Determination of DNA Size, A Post-Lab Activity

DNA fragments in an electric field migrate at rates inversely proportional to the log10 of their molecular weights. For our purposes, molecular weight is expressed as the basepair length of a particular fragment. Distance migrated is the distance, in centimeters, that the fragment moved from the origin (well).

The "molecular weight standard" loaded in our gels a “ladder” with 10 bands of DNA that have been made with the following sizes: 10,000bp , 8,000bp, 6,000bp, 5,000bp, 4,000bp, 3,000bp (this is usually a thick band), 2,000bp, 1,500bp, 1,000bp, and 500bp. Can you find these bands on your gel?

Activity Directions

1. Carefully measure the distance (in cm) on your photograph that each known DNA marker migrated from the origin (well). Measure from the front edge of the well to the front edge of each band. Record your information in **Table I**.

Table 1

Fragment size (bp)	Distance from well (cm)
10,000	
8,000	
6,000	
5,000	
4,000	
3,000	
2,000	
1,500	
1,000	
500	

Example: *5,000 bp piece* *Distance migrated: 3.2 cm*

2. In table 2 record the distance of your DNA fragments for each of your samples

Table 2 Title:

Band	A+	A-	K+	K-	Ligase
Band 1					
Band 2					
Band 3?					

3. Graph your data for your **ladder** on semi-log paper – see example. Find the line of best fit. Use the graph to determine the fragment size in bp of your DNA fragments in your gel. Record this data in table 3

Table 3 Title:

Band	A+	A-	K+	K-	Ligase
Band 1					
Band 2					
Band 3?					

4. Check your work using a computer

Using Linear Regression to Calculate Plasmid and restriction fragment size

By plotting the migration distance versus the log of the fragment size for each band in the 50 bp ladder from your gel photo on a graph, a line-of-best-fit can be generated of which you will determine the equation, a process called **linear regression**. As you may recall from math class, the equation of a line is expressed as $y = mx + b$,

where y = the log of the fragment size
m = the slope of the line
x = the distance from the well (in cm)
b = the point where the regression line intercepts the y axis.

Using this equation, you will be able to determine the y-value (log of fragment size) for your given x-value(s) (migration distance of each of your segments of DNA). Taking the antilog of those y-values will give you your fragment sizes in base pairs. You will be using Microsoft Excel for all your graphing and calculations. Before getting started with these calculations, let's first define some mathematical terms that you'll be using.

An **exponent** is a number written above and to the right of another number, called the **base**, indicating the power to which the base is to be raised.

Examples:

base exponent

a) $10^2 = 10 \text{ X } 10 = 100$
b) $10^1 = 10$
c) $10^3 = 10 \text{ X } 10 \text{ X } 10 = 1000$
d) $10^0 = 1$

The **logarithm** (or **log**) of a number in base 10 is the exponent to which 10 must be raised in order to get that number.

Examples:

1. Log 100 = 2, because $100 = 10 \text{ X } 10 = 10^2$.
2. Log 1000 = 3, because $1000 = 10^3$.
3. Log 10 = 1, because $10 = 10^1$.
4. Log 1 = 0, because $10^0 = 1$.
5. Log 20 = 1.3, because $10^{1.3} = 20$

Notice that the log of 20 is a number between the log of 10 (which is 1) and the log of 100 (which is 2). The log of a number does not need to be a whole integer. Log values can be determined on most calculators by entering the number and then pressing the ***log*** key.

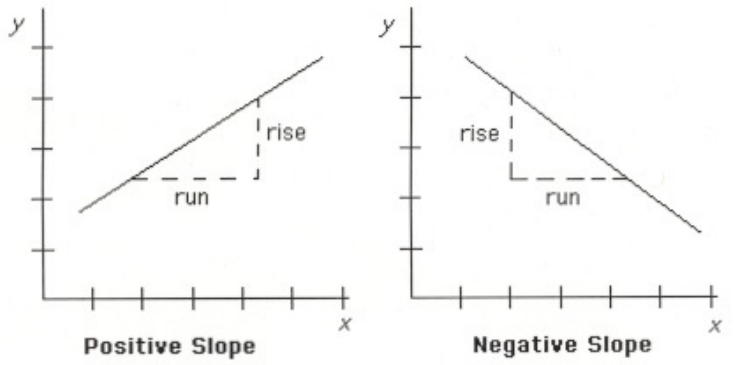
The mathematical opposite of log is antilog. An **antilog** is the number obtained when a given value is used as an exponent of 10. For example, the antilog of 2 is 100 since $10^2 = 100$. Likewise, the antilog of 3 is 1000 since $10^3 = 1000$.

The **slope** of a line (designated ***m*** or ***a***) refers to its inclination; moving from left to right, does the line slant upwards or does it slant downwards. If the line rises, it has a positive slope. If the line falls, the line has a negative

slope (Figure 7). The slope value is frequently stated to be calculated as “the rise over the run” – the amount it rises (or falls) divided by the horizontal distance used to measure that rise (or fall).

Figure 7.

The slope of a line is positive if it rises from left to right and negative if it falls from left to right. Slope is calculated as the "rise” of the line divided by the “run."



Graphing is often used to demonstrate a dependent relationship between two variables. These two variables can be any type of measures that might show some relatedness. Examples might include, distance from the equator and mean temperature, calorie intake and obesity, or exposure to cigarette smoke and incidence of lung cancer. The scientist will then draw a straight line (a "line of best fit" or **regression line**) through, or close to, as many points as possible to see if a linear relationship can be demonstrated between the two variables. Although it is possible to do this “by eye,” computer programs exist that make this a much easier and more accurate task.

The degree to which a regression line represents the data is given by the **correlation coefficient** (expressed either as r or R^2); it is a measure of how closely the points on a graph align with the line of best fit. The correlation coefficient will have the same sign (positive or negative) as the slope of the regression line.

The value of r will always fall between –1 and +1. If all the points lie *exactly* on the regression line and the line slopes upward, r will have a value of +1. If all the points fit exactly on the regression line and the line slopes downward, r will have a value of –1. The closer the r value is to 1 or -1, the better is the correlation between the points on the graph and the line of best fit. Figure 8 shows the how the correlation coefficient relates to the regression line.

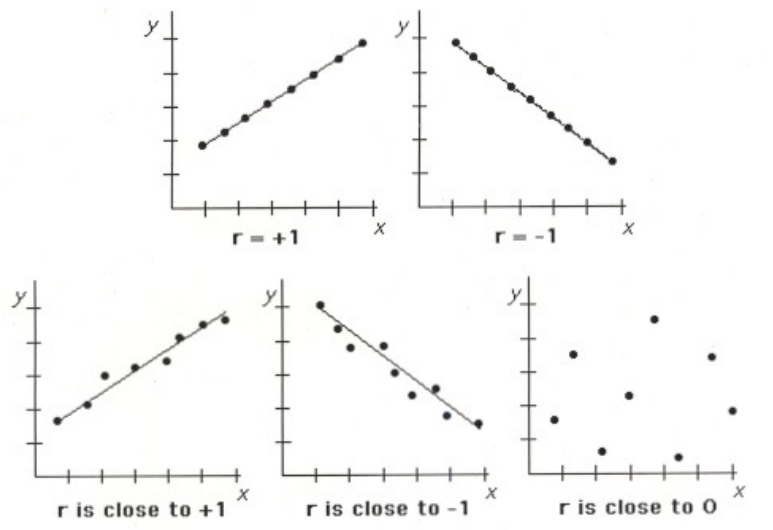


Figure 8.

The correlation coefficient (r) is a measure of how well points on a graph align along a regression line (line of best fit).

Using Microsoft Excel to Calculate Linear Regression

Most computers, whether PC or Macintosh, are equipped with the Microsoft Excel software application. It is generally used to create spreadsheets and graphs. However, Excel is also capable of doing sophisticated mathematical calculations, such as linear regression. To calculate the linear regression equation for the molecular weight standard on the gel, perform the following steps.

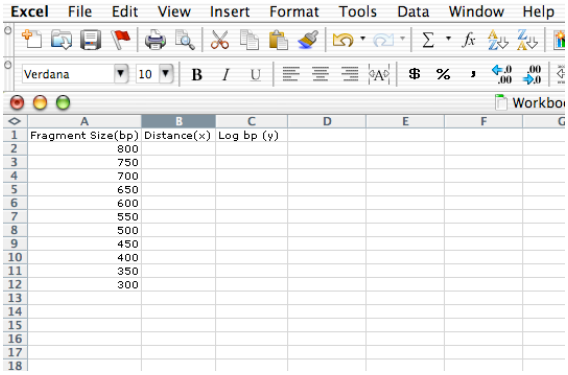
Linear Regression Using Microsoft Excel

Data Entry:

1. Open up a new Microsoft Excel document by clicking on its icon.

2. Make a table for your restriction and ligation data.
   ⇨ Title the first column (A1) "**Fragment size (bp)**" and fill in the size of each fragment you measured from the marker (as in the table 1
   ⇨ Title the second column (B1) "**Distance (x)"** and fill in the corresponding measurements.
   ⇨ The third column (C1), should be titled "**Log of fragment size (y)**".

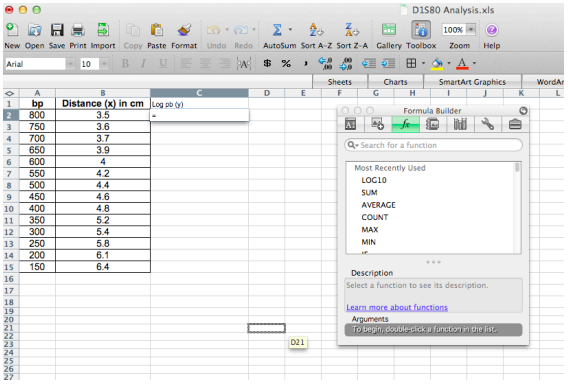
   Note: *Adjust the width of each column to fit the headings.*

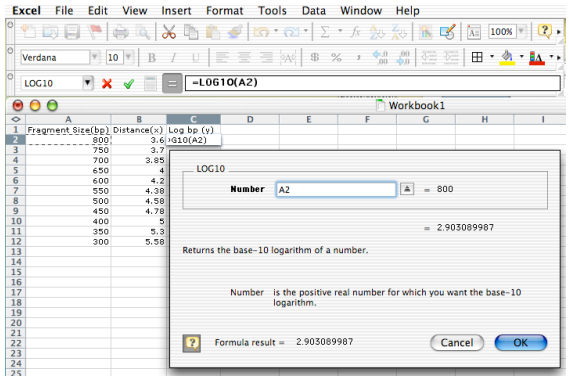
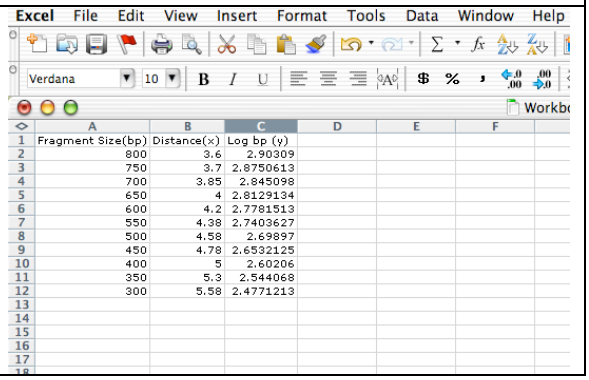


3. You will have Excel automatically calculate the log of all the fragment sizes in the table you just made.
   ⇨ First, click once on the first box **(C2)** under the "Log of fragment size (y)" heading to highlight it.
   ⇨ Click on the **arrow** next to AutoSum **(Σ)** in the tool bar at the top of the screen.
   Choose "more functions"

   A "Formula Builder" box should appear.

   ⇨ Search for "**log 10**" from under the "Function category:" list.
   ⇨ Double click "**LOG 10**" from the

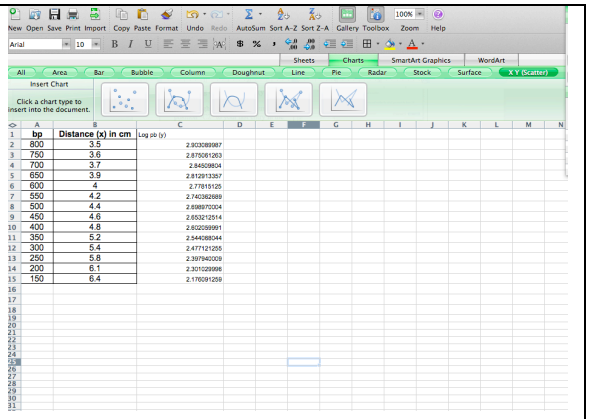


"Function name:" list	
4. Now, at the bottom of the formula builder box you should see "arguments" with a log 10 number box. ⇨ Click on **10,000** in box A2. "A2" should then appear in the "Number" box. ⇨ Click "**OK**." This will enter the log value of 10,000 in the highlighted box in the "Log of fragment size (y)" column. ⇨ **Highlight** the entire C column from position "C2" to your last entry in the column. ⇨ Go to **Edit tab** on the menu bar. ⇨ Select "**Fill down**". **Note:** *Entering "Control" and "D" on the keyboard can also accomplish this fill of all the remaining log values in column C.*	
Graphing Data:	
5. You are now ready to graph your data. ⇨ **Highlight** the values in the "Distance (x)" and "Log of fragment size (y)" columns, **being careful** not to include any column headings or fragment size values.	

6. From the Menu bar strip above,
   ⇨ Click on the **Chart Wizard** button to bring up the Chart Wizard-Chart type window.
   ⇨ Select the "**XY (Scatter)**" chart type and then click "Finish."

A chart will appear over the spreadsheet showing points on a graph.

   ⇨ Under the Chart menu, select "**Chart Options**." Label the x and y axes and create a title for your graph.

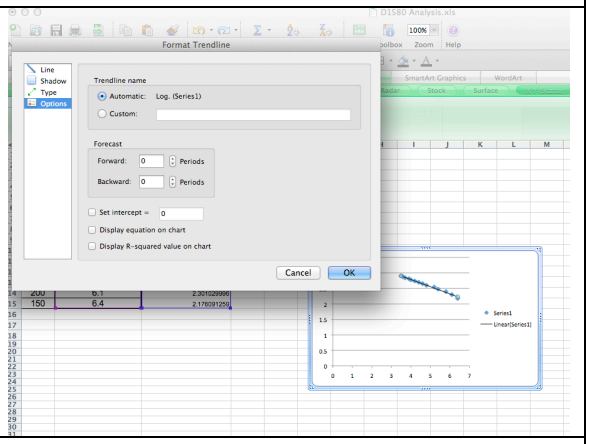


Calculating the Regression Line:

7. Go back to the "**Chart**" on the menu bar.
   ⇨ Select "**Add Trendline**".

The pop up window will display "Type" tab being automatically selected. Make sure the "Linear" box is selected under the Trend/Regression type.

   ⇨ Go on to the step 8 **before** clicking "OK.".

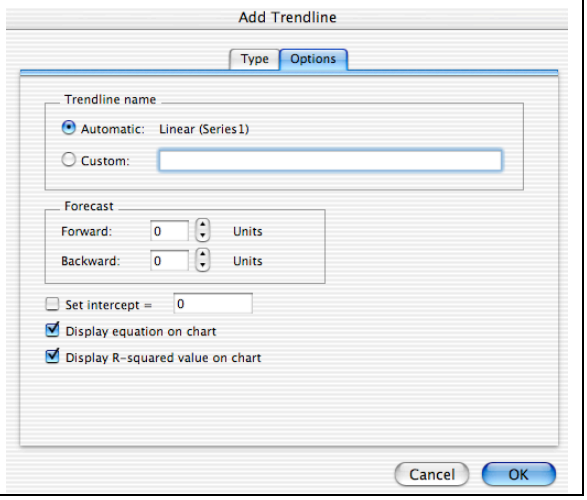


8. Now, select the "Options" tab in the "Add Trendline" window.

   ⇨ **C**heck the last two boxes only; "**Display equation on chart**" and "**Display r-squared value on chart**."

When finished, click "**OK**" and the line of best fit will be drawn on the graph.

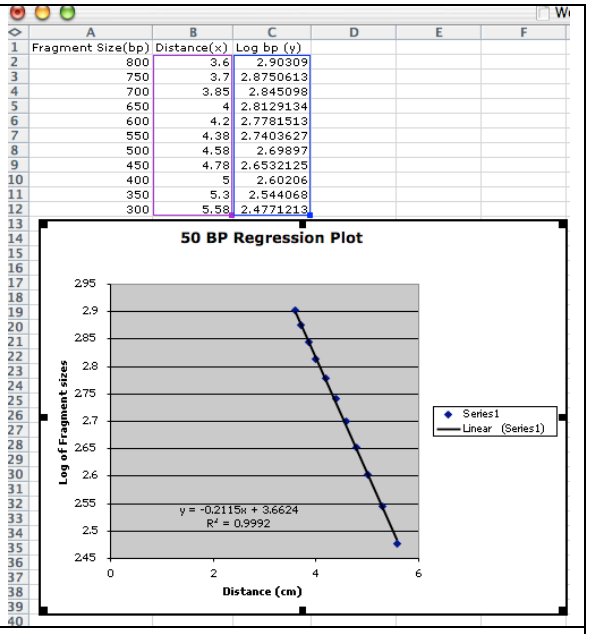
Note: *You will also be given the regression line equation and the* R^2 *value.*



9. Based on the R^2value, does your regression line closely fit the points on the graph? (See page 27)

_____ Yes

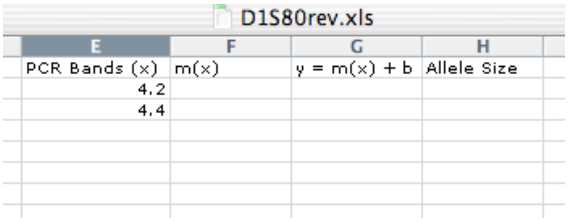
_____ No, because _____



	A	B	C	D	E	F
1	Fragment Size(bp)	Distance(x)	Log bp (y)			
2	800	3.6	2.90309			
3	750	3.7	2.8750613			
4	700	3.85	2.845098			
5	650	4	2.8129134			
6	600	4.2	2.7781513			
7	550	4.38	2.7403627			
8	500	4.58	2.69897			
9	450	4.78	2.6532125			
10	400	5	2.60206			
11	350	5.3	2.544068			
12	300	5.58	2.4771213			

Determining Your Fragment Size Using Excel:

10. In a series of steps, you will calculate the corresponding y-value for each PCR band that you measured (x-value). Taking the antilog of that y-value will give you your allele size.
    - ⇨ Start a **new column (E1)**
    - ⇨ Title it "**PCR bands (x)"**
    - ⇨ Below the title, **enter** the distance each of your PCR bands traveled.
    - ⇨ **Title** cell F1 "***m*(x)"**
    - ⇨ **Title** cell G1 "**y = m(x) + b**"
    - ⇨ Lastly, **Title** cell H1 "**allele size**."



D1S80rev.xls

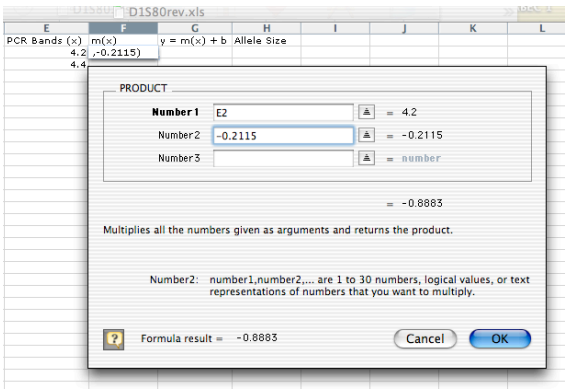
E	F	G	H
PCR Bands (x)	m(x)	y = m(x) + b	Allele Size
4.2			
4.4			

11. Calculate the *m*(x) value by **highlighting cell F2**.
⇨ Click on the function icon **(*f*x)** in the tool bar.

As the "Paste Function" dialogue box that appears,
⇨ Select "**Math & Trig**" under "Function category:"
⇨ Scroll down and select "**PRODUCT**" under "Function name:"
⇨ Next, click the "**OK**" button.

In the Product Dialogue box that appears,
⇨ Enter your regression **equation's *m* value** in the "**Number 1**" box
⇨ Also, enter the ***x* value (distance in cm)** for the band being examined in the "**Number 2**" box.
⇨ Click the "**OK**" button.

The product will be entered in the box you highlighted.

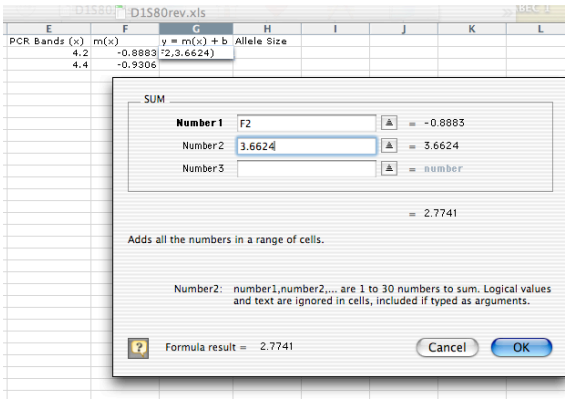


12. Now, let's calculate the value of y, which is *m*(x) + b.
⇨ **Highlight** the next cell **(G2)**.
⇨Click on the function icon **(*f*x)**.

In the "Paste Function" dialogue box,
⇨ Choose "**Math & Trig**" under "Function category:"
⇨ Scroll down and select "**SUM**" under "Function name:"
⇨ Click the "**OK**" button.

In the "SUM" dialogue box that appears,
⇨ Click on **cell F2**, with the m (x) value. *You should see F2 appear in the Number 1 box*.
⇨ Enter your **regression equation's *b* value** in the "**Number 2**" box.
⇨ Click "**OK**."
The sum of the entered numbers will appear in the highlighted box.

If you have a second band, repeat the steps above calculating the y value of G3.



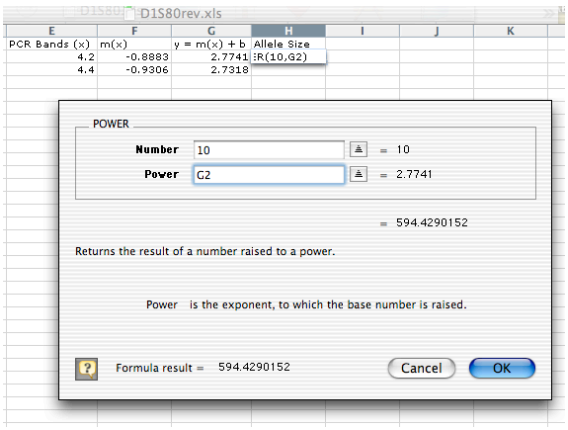
13. Taking the anti-log of the y-values in column G will give you the "Fragment" of each plasmid fragment (column H), in base pairs.
    - ⇨ Click on **cell H1** first.
    - ⇨ Select the function icon **(*fx*)** from the toolbar above.

In the "Paste Function" dialogue box that appears,

- ⇨ Choose "**Math & Trig**" under "Function category:"
- ⇨ Scroll down and select "**POWER**" under "Function name:"
- ⇨ Click the "**OK**" button.

In the "POWER" dialogue box that appears,

- ⇨ Enter "**10**" in the "**Numbe**r" box (since you are operating in base 10).
- ⇨ In the "**Power**" box, enter the **G2 (y value)** you calculated above.
- ⇨ Click the "**OK**" button.



****

Repeat this with H2 if you have a second band. The values in column H are your fragment in base pairs.

Create a table 3 that has your tubes and the bp sizes of each fragment of DNA in your gel