UNIVERSITY OF CALIFORNIA SAN DIEGO

Model Selecting Upon Cell Differentiation Shows Minor Changes in Modularity in the Gene Networks

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

in

Biology

by

Shea B. Summers

Committee in Charge:

Professor Scott Rifkin, Chair
Professor Sergey Kryazhimskiy
Professor Deirdre Lyons

2019

The Thesis of Shea B. Summers is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

Chair

University of California San Diego

2019

DEDICATION

Thank you to Dr. Scott Rifkin, who accepted me into his lab and graciously mentored me through my graduate schooling, I am sincerely grateful.

Thank you to Dr. Deirdre Lyons and Dr. Sergey Kryazhimskiy, for their time and effort in helping me with my thesis and defense.

Thank you to my family, who are always there when I have needed them most and I am sure will be close when I need them again.

EPIGRAPH

“Do you think God stays in heaven because he too lives in

fear of what he's created?”

-Steve Buscemi, *Spy Kids 2*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT OF THE THESIS

Model Selecting Upon Cell Differentiation Shows Minor Changes in Modularity in the Gene Networks

by

Shea B. Summers

Master of Science in Biology

University of California San Diego, 2019

Professor Scott Rifkin, Chair

It is difficult to research the evolution of cell differentiation in nature because of the unavailability of a suitable model organism and the advanced time since the divergence of single and multicellular organisms. Instead of looking at a natural model we decided to explore this issue with a computer model. We based our model on the one proposed by Andreas Wagner in 1994 to look at the evolution of gene networks. We modified the model to account for selection goals that promoted cell differentiation. Because of the prevalence of modular components in more advanced biological systems we expected modularity to form in the evolved gene

networks. Therefore, we tested the gene networks for modularity using the K-Means clustering test. The model was completed after many editions that swapped between concepts introduced in previous scientists' adaptations of the Wagner model. Increased modularity was observed in the evolved populations of the model over the unevolved populations, 0.149 over 0.136 $p < 2.2e-16$. However, this change was so minor there may be no biological significance. Additionally, whether the model evolved towards one fitness goal then the other or both at the same time was tested and found to be both. Finally, whether connectivity increased when selecting for differentiation was tested and the results are inconclusive at best.

Introduction

Life on earth presents itself in a variety of ways. From fish to fowl, animals roam the earth, plants take root, and bacteria survive pretty much anywhere. There are many clear divides in the tree of life, but few are as obvious as the divide between multicellular and unicellular organisms. Somewhere deep into the past, life made a huge jump changing from being one cell holistically doing all the organism needed to live, to a collection of cells that separated functions and combined formed one working organism. How did this happen? No one really knows and there are many problems with studying it. The biggest is that the evolution of multicellularity for the common ancestor of the plants and animals that we have today happened too long ago into the past to study with our DNA analyzing techniques (Miller 2010). This is mostly due to with so many changes happening to the genomes that the development of multicellularity is hard to pinpoint. There are some organisms that could be used to model the development of cell differentiation and multicellularity, Volvox and Dictyostelium discoideum. The volvox are a type of algae that have developed to have only two distinct cell types, somatic cells and reproductive cells (Gilbert 2000). They seem to be the simplest multicellular organism discovered, while Dictyostelium discoideum are unicellular organisms on the cusp of multicellularity. They are a unicellular organism that reproduces with binary fission, but when nutrients are exhausted in their environment thousands of them congregate to form a slug like colony that consists of cells used to move the colony, dying off later, and cells that will go on to reproduce at the new habitat (Gilbert 2000). These organisms seem to be on either side of the differentiation divide, but even using them as models it is hard to find the bridge between the two. Therefore, we turn to computer modeling to do so.

Computer modeling has been important to science since the advent of computing, helping humans recreate things they can't understand or don't see in the real world. Therefore, it is a great solution to a problem of trying to understand a phenomenon that we know to have happened like the development of cell differentiation and multicellularity. Rather than build a completely new model for this project we decided to adapt an existing model, the Wagner model. The Wagner model is a simple mathematical model made by Andreas Wagner that was originally used to model the effects of gene duplications effects on a gene network (Wagner 1994). He would go on to adapt the model for use in an investigation into evolutionary plasticity (Wagner 1996). It is a very easy to understand model that mainly consists of a gene network matrix and an expression vector being used to simulate the development of an organism (Wagner 1996). This simplicity in design and focus on gene networks has caused it to be a staple in evolutionary modeling and to be commandeered and adapted for many unique purposes. It has been used to investigate developmental stability (Siegal & Berman 2002), network robustness (Sevim & Rivkold), genetic assimilation (Masel 2004), and the effect of sexual dimorphism on gene networks (Fierst 2011). Using the Wagner model as a basis and the many offshoot models as inspiration for changes we set out to create a model that would simulate as simply as possible the evolution of a gene network that was under selective pressures to differentiate its cell type according to outside stimuli.

With the goal of evolving differentiation set the next question would be what do we expect to happen to the gene network? For this we looked at somethings held common in evolving more complex organisms, modules (Wagner & Mezey 2001). Modules are the clustering of elements in a network or development that are relatively autonomous in function. A good example of this is the development of limb buds in embryos (Wagner & Mezey 2001).

Complex organisms have a multitude of these modular structures in their genomes and development (Wagner 2007). There is also some mystery around the origin of modularity in biological systems, and whether the mechanisms from which it arises (Wagner 2001, 2007 Kahstan 2005). Therefore, we hypothesis if a model based on the 1996 Wagner model is built that selects for cell differentiation then the resultant gene network will become more modular because of it.

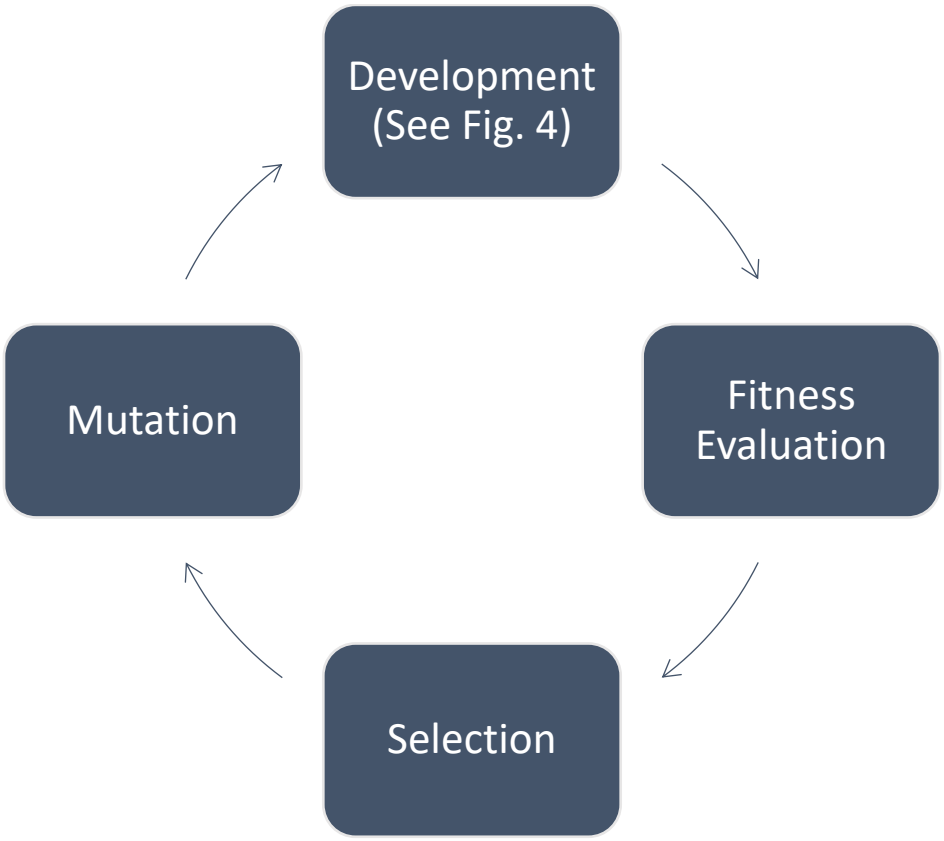


Figure 1: Model Flow Chart

Model:

The model begins with creating a population of ***x*** individuals each having ***n*** genes. The genes can be thought of as all genes that make up the organism's genome or just a small subsect of

n = 4				
Genes	1	2	3	4
1	1 on 1	1 on 2	1 on 3	1 on 4
2	2 on 1	2 on 2	2 on 3	2 on 4
3	3 on 1	3 on 2	3 on 3	3 on 4
4	4 on 1	4 on 2	4 on 3	4 on 4

Figure 2: Representation of a Gene Network, n = 4

genes concerning a function that needs to be differentiated between cell types. in the final cell differentiation model ***x*** = 1000 and ***n*** =20. All individuals are denoted by their gene network. This gene network is a matrix with ***n*** rows and ***n*** columns. Each value denotes the effect of the row gene on the column gene (Figure 1). In our model these values are randomly chosen between -1 and 1. However not all genes affect the expression of other genes therefore some values in the gene network are changed to 0. The percentage of which is denoted by the variable

c, c

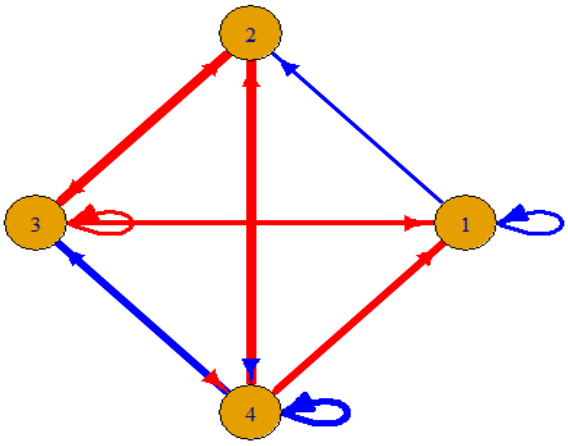


Figure 3: Visual Representation of a Gene Network, n =4

= 0.6 in the final model**.** Each individual has one expression vector, a vector of length ***n*** whose values, either 0 or 1, denote the expression level of each pertinent gene at the start of the organism's life cycle. The individual develops by a series of cross products between their gene network and the expression vector. The starting expression vector is crossed with the gene network and a vector of length ***n.*** All vector values greater than 0 are then changed to 1 and the rest are changed to 0. This vector is the new expression vector of the individual and is then crossed with the individual's gene network, repeating ***I*** ±6 times, ***I*** = 80 in the final model. This model structure is consistent with the original Wager Model and most of its off shoots.

However, to look at cell differentiation one major change was made. Instead of one expression vector there are two. The second expression vector is nearly identical to the first except for one value being changed from either 1 to 0 or vice versa. This is to simulate a change

in environment or an initial difference between two cells that acts as a signal to the organism to have two specific and different expression patterns in two cells at the end of development. This means that after development every individual is defined by three objects, their gene network, the final expression vector resulting from the original starting expression vector (final expression vector 1), and the final expression vector resulting from the modified expression vector (final expression vector 2) (Figure 4).

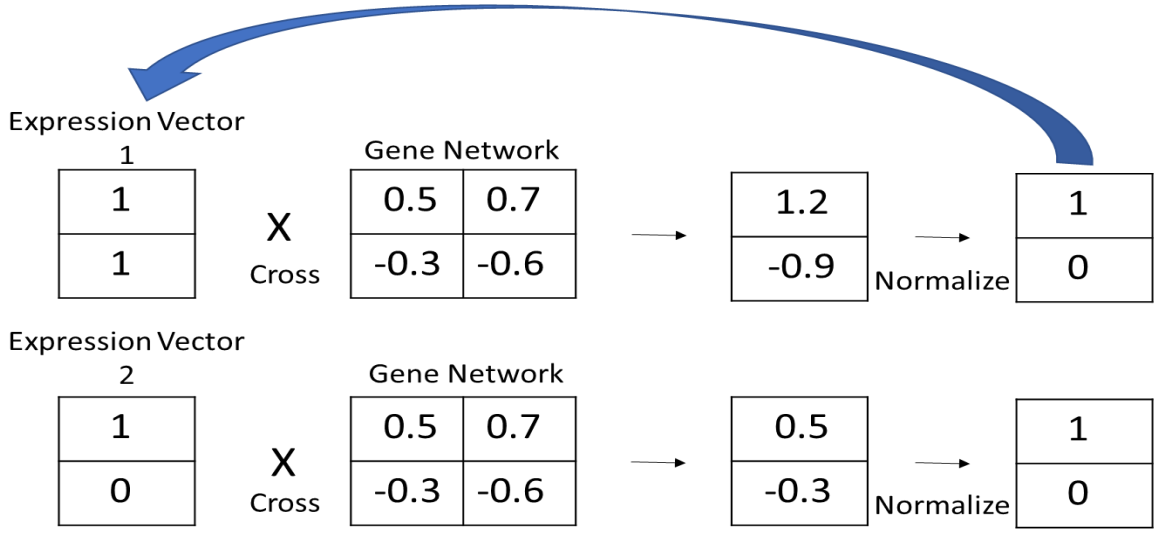


Figure 4: Example of Development n=2

After going through development, the individuals' finesses will be evaluated. Because the model is trying to evolve cell differentiation two different fitness goals are set. The individual is then evaluated by how closely each final expression matches the fitness goals. We used target expression patterns as our fitness goals, (0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1) was the goal for one vector and (1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0) was the goal for the second. The number of matching values between final expression value 1 and fitness goal 1 were added to the number of matches between final expression value 2 and fitness goal 2. These combined matches, ***m***,

were then divided by 2 * ***n***, ***m*** /2 * ***n***. The resulting value is taken as the fitness for that individual in the population.

With the fitness evaluated the population then goes through selection. At this point there is a population of ***x*** individuals each with a fitness value between 0 and 1. The next generation will also have a population of ***x***. To choose which individuals "leave offspring" (have their gene networks moved to the next generation) the fitness must be a strong factor, but some randomness will still be introduced. To accomplish this for each position in the next generation an individual from the current generation is randomly selected and a random number between 0 and 1 is generated. The fitness value is compared to this random number, if the fitness value is higher than the random number then a copy of that individual's gene network is placed in that position for the next generation. If the fitness value is lower than the randomly generated value then the individual from the current generation does not get a copy of its gene network placed at that position in the next generation, instead a new random individual from the current generation is selected and a new random number is generated. Once the entire next generation is filled, it is mutated.

Mutations are simulated in the model by creating an array of values the same dimensions as the run's population, $\boldsymbol{x * n^2}$ or $1000*20^2$, then each value is assigned a randomly generated number from 0 to 1. A mutation rate ***r*** is set (in our case it was 0.0005) and if that randomly assigned value is under ***r***, then that position in the population gets a mutation. The mutation replaces the value at the determined index with a new value between -1 and 1. When all mutations are carried out the population is returned to the start of its life cycle and begins development again. This goes on for a pre-determined number of generations, ***g***, or until any increase in fitness from generation to generation flatlines.

Tests

K-Means

To determine whether a gene network forms modular structures we used the kmeans function in R. This the R function that runs K-Means clustering analysis on a data set. This clustering method tries to separate a data set into k groups such that the sum of squares from any point in that group to the groups center is minimized. Two outputs of this function were important to our analysis, the totss and betweenness of the kmeans groups. The totss is the total sum of squares, and the betweenness is the sum of squares outside of the group. The large that value is in comparison to the totss, the better grouped the data is in the kmeans analysis, the more modular it is. We took this analysis and simplified it into one value betweenness/totss which we will refer to as the k-means value.

Index Weight

To test if there were any indices within the network that were more sensitive to changes in value then others a test was run.400 copies of the network were cloned. In each one, one value within the network was replaced with -0.9. The networks were then evaluated for fitness. The value was then replaced with 0.-8 and repeated with intervals of 0.1 until 0.9.

Past Models

This model has gone through many changes, and this section will review some of the more important changes.

Continuous Expression

The first model that we designed had continuous expression values. This feature differs from the original Wagner model (Wagner 1996) which used -1 and 1 to denote gene expression being either on or off, however when first constructing the model we felt that continuous expression values were more characteristic of life and had seen other successful examples of the model being changed in such a way (Siegal 2006).Therefore we set the gene expression levels to be between 0 and 1. This resulted in a smoother fitness curve and a longer generational time to reach near perfect fitness. We reverted the model back to an on/off expression model for three reasons, first to try and force a more modular structure, second to simplify the output to more easily interpret the data, and third to return to the original form of the Wagner model instead of an offshoot.

No ***c***

Connectivity, or what percentage of the gene networks values are not 0, was a basic component of the Wagner model. However previous studies demonstrated that a higher connectivity or ***c*** will cause higher evolutionary stability and robustness (Siegal & Berman 2002). Therefore, we justified having a network fully connected to increase said attributes without selecting for it. However, we later felt that having a fully connected network would be an assumption too far from reality to justify adding to the model. Additionally, breaking up the gene networks with random 0 values would inherently make it more modular, and hopefully force more noticeable modularity in evolved networks than we had seen.

Summed Fitness/All Zero or All One

The original Wagner model was used to simulate, and test how gene networks changed due to duplication (Wagner 1994) and evolutionary plasticity (Wagner 1996). The original model and many of the variations have a set optimum fitness goals for after development. Other papers in that attempt to evolve modularity into a network note that adding such optimums increases the likelihood for modules to occur (Kashtan 2005). Therefore, we knew we needed clearly defined fitness goals that were centered around differentiation. We first followed the example of Emilia Huerta-Sanchez and her version of the Wagner model and set one fitness goal to be all fully expressed genes (Huerta-Sanchez & Durrett 2007), (1,1,1,1,1….), and have the other goal as no gene expression from any of the genes, (0,0,0,0,0…..). To simplify this further fitness was tested by comparing the sum of an individual's final expressions and a fitness goal of ***n*** or 0. However we worried about the effects of having an all 0-expression optimum on the gene network and the simplicity of achieving all 1 or zeros as expressions. Therefore, the expression goals were changed to be perfectly dissimilar but equally complex expression vectors of alternating ones and zeros, (1,0,1,0…) and (0,1,0,1…).

Exact Development

Originally the model had a fixed end point in development. Instead of having individuals develop over 80 ± 6 iterations, every individual developed for exactly 80. Stability is a large component of both papers explaining the Wager model (Wagner 1994, Wagner 1996) and has been tested in many papers since (Bergman & Siegal 2003, Sevim & Rivkold 2008, Pinho 2012, Masel 2004). Initial testing with our model seemed to show a higher frequency of stable networks than previously seen (Bergman & Siegal 2003). Therefore, stability was not initially tested. However, after many changes to the model that were largely a return to the original proposed by Wagner stability became an issue. To test it we tracked fitness of individuals during development. We found the fitness of the individuals fluctuated constantly during development and the finesses we observed after iteration 80 often bore no resemblance to the fitness at iteration 79 or, if we

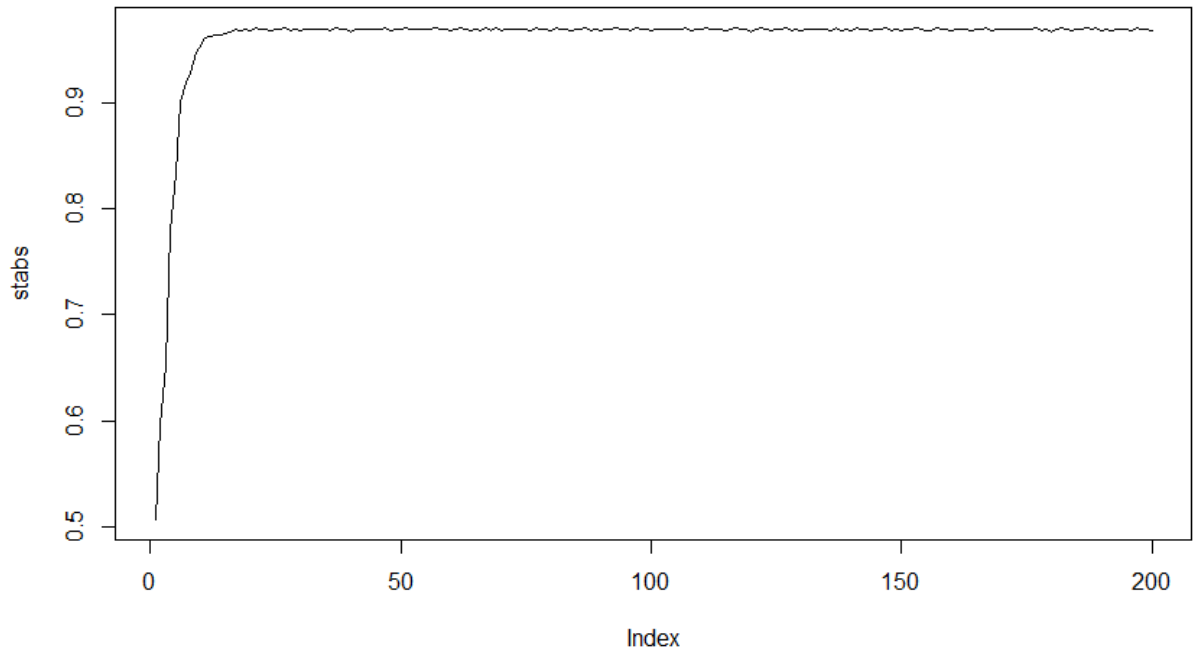


Figure 5: Fitness over Development after Change

extended development, at iteration 81. There needed to be a change in development that would

cause the expression vectors to stabilize by maturity as we added, variation to the development time. This lack of stability clearly showed that the model would do what we told it to do and nothing more. The added variation would force the gene networks to evolve in such a way that the expression values were stable and at highest fitness at all possible maturation points. This resulted in stabilizing the expression vectors to a level we felt was acceptable for the model. We are not as concerned with stability it is not up to the level of (Bergman & Siegal 2003) but as seen in Figure 5 the fitness values seem to be consistent for the average individual in a population, and there seems to be a clear developmental trajectory after the change.

Unique Individuals

The original Wagner model does not have unique individuals in the population at the start of the simulation and this is a common trait to the offshoot models as well. We attempted to add more genetic diversity to the initial population by having every individual in the population be a completely different constructed network. In tracking the progress of these populations, it became obvious that the initial genetic diversity was quickly consumed by one superior individual and their offspring. We felt that because of these results there was little point in adding another variable to the model that would be quickly rendered insignificant after 100 generations. particularly since mutation seemed enough to generate diversity for selection to act upon. The populations in our model start with clones of one individual

Additive Mutations

There are two theories on how to simulate mutations. The first is the way the original Wagner model does, and our current model follows. The mutation is thought to be separate from the gene networks initial value. Therefore, a mutation will replace that value with a new completely random value that is still within the constraints of the initial assignment of value (Hodgin-Davis 2015). If the mutated index has a value of 0.5 the mutated value can be any number between -1 and 1. There is another school of thought that assumes that a mutation builds from the current expression/network value of the gene (Zeng & Cockerham 1993). If the mutated index has a value of 0.5 then the mutated value can be $0.5 \pm (0\text{-}1)$. This thought process was what we ascribed to when initially forming the differentiation model, and the mutations were additive. One major problem came up in the model, with additive mutations the gene network would occasionally get values that were significantly higher than the initial formation allowed. For example, one evolved gene network had an interaction value of approximately 8. This led to the decision to change the way mutations worked in our model to more closely resemble the original Wagner model. Another option to correct this could have been adding more constraints to the additive mutation mechanism so that the gene network values could not go beyond set boundaries, but we felt that a return to form of a model was necessary.

Expressions Represented by -1 and 1

In the original model expression of a gene is noted as being either on or off, in the expression vector this is noted by -1 for off and positive 1 for on. There appears to be no reasoning for these values other than to denote an active an inactive gene as -1, and many papers

change the model to have the inactive/off state be represented by a 0 instead (Masel 2004, Espinosa-Soto 2011). They reason that 0 being the representation of no gene expression is a more biologically accurate interpretation than using -1. We agree with this thought process and have implemented it in our model. However, we have implemented the -1 and 1 notation in past editions and see no real difference in output, so it is an arbitrary decision.

Variable Explanation

Table 1: List of Variables

Variable	Function	Value in Final Model Runs
x	Number of individuals in a population	1000
n	Number of genes selected upon	20
c	Connectivity, fraction of nonzero values in a gene network	0.6
I	Iterations performed during development	80
g	Number of generations carried out in a run	40,000
r	Mutation Rate	0.0005

Results

Evolution

The first result we were concerned with was whether the model would actually evolve a population that reaches the set fitness goals, or at least comes close enough to show that we are evolving populations towards differentiation. This seems to have been accomplished as shown in Figure 6. The average fitness of the population will always lag the best fit because of the high

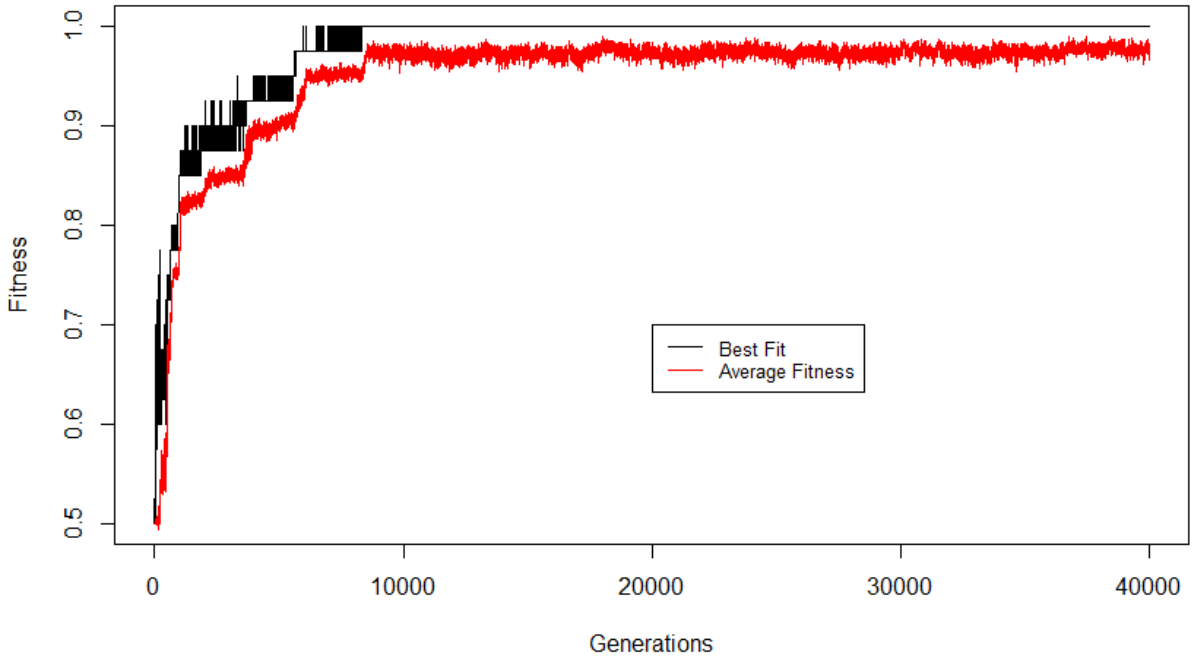


Figure 6: Graph of Complete Model Fitness across Generations

mutation rate, but over all the two lines seem to match each other's progress. There is one individual more fit than the others that survives drift and proceeds to have its offspring take over the population The Best Fit line also has many peaks that fall off by the next couple generations, which shows that some more fit individuals are lost to drift. As that as the fitness increases so

does the number of generations it takes till the next jump in fitness. The graph shows a pattern of evolution expected from the model.

Modularity

Throughout the building and refining of our model we have been searching for the development of modularity within the gene network. To do this we would take a newly generated population and get their k-means value for each individual in that population. The population would then go through 39,000 generations of evolution to meet the differentiation fitness goals. The mean value for both the evolved and un-evolved networks were compared and showed no

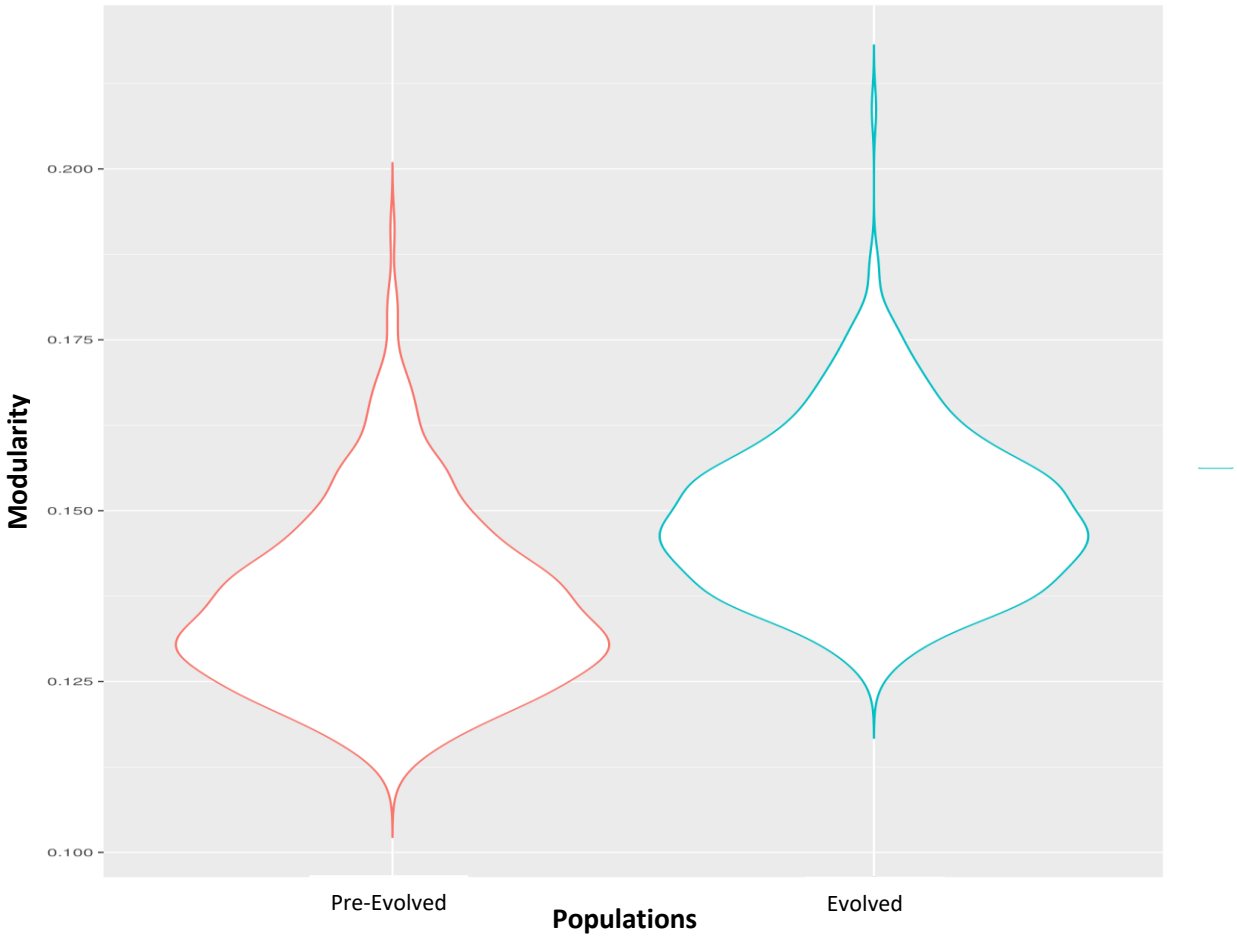


Figure 7: Violin Graph of K-Means for Evolved and Un-Evolved

change in modularity. Figure 7 shows a violin plot of 370 runs using the variables listed in Table

1 showing these results. When a t-test was run on the data the p value <2.2e-16, with mean values of 0.1362998 for the pre-evolved and 0.1496850 for the evolved. Therefore, we can conclude that through the evolution of a population selecting upon cell differentiation modularity does form in the population's gene networks. However, we expected more of a difference in mean than seen, therefore we can only call this a minor change in modularity and have doubts on its biological significance. Though not as numerous as tests done to make the figure above every edition of the model was tested for modularity using k-means clustering and came up with similar values to the means for both evolved and un-evolved individuals. Graphical representations of the gene networks were also made to better visualize the data similar to Figure 3 and use some of the graphical clustering tests in R. Specifically cluster_walktrap was used, which is a function that through random pathing in a graph network tries to find highly connected portions of the graph known as communities. This test yielded no significant difference in modularity of the gene networks from evolution. The index weight tests were then preformed to try to find any other forms of structure in the gene networks besides modularity. The results of this were a seemingly random set of indices that caused major change in fitness if altered. They were not the indices with the most negative or most positive values, nor was the replacement value that caused the most extreme change in the network the one that caused the most extreme change in fitness. The columns were even normalized to see if it was the values with the highest effect on their respective genes that were being highlighted, and that also showed no similarities to the seemingly fitness important indicis.

Development

After seeing minor change in modularity from un-evolved to evolved networks we decided to investigate the evolutionary mechanics to see whether the model evolved to reach one fitness goal and then the other or both at the same time. To do this we took the previous 370 runs of the model and looked at the fitness values for each fitness goal every 100 generations. To more easily analyze the data only 1000 individuals from those runs were formed into a 2-

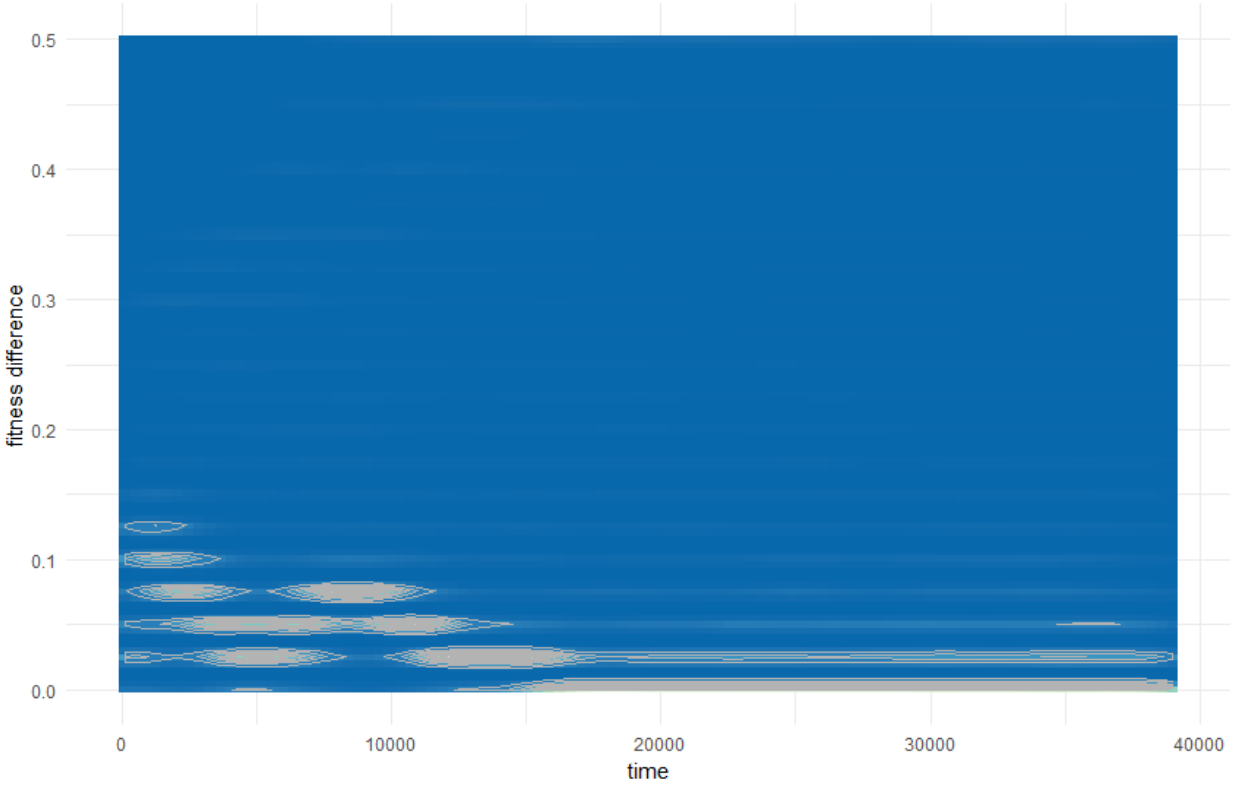
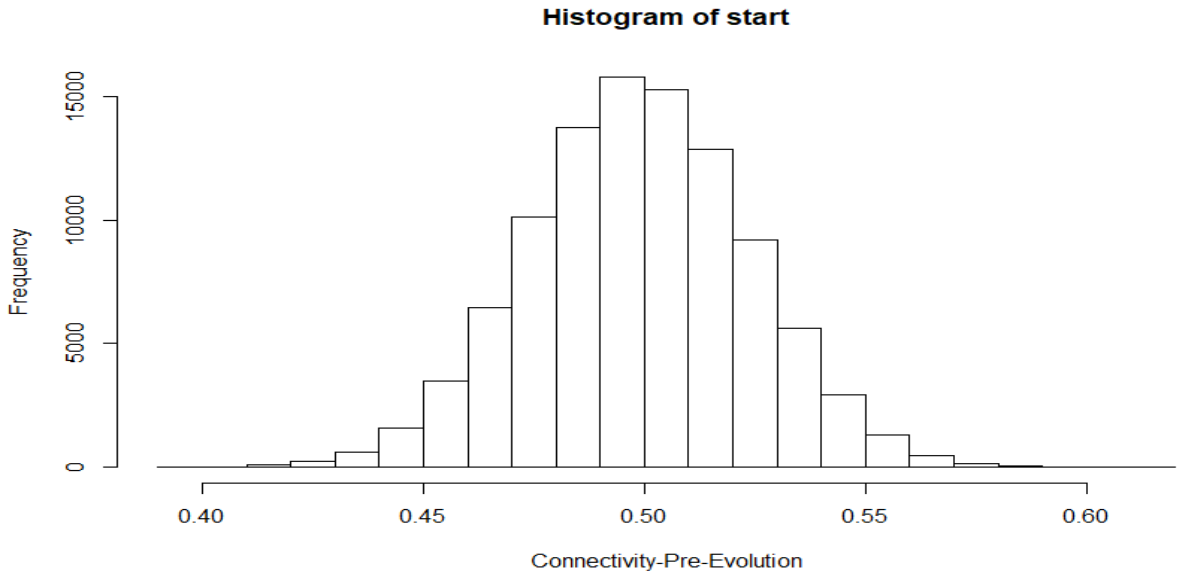


Figure 8: 2D Histogram of Fitness Differences Between the Two Fitness goals over Time in Generations

dimensional histogram. the results of which are in Figure 8. The lighter the area the more individuals that have a difference of fitness of that magnitude in that generation. The figure shows at the start of the runs the difference is at its maximum for most of the population, and as time progresses that difference steadily decreases. This seems to indicate that they seemed to evolve mostly together with one goal being slightly more fit at any given time.

Connectivity

Finally, we decided to look at how selecting for differentiation would affect the connectivity if the starting population had individuals with different connectivity values, ***c***. To do this we edited the model so that the original population created unique individuals with a randomly chosen ***c*** normally distributed around 0.5. This can be seen in Figure 9, in the first graph for a pre-evolved population. The population then went through 1000 generations selection in the model. 100 of these runs were performed. For each individual in the starting populations a ***c*** value was obtained, and a histogram was made to examine the connectivity distribution and mean, which was 0.5. The same was done to for the populations that went through 1000 generations of evolution, with a mean of 0.496. The data shows that there is a noticeable change from the normal distribution at the start of the runs to values skewed towards a lower ***c*** value and a smaller range than the original distribution, as seen in the second graph of Figure 9.



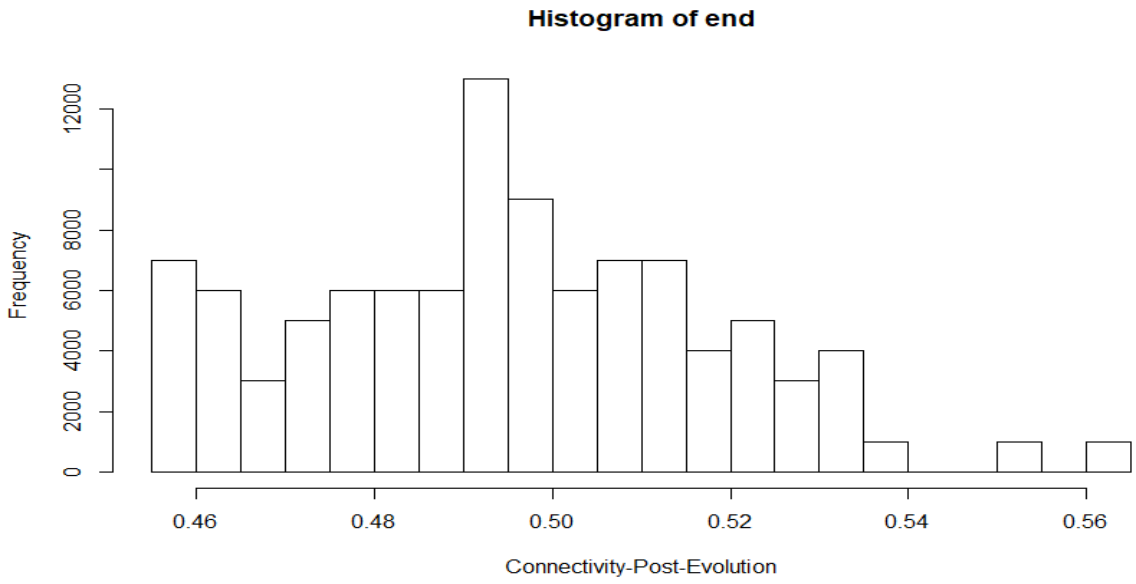


Figure 9: Connectivity Graphs

Future Directions

I believe that this model has further unexplored potential, that will be realized in future experimentation. Though there was little significance to be found in searching for modularity in gene networks of evolved individuals from the simulation, some sort of structure or at least collection of weighted values seems to be present in the networks. A further examination of this aspect could yield more answers on the general structuring of gene networks in more advanced multicellular organisms. Exploration into the effects of selection for differentiation on the development of the individuals would also be an aspect of the model that was not explored in depth but could yield interesting results. Finally, I hope that sexual recombination and gene duplication could be added into the model, to more closely resemble the original Wagner models, and make the cell differentiation model add vital evolutionary components that could bring out the structure we had been searching for throughout our experimentation.

References

Azevedo, R.B.R., Lohaus, R., Srinivasan, S., Dang, K.K., and Burch, C.L. (2006). Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. Nature *440*, 87–90.

Bergman, A., and Siegal, M.L. (2003). Evolutionary capacitance as a general feature of complex gene networks. Nature *424*, 549–552.

Bongard, J. "Evolving Modular Genetic Regulatory Networks." Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No.02TH8600), 7 Aug. 2002, doi:10.1109/cec.2002.1004528.

Ciliberti, S., Martin, O.C., and Wagner, A. (2007a). Robustness Can Evolve Gradually in Complex Regulatory Gene Networks with Varying Topology. PLoS Comput Biol *3*, e15.

Ciliberti, S., Martin, O.C., and Wagner, A. (2007b). Innovation and robustness in complex regulatory gene networks. PNAS *104*, 13591–13596.

Draghi, J., and Wagner, G.P. (2008). Evolution of Evolvability in a Developmental Model. Evolution *62*, 301–315.

Draghi, J., and Wagner, G.P. (2009). The evolutionary dynamics of evolvability in a gene network model. Journal of Evolutionary Biology *22*, 599–611.

Draghi, J.A., Parsons, T.L., Wagner, G.P., and Plotkin, J.B. (2010). Mutational robustness can facilitate adaptation. Nature *463*, 353–355.

Espinosa-Soto, C., Martin, O.C., and Wagner, A. (2011a). Phenotypic robustness can increase phenotypic variability after nongenetic perturbations in gene regulatory circuits. Journal of Evolutionary Biology *24*, 1284–1297.

Espinosa-Soto, C., Martin, O., and Wagner, A. (2011b). Phenotypic plasticity can facilitate adaptive evolution in gene regulatory circuits. BMC Evolutionary Biology *11*, 5.

Fierst, J.L. (2010). Sexual Dimorphism Increases Evolvability in a Genetic Regulatory Network. Evol Biol *38*, 52–67.

Fierst, J.L. (2011). A history of phenotypic plasticity accelerates adaptation to a new environment. Journal of Evolutionary Biology *24*, 1992–2001.

Fierst, J.L., and Phillips, P.C. (2015). Modeling the evolution of complex genetic systems: The gene network family tree. J. Exp. Zool. (Mol. Dev. Evol.) *324*, 1–12.

Gilbert SF. Developmental Biology. 6th edition. Sunderland (MA): Sinauer Associates; 2000. Multicellularity: The Evolution of Differentiation. Available from: https://www.ncbi.nlm.nih.gov/books/NBK10031/

Huerta-Sanchez, E., and Durrett, R. (2007). Wagner's canalization model. Theoretical Population Biology *71*, 121–130.

Kashtan, N., and U. Alon. "Spontaneous Evolution of Modularity and Network Motifs." Proceedings of the National Academy of Sciences, vol. 102, no. 39, 2005, pp. 13773–13778., doi:10.1073/pnas.0503610102.

Leclerc, R.D. (2008). Survival of the sparsest: robust gene networks are parsimonious. Molecular Systems Biology *4*, 213.

Masel, J. (2004). Genetic assimilation can occur in the absence of selection for the assimilating phenotype, suggesting a role for the canalization heuristic. Journal of Evolutionary Biology *17*, 1106–1110.

Miller, S. M. (2010) Volvox, Chlamydomonas, and the Evolution of Multicellularity. Nature Education 3(9):65

Pinho, R., Borenstein, E., and Feldman, M.W. (2012). Most Networks in Wagner's Model Are Cycling. PLoS ONE *7*, e34285.

Pinho, R., Garcia, V., and Feldman, M.W. (2015). Phenotype Accessibility and Noise in Random Threshold Gene Regulatory Networks. PLoS One *10*.

Sevim, V., and Rikvold, P.A. (2008). Chaotic gene regulatory networks can be robust against mutations and noise. Journal of Theoretical Biology *253*, 323–332.

Siegal, M.L., and Bergman, A. (2002). Waddington's canalization revisited: Developmental stability and evolution. PNAS *99*, 10528–10532.

Siegal, M.L., Promislow, D.E.L., and Bergman, A. (2006). Functional and evolutionary inference in gene networks: does topology matter? Genetica *129*, 83–103.

Stearns, S.C. (2002). Progress on canalization. PNAS *99*, 10229–10230.

Wagner, A. (1994). Evolution of gene networks by gene duplications: a mathematical model and its implications on genome organization. PNAS *91*, 4387–4391.

Wagner, A. (1996). Does Evolutionary Plasticity Evolve? Evolution *50*, 1008–1023.

Wagner, Gunter P, et al. "Natural Selection and the Origin of Modules." Modularity, 13 Mar. 2001, doi:10.7551/mitpress/4734.003.0007.

Wagner, Günter P., et al. “The Road to Modularity.” *Nature Reviews Genetics*, vol. 8, no. 12, 2007, pp. 921–931., doi:10.1038/nrg2267.