

**BE/Bi 101: Order-of-Magnitude Biology**  
**Homework 3**  
**Due date: Friday, January 30, 2015**

“Leave it as it is. You cannot improve on it. The ages have been at work on it, and man can only mar it.”

—Theodore Roosevelt speaking of the Grand Canyon

**1. Most abundant protein on Earth.**

In class we mentioned that there are often conflicting views on the most abundant proteins on Earth. Currently, the best guesses that we have been able to come up with are histones, collagen and RuBisCO. In this problem, you will try to come up with estimates for each of these protein counts to draw your own conclusions about protein abundance.

- a) We'll start with estimates of histones. Recall that histones are the proteins that make up nucleosomes, the architectural units that are responsible for wrapping up the DNA in eukaryotic nuclei. To be concrete, how long is the human genome in meters? If 147 bp are wrapped around each histone octamer to make up a nucleosome, and the spacing between nucleosomes is between 25 and 100 bp, how many nucleosomes (and histone proteins) are there for each of our cells? How many cells do we have and hence, how many histones are there per human? Given that all eukaryotes use these architectural proteins, you will have to figure out a strategy to estimate the total number of eukaryotes and thereby, the amount of histone protein. Estimate the number of copies and total mass of histone proteins on Earth.
- b) Collagen is one of the key protein components of the extracellular matrix, the set of molecules that are secreted by cells and are known to provide mechanical support. Estimate the number of copies of collagen proteins on Earth, or alternatively, the total mass of collagen. Write a few sentences about what role collagen plays in animals and then give your rationale for your estimate about the total amount of collagen protein based on the fraction of animals that is extracellular matrix and how collagen fits into that picture.
- c) We are going to do the estimate of the amount of RuBisCO using a method may seem relatively tortuous, but will give us a chance to think about the relationship between photosynthesis and the atmosphere. We begin by trying to figure out how much CO<sub>2</sub> is used in photosynthesis over the course of a year by looking at the Keeling curve (see Figure 1). Begin by estimating the mass of the atmosphere using

$$M_{\text{atmosphere}}g = p_{\text{atmosphere}}A_{\text{earth}}. \tag{1}$$

Explain why this equation is an appropriate way to estimate the atmospheric mass. Next, using the average mass of the molecules in the atmosphere, figure out how many molecules there are in the atmosphere and then using the Keeling curve, figure out how many CO<sub>2</sub> molecules there are. To determine the number of CO<sub>2</sub> molecules fixed each year

in the process of photosynthesis, we make a crude approximation based on the annual variations in the Keeling curve, which reflect an asymmetry in the photosynthetic output of the northern and southern hemispheres. In particular, we note that the variability in atmospheric  $\text{CO}_2$  is roughly ten parts per million and we make the guess that this variability is comparable to the scale of net carbon fixation itself. Work out how much total  $\text{CO}_2$  is fixed on the basis of the Keeling curve, both in number of molecules and in total mass of carbon. We can now make naive estimates about the number of RuBisCOs by using the rough figure that each such RuBisCO on average fixes one carbon per second. Given the total amount of  $\text{CO}_2$  fixed per year (based on the Keeling curve in the way described above) and the rate at which the enzyme works, how many copies of RuBisCO do you estimate there are on Earth?

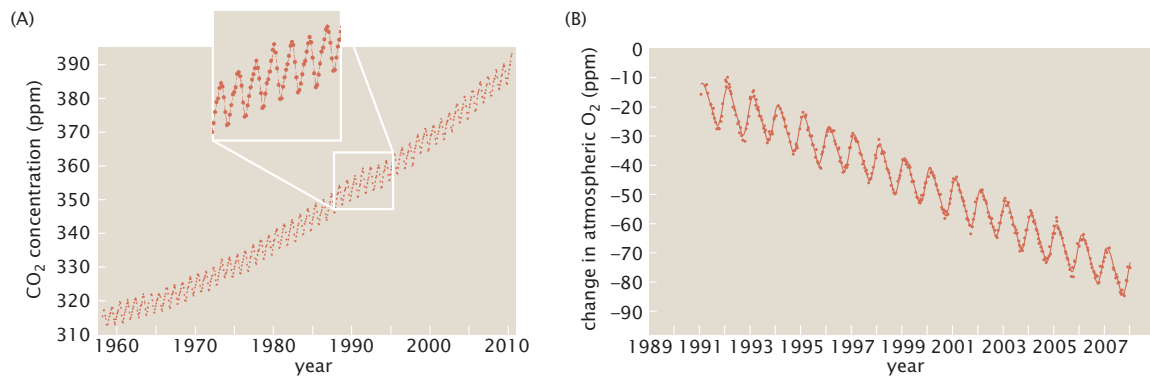


Figure 1: The Keeling curve showing atmospheric (A)  $\text{CO}_2$  concentration and (B)  $\text{O}_2$  concentration in parts per million. Adapted from *Physical Biology of the Cell* by Phillips, Kondev, Theriot and Garcia.

## 2. Estimates on biological information.

Here we continue with the theme begun in homework 1 on the astronomical (or even larger) enormity of sequence space. In homework 1, we considered the number of different possible proteins. Here, let's now think about the total number of genomes that have existed in the history of life. Using the estimates on bacterial diversity highlighted in the Whitman, Coleman and Wiebe article "Prokaryotes: The unseen majority" (which you can download [here](#)) and the virus:bacterium ratio of 10:1, have a go at estimating the total number of genomes over the history of life on Earth. Then, compare this to the total number of possible genomes. Make sure you give your arguments. We don't have a right answer for this one, but are instead trying to understand the relative importance of eukaryotic genomes, prokaryotic genomes and viral genomes in the overall diversity of genomes and also, the size of sequence space.

### 3. Evolution 1.

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, think 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 problem, we will think about such geological processes as they bear on evolution.

We begin by thinking about how fast the Galapagos Islands are moving. You should look up the relevant geographical information about these islands such as their positions and elevation. 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. Espanola is the easternmost island in the Galapagos. Yet, the divergence time for the different species of finch on the islands appears to be more than  $3.5 \times 10^6$  years. How can you make sense of this? 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)?