Grant Messner

Bi 1 Homework 1

1a) Assume that

- Millikan has about 10 floors
- Each floor has an area of about 1000 square meters
- Each shelf occupies about 1 square meter
- Each shelf contains about 100 books
- Each book contains between 100 and 1000 pages
  - Applying the square root rule gives us $\sqrt{100 * 1000} \approx 300$ pages
- Each page contains about 100 words
- Each word contains between 1 and 10 characters
  - Applying the square root rule again gives $\sqrt{1 * 10} \approx 3$ characters
- Each character contains 1 byte

Thus, we can multiply this all out and obtain $10^1 * 10^3 * 10^2 * 3 * 10^2 * 10^2 * 3 * 1 \approx$ $\mathbf{10^{11}}$ **bytes** of data in Millikan

For the *E. Coli* question, assume that

- The *E. Coli* genome contains $5 * 10^6$ base pairs
- Each base pair contains $3 * 10^{-1}$ bytes of data

Thus, we have $5 * 3 * 10^6 * 10^{-1} \approx$ $\mathbf{10^6}$ **bytes** of data in the *E. Coli* genome

For the human genome question, assume that

- The human genome contains $7 * 10^9$ base pairs

- Each base pair contains $3 * 10^{-1}$ bytes of data

Thus, we have $7 * 3 * 10^9 * 10^{-1} \approx \mathbf{2 * 10^9}$ **bytes** of data in the human genome

1b) For the influenza information density question, assume that

- The influenza genome contains 10^4 base pairs
- Each base pair contains $3 * 10^{-1}$ bytes of information
- The virus is a cube of side length 10^2 nm
    - This gives us a volume of $(10^2 * 10^{-9}\text{ m})^3 = 10^{-21}\text{m}^3$

Thus, the total amount of information is $10^4 * 3 * 10^{-1} = 3 * 10^3$ bytes, and the information density is thus $3 * 10^3$ bytes$/10^{-21}\text{m}^3$, which gives us information density of $\mathbf{3 * 10^{24}\ bytes/m^3}$

For the computer drive question, assume that

- We are dealing with a rectangular prism flash drive of dimensions $1\text{ cm} \times 1\text{ cm} \times 1\text{ mm}$
    - This gives us a volume of $10^{-2}\text{ m} \times 10^{-2}\text{ m} \times 10^{-3}\text{ m} = 10^{-7}\text{m}^3$
- Assume that this flash drive holds around 10 GB of information

Thus, the information density is 10^9 bytes$/10^{-7}\text{m}^3$, which gives us an information density of $\mathbf{10^{16}\ bytes/m^3}$

Therefore, the virus can store $\mathbf{3 * 10^8}$ times as much information per unit volume as a flash drive can!

1c) Assume that

- The SRA contains about 10^{16} bases
- Each letter is equivalent to 1 SRA base
- Each word in the Harry Potter books has 10 letters
- Each line of the book has about 10 words
- Each page of the book has 10 lines
- Each book has 10^3 pages
- There are 7 total books

Thus, the total number of "bases" in the entire Harry Potter series is $10 * 10 * 10 * 10^3 * 7 = 7 * 10^6$. Thus, the number of copies that could be written is $\frac{10^{16}}{7*10^6} \approx \mathbf{10^9}$ **copies**.

This is $\frac{10^9}{7*10^7} \approx 10$ times as many copies as there are copies of the whole series ever sold!

2a) The top three matches (for all three Sao Tome frogs) are

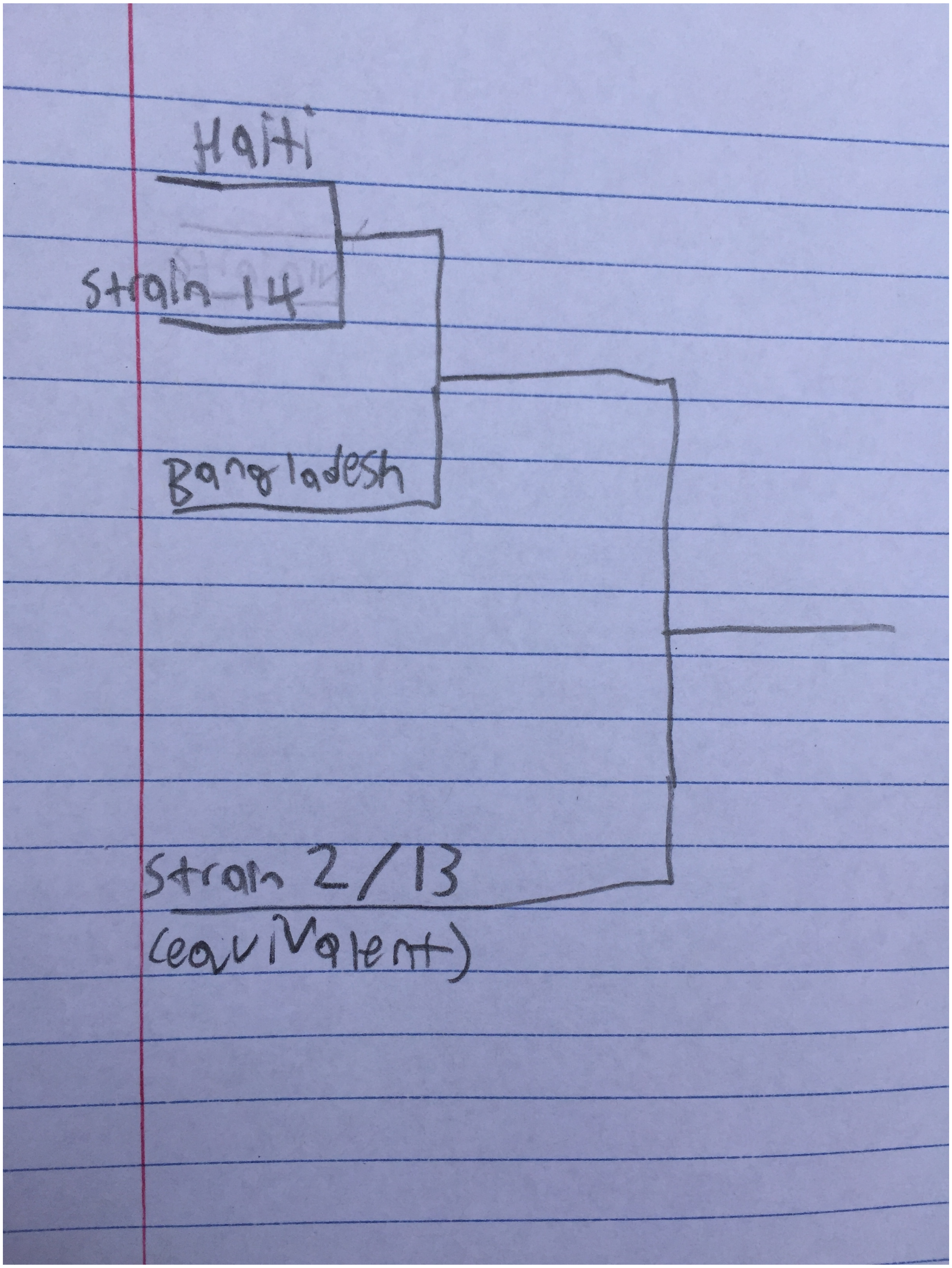
1. Ptychadena nilotica, Tanzania, Kibebe farm
2. Ptychadena nilotica, Uganda, Lake Victoria
3. Ptychadena nilotica, Kenya, Mt Kenya

2b) Assume that

- The frogs likely originated from some part of East Africa (likely either Tanzania, Uganda, or Kenya)
- Most of the entire continent of Africa is flooded in water
- One tree falls into the water every 100 years (i.e., 10^{-2} trees per year)
- One frog is on every 10^{th} tree (i.e., 10^{-1} frogs per tree)
- One in every 100 frogs survive the journey to Sao Tome (i.e., 10^{-2} frogs per journey)
- This all occurred over about 10^{7} years

Thus, we end up with $10^{-1} * 10^{-2} * 10^{-2} = 10^{-5}$ frogs per year, and over the course of 10^{7} years, about $10^{-5} * 10^{7} =$ **10^{2} frogs** will survive the journey.

3a)



3b) Compare the Haiti strain to strain 14 from Nepal. Assume that

- There are 2 SNP's per genome between the two strains
- The size of the genome is about $4 * 10^6$ base pairs
- Strains were collected 3 months apart

Thus, the amount of time that has elapsed is $3 * 30 * 24 * 60 \approx 40{,}000$ minutes. If the bacteria divide every 40 minutes, the number of generations that have passed is $\frac{40{,}000}{40} = 10^3$ generations. This gives us a ratio of 2 SNP's per genome / $4 * 10^6$ base pairs $\approx 5 * 10^{-7}$ SNP's per genome per base pair, and because this occurs over 10^3 generations, we have $5 * \frac{10^{-7}}{10^3} = \mathbf{5 * 10^{-10}}$ **SNP's per genome per base pair per generation.**

4a) The ordering is as follows:

1. Gondwanaland split into two landmasses, one of which would form New Zealand/Caledonia and another that would form Australia, West Antarctica, New Guinea, and Southern South America (80 MYA)
2. New Zealand and New Caledonia split (70 MYA)
3. The landmass that will form Australia, West Antarctica, New Guinea, and Southern South America split into two new landmasses, one which would form Southern South America and West Antarctica, and another that would form Australia and New Guinea (50 MYA)
4. Southern South America and West Antarctica split (30 MYA)
5. Australia and New Guinea split (10 MYA)

4b) Assume that

- The landmasses containing Australia and New Zealand diverged about $8 * 10^7$ years ago
- The mutation rate for these plants is 7 mutations per billion base pairs per generation
- The generation time is 20 years
- The genome length is $5 * 10^8$ base pairs, or equivalently $5 * 10^{-1}$ billion base pairs

The number of generations is $\frac{8*10^7}{2*10} = 4 * 10^6$ generations. Thus, the total number of mutations is $7 * 5 * 10^{-1} * 4 * 10^6 \approx \mathbf{10^7}$ **mutations**

4c) From part 4b, we calculated that about 10^7 mutations occurred since the landmasses split. Since this is very consistent with the actual number of $9 * 10^6$ mutations, we can conclude that the landmasses did in fact diverge about $8 * 10^7$ years ago. This provides evidence for the vicariance hypothesis, as it makes correct predictions in this case.