the experiments performed in the model, one tested how altering the genotype to phenotype mapping affects the evolution of the worms. This mapping value, as seen in Equation 7 and Figure 5, changes how the dauer gene value affects the probability of the worm to choose dauer. A higher mapping value will smooth out the curve and allow more worms of various different genotypes some probability of entering dauer or not, such as in Figure 5B. At the end of simulations, the effect of this parameter can be seen in the types of worms that chose to go into dauer, and the differences between mapping values become apparent in plots like Figure 14B. Higher mapping values allow more worms the opportunity to go into dauer. This seems to give the gene value a bit more flexibility. However, this parameter does not appear to have a strong relationship with the evolution of the average dauer gene of the population (Figure 29A). This plot has no obvious organization, but when only the average dauer value at the last time point in

each simulation is plotted, a slight positive correlation can be detected (Figure 29B). The slope of the best fit line in Figure 29B is 1.460 (95% confidence interval of 0.265 to 2.637) and since this interval does not contain zero, it can be assumed that the genotype to phenotype mapping is driving some adaptation of the gene. Since higher values of this parameter will allow worms with higher dauer gene values the option for dauer, this makes these gene values more desirable. Worms with these gene values will primarily avoid dauer, but can still choose that stage when needed. Additionally, some further testing may be required to see if there is a possible effect on the diversity of results. It is reasonable to expect that higher mapping values may cause more of a variety of possible outcomes, but this remains unclear.

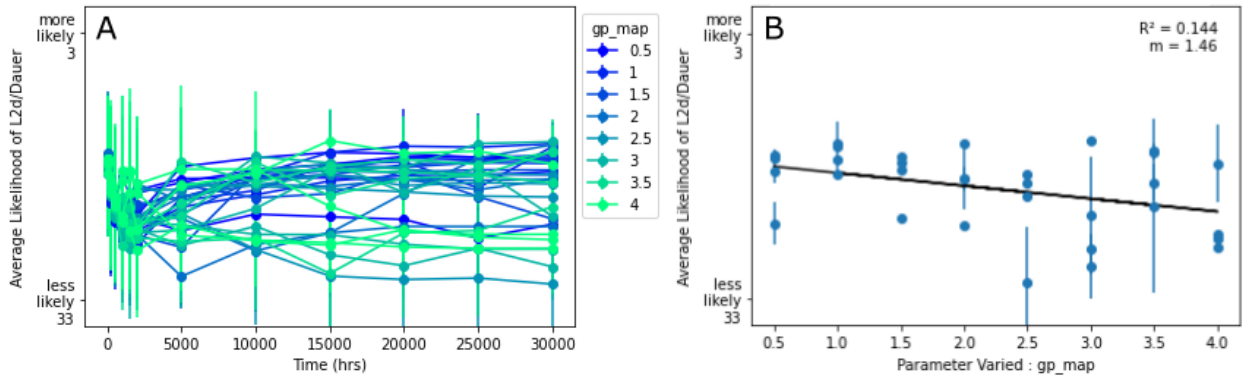

Figure 29: Average dauer value by genotype-phenotype map. These plots combine the results of every simulation run as part of this experiment. The y axis is the average value of all dauer alleles present in the population at a specific time point. Lower dauer values (plotted at the top of the y axis) indicate higher likelihood that a worm will choose dauer and higher dauer values (plotted at the bottom of the y axis) indicate lower likelihood that a worm will choose dauer. In Figure 29A, the average value of all dauer alleles present in the population is plotted as a data point at each time point of data collection. These lines (a separate one for each simulation) are colored based on the value of the input parameter: "gp_map." The value of this parameter affects the probability of all worms to choose dauer. In Figure 29B, only the average value of all dauer alleles present in the population at the last time point is plotted. Each data point in this plot represents an individual simulation from this experiment and the line of best fit is drawn through them. The x axis is the parameter varied: "gp_map." Additionally, R^2 (the fit of the line) and m (the slope) are listed in the top right corner.

On the other hand, this mapping value parameter has no effect on the average travel direction gene (Figure 30). The correlation is very low and has a slope of -0.0016 (95% confidence interval of -0.011 to 0.0089). Since this interval contains zero, it can be assumed that

there is no connection. However, this result is not surprising considering these parameters are practically unrelated.

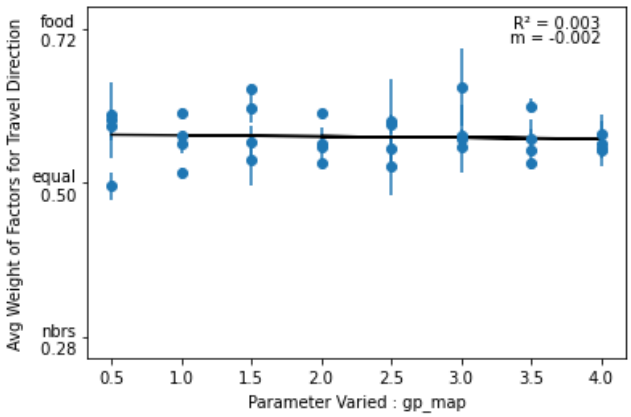

Figure 30: Average resulting travel direction value by genotype-phenotype map. This plot combines the results of every simulation run as part of this experiment. The x axis is the parameter varied: "gp_map." The value of this parameter affects the probability of all worms to choose dauer. The y axis is the average value of all travel direction alleles present in the population at the last time point. Lower travel direction values indicate that worms favor traveling away from their neighbors (or towards neighbors in the case of adult males) and higher travel direction values indicate that worms favor traveling towards their food. These are labeled as "nbrs" for neighbors, "equal" meaning they weigh their decision 50/50 on each factor, and "food." Each data point in this plot represents an individual simulation from this experiment and the line of best fit is drawn through them. Additionally, R^2 (the fit of the line) and m (the slope) are listed in the top right corner.

3.3.5 Seasonality

Among the experiments performed in the model, one tested how adding seasonality to the environmental productivity affects the evolution of the worms. Over the course of time in a simulation, the probability that a new food patch will appear can oscillate according to the sine curve in Equation 3. When the amplitude of this curve is increased until it reaches the same value as the food patch repopulation rate, the result is a curve like in Figure 3. The maximum probability reaches twice the initial rate and the minimum reaches zero. This will create a wildly fluctuating environment and may cause fluctuations in the average dauer gene values as well (Figure 16D). However, this observation is not always consistent across simulations (Figure 31A). The light green lines in Figure 31A represent the simulations with the largest oscillation

and not all of these reflect that in this plot. Conversely, some of the dark blue lines, which represent the simulations with more constant conditions, do show some fluctuating values. These changes are likely due to local environmental conditions and the switching of winning lineages rather than global environmental productivity. Any possible short term adaptations to the environment will need further study. It could be the case that many types of worms are suitable for a variety of conditions. Thus, oscillations do not make much difference, because the optimal adaptation will be to the average rate of productivity. However, this can be studied by changing the midline and/or frequency of the oscillations to see what short term (and possible long term) changes are produced.

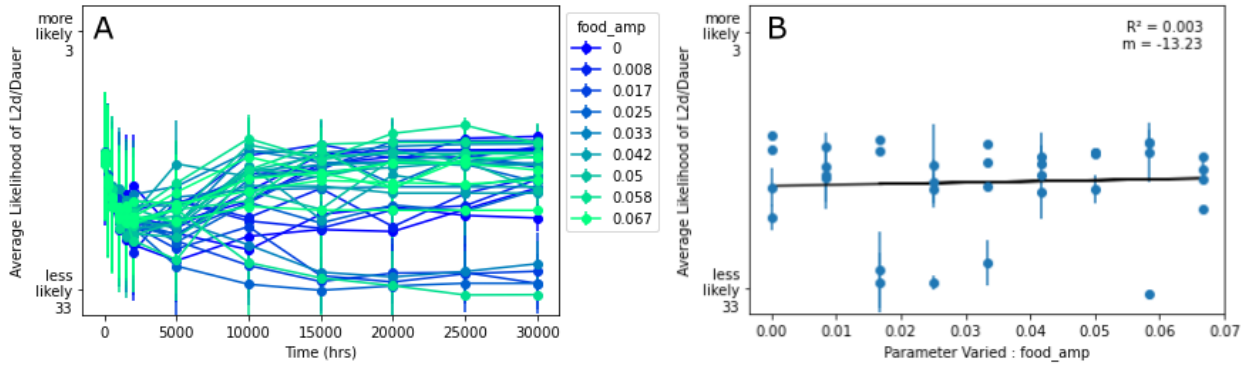

Figure 31: Average dauer value by seasonality. These plots combine the results of every simulation run as part of this experiment. The y axis is the average value of all dauer alleles present in the population at a specific time point. Lower dauer values (plotted at the top of the y axis) indicate higher likelihood that a worm will choose dauer and higher dauer values (plotted at the bottom of the y axis) indicate lower likelihood that a worm will choose dauer. In Figure 31A, the average value of all dauer alleles present in the population is plotted as a data point at each time point of data collection. These lines (a separate one for each simulation) are colored based on the value of the input parameter: "food_amp." The value of this parameter is the amplitude of the oscillations in the food patch repopulation rate. In Figure 31B, only the average value of all dauer alleles present in the population at the last time point is plotted. Each data point in this plot represents an individual simulation from this experiment and the line of best fit is drawn through them. The x axis is the parameter varied: "food_amp." Additionally, R^2 (the fit of the line) and m (the slope) are listed in the top right corner.

Interestingly, the worms do not seem to adapt to this seasonality at all in the long run. Data at the end of each simulation, shown in Figure 31B, have no correlation with the seasonality. There is a slope of -13.233 (95% confidence interval of -88.588 to 61.949) which

contains zero, so no relationship can be assumed. Similarly, there is no correlation between travel direction gene values and seasonality (Figure 32). The slope is -0.345 (95% confidence interval of -0.896 to 0.175) which also contains zero and shows no relationship. Further experimentation is required to draw any concrete conclusions from this study.

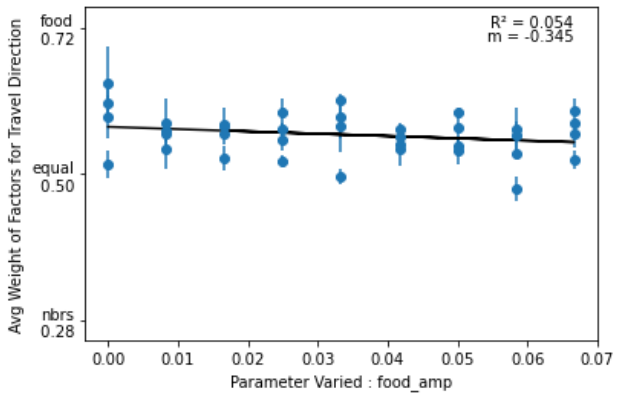

Figure 32: Average resulting travel direction value by seasonality. This plot combines the results of every simulation run as part of this experiment. The x axis is the parameter varied: "food_amp." The value of this parameter is the amplitude of the oscillations in the food patch repopulation rate. The y axis is the average value of all travel direction alleles present in the population at the last time point. Lower travel direction values indicate that worms favor traveling away from their neighbors (or towards neighbors in the case of adult males) and higher travel direction values indicate that worms favor traveling towards their food. These are labeled as "nbrs" for neighbors, "equal" meaning they weigh their decision 50/50 on each factor, and "food." Each data point in this plot represents an individual simulation from this experiment and the line of best fit is drawn through them. Additionally, R^2 (the fit of the line) and m (the slope) are listed in the top right corner.

CHAPTER IV : CONCLUSION

An extensive model in Python has been created to simulate the population dynamics and decision-making strategies of nematodes, such as *C. elegans*, *C. briggsae*, *C. remanei*, and *C. nigoni*, and their responses to various environmental conditions. This program is set up to mimic reality as closely as possible (based on known facts) and contains many aspects of worm life such as their individual attributes and functions (i.e. eating, growing, moving, reproducing, etc.). They also have genetic components which influence their decision to enter dauer and their decision which direction to travel based on their surroundings. All these parts are integrated together to cooperate efficiently and, throughout each simulation, worm information is periodically stored. The goal is to observe how these worms evolve or what patterns arise over time and across different experiments.

There were four experiments conducted, each with multiple versions of the focal variable and multiple replicates of each version as well. In the first experiment, the parameter affecting the likelihood of dying in dauer was altered. This selective property showed strong effects on the average dauer likelihood and average travel direction decision. The stronger the selection, the less worms preferred the dauer stage and the more they avoided their neighbors. In the second experiment, the productivity of the environment was altered. This showed a moderately strong connection to dauer and travel direction. The more often food became available, the less worms preferred the dauer stage and the more they were attracted to the food. In the third experiment, the genotype to phenotype mapping was altered. This parameter changes how strictly the genetic component relates to the likelihood of the worm to choose dauer in any given situation. This showed a weak connection to dauer, but no connection to travel direction. The higher the genotype to phenotype mapping value, the less worms preferred the dauer stage. In the fourth

and last experiment, seasonality in terms of food availability was introduced. Versions of this experiment with large seasonality would go through periods of great environmental productivity and periods of extremely poor conditions. On the other hand, versions with no seasonality had constant conditions. The results of this experiment did not show any connection to the evolution of the worms' genetics in the long run. However, there may be some underlying patterns present that will require further study.

This model is still a work in progress, growing and changing as new facts are discovered and assumptions tested. Many new experiments can arise both in the laboratory and *in silico* as a result of this work. Some projects for the lab include testing more of the assumptions made to improve the model and fill in the gaps in knowledge about these worms. It would be useful to know whether going into dauer differs by gender, exactly how much food is required per life stage and how this translates into energy/growth of the worm, to what extent worms actually avoid their neighbors, will worms evolve to optimize their dauer decision, and more. These kinds of questions can help inform the model and lead to more credible results. Many questions can also arise from the simulations conducted. Other planned experiments include changing the frequency of the sine curve in the seasonality experiment, increasing the probability of mutation, adjusting the *E. coli* carrying capacity, adjusting the initial diversity of the population, using a male/female species, adjusting various parameters in combination, and more. Ultimately, the goal is to be able to predict whether a worm will choose dauer given the specific circumstances.

APPENDIX : LARVAL DIETARY RESTRICTION EXPERIMENT

One assumption made in the model is that worms must consume a specific amount of *E. coli* before they can molt to the next developmental stage. What is unclear is whether the quantity of food available to a worm determines its transition to the next stage of development or if it is simply a matter of time passing. Studies show that life span is food-dependent, specifically adult worms have an increased life span if fed a restricted diet (Klass, 1977; Lee et al., 2006). In addition, worms in the L1 larval stage may survive complete starvation. If an egg hatches into an environment with no food, the L1 worm will arrest in that stage for up to two weeks (Lee et al., 2012). However, it is unknown whether the duration of different larval stages is affected by dietary limitations. To test this, I have performed experiments both *in silico* and in the laboratory.

These experiments were set up with various treatment groups of L1 worms, each fed a different concentration of *E. coli* (6.0, 4.8, 4.5, 3.6, 3.0, 2.4, or 1.5 mg/mL). If the effect is the same on larvae as it is on adults, the expected outcome would be statistically different lengths of time spent in L1 before molting to L2/L2d. To determine when the worms are molting, a specific strain of *C. elegans* (GR1395) with genotype mgIs49[mlt-10p::gfp-PEST] IV and a gfp marker (mlt-10::gfp) on the molting gene was used. This makes the worms fluoresce under an epifluorescent microscope when they are molting and lasts for about three hours, peaking in intensity during the molt (Monsalve et al., 2011). One complication, however, is that if the worms are not completely synchronized in age (i.e. they do not all start from time zero in their life stage), the results may be confounded by this variation. In an effort to answer this question, I performed a power analysis by simulating the data to determine whether unsynchronized L1 worms can produce statistically significant results when comparing between treatment groups.

In the simulation, worms are sorted into four treatment groups with varying *E. coli* concentrations (6 mg/mL, 4.8 mg/mL, 3.6 mg/mL, or 2.4 mg/mL). Each worm is given a starting age and an age they must reach before molting which is experimentally determined (Avery, 2014) and dependent on their treatment, assuming that less food will lengthen the amount of time spent in L1. Subtracting the starting age from the molting age will determine how much longer a worm will remain in L1. Worms will then get placed into bins depending on how long it takes for them to molt and taking into account the length of time they glow. These bins, each three hours wide (how long they glow), represent the observed molting worms for each sample taken. The quantity of bins and amount of time overlap between the bins are calculated based on the sampling frequency (e.g. if samples are collected more frequently, there will be more bins with more overlap between each bin), as shown in Figure A1. Conducting t-tests determines the statistical significance between the four treatment groups and the fraction of worms glowing over time is plotted for visual reference (Figure A2).

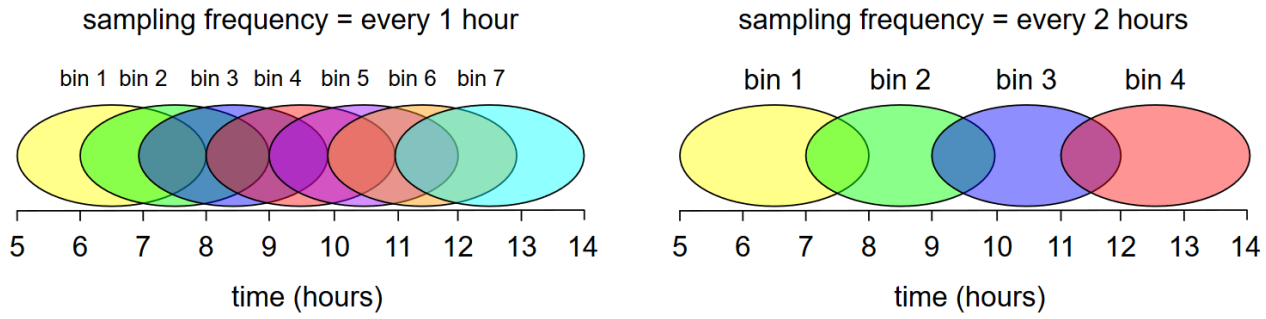

Figure A1: Bin overlap dependent on sampling frequency. Each bin for collecting and counting molting worms encompasses three hours of worms, which is the length of time they will typically glow. The overlap between bins changes depending on how frequently samples are collected. On the left, samples are collected once every hour and there is more overlap between bins than on the right, where samples are collected every two hours.

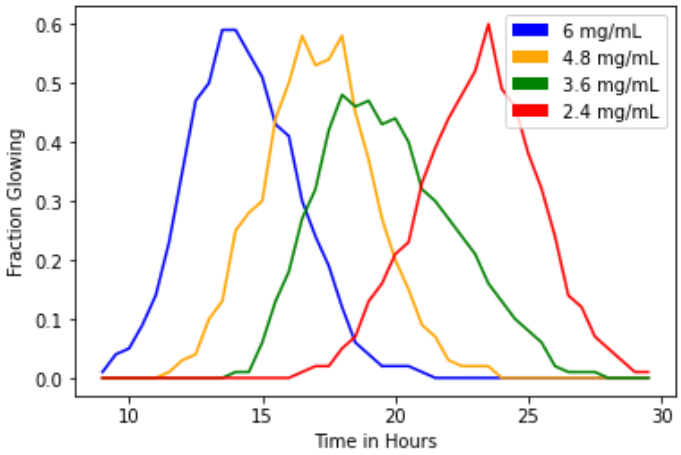

Figure A2: Simulated molting differences between groups. This is an example of what the optimal results from a simulation might look like. There are four treatment groups colored by their *E. coli* concentrations. The parameters were set to the following: samples were taken every half hour, unsynchronized worms were no more than two hours different in age, each treatment group included 100 worms, and the difference in means between groups was three hours. Pairwise comparisons between each of these groups showed statistical significance (all p values << 0.05).

The simulations show how modifying different variables impacts the statistical significance between the treatment groups. The most desirable combination of variables includes sampling every half hour, synchronizing the worms (less than two hours apart in age), and using treatment group sizes of 100 (Figure A2). Also, the larger the difference in average molting time between treatment groups, the easier it will be to differentiate between them (this cannot be controlled in the lab). The variable that seemed to have the biggest effect on t-test significance is synchronicity. Changing the spread of the worms greatly changes the results in comparison to changing the sampling frequency or size of the groups. Thus, it will be crucial to manipulate the worms in laboratory experiments so they are as synchronous in age as possible.

Based on the results of this simulation, I designed an experiment in the laboratory to check the assumption that larval dietary restriction elongates the duration of larval stages. The best results can be obtained when worms are synchronized, meaning that they hatch and then arrest in L1. This will allow all worms in the experiment to start from the same time point in their development. To get a manageable amount of worms, they are first filtered by size and the

adults/larger worms are collected. These worms are then bleached, which kills and dissolves everything except for the unlaid eggs. After these eggs hatch and arrest in S medium, they can then be pipetted into 12 different wells of a 96-well plate. Protocols for making S medium, growing worms in liquid culture, and using a 96-well plate were adapted from documented protocols in literature (Stiernagle, 2006; Solis & Petrascheck, 2011). In every experiment, there are four treatment groups, each repeated three times. Each treatment consists of a different concentration of *E. coli* suspended in S medium and roughly the same amount of worms. Every hour, the number of worms glowing is counted until the amount glowing starts to decline. The fraction of worms glowing over time can then be plotted (Figure A3), similar to the results of the simulation (Figure A2).

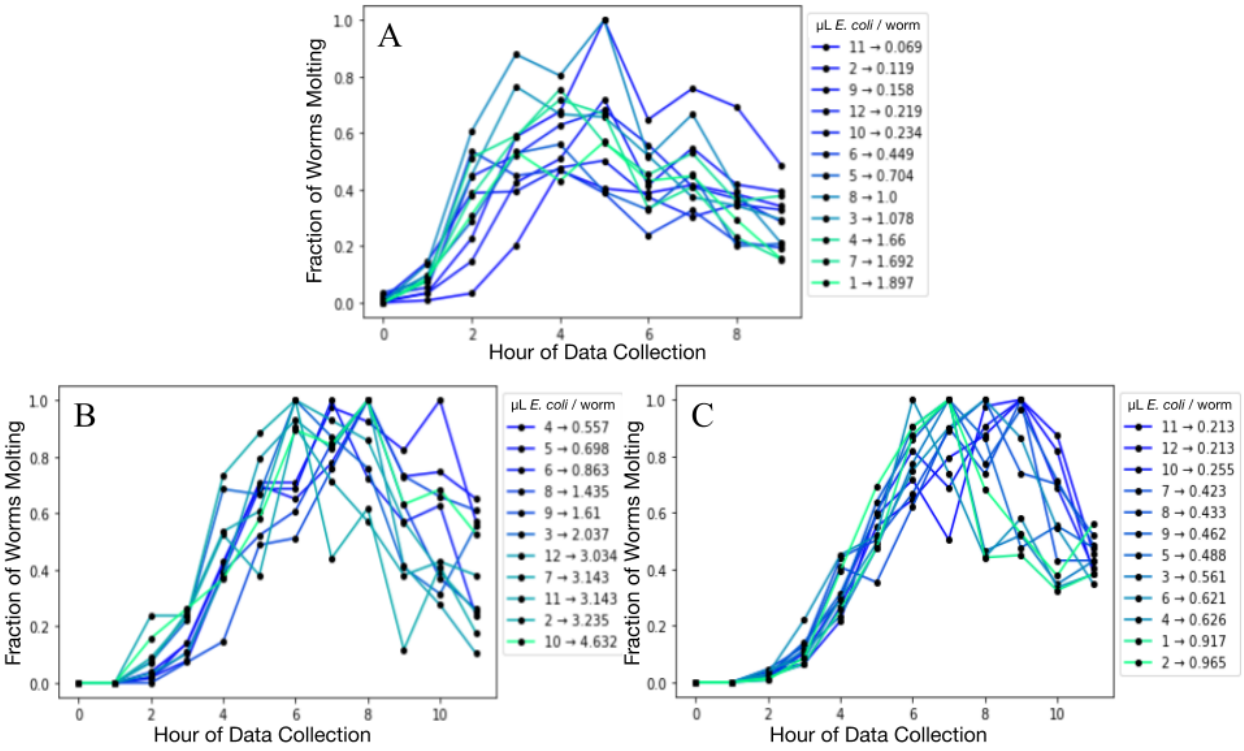

Figure A3: Larval dietary restriction results. These are the results of three larval dietary restriction experiments performed on L1 worms. Each of them shows the fraction of worms glowing throughout time, with data collected every hour for at least 10 hours. The lines are colored by *E. coli* concentration per worm, shown to the right of each graph. The key is labeled as follows: experimental group number → concentration (µL *E. coli*) per worm.

After collecting data from several experiments, statistical analyses were performed. Within each experiment separately, all groups were pairwise compared using a t-test. All significant results (with a p value less than 0.05) were then plotted to determine if an increase in *E. coli* concentration led to an increase or decrease in the amount of time it took the worms to molt (Figure A4). There were 98 significant differences between groups (green dots) that show increasing the food concentration produces a decrease in time spent in L1, while 9 significantly different groups (red dots) show the opposite. A trend line through all these points indicates that there is a negative correlation. Therefore, the conclusion can be made that the more *E. coli* fed to a worm, the faster it will molt to the next stage. These results confirm the assumption made in the model and validate its use. Other similar experimental tests in the laboratory can and should be conducted to check the remaining assumptions for accuracy.

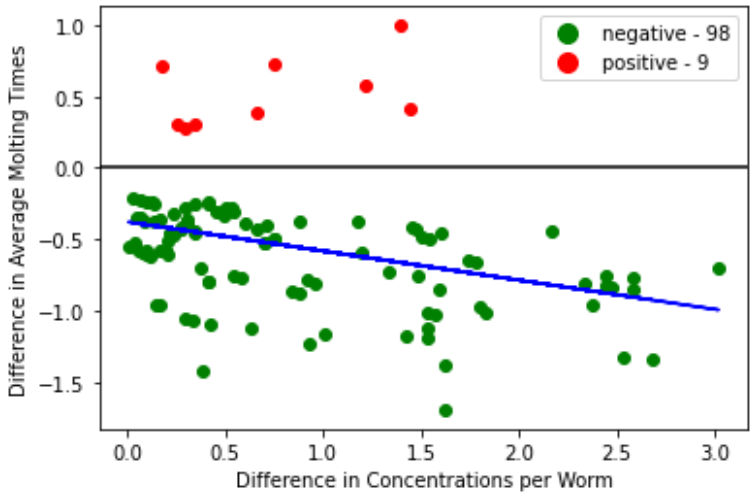

Figure A4: Statistical significance of experimental groups. Each point in this plot represents a pairwise comparison between groups within each of the experiments. Only those with a p value of less than 0.05 are plotted. As the *E. coli* concentration from one group to the next increases (independent variable), the effect on molting time (dependent variable) may be positive or negative. Green points (98 in total) indicate a negative relationship, while red points (9 in total) indicate a positive relationship. The line of best fit is plotted in blue and shows a negative trend.

REFERENCES

Ailion, M., & Thomas, J. H. (2000). Dauer Formation Induced by High Temperatures in Caenorhabditis elegans. *Genetics Society of America*.

Altun, Z. F., & Hall, D. H. (2009). INTRODUCTION TO C. elegans ANATOMY. In L. A. Herndon (Ed.), *WormAtlas*. doi:10.3908/wormatlas.1.1

Anderson, J. L., Morran, L. T., & Phillips, P. C. (2010). Outcrossing and the maintenance of males within C. elegans populations. *Journal of Heredity*, *101*(Supplement 1), S62-S74.

Avery, L. (2014). A model of the effect of uncertainty on the C elegans L2/L2d decision. *PLoS ONE*, *9*(7).

Ben Arous, J., Laffont, S., & Chatenay, D. (2009). Molecular and sensory basis of a food related two-state behavior in C. elegans. *PLoS ONE*, *4*(10).

Brenner, S. (1974, May). The Genetics of CAENORHABDITIS ELEGANS. *Genetics*, *77*(1), 71-94.

Chasnov, J. R., So, W. K., Chan, C. M., & Chow, K. L. (2007, April 17). The species, sex, and stage specificity of a Caenorhabditis sex pheromone. *PNAS*, *104*(16), 6730-6735. https://doi.org/10.1073/pnas.0608050104

Dilberger, B., Baumanns, S., Schmitt, F., Schmiedl, T., Hardt, M., Wenzel, U., & Eckert, G. P. (2019). Mitochondrial Oxidative Stress Impairs Energy Metabolism and Reduces Stress Resistance and Longevity of C. elegans. *Oxidative Medicine and Cellular Longevity*, *2019*.

Dorado-Morales, P., Iglesias, A., Zafrilla, G., Valero, A., Torres, A., Miravet-Verde, S., de Loma, J., Mañas, M., Ruiz, A., Corman, A., Morales, L. J., Peretó, J., Vilanova, C., & Porcar, M. (2014, Feb 4). Engineering bacteria to form a biofilm and induce clumping in Caenorhabditis elegans. *ACS Synth Biol*, *3*(12), 941-3. 10.1021/sb4001883

Ewens, W. J. (2016, April 1). Motoo Kimura and James Crow on the Infinitely Many Alleles Model. *Genetics*, *202*(4), 1243-1245. https://doi.org/10.1534/genetics.116.188433

Félix, M. A., & Duveau, F. (2012). Population dynamics and habitat sharing of natural populations of Caenorhabditis elegans and C. briggsae. *BMC Biology*, *10*(59).

Frézal, L., & Félix, M.-A. (2015). The Natural History of Model Organisms: C. elegans outside the Petri dish. *eLife*, *4*. https://doi.org/10.7554/eLife.05849.001

Hodgkin, J., & Barnes, T. M. (1991, October 22). More is not better: brood size and population growth in a self-fertilizing nematode. *Proc. R. Soc. Lond. B*, *246*(1315), 19-24. https://doi.org/10.1098/rspb.1991.0119

Hu, P. J. (2017). Dauer (D. L. Riddler, Ed.). In *WormBook* (pp. 1-19).

Kimura, M., & Crow, J. F. (1964, April 1). THE NUMBER OF ALLELES THAT CAN BE MAINTAINED IN A FINITE POPULATION. *GENETICS*, *49*(4), 725-738.

Klass, M. R. (1977). AGING IN THE NEMATODE CAENORHABDITIS ELEGANS: MAJOR BIOLOGICAL AND ENVIRONMENTAL FACTORS INFLUENCING LIFE SPAN. *Mechanisms of Ageing and Development*, *6*, 413-429.

Lee, G. D., Wilson, M. A., Zhu, M., Wolkow, C. A., De Cabo, R., Ingram, D. K., & Zou, S. (2006). Dietary deprivation extends lifespan in Caenorhabditis elegans. *Aging Cell*, *5*(6), 515-524.

Lee, I., Hendrix, A., Kim, J., Yoshimoto, J., & You, Y. J. (2012). Metabolic Rate Regulates L1 Longevity in C. elegans. *PLoS ONE*, *7*(9).

MacNeil, L. T., Watson, E., Arda, H. E., Zhu, L. J., & Walhout, A. J.M. (2013, March 28). Diet-Induced Developmental Acceleration Independent of TOR and Insulin in C. elegans. *Cell*, *153*, 240–252. http://dx.doi.org/10.1016/j.cell.2013.02.049

McGrath, P. T., & Ruvinsky, I. (2019, Feb). A primer on pheromone signaling in Caenorhabditis elegans for systems biologists. *Current opinion in systems biology*, *13*, 23-30. https://doi.org/10.1016/j.coisb.2018.08.012

Monsalve, G. C., Van Buskirk, C., & Frand, A. R. (2011). LIN-42/PERIOD controls cyclical and developmental progression of C. elegans molts. *Current Biology*, *21*(24), 2033-2045.

Rodríguez-Palero, M. J., López-Díaz, A., Marsac, R., Gomes, J.-E., Olmedo, M., & Artal-Sanz, M. (2018, February 26). An automated method for the analysis of food intake behaviour in Caenorhabditis elegans. *Scientific Reports*, *8*(3633). https://doi.org/10.1038/s41598-018-21964-z

Serena Ding, S., Schumacher, L. J., Javer, A. E., Endres, R. G., & Brown, A. E. (2019). Shared behavioral mechanisms underlie C. elegans aggregation and swarming. *eLife*, *8*, 1-32. https://doi.org/10.7554/eLife.43318.001

Shi, Z., Yu, H., Sun, Y., Yang, C., Lian, H., & Cai, P. (2015). The energy metabolism in caenorhabditis elegans under the extremely low-frequency electromagnetic field exposure. *Scientific Reports*, *5*(847), 1-11.

Solis, G. M., & Petrascheck, M. (2011). Measuring caenorhabditis elegans life span in 96 well microtiter plates. *Journal of Visualized Experiments*, (49), 1-6.

Stewart, A. D., & Phillips, P. C. (2002). Selection and Maintenance of Androdioecy in Caenorhabditis elegans. *Genetics Society of America*, (160), 975-982.

Stiernagle, T. (2006). Maintenance of C. elegans (D. Fay & V. Ambros, Eds.). In *WormBook* (pp. 51-67).

Uppaluri, S., & Brangwynne, C. P. (2015). A size threshold governs Caenorhabditis elegans Developmental Progression. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1813).

Van Raamsdonk, J. M., Meng, Y., Camp, D., Yang, W., Jia, X., Bénard, C., & Hekimi, S. (2010). Decreased energy metabolism extends life span in Caenorhabditis elegans without reducing oxidative damage. *Genetics*, *185*(2), 559-571.

Viney, M. E., Gardner, M. P., & Jackson, J. A. (2003, August). Variation in Caenorhabditis elegans dauer larva formation. *Development, growth & differentiation*, *45*(4), 389–396. https://doi.org/10.1046/j.1440-169x.2003.00703.x

White, J. Q., Nicholas, T. J., Gritton, J., Troung, L., Davidson, E. R., & Jorgensen, E. M. (2007, November 6). The Sensory Circuitry for Sexual Attraction in C. elegans Males. *Current Biology*, *17*(21), 1847-1857. https://doi.org/10.1016/j.cub.2007.09.011

Yeh, S.-D., Saxena, A. S., Crombie, T. A., Feistel, D., Johnson, L. M., Lam, I., Lam, J., Saber, S., & Baer, C. F. (2018, Jan). The mutational decay of male-male and hermaphrodite-hermaphrodite competitive fitness in the androdioecious nematode C. elegans. *Heredity*, *120*(1), 1-12. 10.1038/s41437-017-0003-8

Yin, D., & Haag, E. S. (2019, June 25). Evolution of sex ratio through gene loss. *PNAS*, *116*(26), 12919-12924. https://doi.org/10.1073/pnas.1903925116

Zanni, E., Laudenzi, C., Schifano, E., Palleschi, C., Perozzi, G., Uccelletti, D., & Devirgiliis, C. (2015). Impact of a complex food microbiota on energy metabolism in the model organism caenorhabditis elegans. *BioMed Research International*, *2015*.