THE NEED FOR A QUANTIFIABLE MODEL OF EVOLUTION

Elizabeth A Siewert*[1] and Thomas A Siewert[2]

[1]Retired from University of Colorado & [2]Retired from NIST

(Received on Date: 6th August 2017 Date of Acceptance: 18th September 2017)

ABSTRACT

When Charles Darwin first published his book, the origin of the species, he explained his theory by appealing to analogy. He observed that man can develop desirable traits in domestic plants and animals by selectively breeding for these traits. In a similar way, nature can develop desirable traits in plants and animals – not by intention, but by an *undirected process* of natural selection acting on random variation. Unfortunately, much of evolutionary theory today is still explained using only descriptive arguments. There have been few attempts to develop a quantitative model of this undirected process. We describe two attempts at a quantitative model – one by Sir Fred Hoyle and the other by Dr. Richard Dawkins. While most models modify a previously-existing, functional, complex structure, both of these models are attempts to quantify the ability of evolution to innovate rather than modify. We critique these two models, and present our quantitative model. Specifically, we hypothesize a primitive version of *Chlamydomonas reinhardtii*, a simple single-celled green alga, which does not yet have a functional eyespot. We then calculate the probability of developing an eyespot that is functional enough to confer an evolutionary benefit assuming all components are pre-existing, except a few proteins. We conclude by encouraging others to become involved in developing quantitative evolutionary models

Keywords: evolution, quantitative model, eyespot, *Chlamydomonas*

No: of References: 8

INTRODUCTION

In his book, The Origin of the Species, Charles Darwin speculated that the wide variety of beaks seen on birds of the Galápagos Islands developed their different shapes and sizes (depending upon the type of food available to them on each island) through a series of incremental changes in much the same way that desirable traits of domesticated plants and animals are developed through husbandry. He proposed that while man breeds and selects for traits that he desires in his plants and animals, the variation of traits seen in nature is due to an *undirected process of natural selection acting on random variation*. He further extended this undirected process of developing variation within a species to the development of the various species, themselves – thus the title of his book, The Origin of the Species. In his concluding remarks, he states, "Thus the small differences distinguishing varieties of the same species steadily tend to increase, till they equal the greater differences between species of the same genus, or even of distinct genera." (Darwin, 1859)

It's important to understand that in 1859 when Darwin published his theory, the cell was thought to be a fairly simple bag of protoplasm; its internal structure was not discovered until the advent of the electron microscope in the 20th century. The mechanism for inheritance was also unknown until the structure of DNA was discovered in 1953. In the early 1800s, Lamarck had proposed that acquired traits (e.g., being able to run faster if you're a wolf or, for that matter, a deer) might be inherited by some unknown mechanism – thus allowing a continuous gradual change from one generation to the next. Darwin praised Lamarck in his Origin of the Species, saying "Lamarck was the first man whose conclusions on the subject excited much attention" and calling him a "justly celebrated naturalist." (Darwin, 1859)

So, considering the state of scientific knowledge at that time, Darwin could only argue intuitively from what he knew. He saw the adaptations in organisms that favored survival (eg., green leaf-eating insects, mottled-grey bark-feeding insects), he observed the obvious relationships among the species based on gross anatomies, and he understood the practice of selective breeding to obtain desirable characteristics in domestic plants and animals. On these observations, he proposed his theory of the origin of the species through an undirected process of natural selection acting on random, incremental (and at that time, not understood) variation among individuals.

Today, we know much more about the complex structure of the cell, the source of variation among individuals, and the genetic relationship among the species. However, Darwin's theory is still mostly presented using only descriptive arguments. Although this gives some intuitive understanding of his theory, it lacks

rigor. What is needed is a quantifiable model to provide mathematical precision to his theory. There have been multiple attempts to quantify the probability that complex structures can develop via a gradual process of random variation and selection. We provide two of these attempts – a very large-scale one from the astronomer, Sir Fred Hoyle, and a small-scale one from evolutionary biologist, Dr. Richard Dawkins. While most models modify a previously-existing, functional, complex structure, both of these models are attempts to quantify the ability of evolution to innovate rather than modify.

Fred Hoyle, along with many scientists of the mid 1900s, was a proponent of the Steady State Theory, proposed by Sir Isaac Newton in 1687, which postulated an infinite age of the universe. He is credited with giving the Big Bang theory its name as a pejorative because he objected to the implication that a finite universe put too much restriction on the ability of life to occur by chance. In his book, Evolution from Space, he states, "Life cannot have had a random beginning … The trouble is that there are about two thousand enzymes, and the chance of obtaining them all in a random trial is only one part in $10^{40,000}$, an outrageously small probability that could not be faced even if the whole universe consisted of organic soup." (Hoyle and Wickramasinghe , 1981)

Hoyle's model dramatically illustrates the "outrageously small" probability of the "origin of life by chance" problem. However, evolutionary theory states that complexity is built up slowly step by incremental step–not all at once and suddenly. This model, although quantifiable, is too complex to represent a single, incremental evolutionary step.

Richard Dawkins, on the other hand, has presented a much simpler model. In his book, The Blind Watchmaker, he proposed a model of developing the phrase, "METHINKS IT IS LIKE A WEASEL" in incremental steps. (Dawkins, 2006) In this case, at each iteration he randomly generated letters and spaces for each of the 28 positions, kept the letters and spaces that matched the target phrase, and regenerated the letters and spaces that did not match the desired phrase until all matched. He calculated that if he had tried to generate this phrase in a single step, the probability would be one chance in 27^{28} or 10^{40} – an extremely low probability. However, by keeping any letters and/or spaces that agreed with the target phrase, "METHINKS IT IS LIKE A WEASEL" at each step, he showed that he could reach the 100% match in only a small number of iterations.

Generation 01: WDLTMNLT DTJBKWIRZREZLMQCO P
Generation 02: WDLTMNLT DTJBSWIRZREZLMQCO P
Generation 10: MDLDMNLS ITJISWHRZREZ MECS P
Generation 20: MELDINLS IT ISWPRKE Z WECSEL
Generation 30: METHINGS IT ISWLIKE B WECSEL
Generation 40: METHINKS IT IS LIKE I WEASEL
Generation 43: METHINKS IT IS LIKE A WEASEL

While this example illustrates the process of achieving complexity via random, incremental steps, it does not correctly model an *undirected* process. Dawkins had a predefined goal and retained letters until they all matched that defined goal. He knew, as the programmer, what phrase he wanted to reach and defined each incremental "success" with respect to that predefined goal. In other words, in this model, Dawkin's blind watchmaker had the resulting phrase firmly in sight.

Methods and Results

Neither of the two previous models fairly represents Darwin's theory of evolution. While Hoyle's model is much too complex and thus fails to be an incremental step in evolution, Dawkins's model pre-defines the final goal, and thus fails to be undirected. We need an example that is simple enough to depict the small *incremental* steps of evolution, and undefined enough to model an *undirected* process of evolution. One possibility would be to improve Dawkin's "METHINKS IT IS LIKE A WEASEL" program by restricting each success to be independent of the final goal. For example, one approach might be to define success as matching any word in the dictionary, and possibly including grammar and syntax rules to determine whether a coherent sentence or phrase could be formed.

A. Proposed Quantifiable Model of Darwin's Theory

With today's knowledge of molecular biology, and the development of mathematical modeling of biological systems, it is no longer justifiable to present Darwin's theory using only descriptive arguments. This is an important issue and should be modeled quantitatively. In this paper, we propose a biological model for one incremental, undirected, random step in evolution that is quantifiable.

Specifically, we propose to quantify the occurrence of one small component of a primitive eyespot assuming all other components already exist. We chose the eye, as it seems to be a popular example by both the critics and the proponents of Darwin's theory. In fact, Darwin, himself, presented the eye as a testable example

of his theory. He stated, "If numerous gradations from a simple and imperfect eye to one complex and perfect can be shown to exist … then the difficulty of believing that a perfect and complex eye could be formed by natural selection, though insuperable by our imagination, should not be considered as subversive of the theory."(Darwin, 1859)

Critics of Darwin cite the human eye as an example of a complex structure to illustrate the falsehood of Darwin's theory. However, these critics make the same mistake that Hoyle made in that the fully developed human eye is too complex. In contrast, proponents of Darwin respond by stating that the human eye didn't happen all at once. It developed slowly over a great many years, one incremental step at a time, with each step giving the organism a slight survival advantage. They hypothesize that the earliest eye might have been just a patch of light-sensitive cells on the skin. Then, through random mutations, a depression in the light-sensitive spot developed and slowly evolved into a retina, and over more time, perhaps a lens developed. (Nilsson and Pelger, 1994) These, however, are still only descriptive rather than quantitative models.

B. Proposal of the Eyespot as an Incremental Step in Evolution

We propose to use this hypothesized patch of light-sensitive cells as our model. Specifically, we propose a primitive version of the eyespot of *Chlamydomonas reinhardtii* which is a motile, single-celled green alga. This organism is a facultative heterotroph meaning that it can make its own food (glucose) using its chloroplast and sunlight as an energy source, or it can use acetate as a food, if it's in the dark. So, evolutionarily, it can survive while a functional eyespot develops, which would furnish a survival benefit over those that do not have an eyespot. The modern eyespot of *Chlamydomonas* is located in the chloroplast membrane and consists of lipid (fat-like) globules and proteins arranged in highly ordered layers which can initiate a signaling cascade in response to light. This signal is transmitted to the flagella (the means of motility for the organism) so that it can maintain the proper light environment for the organism.

The question we asked is, "How difficult would it be for a *Chlamydomonas* without a functional eyespot to develop a primitive eyespot that is functional enough to confer an evolutionary benefit? "The organism would need to solve several problems to gain this minimum functionality. It would need to be able to detect the presence of light (the eyespot), the ability to swim toward that light (the flagella), and a signaling system from the primitive eyespot to the flagella. Otherwise, given that it could swim, and could detect where it wanted to swim, if there are no means of communication between the eyespot and the flagella, there is not a functional advantage.

C. Assumptions for Calculation

In order to model an incremental step, we will make some simplifying assumptions in calculating the difficulty of developing a minimally functional eyespot.

- The *Chlamydomonas* already has the ability to swim toward the light (if it could detect that light) and the signaling system from the (yet non-functioning) eyespot to the flagella.
- All other components for the patch of light-sensitive cells exist. Only 10 more proteins are needed for the eyespot to be functional. (Proteomic analysis identified 202 different proteins from the modern day eyespot of *Chlamydomonas* (Schmidt, et al ., 2006), so this is a reasonable assumption to model an incremental step.)
- The chloroplast is already functioning – presumably being used when the 'blind' *Chlamydomonas* happens to be in sunlight.
- All the amino acids are available for use and are of the same optical isomerization (the 'left-handed' type). Since proteins made out of a mixture of optical isomerizations may not fold properly, we assume we only have the "correct" type of isomer.
- A mechanism for building proteins out of the amino acids is available, i.e., gathering the amino acids and forming the peptide bonds.
- The amino acids are completely interchangeable within each of the four broad types of amino acids (basic, acidic, neutral and hydrophobic, neutral and hydrophilic). *At times*, one amino acid might be able to substitute for another amino acid without affecting the folding properties or other characteristics of the protein. In our model, we make the assumption that this is *always* true. In other words, rather than building the proteins out of the 20 amino acids common in today's proteins, we will assume that we only need to choose from 4 generic amino acids.
- Each of the 10 proteins is 100 amino acids long. (The photoreceptor proteins in the modern day *Chlamydomonas* eyespot are ~700 amino acids long (Sineshchekov, et al.,2002), so this, also, is a reasonable assumption.)

D. Probability Calculations

It is important to remember that we are not calculating the number of ways to make a patch of light-sensitive cells. Our assumptions allow for already existing components and capabilities for the eyespot to be functional enough to provide some survival advantage – if only it had these 10 more proteins. We are also not calculating the number of ways to make 10 specific proteins. Our only requirement is that these proteins have a more or less similar conformation as the specific (larger) proteins so that it can have *some* minimal light-detecting capability. Our assumptions also allow for

substitutions of amino acids as long as they are within one of the four general types of amino acids. Therefore, we submit that these assumptions model an undirected, incremental step in evolution.

There are 4^{100} ways to make one 100-unit-long protein using 4 types of amino acids. But, this one protein has no value by itself, as we need all 10 proteins to make this primitive eyespot functional. Therefore, we need to determine how many ways we can make 10 proteins. To do this we raise the number of ways to make one protein to the 10th power. Thus, there are $(4^{100})^{10}$ or 10^{602} possible ways to make these 10 proteins. So, the probability to make 10 proteins that have some light-detecting capability is 10^{-602}.

E. Resources

This calculation, however, is too conservative because it doesn't take into account the probabilistic resources. It is only the probability to make the 10 proteins given *one* chance. But, evolution offers lots of chances for this to occur. How many chances could there be if many reactions were occurring in parallel?

Since the *Chlamydomonas* lives in water, it is logical to calculate the probability of this eyespot developing by chance somewhere in the oceans of the world. The earth contains about 1.3 billion cubic kilometers of water. We can convert this volume of water to molecules of water by using Avogadro's number (6.022×10^{23} molecules/mole) and the definition of one mole (18 ml of water = 1 mole) to give 4.3×10^{46} molecules of water in the ocean. We will assume that 10% of these molecules in the oceans are amino acids instead of water which results in 4.3×10^{45} ($\sim 10^{46}$) amino acids.

Next, we assume each amino acid and/or polypeptide can interact with every other one at the molecular vibration rate of 10^{14} Hz – the highest vibration frequency for a diatomic molecule – that of the hydrogen molecule. This gives $10^{46} \times 10^{14} = 10^{60}$ attempts to form the functional proteins every second. We can also give these reactions more time to occur, by incorporating the age of the earth into our resources. Although most would say that the water on the early earth was present as steam, we will assume that it was in a liquid state since the earth was formed 4.5 billion years ($\sim 10^{17}$ seconds) ago. This gives a total of $10^{46} \times 10^{14} \times 10^{17} = 10^{77}$ possible attempts. Given these resources, we can calculate the probability of making 10 proteins that have some light-detecting capability somewhere in the oceans since the formation of the earth.

$$1 - (1 - 10^{-602})^{10\wedge 77}$$

$$\sim 10^{-602} \times 10^{77}$$

$$= 10^{-525}$$

This extremely low probability of forming the appropriate proteins should cause us to re-evaluate our assumptions. One possibility proposed by some is that proteins within an already functioning

organelle in our primitive *Chlamydomonas* could be "recycled" to create a set of photosensitive proteins for the eyespot. However, this just transfers the probability of making the proteins to the probability of making the previously existing proteins, with the additional problem of modifying those existing proteins to give the capability of detecting light. So, that model would result in an even smaller probability than we already have.

There is a better way to increase the probability of the primitive eyespot. We don't need to limit the origin of life to earth. It could have originated in any number of likely planets throughout the universe. Rather than calculate the probability using the number of molecules in the ocean and the age of the earth, we will substitute the number of particles in the universe and the age of the universe since the Big Bang. The universe contains about 10^{80} particles and the age of the universe gives us an additional 10 billion years of time (for a total of 10^{18} sec) to randomly generate the 10 light-sensitive proteins. With these additional resources, we have $10^{80} \times 10^{14} \times 10^{18} = 10^{112}$ chances to produce the amino acid combinations. The probability in this case is only slightly better.

$$1 - (1 - 10^{-602})^{10\text{^}112}$$

$$\sim 10^{-602} \times 10^{112}$$

$$= 10^{-490}$$

This is still an extremely small probability to produce one piece of a primitive eyespot somewhere in the universe during the life of the universe – assuming all the other components are available and can be assembled correctly to produce a sufficiently-functioning organelle to confer a survival benefit upon which natural selection can operate. We submit this quantitative model as one incremental, undirected, random step of evolution.

Discussion

Evolution is the theory that the development of the species is best explained by a series of many incremental steps – each step an undirected process of natural selection acting on random variation. This is a useful theory in many, wide-ranging areas of biological research. Examples range from utilizing the evolutionary distances among the species to extract genomic information using sequence and expression data (Siewert and Kechris, 2013) to studying alcohol susceptibility in fruit flies to understand the genetics of alcoholism in humans. (Rodan and Rotherfluh, 2010)

Darwin explained his theory by appealing to the obvious relationship among the species and the common practice of selective breeding in domestic plant and animal husbandry. Today, most evolutionary models still are only descriptive in nature. Some use genetic sequences in much the same way that Darwin used gross anatomies to show that the species are related. Others propose a

sequence of events to describe how an organism could co-opt another already functioning organelle or proteins from other functioning organelles to acquire a novel function, but again, these are only descriptive models – not quantitative. Since our knowledge of molecular biology has dramatically increased, it is no longer justifiable to present evolutionary models using only descriptive arguments. This is an important issue and should be modeled quantitatively. There are a few quantitative evolutionary models, but so far most have been limited to methods for determining the best topology of evolutionary trees and the lengths of their branches (to show the relationship among the species), simulations of speciation assuming an original species, and calculations for the time it takes one functioning structure to evolve into another functioning structure.

Rather than modeling the modification of a previously functioning organelle to make a better functioning organelle, we attempted to quantitate how likely it is for evolution to innovate rather than modify a previously-existing, functional, complex structure where only a few more proteins are required. We proposed a model assuming all components, except one, were pre-existing, which needed to occur to obtain a minimally functioning organelle. Both Hoyle and Dawkins attempted to answer this question. However, Hoyle's model was not an incremental step. And while Dawkins's model shows the power of an incremental process to innovate a novel function rather than modifying a previously existing function, it was not an undirected search. Ours, on the other hand, is an undirected search because our simplifying assumptions 1) do not require specific proteins and 2) require only a minimal length of the proteins. It is also incremental because we have assumed all necessary components to make a minimally-functioning eyespot, except for a few proteins.

Although we used a primitive eyespot based on *Chlamydomonas* structure, nothing in our calculations requires this. Our calculations are applicable to any novel proteins; and, even making reasonable assumptions, it would seem that the event of even a few small proteins occurring by chance somewhere in the universe at some time during the life of the universe is nearly impossible. These results are rather unsatisfying and we can empathize with Fred Hoyle's rejection of the Big Bang Theory because of the "outrageously small probability." Nonetheless, this is our attempt to develop a quantifiable, biological model of an undirected, random, incremental step in evolution. Note that this is not a model of evolution in general, but a model for evolution to innovate rather than to modify. Our hope is that this paper would encourage others to become involved in either modifying this model or developing other quantitative models of an incremental, undirected, random evolutionary step. There obviously needs to be more work done in this area.

REFERENCES

Darwin, C. (1859): The Origin of the Species by means of Natural Selection or, The Preservation of Favoured Races in the Struggle for Life; 6th ed. Project Gutenburg.

Dawkins, R. (2006): The Blind Watchmaker. 5thed. UK: Penguin.

Hoyle, F. and Wickramasinghe, N. C. (1981): Evolution from Space; London: J.M. Dent & Sons.

Nilsson, D-E. and Pelger, S. (1994): A pessimistic estimate of the time required for an eye to evolve. *Proceedings: Biological Sciences*, **256**:pp. 53-58.

Rodan, A. R., and Rotherfluh, A. (2010): The genetics of behavioral alcohol responses in drosophila. *Int Rev Neurobiol*, **91**:pp. 25-51.

Schmidt, M., Geßner, G., Luff, M., Heiland, I., Wagner, V., Kaminski, M., Geimer, S., Eitzinger, N., Reißenweber, T., Voytsekh, O., Fiedler, M., Mittag, M., and Kreimer, G. (2006): Proteomic analysis of the eyespot of *Chlamydomonas reinhardtii* provides novel insights into its component and tactic movements. *The Plant Cell,* **18**:pp. 1908-1930.

Siewert, E. and Kechris, K. J. (2013): Modeling considerations for using expression data from multiple species. *Statist Med*, **32**:pp. 4057-4070.

Sineshchekov, O.A., Jung, K-H., and Spudich, J. L. (2002): Two rhodopsins mediate phototaxis to low- and high-intensity light in *Chlamydomonas reinhardtii*. *PNAS*, **99**(13):pp. 8689-8694.