Bi 1x
The Luria-Delbrück fluctuation test

1 Overview

In this module, we will explore one of the most clever and intriguing biology experiments performed in the last century. In 1943, Salvador Luria and Max Delbrück showed how two different theories of evolution, a unifying theme in biology, could be distinguished through a potent combination of theory and experiment. The lab focuses on computing the mutation rates of microorganisms and examines this result in light of experimental findings. After many decades, this experiment is still used in molecular biology labs.

2 Background

When mutations occur in nature they are often deleterious to an organism, since it is easier to break a complex system refined by hundreds of years of evolution than to serendipitously land upon a beneficial change. At the same time, mutations are a critical part of the genetic heritage of living organisms, arising in every type of creature and allowing life to evolve and adapt to new environments. In 1943, the question of how microorganisms acquire mutations was described in a famous article by Luria and Delbrück (S. E. Luria and M. Delbrück, *Genetics*, **28**, 491–511, 1943). At the time, there were two prominent theories of genetic inheritance. Scientists did not know if mutations arose randomly in the absence of an environmental cue, the "random mutation hypothesis," or whether they occur as an adaptive response to an environmental stimulus, the "adaptive immunity hypothesis." See Fig. 1.

To test these two hypotheses, Luria and Delbrück grew many parallel cultures of bacteria and then plated each culture on agar containing viruses known as phages (which infect and kill nearly all of the bacteria). Although most bacteria are unable to survive in the presence of phages, often mutations could enable a few survivors to give rise to resistant mutant colonies. If the adaptive immunity hypothesis is correct, mutations occur only after bacteria come in contact with phages, thus only after plating the bacteria on phage-agar plates. Under this hypothesis, we would expect a low variance in the number of resistant colonies that appear on each plate.

However, if the mutations arose randomly prior to phage exposure as bacteria were growing in the liquid culture, the number of mutations in each culture would vary wildly, as mutations could occur at any time during the liquid culture phase and accumulate exponentially. Mutations that arise early in the culture will give rise to an exponentially growing population of mutant cells, which will result in large number of resistant colonies after plating. In contrast, mutations that occur at later times will result in fewer colony counts after plating. Hence, the random mutation hypothesis

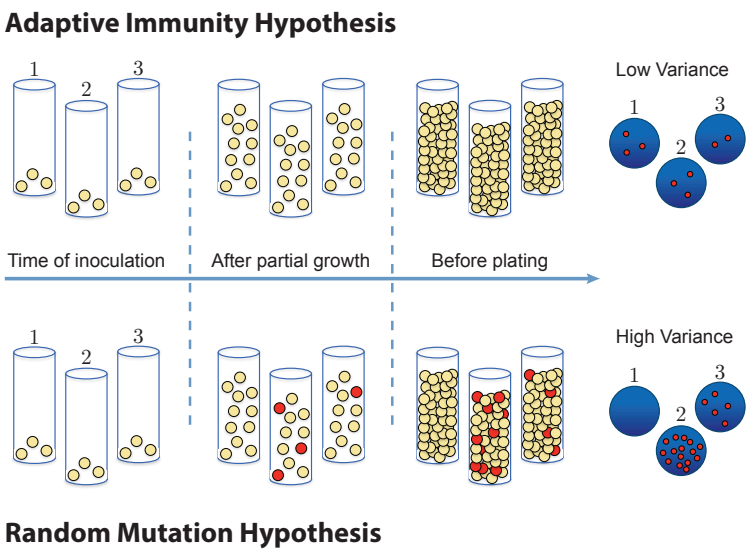

Figure 1: A schematic for the Luria-Delbrück fluctuation experiment. For the "adaptive immunity hypothesis," we expect to see a low variance in the number of cells immune to the selection factor per trial. The "random mutation hypothesis," on the other hand, predicts a high variance in the number of immune cells per trial.

makes the prediction that there will be a large variance in the number of resistant colonies coming from different cultures.

Your task is to test the two hypotheses of inheritance using cultures of the yeast *S. cerevisiae*. Instead of using phage, we will expose the yeast to canavanine, a toxin. Using the mean and variance obtained from number of mutant colonies in each culture, you will be able to deduce which one of the two hypotheses more accurately describes the mechanism underlying the rise of mutations. Additionally, we will calculate the mutation rates in the wild type and a mutator strain. The mutator strain contains a deletion in one of its DNA repair genes. The logic is that the mutator strain should have a higher mutation rate and hence should result in more colonies reflecting a higher proportion of resistant cells.

2.1 Theoretical predictions for the Luria-Delbrück experiments

How can we determine which hypothesis is correct from a Luria-Delbrück experiment? As we now show, the two hypotheses predict significantly different statistics so that by comparing theory with experiment we can deduce whether the random mutation hypothesis or the adaptive immunity hypothesis is more likely.

2.2 A toy problem

Before doing a more involved calculation, it is generally a good idea to test your intuition with toy problems. Let us illustrate this approach with an example. The beauty of this approach is that using just three generations, we can see how statistical metrics of mutations vary between the two hypotheses.

Fig. 2 shows the trajectories from a single wild type cell in which mutations may occur in every generation. The probability of a mutation occurring for a given cell division is a. We can sketch out all possible trajectories for the random mutation hypothesis and the adaptive immunity hypothesis. Note that the adaptive immunity hypothesis does not allow for the trajectory in which a cell is mutated prior to exposure to toxin.

For the random mutation hypothesis, the expected number of mutated cells present at time $t = 1$, is calculated as

$$M_{\text{mut}} = \frac{0 \cdot 1 + 2 \cdot a + 1 \cdot a + 1 \cdot a + 2 \cdot a^2}{1 + a + a + a + a^2} = \frac{4a + 2a^2}{1 + 3a + a^2} \approx 4a, \tag{1}$$

were we have used the fact that $a \ll 1$. In prelab problem 1, you will compute the mean for the adaptive immunity hypothesis, and the variance for both hypotheses. From there, you can compute the **Fano factor**, the ratio of the variance to the mean (expected number of mutated cells), as a metric of the expected variation in the number of mutants. The Fano factor is decidedly different for the two hypotheses, and can therefore be used to determine which hypothesis is more likely to be true.

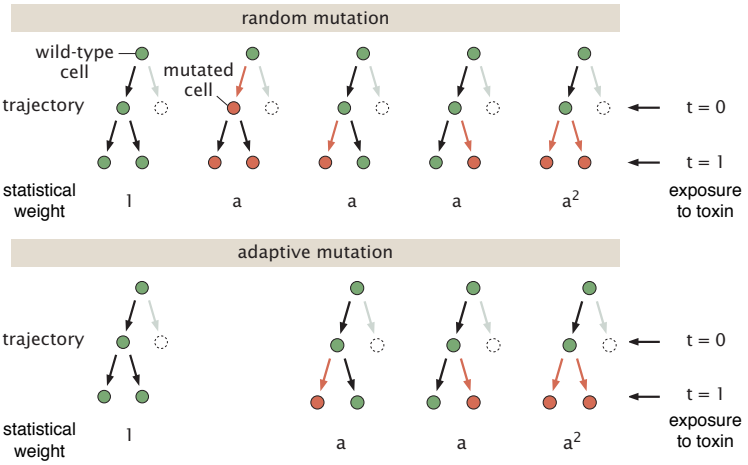

Figure 2: A theoretical approach to testing the Luria-Delbrück mutation hypothesis. Adapted from Fig. 21.30 of *Physical Biology of the Cell, 2nd Ed.*, by Phillips, Kondev, Theriot, and Garcia, Garland Press, 2012.

2.3 Statistics under the adaptive immunity hypothesis

Under the adaptive immunity hypothesis, no cells undergo mutations prior to exposure to the toxin. Upon exposure to toxin, the probability that a given cell will mutate in the next generation is a. What is the probability that one generation after toxins are introduced, n out of a total of N cells are mutated? ($t = 1$ at the bottom of Fig. 2 shows this for $N = 2$ cells.) There are two components to this probability. First, there is the chance that cells 1 through n are mutated and cells $n+1$ through N are not. Since the cells have independent chances of mutation, this is $a^n(1-a)^{N-n}$. But, we arbitrarily labeled the cells. We have to also take into account the number of ways to label the cells that are mutated. There are $\binom{N}{n}$, pronounced "N choose n" ways to do that, where

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}. \tag{2}$$

Thus, we have

$$P(n;N) = \frac{N!}{n!(N-n)!}\, a^n(1-a)^{N-n}, \tag{3}$$

the probability of having exactly n out of N cells be mutated. This probability distribution is called the **binomial distribution**.

What is the most likely value of n? Answering this question requires computing the first moment of $P(n;N)$. We recall the formula for computing the **moment** of a probability distribution.

$$\langle n^m \rangle = \sum_{n=0}^{N} n^m\, P(n;N), \tag{4}$$

where $\langle n^m \rangle$ is the **expectation value** for n^m, or the mth moment of the distribution $P(n;N)$. In particular, the **mean** is $\mu = \langle n \rangle$ and the **variance** is $\sigma^2 = \langle n^2 \rangle - \langle n \rangle^2$. As derived in the appendix (section 5), the mean and variance are

$$\mu = Na; \tag{5}$$

$$\sigma^2 = Na(1-a) \approx Na \text{ for } a \ll 1. \tag{6}$$

Therefore, if the adaptive immunity hypothesis is correct, the mean number of surviving colonies should be equal to the variance in the number of surviving colonies. This means that the Fano factor, $F = \sigma^2/\mu$, is approximately unity.

2.4 Statistics under the random mutation hypothesis

Under the random mutation hypothesis, mutations can occur at all times before exposure to toxin. Computing these statistics is more involved than for the adaptive

immunity hypothesis, which involved only a single generation. We will not derive the statistics here, but encourage you to read Luria and Delbrück's original paper and/or section 21.4.3 of *Physical Biology of the Cell, 2nd Ed.* by Phillips, Kondev, Theriot, and Garcia for a discussion. The important result is that the mean scales with N (as we would expect), but the variance scales as the square of the number of cells present, or $\sigma^2 \sim N^2$. Contrast that to the adaptive immunity hypothesis, where $\sigma^2 \sim N$. Thus, the Fano factor F scales like $F \sim N$, which is much bigger than that for the adaptive immunity hypothesis, where $F \approx 1$. So, if we wait enough generations such that N becomes substantial, we can check:

- The Fano factor is close to unity $\Rightarrow$ the adaptive immunity hypothesis is more likely true.
- The Fano factor is large $\Rightarrow$ the random mutation hypothesis is more likely true.

To compute the Fano factors, we will need to compute the mean and variance from our data.

2.5 Deducing the mutation rate

Regardless of whether the adaptive immunity or random mutation hypothesis is correct, there is a clever way to determine the mutation rate. Say we have N opportunities for mutations to occur (i.e., if we take a time interval for mutations to occur to be the time of cell division, there are N such time intervals). The probability that mutations occur n times is given by the binomial distribution, equation (3). The probability that zero mutations occurred is then

$$P(0;N) = \frac{N!}{0!(N-0)!}\, a^0 (1-a)^{N-0} = (1-a)^N. \tag{7}$$

Therefore, we can measure the fraction of spots that have zero colonies to obtain an estimate of $P(0;N)$. We can compute N from counting cells and then use equation (7) to determine the mutation rate a.

For the adaptive immunity hypothesis, the number of mutations equals the number of cells carrying mutations, so both are binomially distributed. Conversely, under the random mutation hypothesis, the number of mutations is binomially distributed and *not* the number of cells carrying mutations. The probability that, e.g., 16 colonies are found on a spot does not equal the probability that 16 mutations occurred. A single cell could have had a mutation and divided four times, giving 16 daughter cells, or there could have been 16 mutations in different cells. However, in the case where there are zero colonies, there are no mutations, so the probability that there are no mutations equals the probability that no colonies form.

3 Protocol

This experiment will take place over two lab periods. Parallel cultures of wild type and mutator strains of *S. cerevisiae* have already been started for you. To begin the cultures, the TAs grew each strain in 96-well plates and incubated them. During this lab period, you will plate these cultures, and after two or three days you will be able to count the number of mutant colonies that arise from each culture. The culture conditions have already been optimized to ensure that a portion of the plated cultures bear zero mutations, since this information will be used for calculating mutation rates.

3.1 Counting cells (Week 3)

You will be given a 96-well plate containing saturated cultures. In order to calculate the mutation rate of each strain, you will need to determine the total number of cells. You will count the yeast cells using a hemocytometer. This device contains a specialized cell counting chamber, which has a grid of known area and depth. Three separate cultures of each strain of yeast will be counted to determine the reproducibility of the counting method. In each pair, one person will be responsible for the wild type strain, and the other for the MSH2 mutator strain.

For the wild type strain only, dilute 10 µL of the strain into 90 µL of water before counting in the hemocytometer because this strain is much more numerous than MSH2. Therefore, the cell density you will measure in the hemocytometer for wild type will be 1/10 of its actual density.

1. **Sample preparation**: Label one of your hemocytometer's sample injection areas with "WT" and the other with "MSH2." Choose a culture (WT or MSH2) and re-suspend cells thoroughly by pipetting up and down. You should have your pipette volume set to 10 µL.
2. **Sample injection** With a steady hand, pipette 10 µL of the appropriate culture in the corresponding injection area (see Figure 3), making sure to fill the hemocytometer chamber.
3. **Cell counting**: When you are ready to view your yeast cells, place the hemocytometer in a microscope. Use the Brightfield setting with a $10\times$ objective lens and appropriate phase contrast. As shown in Figure 3, you will see a 3×3 grid of 1mm$\times$1mm squares. Each square is further subdivided into smaller regions, since the field of view is not large enough to view an entire 1mm$\times$1mm square. Take pictures in the middle square and the four corner squares so you can calculate the cell density in units of cells/mL. To calculate density, you will need the volume of the detection area which can be computed using Figure 3. Remember to change the prefix of your filenames when you switch from wild type to MSH2.

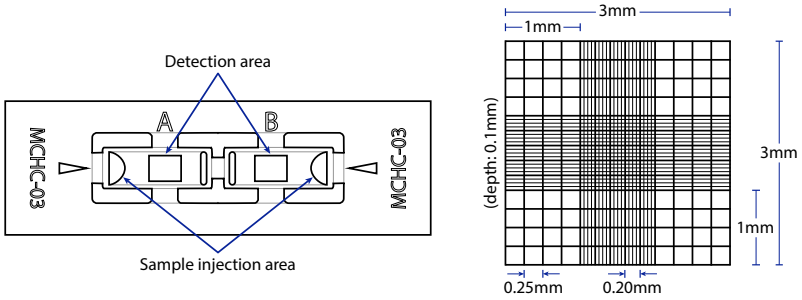

Figure 3: Left, loading a hemocytometer. Right, cells on a grid used for cell counting.

3.2 Plating cells (Week 3)

Once the yeast cells are counted, the next step will be to plate 27 cultures of each strain on dry agar plates containing canavanine, a selective agent that is toxic to yeast. The dry plates have a textured surface and reduced moisture content that will provide a good surface for pipetting large-volume spots onto a plate. In each pair, one person will be responsible for plating wild type cells while the other is responsible for plating mutator (MSH2) cells. There will be a total of 3 plates per person and 6 plates per pair.

1. Label your 3 plates. Indicate which strain of yeast is being plated (WT or MSH2), your full name, and the date on the edge of the plate. Be sure to label the bottom of the plate, as lids can get inadvertently switched between plates.
2. Using the spotting template at your station, draw 9 squares on the bottom of each plate to mark where your spots will go. This will help you count your colonies.
3. Pipette 100 µL of culture from a chamber in the 96-well plate containing the culture you are assigned to do. Place the 100 µL onto a spot in the top left corner on one of the plates.
   (a) Be sure to **pipette up and down** to mix before plating cells.
   (b) Try to place the spots far enough apart that the individual spots do not merge but avoid the edge of the plate.
   (c) Caution: Do not tilt the plates after the 100 µL spots have been placed. Keep the plates in a horizontal position until the spots have dried. Otherwise, the liquid spots will run!
4. Repeat for another 8 cultures, placing spots on the plate such that they form a 3×3 grid. (See Fig 4.)
5. Repeat steps 2 through 3 for each of your plates. You should have placed a total of 27 spots by the end.

6. Give your plates to the TA for incubating at 30°C for 2-3 days to allow cells to grow.

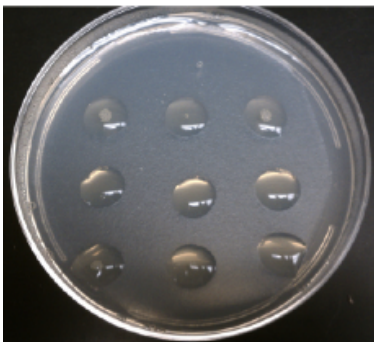

Figure 4: Spotting pattern for canavanine plates.

3.3 Counting colonies (Week 4)

After 2-3 days, colonies will grow where you have placed spots on the canavanine dry plates. Cells that contain a mutation conferring resistance to canavanine will grow up to form colonies. You will now examine the numbers of resistant colonies per spot for the wild type and mutator strains. □

1. Obtain your plates and count the colonies in each spot for your strain. Record these numbers in your lab notebook (each person in each group should have record 27 numbers: 3 plates $\times$ 9 spots/plate).
2. Make a note in your lab notebook if spots contain too many colonies to count accurately by eye.
3. Be sure to record zero-colony spots as they are important in determining mutation rate.
4. Your TAs will provide you a spreadsheet in which to enter your cell counts. Do this before leaving the lab!

3.4 Sequencing the CAN1 region of mutated yeast cells (Week 4)

The CAN1 gene produces a protein that carries canavanine across the cell barrier. Mutations in this gene are what allow yeast to survive on the selective media. You will perform PCR in several weeks on survivor colonies and have the CAN1 region of the genome sequenced. This may show what mutation(s) allowed survival of these colonies in the face of canavanine. The protocol will be performed during the DNA extraction portion of the Winogradsky column experiment. The details of the protocol are contained in the handout of that module.

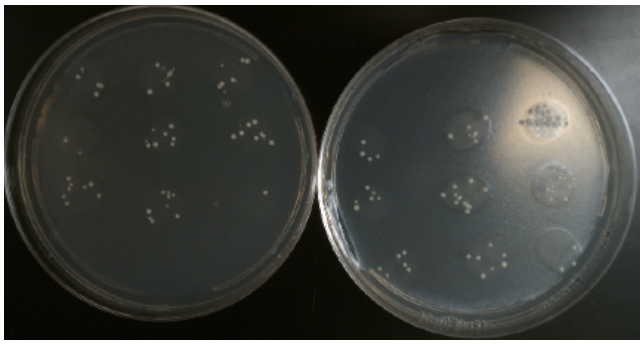

Figure 5: Mutator yeast colonies after 17 hours of incubation. Note that there are no zero-colony spots in on these plates, so this data set could not be used to compute mutation rates.

3.4.1 Colony PCR (Week 4)

1. Before combining your colony PCR reaction mixes, you will need to suspend colonies in 50 µL of sterile water (provided). Suspend one colony of wild type cells in 50 µL of water. Repeat this process for one colony of mutator cells.
2. You will have two "experiment" PCR reactions (one for wild type and one for the mutator strain) and one no-template control. The recipes for the PCR mixes are shown below. Note that the "Master mix" contains the DNA polymerase, free nucleotides, and salts at twice the necessary concentration for our PCR reaction. CAN-1 primers are initially at 10 µM.

Table 1: Colony PCR reaction mixes

Reagent	**No-template control**	**Experiment**
sterile water	10 µL	19.5 µL
Phusion 2× Mastermix	12.5 µL	25 µL
forward primer (0.5 µM)	1.25 µL	2.5 µL
reverse primer (0.5 µM)	1.25 µL	2.5 µL
colony suspension	—	0.5 µL
total	**25 µL**	**50 µL**

3. Mix gently, spin down, and place tubes in a thermocycler. Clearly label your tubes! (This may be easier if written on the side of the tube.) The PCR reaction conditions are given in the following table.

Table 2: L-D colony PCR: thermocycler conditions using CAN1 primers

Step	Temperature	Time	Cycles
initial denaturation	98°C	30 sec.	1
denaturation	98°C	8 sec.	30
annealing	51°C	20 sec.	
extension	72°C	45 sec.	
final extension	72°C	5 min	1

4. After cycling, the reaction will automatically be maintained at 4°C until you retrieve it for gel electrophoresis.

When the PCR reaction has been completed, the TAs will take your samples to send away for sequencing. **Make sure your PCR tubes are labeled with your and your partner's initials and the tubes' contents!**

4 Assignment

Problem 0 (Summary)**.**
Write a summary of this experiment, its objectives, and conclusions between one paragraph and one page in length.

Problem 1 (Fano factors in the toy problem (**prelab**))**.**
Consider the toy problem discussed in section 2.2. Compute the expected number of mutant cells under the adaptive immunity hypothesis. (We already computed this mean for the random mutation hypothesis in section 2.2.) Compute the variance in the number of cells for both the hypotheses. Finally, compute the Fano factor for both hypotheses. How do they compare?

Problem 2 (Hemocytometer volumes (**prelab**))**.**
Consider the central (1mm $\times$ 1mm) region of a hemocytometer in Figure 3. Calculate the volume in microliters of a large (0.2mm $\times$ 0.2mm) square and small (0.05mm $\times$ 0.05mm) square in the hemocytometer.

Problem 3 (Simulating mutations)**.**
In this problem, we will simulate mutations among populations of cells for the adaptive immunity hypothesis and for the random mutation hypothesis.

a) Starting with 1000 non-mutated cells and a mutation rate of 10^{-5} mutations per cell division, run a computer simulation of mutations over the course of

several generations. Do this at least 100 times each for both the adaptive immunity hypothesis and for the random mutation hypothesis.

b) Plot the probability distributions ($P(n; N)$ vs. n) for each hypothesis, where n is the number of mutant cells present at the end of the simulation and N is the total number of cells.

c) Calculate the mean, variance, and Fano factors for each hypothesis.

d) For which hypothesis is the distribution binomial? Does this jibe with the discussion in the Background section?

Problem 4 (Data analysis)**.**

a) Build a histogram of the number of colonies per spot for wild type and MSH2 strains.

b) Calculate the mean, variance, and Fano factor for each distribution.

c) Based on your results from data collected in class, is the distribution of the number of mutants binomial? Looking back at problem 3, which hypothesis results in a simulated distribution of mutants most similar to what is observed experimentally? What do you conclude about how mutations arise?

d) Using the fraction of spots that have zero colonies, estimate the mutation rate a from equation (7) for both the wild type and MSH2 strains.

e) What are some of the sources of error in this experiment?

Problem 5 (Sequence analysis)**.**
During the Luria Delbrück experiment, we exposed *S. cerevisiae* cells to the toxin canavanine to obtain information about the distribution of the numbers of mutants that arise under selective pressure. When grown on canavanine plates, *S. cerevisiae* take up canavanine, which takes the place of the amino acid, arginine, in polypeptide chains. Yeast cells that incorporate canavanine into polypeptide chains instead of arginine cannot survive.

The CAN1 gene codes for a membrane protein that is responsible for the uptake of canavanine. Yeast cells with the CAN1 mutation that renders this gene nonfunctional would not introduce canavanine into the cell. Consequently, these cells that contain CAN1 mutations survive when grown on canavanine plates.

The colony PCR reactions you performed using the MSH2 and wild type *S. cerevisiae* strains were sequenced by Laragen. Since the CAN1 gene is too long to sequence in one single reaction, five primers were used to perform five sequencing reactions. The sequencing results can be downloaded from the course website. The naming convention of the sequencing results, as we received them from Laragen, is

XXXX_STRAIN_P.ab1, where XXXX are the initials of the experimenter, STRAIN is the name of the strain (either WT or MSH2), and P is the number of the primer, taking values 1, 2, 3, 4, and 5. These ab1 files have the sequences, as well as information about the quality of the sequence; that is how sure we are of the identity of each base. There are also files names XXXX_STRAIN_P.Seq that have just the sequencing data. Note that a if a base is reported as N, the identity of the base could not be ascertained.

As you are analyzing your sequences, you may find that you have failed reads, or very low quality sequences. This can result from failed PCR. IF this is the case, note it in your assignment and you may randomly choose another group's set of sequences. Please be sure to note which sequences you used.

The first step in analysis is to align the sequences with the reference wild type CAN1 gene. To get the wild type sequence, you can query the Yeast Genome Database for CAN1. We can then download a FASTA (.fsa) file containing the sequence.

To perform the alignment, we will use Benchling. You should get an account at benchling.com. To assemble the sequences using Benchling, do the following.

1. Create a new project folder (call it something like LD_sequence_assembly) using the + button in the left panel.
2. In the new project folder you can then click the + button in the Inventory panel (again in the left panel). Select Import Sequences. You can now import your FASTA file of the CAN1 gene.
3. Perform the alignment: Select the CAN1 sequence that you imported. On the right most panel, there are several icons with different tools. Hover over them and click the Align tool (5th one down). Select CREATE NEW ALIGNMENT. Select CHOOSE FILE(s) and find your sequencing files. You should use the ab1 files because you will also be able to see the quality of the sequences. Finally select CREATE ALIGNMENT.
4. After the alignment is complete, you should be able to see whether the colonies you selected have any mutations. Nucleotide positions that are highlighted red are mismatches between the sequence and the CAN1 template. Keep in mind that sequences near the start of a sequencing run and near the end will have low quality scores and be unreliable (that's why we've performed 5 staggered sequencing runs).

Use the alignment to find discrepancies between the wild type CAN1 gene and those that you sequenced from your colonies that survived Canavanine exposure. You should only consider a mismatch between your sequences and the reference sequence to be a mutation if all (and at least two) of the sequences covering the region of the reference sequence show the same discrepancy. A snapshot (using print screen or similar) is sufficient when you submit your assignment.

What types of mutations (single base pair substitution, insertion, deletion, etc) are there? Where are the mutations located? Do you find these results surprising? How do the mutations compare between the WT and MSH2 strains? Comment on the differences you observe. Why do you think you observe these differences?

Note: In this problem, we have left the algorithmic details of sequence alignment to the built-in algorithms in Benchling. If you are interested in the algorithms for alignment, there is a vast literature to explore. You might start by looking into the Needleman-Wunsch and Smith-Waterman algorithms.

Problem 6 (Data and code)**.**
Attach all code, data, and sample images not specifically asked for in the other problems.

5 Appendix: moments of the binomial distribution

We begin by computing the first moment, the mean.

$$\langle n \rangle = \sum_{n=0}^{N} n \frac{N!}{(N-n)!n!} a^n (1-a)^{N-n} = \sum_{n=1}^{N} \frac{N!}{(N-n)!(n-1)!} a^n (1-a)^{N-n}$$

$$= Na \sum_{n=1}^{N} \frac{(N-1)!}{(N-n)!(n-1)!} a^{n-1} (1-a)^{N-n}. \tag{8}$$

We make the substitutions $m = n-1$ and $M = N-1$ to get

$$\langle n \rangle = Na \sum_{m=0}^{M} \frac{M!}{(M-m)!m!} a^m (1-a)^{M-m} = Na, \tag{9}$$

where the sum evaluates to unity because the binomial distribution is normalized. To compute the second moment, it is easier to compute

$$\langle n(n-1) \rangle = \sum_{n=0}^{N} n(n-1) \frac{N!}{(N-n)!n!} a^n (1-a)^{N-n}$$

$$= \sum_{n=2}^{N} \frac{N!}{(N-n)!(n-2)!} a^n (1-a)^{N-n}$$

$$= N(N-1)a^2 \sum_{n=2}^{N} \frac{(N-2)!}{(N-n)!(n-2)!} a^{n-2} (1-a)^{N-n}$$

$$= N(N-1)a^2 \sum_{m=0}^{M} \frac{M!}{(M-m)!m!} a^m (1-a)^{M-m}$$

$$= N(N-1)a^2, \tag{10}$$

where this time we have made the substitutions $m = n - 2$ and $M = N - 2$. Thus, we have

$$\langle n(n-1) \rangle = \langle n^2 \rangle - \langle n \rangle = N(N-1)a^2$$

$$\Rightarrow \ \langle n^2 \rangle = Na(1 - a + Na). \tag{11}$$

Therefore, the variance is

$$\sigma^2 = \langle n^2 \rangle - \langle n \rangle^2 = Na(1-a). \tag{12}$$