

Bi/GE105: Evolution  
Homework 1  
Due Date: Tuesday, January 14, 2013

“In Science, self satisfaction is death. Personal self satisfaction is the death of the scientist. Collective self satisfaction is the death of research. It is the restlessness, anxiety, dissatisfaction, agony of mind that nourish Science”. - Jacques Monod

**1. A feeling for the numbers in evolution**

The processes of evolution take place at many different scales in both space and time. The goal of this first problem is nothing more than to “play” with some of the characteristic scales associated with a broad range of processes in evolution ranging from the very small (e.g. number of mutations per cell in a bacterium after one round of replication) to the very large (e.g. how far do the Galapagos islands travel in a million years). These estimates are intended to be done using simple arithmetic of the “one-few-ten” variety (i.e. few times few is ten) and to give an order-of-magnitude picture of the phenomenon of interest. Take pride in your results and state and justify (with citations) the assumptions you make carefully and give a simple, intuitive description of how you came to your results. Please don’t report rough estimates with long lists of “significant” figures.

a) Genomes are one of the most interesting features of “living matter”. One question of interest is the extent to which the “space” of possible genomes and gene products has been explored over the history of life. In a very interesting article by Whitman *et al.* called “Prokaryotes: the unseen majority”, we learn of the vast numbers of bacteria on Earth, with the current estimate coming in at something like  $10^{30}$  bacteria. The number of viruses is even greater with the so-called “virus to bacterium ratio” having a value of roughly 10, implying something on the order of  $10^{31}$  phages on earth. If we assume that over more than 3 billion years, these viruses have been steadily replicating in their cycle of infection and lysis, how many total genomes have there been in the history of life? (Obviously, this is a very rough estimate). Now, compare this number to the number of *possible* viral genomes, assuming that each viral genome is 50,000 bp in length. What

does this estimate tell you about the extent to which sequence space has been explored? Note: do the approximations, errors and uncertainties in our estimates have any bearing on our conclusions here?

b) In a 2001 Bioengineering seminar, Professor Frances Arnold made a startling remark regarding the astronomical number of possible protein sequences. In this problem we would like you to generate some intuition for the **HUGE, ASTRONOMICAL** number of ways of choosing amino acid sequences. To drive home this point, Prof. Arnold noted that if we consider a protein with 300 amino acids, there will be a huge number of different possible peptide sequences. Compute how many different sequences there are for a 300 amino acid protein?

While interesting, this wasn't the shocking part. Prof. Arnold's provocative remark was that if we took only one molecule of each of these different possible proteins, it would take a volume equal to five of our universes to contain all of these different *distinct* molecules. Estimate the physical size of a protein with 300 amino acids. Justify your result, but remember it is an estimate. Next, find an estimate of the size of the universe and figure out whether Prof. Arnold made a reasonable statement. Was her estimate of the volume of all of these possible proteins too large or too small?

c) Mutations are thought to be one of the main genomic ingredients of evolution. Given that the mutation rate is of order  $10^{-9}$  per bp per replication, how many single base pair mutations do you expect to see in a 5 mL tube of bacteria that is saturated after an overnight culture? First, given a roughly 20 minute doubling time, figure out how many cells you expect in such a culture after 12 hours given that you started with only a single cell. Then, use the replication error rate quoted above, and make an estimate of how many times each possible point mutation in the bacterial genome will be found in that culture.

d) One of the reasons that evolution is hard to think about is because of the vast times involved in evolutionary processes. One of the more interesting things we have to remember is the interplay between the dynamics of the earth and the dynamics of the living organisms on earth. For example, we will later learn about the consequences of the closing off of the Isthmus of

Panama for evolution. Similarly, the collision of India with the Asian continent brought an end to the Tethys Sea. In this part of the problem, we will think about such geological processes as they bear on some of our favorite topics from the course.

We begin by thinking about how fast the Galapagos Islands are moving. Note that these islands, like the Hawaiian Islands, are being produced by a “hotspot” that is near the current islands of Isabella and Fernandina. The islands then move in a southeasterly direction towards the coast of South America. Given that the island of Espanola is roughly  $3.5 \times 10^6$  years old, make an estimate for the mean rate at which these islands are moving to the southeast per year. Give your answer in cm/year. Also, notice that as the islands age, their height is reduced with the new islands of Fernandina and Isabella with volcanoes over 1500 m in height while Espanola has a height of only roughly 200 m. Assuming that the speed you found for the Galapagos is typical for island chains, make an estimate of the age of the island Kauai using the same kind of logic. What factors might complicate this comparison, and how would they change this estimate (qualitatively)?

e) One of the great controversies in the history of the development of our understanding of evolution had to do with the question of whether or not the Earth was old enough to accommodate the “slowness” of evolution. However, it is not at all clear how people knew how to assign any numbers to the debate. Lord Kelvin was able to make an estimate for the age of the earth and found an answer in the millions of years which was claimed to be too short. Let’s examine the timing question by making some estimates about one of the key case studies from this course which is the evolution of whales. For simplicity, assume that at the time of the extinction of the dinosaurs (65 Ma), the mammalian ancestor of whales was  $\approx 10$  cm in length. Using *Rodhocetus* and *Basilosaurus* as examples, figure out how much change in length there was per generation in the overall body plan in going from the post-dinosaur ancestor to these early whales. The logic of your estimate should involve figuring out when these whale ancestors lived, how big they were and estimating the typical generation time. It is not entirely clear that this estimate provides any insight into how whales actually evolved, but the numbers provide an interesting sense of how little it would take on a generation by generation basis to result in enormous structural changes over geological time

scales. Also, it raises the question again of how and why scientists felt that the time scale proposed by Kelvin for the age of the earth was “too short” for evolution to have produced the world we see.

## **2. The place of evolution in biology**

In the title to a famous article, Theodosius Dobzhansky noted that “Nothing in Biology Makes Sense Except in the Light of Evolution”. In class we discussed a variety of examples of how Darwinian thinking and evolutionary principles influence other areas of the life sciences (e.g. agriculture, medicine, biotechnology). Research and describe a modern day example of how evolutionary theory has been incorporated into other biological disciplines. Explain the topic/problem and your thoughts on how evolutionary principles/thinking were applied. Your brief essay should be 1-2 paragraphs. Please submit by email in pdf form to Profs. Orphan, Phillips and the TAs.