Bi/Ge 105: Evolution
Homework 2
Due Date: Wednesday, January 17, 2024

> "I am losing precious days. I am degenerating into a machine for making money. I am learning nothing in this trivial world... I must break away and get out into the mountains to learn the news."
>
> — John Muir

1. A feeling for the numbers in evolution, continued

The processes of evolution take place at many different scales in both space and time. Like last week, 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. the age of the Grand Canyon).

These estimates are intended to be done using simple arithmetic of the "one-few-ten" variety (i.e. few (f) 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.

Viruses: everything everywhere all at once

Genomes are one of the most interesting features of living matter. To what extent has the space of possible genomes and gene products been explored in the history of life? In *Prokaryotes: the unseen majority*, Whitman et al. give a sense for 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, since there are thought to be roughly 10 viral particles for every bacterium; this implies something on the order of 10^{31} viruses on Earth, making them the largest genomic reservoir on our planet.

Question 1a

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 viral genomes have there been in the history of life? (This can be 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?

Comment briefly on whether the approximations, errors and uncertainties in our estimates have any bearing on our conclusions here.

Astronomical proteins

Let's continue with the theme of the enormity of sequence space in biology. In a 2001 bioengineering seminar, Professor Frances Arnold made a startling remark regarding the astronomical number of possible protein sequences. The provocation? She claimed that if one were to synthesize exactly one copy of every possible 300 amino acid protein, the resulting products would occupy a volume five times greater than that of the known universe! Could that really be possible?

Question 1b

How many different 300 amino acid proteins are possible? Estimate the volume of a single one of these proteins. Now look up a value for the size of the universe. Use your calculations to check Professor Arnold's claim, and comment briefly on the result.

Mutations at the large N limit

Mutations are thought to be one of the main genomic ingredients of evolution. Let's get a sense for just how frequently they can occur in large populations.

Question 1c

Consider a 5 mL tube of *E. coli* bacterial culture that was grown overnight until saturation. Assuming we started with a single cell that doubles every 20 minutes, how many cells are in the culture after 12 hours? If the bacterial mutation rate is of order 10^{-9} per bp per replication, how many single base pair mutations do you expect to find here? Thus, how many times is each distinct possible point mutation represented? It may be helpful to know that the *E. coli* genome is 4 million bp long.

Nanoscale expeditions

The history of biology is filled with stories of great naturalist voyages: Humboldt, Darwin, Wallace, Huxley, Hooker, and others all traveled the world in search of biological discovery. But what about our exploration of the nanoscale?

Humanity's first forays into the microscopic world were made possible by (you guessed it) microscopy, which was invented in the 17th century by pioneers such as Robert Hooke and Antonie van Leeuwenhoek. Later, the development of electron microscopes — fantastically powerful tools with nanometer resolution — pushed the envelope of what we can see even further.

Question 1d

A typical electron micrograph corresponds to an area of roughly $1\mu\text{m} \times 1\mu\text{m}$. Estimate the total area we've viewed using electron microscopes in the entire history of science, especially for biological samples. How does this correspond to the area of the Earth? What do you conclude about the extent to which we have "explored" our planet at the nanoscale?

Dating Earth

A great controversy in the history of evolutionary theory was the question of whether or not the Earth was old enough to accommodate evolution. For example, Lord Kelvin argued that the Earth was only millions of years old, and that this was not enough time for the slow process of evolution to have occurred. Although we now know Kelvin's estimate was wrong, it is interesting to consider how long it might take for a new species to evolve from an existing one.

Question 1e

Rodhocetus and *Basilosaurus* are two early whale species, now extinct. By figuring out when these species lived, how big they were, and their typical generation times, make a rough estimate of how much change in body length occurred per generation as they evolved from a 10 cm-long ancestor at the time of the extinction of the dinosaurs, 65 million years ago.

Note that while this is a simple estimate (and by no means a conclusive history of the trajectory by which whales actually evolved), the numbers provide an interesting sense for how little it can take (on a generation-by-generation basis) to create big structural changes over geological time scales.

How does your answer affect your reaction to Kelvin's claim, even assuming he had been right about the age of Earth?

2. On phosphorus

Phosphorus is an element essential for all living things. Its story spans length scales ranging across nearly a dozen orders of magnitude, from the molecular building blocks of bacterial cells, to planetary-scale biogeochemical flows. The goal of this problem is to gain an appreciation for the importance of phosphorus at both extremes, and to comment briefly on some of its more surprising roles in our daily lives.

Zooming in: a brief bacterial census

In addition to the big ticket chemical elements in cells (carbon, hydrogen, nitrogen, and oxygen), phosphorus is similarly essential. Think of the ubiquity of nucleic acids, which are linked via phosphodiester bonds, and of the phosphates in ATP, the "energy currency" of the cell. Let's first delve into phosphorus at the molecular scale by considering the elemental makeup of a bacterial cell.

Much like the day-to-day work in our lab, here we will use the humble bacterium *Escherichia coli* as our model organism.[1]

Question 2a

Estimate the total number of phosphorus atoms in a typical *E. coli* cell. What fraction of its total mass does this represent?

To figure this out, we will need to consider all the biomolecular contexts in which we might find phosphorus. For example, we know that in both RNA and DNA, every nucleotide carries its own phosphate group. There are also important nucleoside triphosphates in the cell, such as ATP. Many lipids are phospholipids, with polar heads containing phosphate groups as well. At times, some proteins can be phosphorylated. And don't forget ribosomes, which are about 2/3 RNA by mass. Other potentially-relevant quantities are the length of the *E. coli* genome (4 million base pairs) and the volume of an *E. coli* cell (1 cubic micron).

As usual, the Bionumbers database may be a helpful resource, but for best results you should try to minimize looking up external quantities (and cite clearly if you do).

Zooming out: tenuous appetites

Beyond bacteria, phosphorus is also an important plant nutrient. It is a crucial component of modern fertilizers, which are typically labeled with "NPK" values reporting the relative amounts of nitrogen (as N), phosphorus (as P_2O_5), and potassium (as K_2O) they contain by weight.

[1] While you may be familiar with *E. coli* because of a few virulent strains which can sometimes cause food poisoning, the majority of strains are perfectly harmless. Instead, it deserves to be celebrated as a tireless laboratory workhorse for molecular biology, and the foundational testbed for much of our current understanding of genetics and evolution.

Question 2b

Imagine a world in which everyone gets all 2000 kcal of their daily diet from wheat (i.e. everyone is vegetarian and gluten-tolerant). How much wheat is needed to feed Earth's population?

Using the biomass P fraction you derived in the previous part, calculate how much phosphorus all the wheat contains. Hence, write a brief statement about the amount of phosphate fertilizer that might be needed to cultivate it.

Now that we have a rough sense for fertilizer demand, we turn to the natural follow-up: how do we get it? Until surprisingly recently, the answer was almost unanimously guano[2]: the accreted excrement of seabirds!

The Guano Age

Guano may seem like an unlikely source of wealth, but at the height of the global trade in the 19th century, wars were being fought for it. At the time, the global fertilizer supply chain was dominated by the export of Guano from a tiny handful of islands in the Pacific — such as the Chincha Islands off the coast of Peru — which were home to enormous colonies of nesting seabirds.

Question 2c

Consider Central Chincha Island, shown in Figure 1B. Assuming they cover the island completely, how many Guanay cormorants (Figure 2) could nest there?

Estimate the mass of guano this colony produced between the years 1200-1800 assuming they maintained a constant population.

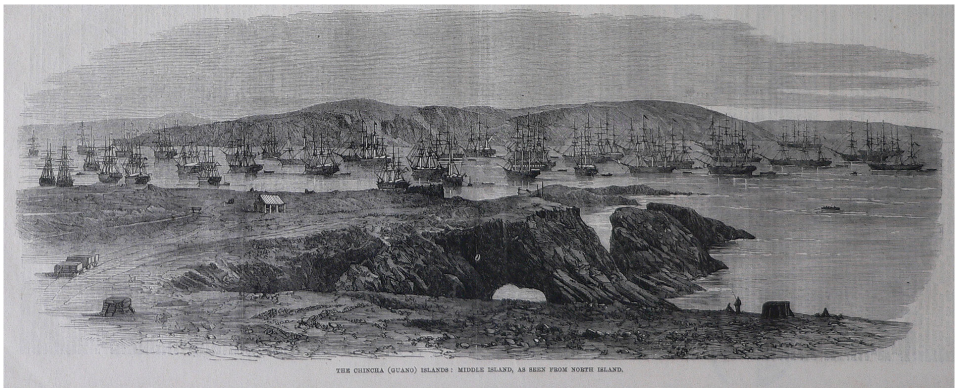
Although the world once ran on guano, today those deposits are largely exhausted, and the seabird colonies are a fraction of their former size.

[2]It is also the namesake for the nucleobase guanine (G), which was first isolated from guano!

Question 2d

At the height of the guano boom, mining was so extensive that the Chinchas' reserves, mostly intact in 1820, were largely exhausted by around 1880. Use your answer from the previous part to estimate how much guano Peru would have had to export every year to accomplish this.

Consider the scene of ships waiting to export guano shown in Figure 1A. How often would they have had to be loaded and sent back to Europe in order to sustain this export rate?



(a) View of Central Chincha Island at the height of the guano trade, circa ~1860.

(b) View of Central Chincha Island, circa 2024.

Figure 1: Past and present views of Peru's Chincha Islands. Once home to enormous seabird colonies, and prized since Incan times for their vast guano deposits, today their wealth is essentially exhausted.

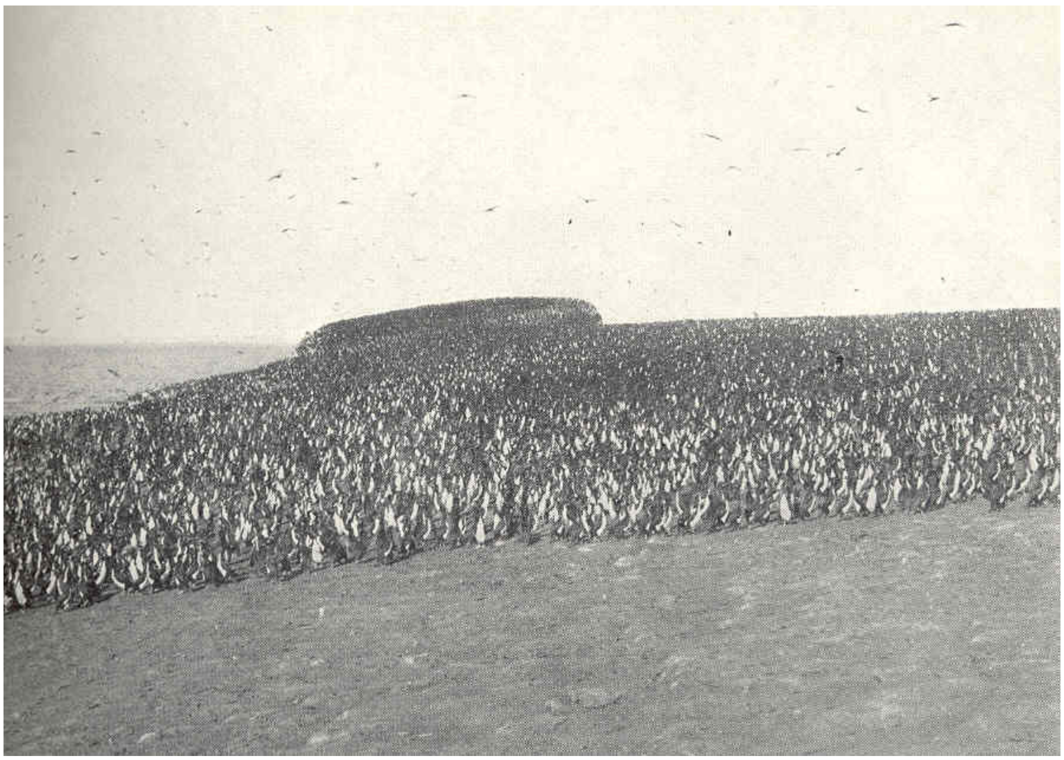


Figure 2: Guanay cormorants on the Chincha Islands in 1910. Only a portion of the whole flock is visible here. Note that even by this point, populations had already declined compared to their historical highs.