Hints for Yeast Showdown Rematch: Growth Curves

*You are encouraged to complete the exercise first (original text **in red**) without reading these hints (everything **in black**). Once you are done, or if you are having trouble, this will break down the exercise into more manageable chunks. It is ok if what you did looks very different from this. In fact, in real life, you will not get hints, so as you read this, ask yourself why the exercise has been broken up into these steps, and what are the advantages (and disadvantages!) of doing the analysis the way we've presented here.*

1. Explore the data by plotting growth curves (OD versus time).
The first step in any data analysis task is to explore the data and make sure you see what you expect to see. And "explore" almost always means to plot in some way. However, it's hard to know where to start if you're dealing with data containing many dimensions. To get a clue, try to see if you can just look at all your data in summary form, even if you suspect only a subset of the data will be useful. This is why the first step in this exercise is to simply look at all the data on the 96-well plates.

a. Warm up with one well. Before visualizing all the data, you probably want to start to play with data for one single well (for example, A1). How does OD change versus time? Convince yourself that there is an exponential growth phase. Try to overlay data from two replicate plates, and check how reproducible they are.

i. Load the data. Before we can plot, we have to load the data. Load the `growth_curves_data.mat` file into MATLAB.

ii. Plot the growth curve from a single well. When programming, it's always good to start by coding up simple operations on the smallest possible subset of the data. Once you get a snippet of code working you can scale it up to larger amounts of data. In this case, let's ignore the replicate plate and focus only on plate 1, and also focus on just a single well.

An immediate challenge here is telling MATLAB which wells to plot. In real life we name the wells A1, B1, etc., which, when the letters are enumerated, translate to coordinates like (1,1) and (2,1). But in order to avoid having 4-dimensional arrays we've compressed the 96 wells into a single dimension, counting them from top to bottom, and then left to right. This means that A1 is index 1, B1 is index 2, and A2 is index 9 (count down through the first column and then add 1). Write code to plot the growth curve (OD versus time) in A1 (mutant in rich media).

Of course, later you will want to look at more than just well A1.

b. To get a sense of what the data looks like, plot growth curves for all the data, including the wells that aren't specific to this experiment. In other words, make a grid of sub-plots in the same layout as the 96-well plate, and plot the growth curve corresponding to each well in the corresponding sub-plot. Which wells are the ones we want? How well do the replicates agree with each other? Do all the wells on the 2 plates appear to contain growing cells?

i. Write the index `wellIndex` of a well in row `r` and column `c`. In other words, use the explanation above to come up with the appropriate the MATLAB expression involving variables `r` and `c` and assign it to a variable named **`wellIndex`**. Now you're ready to plot growth curves on a mass scale.

ii. Using a loop (or two), plot an 8-by-12 grid of growth curves corresponding to the wells on replicate plate 1. You'll probably want to open up a new **`figure`**, and then use the **`subplot`** command many times. Where does the **`wellIndex`** from the previous part come in?

Next, plot both replicates (in different colors, if you can) on the same grid of subplots. You need to add no more than one or two lines of additional code.

At this point, go back and revisit the big-picture questions posed by part a.

c. Now plot growth curves only for the 4 wells that you are concerned with for the current experiment. Compare the growth rates of the strains *qualitatively*. What do you see? What do your media control "growth curves" tell you?

Once you do your initial survey of the plates, it's time to zoom in on the data you care about. Although you could in principle just focus on the 4 wells you care about in grid of plots from part a, they are small and not conveniently arranged to facilitate comparison.

i. Use the subplot command to compare growth curves for mutant and wildtype in both media. Let's arrange the 2x2=4 growth curves we want to compare, and two control wells as follows:

WT/rich	mut/Gly+
Mut/rich	Mut/Gly+
rich control	Gly+ control

Look at the plots. What do you see? How do WT and mut compare in rich media? How do they compare in Gly+ media? What can you conclude about the effect of the mutation?

What can you conclude from looking at the data from the media control wells? Is this what you expected? Why did we need to include this in the first place? What other outcomes could you have seen in the control wells?

2. Compare the growth rates *quantitatively*. Does the mutant have a lower growth rate than the wildtype? Do you think this is due to a defect in glycine metabolism?

In other words, calculate the growth rate of each strain in each condition. How you do this is the major conceptual challenge of this section. We suggest a method in the steps below, but feel free to come up with your own method, and compare it to the result in the solution file.

a. Replot everything, but log-transform the OD data (i.e. plot the log of the OD versus time). What changed about the growth curves? Is this what you expected?

The base of the logarithm isn't important for comparisons between different strains -- a strain that grows twice as fast as another will still show the same relative difference no matter what log we use. (Why is this?) However, if we use log base 2, the units of growth rate come out to the doublings / time, which is very intuitive to think about.

Quantifying the growth rate essentially means calculating the slope of the middle (linear) part of each log-transformed growth curve. The problem is that the linear phase happens at a different window in time for each plot. Can you think of a quick and dirty way to do this calculation?

One possibility is just finding the coordinates of two points in the middle of the growth curve with the data cursor () on each plot and taking the slope between them. A slightly more sophisticated way is to somehow cut out the middle part of each data series and fit a line to it.

Either of these methods is acceptable, but both require a bit of manual intervention. There is another way which is easier to scale up to large datasets. This involves noticing that the exponential phase of each (log-transformed) growth curve is simply the part with the steepest slope. If we have a way of calculating the derivative of that entire curve, then the maximum value of the derivative is exactly the slope, and thus growth rate, we want! You already know how to calculate a maximum, so it just remains to figure out how to get a derivative. Using MATLAB's help commands, can you come up with a way of calculating a derivative on your data?

It turns out we can calculate a derivative using MATLAB's **`diff`** command, which calculates successive differences. Look up the **`diff`** command, and use it to calculate the derivative of each growth curve. Keep in mind you'll need to use both an array of OD values as well as an array of time values for this (why?), and that the result will be 1 element shorter than what you started with (why?).

b. Plot the derivatives of each growth curve. Also calculate growth rates (i.e. the maximum derivative) values for all 4 strain/condition pairings.

Now you are ready to think about your conclusions.

c. Make a table or graph of the growth rates. What do you think is the best way to compare the effect of the mutation, and to tell if it caused a defect specific to glycine metabolism? You may want to check out MATLAB's **`bar`** command. How do these results compare with your qualitative conclusions from part 1?

3. (BONUS) Distinguish between technical and biological explanations for the 'lag' phase. For the growth curves you plot, you might find that they have an 'S' shape, i.e. there is a 'lag' phase before exponential growth. This could be an interesting observation, but we need to make sure that the observation is due to the actual behavior of the cells and not an artifact of our measurement technique or data analysis.

a. **Is the lag phase an artifact of how you are scaling the OD axis (linear or logarithmic)? If not, what are possible explanations for this lag phase?**

   One obvious reason for the "lag" phase might simply be that you are plotting the OD in linear axes. However, once you plot log(OD) versus time, you'll see that there still seems to be a phase where the OD isn't increasing. This is inconsistent with exponential growth, which would yield a straight line with positive slope when plotted on log axes. Either the cells are initially in a non-growing state, or there is some other technical problem.

b. **Try to think of one biological reason and one technical reason that we might observe a S-shaped plot of log(OD) versus time.**

   For a biological explanation, see the explanation above. For a technical reason, look up optical density on Google or Wikipedia and think about how this measurement is obtained. When might this measurement not be a good indicator of the number of cells in the media? Look at the OD of the media control wells. What does this tell you about our OD measurements?

c. **Using the given data and controls, think of a way to rule out or correct for one possible technical explanation for the lag phase. Perform this correction, and plot the growth curves. Do you think the lag phase is due to technical or biological factors?**

   In theory, OD should be proportional to the number of cells, but in practice it is affected by dissolved solutes in the media, the transparency of the culture plates, etc. Additionally the plate reader itself may have a limited range of sensitivity or a detection limit (i.e. a threshold below which it cannot yield reliable readings). In other words, there may be a "background" contribution to the OD measurements. Why might this result in the perception of a lag phase, even if cells are growing exponentially?

   To see if background OD is affecting the shape of our growth curves, recall that there are control wells in our experiment that do not contain cells. Assuming that only the OD contribution from cells is changing during the experiment, how can you use the data from control wells to correct for background? What about if you didn't have control wells, but

knew that the cell density at the beginning of each culture was below the detection limit for the plate reader?

Using either or both of these ideas, subtract background for your OD readings, and replot the growth curves. How do they differ now from plots of the raw data? Can you conclude if the lag phase was due to delayed cell growth or measurement artifact?

d. Now recalculate the growth rates of each of the cultures. Does correcting for technical artifacts affect your conclusion from part 2?

Use the background-corrected data from c. above to re-calculate growth rates.

Created 8/9/2011 by JW.
Updated 8/9/2011 by RJ, 5/24/2012 by BH, 5/28/2012 by JW.