

Bi 1: The Great Ideas of Biology

Homework 3

Due Date: (2023-04-27)

1. Deep time and earth history

One of the most interesting topics in science is how we have learned to probe deep time. In this course, the subject of deep time will appear repeatedly and we will spend a lot of time examining how DNA sequence has permitted us to explore deep time in the biological setting. Of course, biology and the dynamics of the Earth are not independent phenomena and the point of this problem is to better understand the details of how scientists figure out how old the Earth is as well as how old various fossil-bearing strata are. Specifically, we will be interested in two great biological mysteries and treasures: evolution of humans in East Africa and the evolution of the fauna of the Galapagos Islands. To that end, we will first consider a simple model of the radioactive decay process for potassium-argon dating methods, recognizing that there are many other dating methods that complement the one considered here.

Potassium-Argon Dating

Potassium-argon dating is based upon the decay of ^{40}K into ^{40}Ar . To a first approximation, this method can be thought of as a simple stopwatch in which at $t = 0$ (i.e. when the rocks crystallize), the amount of ^{40}Ar is zero, since it is presumed that all of the inert argon has escaped. We can write an equation for the number of potassium nuclei at time $t + \Delta t$ as

$$N_K(t + \Delta t) = N_K(t) - (k_K \Delta t) N_K(t). \quad (1)$$

Stated simply, this means that in every small time increment Δt , every nucleus has a probability $k_K \Delta t$ of decaying, where k_K is the decay rate of ^{40}K into ^{40}Ar . We also employ the important constraint that the number of total nuclei in the system must remain constant, so that

$$N_K(0) = N_K(t) + N_{Ar}(t), \quad (2)$$

where $N_K(0)$ is the number of ^{40}K nuclei present when the rock is formed, $N_K(t)$ is the number of ^{40}K nuclei present in the rock at time t , and $N_{Ar}(t)$ is likewise the number of ^{40}Ar nuclei present in the rock at time t . In this part of the problem you will use equations 1 and 2 to construct differential equations

to find the relationship between $N_K(t)$, $N_{Ar}(t)$, and t .

Question 1a:

Using equations 1 and 2 as a guide, write differential equations for $N_K(t)$ and $N_{Ar}(t)$. How do these two expressions relate to one another?

Question 1b

Next, we note that the solution for a linear differential equation of the form $\frac{dx}{dt} = kx$ is given by $x(t) = x(0)e^{kt}$. Use this result to solve for $N_K(t)$.

Question 1c

Use the constraint encapsulated by equation 2 to write an equation for the lifetime of the rock, t , in terms of the ratio $\frac{N_{Ar}}{N_K}$.

Determining Lucy's Age

Unfortunately for us, real-world K-Ar dating data is generally not neatly presented in the form of N_{Ar} and N_K . Instead, geologists will measure a concentration of ^{40}Ar in mol/g and a weight percent of K_2O . These data must be used to identify the number of ^{40}Ar and ^{40}K nuclei in the sample. In this part of the problem, we will look at such measurements from an actual paleontological specimen as reported in Aronsen (1977) in order to determine its age.

In 1974, a fossil of *Australopithecus afarensis* was discovered in Ethiopia. This specimen shown in Figure 1, which was dubbed "Lucy," marks an important step in understanding human evolution because it was the earliest known species to show evidence of bipedal locomotion. Because Lucy was found in an area that was rich in volcanic rock, which is ideal for potassium-argon dating, potassium-argon dating was used to determine Lucy's age (Aronsen 1977). Later in the course, we will reflect further on the story of human evolution,



Figure 1: The fossil remains of Lucy, a specimen of *Australopithecus afarensis*.

but this first problem gives us a chance to think about the challenge of understanding human origins and the intimate interplay between the history of Earth and the history of life on our planet.

Question 1d

Using the table of ^{40}Ar and K_2O measurements below (Aronsen 1977) and your answer to part (1c), obtain an estimate for Lucy's age. Since each sample is taken from the area in which Lucy was found, we expect each sample to give you roughly the same answer; you will need to take the mean of the ages of each sample to obtain an estimate for Lucy's age.

Assume that each sample has a total mass of 1 g. Also, note that not all of the potassium in the sample will be the isotope ^{40}K , so you will need to use the ratio of ^{40}K to total potassium, $\frac{^{40}\text{K}}{\text{K}_{\text{total}}} \approx 1.2 \times 10^{-4}$. Additionally, use the decay constant $k_K \approx 5.8 \times 10^{-11} \text{ yr}^{-1}$.

Sample Number	$^{40}\text{Ar} \times 10^{-12} \text{ mol/g}$	wt. % K_2O
1	2.91	0.657
2	3.18	0.755
3	3.08	0.680

Age of the Galapagos Islands

One of the most important natural laboratories of evolution that remains as relevant now as it was at the time of Darwin's Voyage of the Beagle is the Galapagos Islands, roughly 1000 km off the coast of Ecuador in South America. The Galapagos Islands are known as oceanic islands because they are of volcanic origin and have never been in contact with any larger continent (as was New Zealand, for example). The islands at the western side of the archipelago are the youngest and have currently active volcanoes while those to the east are older, harbor more plant life and are of lower altitude. Because of the motion of tectonic plates, the islands are traveling in a southwesterly direction.

These islands are home to a panoply of bizarre animals and plants including equatorial penguins, the only species of marine iguanas in the world, the only species of flightless cormorant in the world, more than a dozen species of finches and the famed giant tortoises that will be the subject of this part of the problem. One of the observations made by Darwin about both the finches and the tortoises is that different islands have different species. A question

of great interest was to figure out the rate of evolution, a question all of us watched play out in real time as new variants of the SARS-CoV-2 virus worked their way through populations across the globe. In the context of both the finches and the tortoises of the Galapagos, it is of great interest to figure out how long ago they diverged from their common ancestors on the mainland of the Americas, and what the divergence time is between the different extant species on the islands now. As part of an answer to those questions, we need to figure out the age of the islands.

This part of the problem repeats the use of the methods we considered in the analysis of the age of Lucy with the goal here not being repetition, but rather, just a chance to look at the real world data and to get a sense for the age of the islands.

Question 1e

The table below contains ^{40}Ar and K_2O measurements from samples taken from three different Galapagos Islands. Use the same procedure as in part (1d) to estimate the maximum age of the Galapagos Islands.

As before, use the ratio of ^{40}K to total potassium, $\frac{^{40}\text{K}}{\text{K}_{\text{total}}} \approx 1.2 \times 10^{-4}$ and the decay constant $k_K \approx 5.8 \times 10^{-11} \text{ yr}^{-1}$. Note that, unlike in the case of Lucy's fossil, here we are given the total number of moles of ^{40}Ar instead of the number of moles per gram. Additionally, we are also given the total weight of the samples.

Island	Weight (g)	Total ^{40}Ar ($\times 10^{-11}$ mole)	wt. % K_2O
Plazas	8.50	0.2180	0.0412
Santa Fe	21.6	5.79	0.645
Española	6.31	2.79	0.986

2. The molecular clock

Thus far in this homework, we have seen how to date the rocks surrounding fossils of interest and the islands upon which the evolutionary drama is unfolding. But how do we figure out how long ago two related species diverged

from their common ancestor? The answer to that question is tied to one of the most brilliant and tricky ideas of modern science, namely, the molecular clock, which we had introduced in the previous homework. To recap, the simplest way to state the idea is to remember that all organisms harbor a genome, the set of As, Cs, Gs and Ts that make up its genetic heritage. The fundamental idea is that the more closely related the DNA sequences of two organisms are, the more closely related they are. Further, the number of differences between two genomes of interest can be used as a clock to measure how long ago they shared a common ancestor. For example, all of the finches in the Galapagos are the result of an adaptive radiation that followed an initial colonization from ancestral finches from Central America. Similarly, the tortoises of the Galapagos also all share a common South American ancestor.

We first consider the giant tortoises found on the island of Isabela, an island built up from five distinct volcanoes with an age of less than a million years. One of the most fascinating things about this island is that the different volcanoes are separated by barren, black lava fields which mean that tortoises on the separate parts of the island are reproductively isolated. As a result, they have suffered what is known as an adaptive radiation, the successive divergence of traits that occur over time due to such reproductive isolation.

In the case of the tortoises on Isabela, an early approach to examining the divergence of the populations was to use the mitochondrial DNA (mtDNA). Mitochondrial DNA mutates more quickly than genomic DNA, acting as a faster molecular clock that is useful for comparing sequences over shorter time scales or between species that are closely related. In this exercise, we will compare the mtDNA sequences of five giant Galapagos tortoise subspecies obtained by Beheregaray *et al.*. Each of these subspecies inhabits the areas surrounding one of the five volcanoes on Isabela. The sequences are around 700 nucleotides long and are sequenced from the mtDNA control region, which contains the origin of replication and regulates gene expression in the mitochondria.

Question 2a

To perform sequence comparisons, the first step is to align the sequences. That is to say, we would like to arrange the DNA sequences in such a way that the sites that are identical or similar are vertically aligned, so that we can easily identify regions where mutations have occurred.

In the file named *ggt_beheregaray.fasta*, you will find the sequences of the mtDNA control region from the five giant tortoise subspecies. Align these sequences using the Clustal Omega multiple sequence alignment program. Make sure to specify that the input sequences are from DNA and select ClustalW as the output format.

Take a look at the alignment and the Guide Tree. Which subspecies of giant Galapagos tortoises appears to have diverged the most from the other subspecies?

Question 2b

Next, we quantify the difference between each pair of sequences by counting the number of substitutions. In order to do this, let's first re-run the Clustal Omega alignment program. This time, select Pearson/FASTA as the output format as it is easier to process using Python. Download the alignment file and write a Python function to count the number of substitutions between each pair of subspecies. If there is a gap ('-') in either of the two sequences, ignore that position.

Question 2c

Evolutionary time can be estimated by taking the product between the number of substitutions and the rate of substitution. On average, the rate of substitutions in mtDNA is roughly 2% every million years. Calculate the rate of substitution between each pair of sequences by taking the number of substitutions that you counted in Question 2b and dividing it by the length of the sequence. Then estimate the divergence time between each pair of giant tortoise species and report it in a similar table as what you used in Question 2b. Compare the divergence times with the age of the Isabela islands (0.7 Myr). What do you notice?

One hypothesis that explains the observation above is that there are two separate colonization events that happened around volcano Wolf and the other volcanoes on the Isabela Island. It is of interest to us to find the closest relative of the Galapagos tortoises in order to determine which species of tortoises from which part of the world colonized the Isabela Island.

Here, we will examine another set of sequences from the 16S ribosomal and *cytb* mtDNA, which have a length of around 970 nucleotides. We will look at the sequences from Volcano Wolf and Volcano Darwin, as well as the sequences from three extant species of tortoises from South America, namely the yellow-footed tortoise *Geochelone denticulata* from Brazil, the red-footed tortoise *Geochelone carbonaria* from northern South America, and the tortoise *Geochelone chilensis* from Chaco, Bolivia.

Question 2d

In the file named *ggt_caccone.fasta*, you will find the sequences of 16S ribosomal and *cytb* mtDNA obtained by Caccone *et al.*. Repeat the analysis performed in Question 2a and 2b to align the mtDNA sequences and count the number of substitutions between each pair of species. Which species is most closely related to the subspecies at Volcano Wolf and the subspecies at Volcano Darwin? Did the same tortoise species colonize both areas of the Isabela island?