

Biology Boot Camp LABORATORY MANUAL

Elsbeth Walker
Katherine Dorfman
Sam Hazen

UMass Summer College 2014



**UMASS
AMHERST**

Table of Contents

General Information	3
Laboratory Basics	3
Example Lab Notebook Entry	6
Laboratory Report Format.....	7
Presenting your Findings	14
Comparing Means: The t-Test	17
The Chi Square Test for Counted Data	27
The Q-test for the Outlying Data Point.....	31
DNA, Amino Acids, and the Genetic Code	34
DNA Molecular Weight Standard	39
Metric Prefixes	40
General Lab Techniques	41
Introduction to micro-pipetting.....	41
Dilutions	51
DNA Extraction and Quantification	55
Lab 1: DNA Extraction	55
Lab 2: DNA Quantification	59

GENERAL INFORMATION

Laboratory Basics

Your station

You will be provided with a variety of materials and equipment for use during the semester, some for exclusive use by you and your lab partner, and some to be shared with another pair at the same bench. You will be responsible for keeping track of these things.

Your station should have the following items:

- 3 micropipettors: 20 μ L, 200 μ L, 1000 μ L
- 3 boxes pipet tips: 20 μ L, 200 μ L, 1000 μ L
- spatula
- scissors
- forceps
- 1 1L beaker of microfuge tubes
- 1 roll label tape
- 1 alcohol-resistant marker
- scotch tape
- 1 package 10 mL pipettes
- 1 10mL pipet filler
- 1 microfuge racks
- 1 freezer box
- 2 test-tube racks
- 1 waste beaker
- ruler
- timer
- ice bucket

You will have to share the following with another pair:

- 2 gel rigs
- 2 250 mL flasks
- 2 gel casting trays
- 2 levels
- 2 15-well combs
- 2 8-well combs
- 2 gel trays
- 1 box Kimwipes
- Parafilm
- 3 boxes gloves: sm, med, lg
- *Biology of Plants* by Peter Raven
- microfuge
- DNA EngineTM Thermocycler
- Dissecting microscope

Organization and Care of Reagents

You will be provided storage boxes to hold tubes containing reagents and biological samples generated in laboratory experiments. Keeping these boxes organized is important to help locate reagents and samples as needed in performing experiments. Thus, you will be expected to keep your tube storage boxes organized and to have all of your tubes clearly labeled (either on the top of the lid or the side). Correctly labeled tubes are ***legible to anyone***, and state the contents of the tube, the date the tube was generated, and the initials of your lab group. ***The biological samples you generate the first day will be used again the next, so the labels should be clear.*** Write on disposable tubes with permanent marker; use label tape for permanent items such as racks. Do not use label tape on microfuge tubes, or permanent marker on glassware.

For these experiments, we will provide you with just enough reagents to carry out the experiments described in the laboratory manual. Often these are very small volumes. It is extremely important that you manually mix (by shaking or vortexing) reagents before using them. After mixing, always spin the tube briefly in the microfuge to settle the contents at the bottom. This is especially important when you use frozen samples, which first must be **thawed completely**, and then thoroughly mixed.

Due to condensation and handling, droplets of liquid reagents often adhere to the lids or sidewalls of microcentrifuge tubes. Thus, it is important that you **always** briefly spin your reagent tubes in the microcentrifuge to collect reagents at the bottom of the tube **before** opening the tubes. Having all of the reagent at the bottom of the tube will increase pipetting accuracy and ensure you do not run out of reagents.

It is also very important that you never contaminate your stock reagents. This means always using a fresh sterile pipette tip—never one that was used previously, and never one that you have accidentally touched. Even if you can't see liquid in a used tip, microdroplets are in there, and would cross contaminate your solutions if the tip were used more than once.

Laboratory Safety

Personal protection

We do very little in this course that could result in a fire or an explosion. We will, however, be using a few reagents (in very small amounts and low concentrations – be reassured) that have the potential to cause harm. We will alert you to whatever dangers we see in the lab. Your best bet for protecting yourself is not to get anything in your mouth, eyes, or breaks in your skin. Do not eat or drink or apply lip balm or fix your contacts in the lab. **Wash your hands:** after you handle reagents, before you touch your face, before you leave the room, before you have a snack. **Wear gloves:** while you handle anything that might hurt you, or that you might contaminate. **Take your gloves off:** whenever you move from wet bench activities to computer keyboard, or whenever you have touched something you don't want to spread around the room. **Wear your safety goggles:** whenever you handle liquids that could splash or anything that might shatter.

Move slowly to reduce spills and breakage.

Cleaning up

We will alert you to the proper disposal techniques for the reagents and materials we use in this lab. Some are quite benign and can go right in the trash or down the drain, and others require special handling. Check the **Clean up** section for each lab for specific instructions.

Discarding Waste

There are many different types of waste in this lab, each disposed of separately:

- **recyclable paper or containers**
- **broken glass** (unrecyclable, dangerous to handle)
- **sharps** (small sharp objects like razor blades)
- **biohazard waste** (biological specimens that might be infective or invasive)
- **hazardous waste** (toxic, corrosive, or flammable waste)
- **regular trash** (solid waste that is not any of the above)
- **benign liquid waste** (liquid waste that is not hazardous or infective)

Please ask whenever you have any doubt about disposing of something.

Printing in the Biology Labs of the ISB

After you have set up your document the way you want it, you are ready to send it to the printer.

Warning: there are many steps to this process. Much frustration, heartbreak, and wasted paper can be avoided by mastering these steps now.

- Select “print” from the file menu. ***This does not send your job to the printer!***
- Notice the name of the printer, as this will tell you where to look for your printout. (The printers in ISB biology space are named after Pacific islands.)

Printer name	room
Nifiloli	264
Nupani	360
Nukapu	364
Ngawa	368

- You may be able to get a menu that lets you change image quality by looking at your options under “copies and pages”.
- Use the preview function to make sure you will be printing the right thing. Check which pages will be included.
- Click print. ***This does not send your job to the printer!*** This will produce an authentication dialog box. Sign in with your BCRC username and password, or ask your instructor to do so for you. ***This does not send your job to the printer!***
- Go to the print release page (<https://wahoo.nsm.umass.edu/printrelease/>), which is linked to the BCRC web page, and in the pull-down list under “undergraduate” on the Biology Department web page. Enter your name and password again, if asked. ***This does not send your job to the printer!***
- You should be directed to the print queue page. Refresh repeatedly until your job shows up. Check the box for your job and click “release”. (If you have made a mistake, this is a good time to cancel a job.) ***This does not send your job to the printer!***
- After you release your job, you get a confirm release page. Click “really print”, and ***this will finally send your job to the printer!***

Example Lab Notebook Entry

2/10/11

Lab X.Y: Microscopy of rice tissue samples

Purpose/Goal:

To prepare samples of rice flowers for sectioning and microscopy; to take images of rice flowers using a microscope.

Methods:

We followed the methods on pages A-F of the lab manual. In summary, tissue was fixed in standard fixative, dehydrated in an ethanol series (10% - 95%), and then acetone. We embedded the tissue in plastic resin and sectioned it 1mm thick using microtome.

Sections were placed on glass slides, stained with methylene blue dye and viewed under the microscope.

Calculations: $V_i = \frac{V_f \times C_f}{C_i} = \frac{10\text{mL} \times 1\%}{5\%} = 2\text{mL}$

Annotations:
 - Need this much to stain slide (points to 10mL)
 - Need this concentration (points to 1%)
 - use this much stock, plus H₂O to 10 mL (points to 2mL)
 - Stock solution concentration (See Errors, below.) (points to 5%)

Observations:

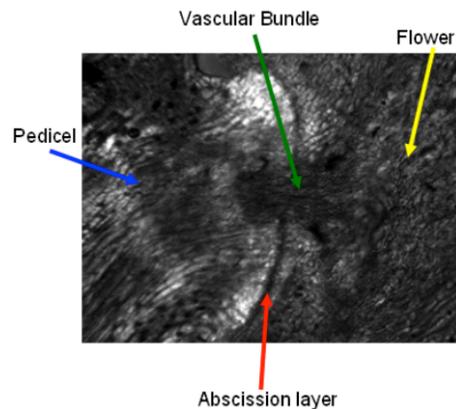
- Tissue was floating on top of the fixative at first, but after 2 hours had sunk to the bottom of the tube.
- Tissue was green to start but by the end of the fixing and dehydrating was much paler.

Errors:

Stock solution was 2.5%, not 5%. Should have used 4mL stock. Needed to leave slide in dye longer than time given in manual.

Results:

labeled image of my rice flower:



Conclusions:

The presence of a dark band (abscission layer) between the flower and pedicel suggests that this flower can easily be shed from the panicle. Easy shedding flowers are a hallmark of weedy or wild rice relatives thus I believe my flower comes from either weedy or wild rice.

Laboratory Report Format

Scientific writing

For additional reading, there is a very nice web site “How to Write Lab Reports for Biology” at the Union College Biology Department web site:

http://www.union.edu/academic_depts/biology/ResearchReports.php.

Pay special attention to the sample lab report on Jumpamine that is given there. It is an excellent example to use both in terms of content and length.

The basics:

- Double space all text
- Use a 12 point font
- Run the spell checker before printing
- Print lab reports—e-copies will NOT be accepted

Plagiarism

DO NOT PLAGIARIZE! Even the use of a *phrase* from a publication or the internet without quotation marks and citing the reference is considered plagiarism. Be especially careful in cases where you write your passages immediately after reading information on the topic in a book or on the Internet. Also, be careful when you and your lab partner are working together. While you both have the same data, and will have identical ‘Results’ sections, the ideas presented in the rest of the lab report must be your own.

Cite your sources. It is imperative to acknowledge where ideas or knowledge not originally your own come from, *even though* you have stated your understanding of the idea in your own words. This is done by putting the first author’s last name and the date of the paper, or the page of a textbook, in parentheses at the end of the sentence containing the idea (as shown on page 10). Then, all sources (books or articles or web sites) you mention in the text of your paper are included in a list of references in a separate section at the end of the paper, as illustrated on page 13.

Although it is technically permissible to use brief quotations (a phrase or one sentence at most), these are very rarely used by practicing biologists. It is *never* appropriate to borrow extensive passages (more than two sentences) from a text or web site. The only reason to use a direct quotation is that the bizarre or amazing wording in the source must be shared with your reader. If you ever did have to use a quotation, you would use quotation marks and cite a reference at the end of the sentence. However, it is very unlikely that that it would be appropriate to make a direct quotation in a lab report.

Clarity, grammar, syntax and spelling: the mechanics of writing

The actual mechanics of your writing *do* matter. This includes spelling, grammar, logic, the flow of the writing and general understandability of the paper. Therefore, you should make sure you proofread your reports, and edit them as necessary. There are two ‘tricks’ that almost always give marked improvement in the clarity of your writing. The first is to ***make an outline before beginning to write***. You learned to do this in grammar school (sometimes they called it a web), then again in middle school, and again in high school. Even though you may have learned the technique in English class, you should use it for all of your writing. Another excellent way to be certain that your writing is clear is to ***read it out loud to***

yourself. Students always think this is stupid/corny/something I made up to embarrass them. In fact, I always read my own writing out loud in order to catch the unclear bits.

There is help available from the University Writing Center (<http://www.umass.edu/writingcenter/>). You paid for it (with your tuition and fees); you should take advantage of it. Your grades will thank you.

Scientific terminology

It is worth learning the correct singular and plural forms of scientific terms. A selection of the most common follows:

one	two
bacterium	bacteria
chiasma	chiasmata
cilium	cilia
criterion	criteria
datum	data
flagellum	flagella
genus	genera
helix	helices
index	indices
larva	larvae
medium	media
mitochondrion	mitochondria
nucleus	nuclei
phenomenon	phenomena
species	species (not a typo! – singular and plural are the same)
stigma	stigmata

Learn the correct format for abbreviations, as well:

- Many scientific acronyms are capitalized, but the phrases they stand for are not. For example, *DNA*, *OD*, and *GFP* stand for *deoxyribonucleic acid*, *optical density*, and *green fluorescent protein*, respectively. The names of the elements are capitalized when abbreviated, but lower case when written out: *NaCl* stands for *sodium chloride*.
- Scientific names of organisms take this form: *Genus species*. The genus name is capitalized, and the species name is lower case. Both are italicized. For very common species, or after you have first written out the entire name, you may abbreviate the genus name with just its first initial. E.g., *Escherichia coli*, *Drosophila melanogaster*, *Homo sapiens*, *Mus musculus*, can be abbreviated *E. coli*, *D. melanogaster*, *H. sapiens*, *M. musculus*. (You do have to be careful, however, that the reader wouldn't think you were referring to the intestinal amoeba *Entamoeba coli* instead of the bacterium *Escherichia coli*.)

Use the most specific term you can, even if it is repetitious and your English teacher told you not to repeat the same word too often. For example, “hydrochloric acid” is more specific than “acid”, which is more specific than “chemical”. Write “lactose” when you mean lactose in particular, and “sugar” when the reader couldn't possibly think you mean any other sugar in that sentence, or when you do mean any sugar.

Miscellaneous good writing

Not scientific exactly, but important nonetheless, clear writing represents clear thinking (and makes a favorable impression on the reader). Watch out for common errors; if you're not

sure, look it up. The OWL at Purdue has a good selection of brief explanations of English grammar and mechanics: <http://owl.english.purdue.edu/owl/> Some of my favorites:

- *it's* is short for *it is* or *it has*. There is no apostrophe in *its* (meaning belonging to it) any more than there is in *his* or *hers*.
- The whole *comprises* its parts; the parts *compose* the whole.
- To decide whether to use *me* or *I*, think about what you would write if only one person were involved. For example, since you would never write “Me did an experiment,” (or at least I *hope* you wouldn't), you shouldn't write “Me and my partner did an experiment.” In addition, for the sake of politeness, put yourself second: “My partner and I did an experiment.” Do the same mental analysis for she/her, he/him, they/them, we/us.
- If you count them, you have a *number* of them. If you weigh it, or measure its volume, you have an *amount*.

Reporting numbers

Decimals. Write the zero in front of a decimal, e.g., 0.05.

Significant figures. Don't report more significant figures than are found in the least precise of your measurements. Put your least precise measurement into scientific notation to find the appropriate precision. If you report DNA concentration as 250.139652 $\mu\text{g/mL}$, you are suggesting that you can distinguish between 250.139653 $\mu\text{g/mL}$ and 250.139652 $\mu\text{g/mL}$; it is more honest to give the concentration as 250 $\mu\text{g/mL}$ or $2.50 \times 10^2 \mu\text{g/mL}$.

Just because your calculator or Excel calculated a number to 9 decimal places doesn't justify reporting all those significant digits. In Excel, use the Format/Cells/Number menu to adjust the number of decimal places. Do it yourself if you are copying a number from your calculator display.

It is very hard to read a list of numbers embedded in a sentence. If you want to report more than two numerical values, put them in a table so that the reader can carefully compare them. In the written part of your paper that refers to the table, you need only point out the most important values, or the range, or how many times larger the maximum is than the second largest, or whatever else you think is important.

Choose your units in a way that will make it easiest for your reader to understand your measurements. In your introductory chemistry and biology labs, you probably used only a few of the most common metric prefixes, such as milli-, centi-, and kilo-. Various textbooks and lab manuals contain longer lists of prefixes, but not many contain a complete list. There is no point in memorizing all of them, but it is nice to be able to look them up when you encounter them as you read the scientific literature. The smaller prefixes, such as nano-, pico- and femto-, are becoming increasingly common as analytical chemistry and biotechnology develop more sensitive methods. The larger prefixes, such as mega-, giga-, and tera-, are becoming more common in computing as engineers find ways to pack more and more memory into tiny devices. Table 2 on page 40 presents prefixes ranging from 10^{24} to 10^{-24} . To help you visualize the effect of these prefixes, the column labeled “a sense of scale” gives some examples of the magnitudes represented.

Specifics

Title

Scientific research reports usually have very specific titles. “Lab 1” is uninformative (and dull). A better title refers to the particular investigation you are writing about. Below are some titles collected at random from various scientific journals.

When do domestic dogs, *Canis familiaris*, start to understand human pointing?

The role of ontogeny in the development of interspecies communication. Dorey, Nicole R., Udell, Monique A.R., Wynne, Clive D.L., 2009. *Animal Behaviour*, 79 (1), 37-41.

Ear and eye in subterranean mole-rats, *Fukomys anselli* (Bathyergidae) and *Spalax ehrenbergi* (Spalacidae): progressive specialisation or regressive degeneration? Burda, Hynek. 2006. *Animal Biology*, 56 (4), 475-486

Structure of a Bacterial Dynamin-like Protein Lipid Tube Provides a Mechanism For Assembly and Membrane Curving. Low, Harry H., Sachse, Carsten, Amos, Linda A., Löwe, Jan, 2009. *Cell*, 139,(7), 1342-1352.

Inbreeding depression on beef cattle traits: Estimates, linearity of effects and heterogeneity among sire-families. Carolino, Nuno, Gama, Luis T., 2008. *Genetics of Selection and Evolution*. 40(5): 511–527.

Discovering genetic ancestry using spectral graph theory. Ann B. Lee, Diana Luca, Lambertus Klei, Bernie Devlin, Kathryn Roeder, 2010. *Genetic Epidemiology* 34 (1): 51-59.

Genomic profiling of breast tumours in relation to BRCA abnormalities and phenotypes. Olafur Andri Stefansson, Jon Gunnlaugur Jonasson, Oskar Thor Johannsson, Kristrun Olafsdottir, Margret Steinarsdottir, Sigridur Valgeirsdottir, and Jorunn Erla Eyfjord. 2009. *Breast Cancer Research* 11(4): R47.

First section: Introduction

The point of this section is to give background information on the general topic of the laboratory experiment, and to describe the purpose of the experiment(s) that you did. We want you to exhibit some independent understanding of what you did in lab, and why you did it. Generally, there is one major idea that a lab will investigate. Please look at the example introduction given on the Union College Biology Website (see above; http://www.union.edu/academic_depts/biology/Jumpamine.php). Note the length of this introduction, and the type of information it contains. Please refrain from writing long detailed introductions that reiterate your readings—these are not needed. Remember that when you write a lab report, you will have already performed the experiment(s), so use the past tense.

Citations:

As indicated in the section on plagiarism on page 7, when you use published information to provide context for your experiment, you must include a reference (Author, year) to the work in situ when you refer to it, for example:

Amylase activity peaks at about pH 5.5 (Chung, et al., 1995).

The complete reference goes at the end of the paper (see p.13 for details).

Second section: Materials & Methods

In general, this is the section where you would explain exactly the experiment you did, and indicate the equipment and any specialized chemicals or reagents you used. The purpose is for another experimenter to be able to exactly repeat your experiment in order to compare his/her results with yours. This section (as in all other sections) is written in full paragraphs—*never as a list*.

Since this is a teaching lab, you are sometimes repeating previously described experiments. Therefore, the materials and methods are sometimes not unique, and there is no point in your describing what we just finished describing to you in the lab manual. ***It is not necessary for you to re-iterate the lab manual.***

At other times, you will truly be designing your own materials and using your own protocols. Some examples are when you choose primers, or when you decide to use specific restriction enzymes, or you decide on a specific set of PCR cycling conditions. In this type of situation, you must give a detailed account (*e.g.*, tell the name and sequence of the primers) of what you did. You might also have changed a protocol from the lab notebook—maybe because your instructor told you to, or maybe because you made a mistake. If so, the change should be noted in the materials and methods. Remember, the purpose is for another experimenter to be able to repeat your experiment exactly.

Third Section: Results

In this section, you will present your data, but will omit any description of what your data mean. The results section is written using a series of figures (pictures and graphs) as a skeleton. The particular figures and tables that you will present should be the first elements in the outline of the chapter. Note carefully that the text must include a prose of the results, including results that you also showed in your figures. This prose description is written in full paragraphs—not single sentences, and not as lists. For example, you might write, “Each dilution of genomic DNA was run on an agarose gel (Figure 1). For the 1/50 and 1/100 dilutions, we did not observe any DNA (Figure 1, lane 3 and 4)...” Notice that I have described what was done, and what I observed, but I did not make any qualitative statement about what the result showed—that is, I did **not** say, “This implies that these dilutions contained so little DNA that it could not be observed.” Also notice that I have used a figure (a picture of a gel, appropriately labeled with lane numbers) and referred to it in the text. I used the text to alert your attention to the most relevant parts of the figure. Remember, make sure that you ***show and state the results you actually found***, even if they are not what you expected or predicted.

You must refer, in the text, to every figure that you include. Explain in your own words the salient features of every figure or table. The choice of what you talk about in the text reflects your understanding of what is important about your results.

Figures and tables are a crucial and integral part of the Results section, as mentioned above. Please realize that a figure is defined as a graph, picture, or diagram. A figure will ***always*** have a detailed legend ***at the bottom***. The legend will have a title, and then will describe what is shown. **This is a formal requirement, and cannot be ignored.** For example, I could show a picture of an agarose gel that has 4 lanes. The lanes would be labeled (1-4 or A-D). The legend would read as follows:

“Figure 3. Agarose gel electrophoresis of *Arabidopsis* genomic DNA. Lane 1: 10 μ L of Fermentas MassRuler DNA Ladder Mix. Lane 2: 5 μ L of 1/20 dilution of *Arabidopsis*

genomic DNA. Lane 3: 5 μ L of 1/50 dilution of *Arabidopsis* genomic DNA. Lane 4: 5 μ L of 1/100 dilution of *Arabidopsis* genomic DNA.”

Tables contain lists of numbers (and sometimes other symbols or words.) By convention, tables have a detailed legend at the top. For figures, the legend is on the bottom. This convention for presentation of figures and tables is illustrated in Figure 1.

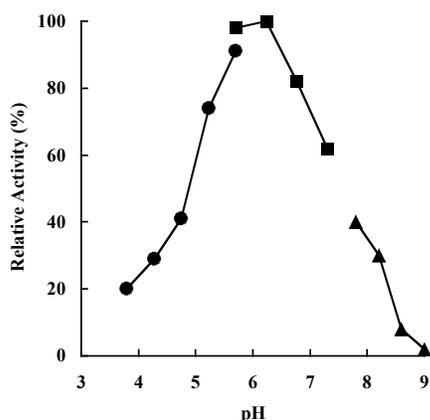


Fig. 2. Effect of pH on the amylase activity. The following buffers were used at 50 mM: sodium acetate buffer between pH 3.8 and 5.7 (circles), sodium-potassium phosphate buffer between pH 5.6 and 7.3 (squares), and glycine-NaOH buffer between pH 7.7 and 9.0 (triangles). All the pH values were adjusted at 70°C.

TABLE 1. Summary of purification of amylase S

Treatment	Total activity (U)	Protein (mg)	Sp act (U/mg)	Yield (%)	Purification (fold)
(NH ₄) ₂ SO ₄	608	434	1.4	100	1
DEAE	582	3.4	171	96	122
Superdex	160	0.14	1,143	26	816

Figure 1. The difference between a figure and a table in a scientific publication (from Chung, Y.C., et al, 1995. *Applied and Environmental Microbiology* 61(4), p. 1502-1506). The figure’s caption is below (like the plaque beneath a painting in a museum); the table’s heading is above (like the page header in a book of mathematical tables).

Occasionally, you may have a result that is not shown in a figure or a table, but this is rare. One example would be a simple calculation, “We calculated the concentration of our DNA sample as conversion factor \times OD₂₆₀ \times dilution factor (0.5 \times 0.98 \times 2) to arrive at a concentration of 0.98 μ g/ μ L.” It is a result, and needs to be included, but is not something that (by itself) warrants a whole figure or table. Please note that, if several such calculations were made, you would want to put them into a table. Another good example of a result that may not need a figure is a statement of statistical significance. These statistics are descriptions of data, and thus belong in the results section, even though they may not require a figure or table.

To reiterate, the Results section consists of a complete prose description, written in full sentences and paragraphs, that describes what you did and observed after doing it. In addition to the text, the Results section will include figures and tables that show the actual data. Each figure and table will have a legend, and figures will be appropriately labeled.

Fourth Section: Discussion

In the discussion, you will talk about *what the results actually mean*. Explain any cause-and-effect relationships. How much variability is there in the results (if you did it twice, did you get the exact same answer each time)? What are some possible causes of this variability? How reliable are the data (can you believe them, or do you think there was an error in carrying out the experiment)? What general conclusions can be drawn? What is the importance of the discoveries you have made? You should also make a point of answering

any questions that are posed in the lab manual. **Do not make a separate section of the answers, but include them in the content of the discussion**, when the topic comes up, as you discuss the results.

Many students, when beginning to write lab reports, feel that they contain a lot of repetition. They feel that they have said the same things in the M&M section, the Results section, and then the Discussion. If done correctly, and with attention to the ‘rules’ given above, the amount of repetition in a well-written lab report is minimal. Most often, a single phrase or sentence can be used to re-orient your reader to a particular idea that was discussed earlier. Consider the following examples:

Methods: Primers oHTJ_F (ATTTAGGGATAGACTT) and oHTJ_R (ATTGTAGGATTGATCCCC) were chosen to match positions 3456 and 3822, respectively.

Results: Primers oHTJ_F and oHTJ_R amplified a 132 bp amplicon, not the 366 bp spanning the third intron of the UNK_A gene, as illustrated in Figure 1.

Discussion: The amplicon produced from cDNA using primers oHTJ_F and oHTJ_R was 132 bp and not 366 bp long, indicating that the gene model we predicted was correct, and that these primers flank an intron.

In these examples, I have oriented you, e.g., by mentioning the primer names, or by reminding you of the amplicon’s expected size, but I have not repeated information extensively.

Final Section: Literature cited

Here you list any books, papers, internet sites, etc. in which you found information that you cited by author and year (see p. 10) in the text of your report. These references are usually listed in alphabetical order by the first author’s last name (this is the Harvard system of citation; an alternative system cites papers by number in the text of the paper and lists them in the reference section in the order of appearance in the paper). Make sure all the authors of a paper or book are listed, and include the title of the book or article, the journal or publisher (and place), and the date. If you used just part of a book, indicate the chapter or pages used. For web sites, give the exact electronic address and any other information you have about it. (It is best if you have an author (if one is listed) and the name of the organization that sponsors the site.) Again, you want to give any information necessary for another researcher to find the facts you did. **Remember—do not include references that you did not cite.**

Examples:

Book: Author(s). Year. *Title*. Location: Publisher. Number of pages, or pages cited, e.g., Hille, Bertil. 1992. *Ionic Channels of Excitable Membranes*. Second Edition. Sunderland, MA: Sinauer Associates, Inc. 607p.

Article: Author(s). Year. Title of Article. *Journal*, volume number, pages. Huxley, A.F. and R. Stämpfli. 1949. Evidence for salutatory conduction in peripheral myelinated nerve fibres. *J Physiol. (Lond.)* 108: 315-339.

Web page: Web page title. Creator or publisher. Subject. URL, date accessed, e.g., The Animated Brain. Brainviews, Ltd. Saltatory conduction. <http://www.brainviews.com/abFiles/AniSalt.htm>, accessed 8/14/07.

It will not be necessary to cite the lab manual for this course—your reader knows that you have used it. (Other professors’ opinions on this vary...)

Presenting your Findings

Prepare a final presentation. Give a broad overview of your gene: what it might do, how you discovered this function, etc.

Follow the guidelines below for making an engaging and informative presentation.

Your talk will be evaluated on

- Content
- Communication (whether it was engaging and clear)
- Organization
- Slide quality
- Professionalism (accuracy and reference to sources)
- Length (did it fit into the allotted time?)
- Presentation style (audibility, enthusiasm, gestures, posture)
- Handling of audience questions

How to make a great presentation:

The tldr version:

Pictures! Short labels! Big font! No reading from the slide!

The longer version:

Show pictures:

The main advantage to using slides over simply giving a speech is the power of pictures. The best talks are spontaneous explanations of *pictures* you show your audience.

Make the picture fill the screen! Don't waste valuable slide real estate with a written explanation of the image – it detracts from the power of the image and your ability to tell the story.

If you have a complicated experimental design, draw a diagram and walk your audience through it. Harness the power of your presentation software to bring one element up at a time, as you get to it in your explanation.

Tables full of multi-digit numbers are hard enough to read on the page right in front of you. They are nearly impossible to make out during a presentation, especially if the audience has to listen to you while simultaneously trying to decipher how many times bigger 431078.009652 is than 3967.031458. Make a figure!

Use only enough text to make your image understandable:

Short, well-placed, easy-to-read labels make an image comprehensible.

Make the labels short and legible. Most of the audience will be a lot farther away from the screen than you are. If your label has to wrap around to the next line, shorten it.

Short, title-like phrases (or even single words!) are sufficient to help the audience follow your reasoning.

Avoid taking up a huge portion of your slide with a not-so-specific word like “RESULTS”.

The detailed figure titles necessary in print are *completely inappropriate* in a presentation – the audience can't read one and listen to you at the same time.

Make it big enough to read from the back of the room:

Project your slide, walk to the back of the room, and check legibility. If you can't do a trial projection, maximize the image on your computer screen and step back from it about 6 feet.

Don't write out your talk on your slide:

First of all, if you can't give your talk without reading from your slides, *you aren't ready!* What does this mean? It means you need to know the material inside and out, and be able to tell the audience about it (while facing them instead of the screen).

Second, if your whole talk is written out on your slides, either you read the slides out loud to the audience, which is boring, or you say different words while they are trying to read the ones on the slide, which is distracting. Don't force your audience to choose between paying attention to you and reading the text.

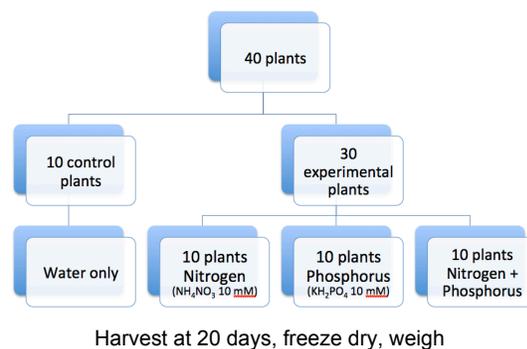
Some Examples:

Compare the following pairs of slides: The left slide in each pair has too small a picture (or none at all) and too much text. The slides on the right, however, contain an image large enough to be seen from a distance, and text (including headings) that enhances audience understanding of the experiment.

EXPERIMENTAL DESIGN

- We obtained 40 plants, all at the early seedling stage, where the cotyledons were first visible above the soil.
- We used 10 of these as controls. They got only water.
- 30 plants were experimental plants. 10 of these got nitrogen supplementation in their water (as NH_4NO_3 – ammonium nitrate – at a concentration of 10 mM).
- Another 10 got phosphorus supplementation in their water (as potassium phosphate monobasic- KH_2PO_4 – at a concentration of 10 mM).
- The last 10 got nitrogen *and* potassium supplementation.
- We harvested all the plants (the above-ground structures) when they were 20 days old, ground them to a powder in liquid nitrogen, and then weighed that powder to get the dry weight in grams.
- We recorded the weight of each plant, and calculated the averages and standard deviations, and ran a t-test.

Fertilizer Experiment



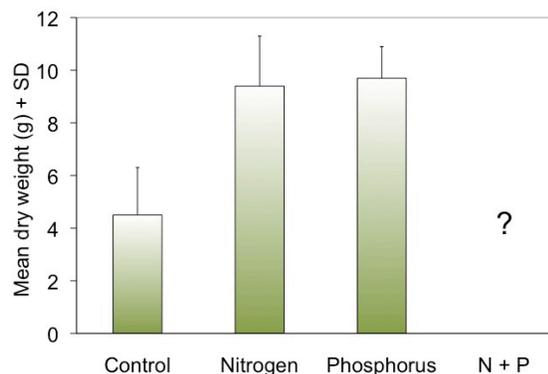
RESULTS:

The control plants had an average dry weight of 4.4967841 ± 1.801024 ; the plants supplemented with nitrogen had an average dry weight of 9.402568 ± 1.899573 ; the plants supplemented with phosphorus had an average dry weight of 9.720315 ± 1.19730487 . There was a significant difference between the dry weight of the control plants and the dry weight of the supplemented plants ($p = 0.019953$ and 0.020396 , respectively). The nitrogen supplemented plants were 2.090954 times heavier than the control plants and the phosphorus supplemented plants were 2.161615 times bigger.

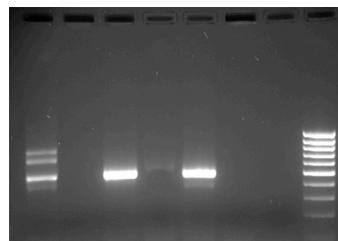
dry weight (g)	control	nitrogen	phosphorus
average	4.4967841	9.402568	9.720315
standard deviation	1.801024	1.89957	1.19730487
fold change =		2.090954	2.161615
p =		0.019953	0.020396

Unfortunately, we were unable to test the simultaneous application of nitrogen and phosphorus due to a growth chamber mix-up. We would have predicted, based on the same papers, that such an application would have produced even greater increases in plant dry weight.

Fertilizer increases plant growth



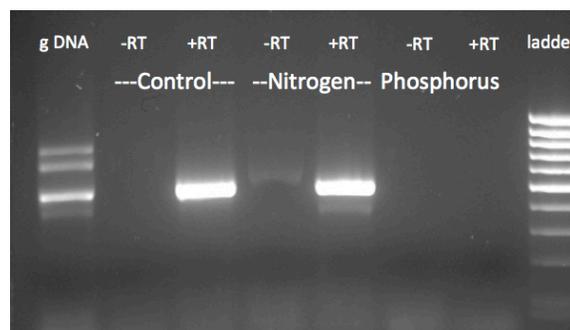
RT-PCR



Lane 1: gDNA; Lane 2: Control -RT; Lane 3: Control +RT; Lane 4: Experimental (Nitrogen) -RT; Lane 5: Experimental (IAA) +RT; Lane 6: Experimental (Phosphorus) +RT; Lane 7: Experimental (Phosphorus) -RT; Lane 8: Fermentas MassRuler, LMW

As can be seen in this gel, our RT-PCR worked, because the lanes without RT have no bands, and the lane with nitrogen treatment and RT have cDNA bands. We did not get an RT band for the phosphorus treated plants. Also, there is approximately the same amount of cDNA in the two RT lanes, showing that our nitrogen treatment did not increase or decrease the amount of gene expression. This is a surprise, because we designed this experiment based on several papers that suggested an effect of nitrogen on the expression of our gene (Brown & Green, 2001, Farmer & Cook, 2009, Gardner & Chamber, 1999).

N does not affect gene expression



Comparing Means: The t-Test

The Mean (Average).

This is probably the most common measure of central tendency. It is calculated by dividing the sum of all the data values by the number of such values:

$$\text{mean} = \bar{x} = \frac{x_1 + x_2 + x_3 \cdots x_n}{n}$$

Excel will calculate the mean for you with the following formula:

=average(data array).

data array: You type in (or click and drag over) the address of the data you want averaged.

Function wizard: You can also use the function wizard (f_x) to calculate the average: pick a cell to contain the average, click on f_x , choose AVERAGE (you may have to hunt for it in the statistical menu), highlight the values to average, and click OK. The cell will contain the formula given above.

The Standard Deviation

The standard deviation is one of the most commonly used and easiest to understand measures of spread. It also has some nice properties that will be described below.

The standard deviation is something like the average of all the individual deviations from the mean. This is a tedious calculation to do, so we usually ask a computer to do it for us (although generations of students before you managed with nothing more than calculators, and slide rules before that).

The calculation is done as follows: each datum is subtracted from the mean of all data (these are the individual deviations). About half of these deviations will be negative, and half positive, and if you add them together, they cancel each other out. To correct this, the deviations are squared so they will all be positive. These squares are added together, and their sum is divided by the number of data (actually, one less than the number of data – sorry) to get the “average”. Finally, the square root of this average is taken to correct for the squaring done earlier.

$$sd = \sqrt{\frac{\sum (\bar{x} - x_i)^2}{n - 1}}$$

Where

sd is the standard deviation,

n is the number of data,

x_i is each individual measurement,

\bar{x} is the mean of all measurements, and

Σ means the sum of

Excel will calculate the standard deviation for you with the following formula:

=stdev(data array).

One nice feature of the standard deviation alluded to above is that it is measured in the same units as the mean, so it is meaningful to add it to or subtract it from the mean.

A second nice feature is that when the data are distributed symmetrically around the mean (that is, when they fall into the famous bell curve of song and legend), between one standard deviation above the mean and one below are found 68% of the data, and between two deviations above and two deviations below the mean are found 95% of the data. This property is illustrated in Figure 2.

The bigger the standard deviation, the wider the spread of data around the mean, as illustrated in Figure 3.

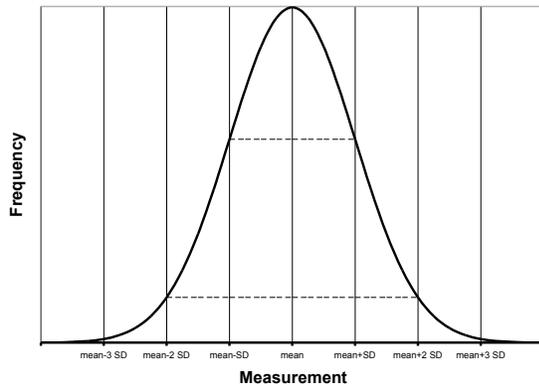


Figure 2. A “bell curve”, showing the symmetrical distribution around the mean. Horizontal lines indicate 1 and 2 standard deviations from the mean.

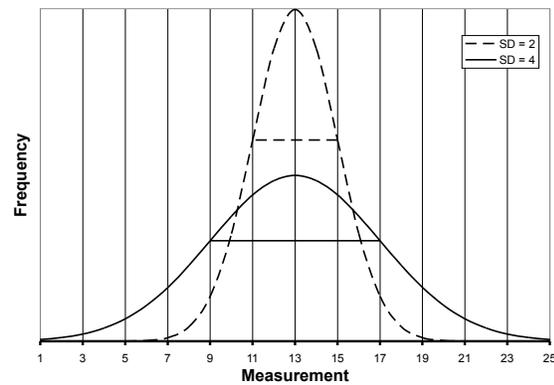


Figure 3. Two “bell curves”, with the same mean (13), but different standard deviations. Horizontal lines indicate 1 standard deviation from the mean.

Comparing The Means.

Usually, even when the means of two groups differ, there is some overlap between the two distributions. How different the two groups “really” are depends, therefore, not only on the difference between their means, but also on the extent of the overlap between their distributions.

Illustrating the relationship.

The standard deviation can help to make a more complete comparison between two sets of data. (Here’s where another nice feature of the standard deviation comes into play: it is in the same units as the mean, so they can be added together.) Figure 4 shows why it is important to report not only the difference in the means between your two groups, but also some measure of the variation in each one. The means of groups 1 and 2 differ from each other by the same amount as do 3 and 4, yet the error bars, illustrating the size of the standard deviation, indicate a much greater degree of overlap between groups 3 and 4.

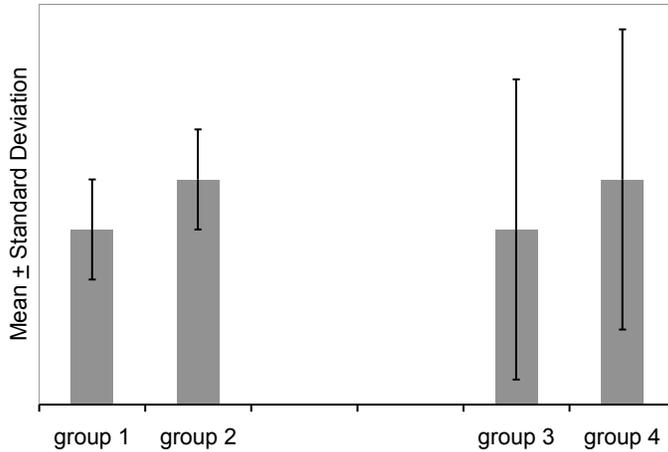


Figure 4. The effect of standard deviation on a comparison between means. Approximately two thirds of the values in each group lie within the one standard deviation error bars. There is much greater overlap between the measurements in groups 3 and 4 than between those in 1 and 2.

The easiest way to create a histogram comparing the means, with standard deviation as the error bars, requires making a little table in Excel that looks like this:

	mean	St dev
control		
experimental		

Choose appropriate descriptive labels for your two groups, as these will appear in your graph. Type the formula for the mean and standard deviation in the appropriate cells, i.e., =average(data array) and =stdev(data array), respectively.

Highlight the cells containing the labels and the averages, shown above with a double outline, and use the chart wizard to make a bar graph. Then highlight the bars and format the data series. Make *custom* Y-error bars, using the two cells in which you have calculated standard deviation. **Note: the “standard deviation” choice inside the Y-error bar dialog box is not what you think it is! Don’t use it! You must calculate the standard deviation in a cell.** (The dialog box has no idea which data went into making the means you are graphing, and cannot possibly calculate their standard deviation. Instead, it calculates the standard deviation of the values you are plotting – in this case the control mean and the experimental mean, and plots an error bar that length starting from the average of those values. I don’t know when this would be useful.)

Quantifying the relationship: calculating “t”.

The statistic “t” is a measure of the difference between two means, divided by the geometric mean of the standard errors of the population means (a sort of average of the standard deviations of the two populations). (The manner of calculating t depends on various characteristics of the experiment and the data, so this is why you must specify a “test type” before you ask Excel to calculate t for you.)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}}}$$

The value of t gets larger as the difference between the means gets larger; but this is counterbalanced by this measure of spread in the denominator. The greater the standard deviations, or the smaller the sample sizes (n), the bigger a difference in means is required to make t large. You can think of this as a ratio of signal (the difference between the means) to noise (the variation within the population).

Figure 5 illustrates the effect of standard deviation on the t statistic. The same difference between means can be significant or not, depending on the amount of variation in the populations being compared.

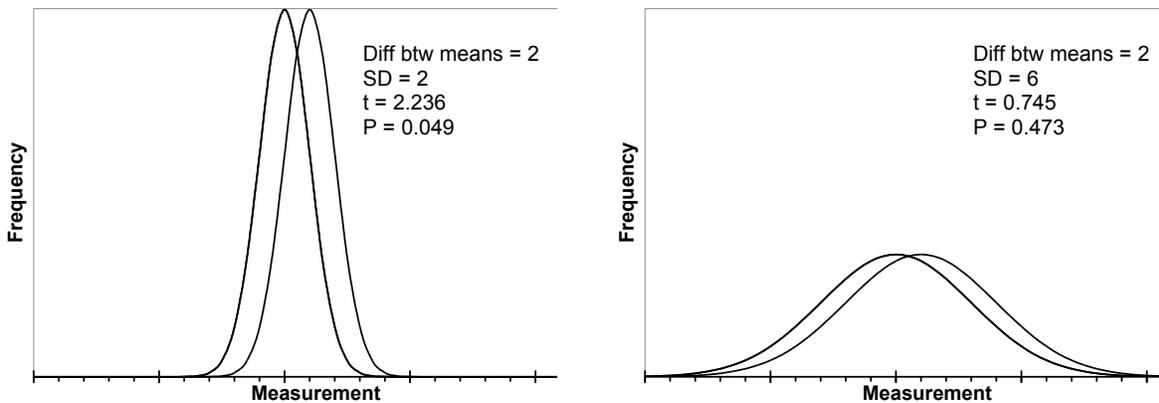


Figure 5. The effect of standard deviation on the t statistic. The means of both pairs of distributions differ by the same amount, yet the t statistic is 2.236 for the pair with the smaller standard deviations, and only 0.745 for those with the larger standard deviations. Notice the greater overlap between the curves on the right. (T values calculated for population size of 10 in each group.)

Interpreting the relationship.

Tests of significance always set up a straw man, the *null hypothesis*, which is that there is no difference between the groups, or no relationship between the variables. Then we find out how likely we are to get a result like our actual result if the null hypothesis were true. From this, we decide to accept or reject the null hypothesis.

If the category in question (*e.g.*, control or treatment) had no effect on the variable being measured (*e.g.*, number of cells in a suspension), you might as well be assigning categories at random. Imagine putting stickers on tubes of cells at random: some get labeled “group 1”, others get labeled “group 2”. Occasionally, such a process would result in all the tubes with dense populations being labeled “group 1”, and all those with very few cells, “group 2”. Such a result would give a rather large t.

How often would a t as large as the one you got in your actual experiment be expected to occur if the stickers were put on at random? The probability of getting a t as large as or larger than yours is called ‘p’, and fortunately, we can look up the p for a given t and number of measurements in any statistics book, on the internet, and inside of Excel.

If a t like yours is rather likely to occur by the chance procedure, you cannot rule chance out, and you must accept that there is no difference between the means of the two groups.

However, if such a t as yours would be exceedingly *uncommon* in a chance process, you may reject the idea that there is no difference between the means, and conclude that there is a *statistically significant difference* between the two groups. The statistic tells you how often a result like yours might occur by chance alone; *it cannot tell you the probability that chance actually caused the difference you observed.*

Conventionally, if a t as big as or bigger than yours can be expected to occur less often than 5% of the time ($p < 0.05$) when there is nothing other than chance acting, we reject the null hypothesis, and conclude that there is a difference between the groups.

By this criterion, the two distributions illustrated on the left in Figure 5 are statistically significantly different from each other, while those on the right are not.

Calculating T With Excel (the quick and dirty method)

If you have both sets of data to compare, use the built-in t-test in Excel (which you can find the statistics category of the function menu:

=ttest(*data array 1, data array 2, number of tails, test type*)

data array 1: the first set of data (enter the addresses, or click and drag over them)

data array 2: the second set

number of tails: 2 if you can't predict how one group will differ from the other, and think the means might vary in either direction, 1 otherwise. (Use 2 most of the time.)

test type: 1, 2, or 3

- 1 for paired data *i.e.*, two measurements on the same thing. In this test type, Excel takes the differences between the paired measurements, rather than between each one and the mean of the population. This works if you are comparing, for example, the absorbance at two different times *of the very same culture* (as opposed to the absorbance at the same time in two different cultures).
- 2 for unpaired data, but where both sets have the same standard deviation (don't use this one – it is unlikely to have two sets of data with identical standard deviations).
- 3 for unpaired data, with unequal standard deviations (two separate sets of measurements, such as number of cells in treated cultures vs. number of cells in control cultures). This is the most likely situation.

Excel will return the value of p that goes with this, but not tell you the actual t. In other words, Excel gives you just the punch line: what is the probability of grabbing two handfuls of data at random from a single population and winding up with two subsets as different from each other as your two data sets are from each other.

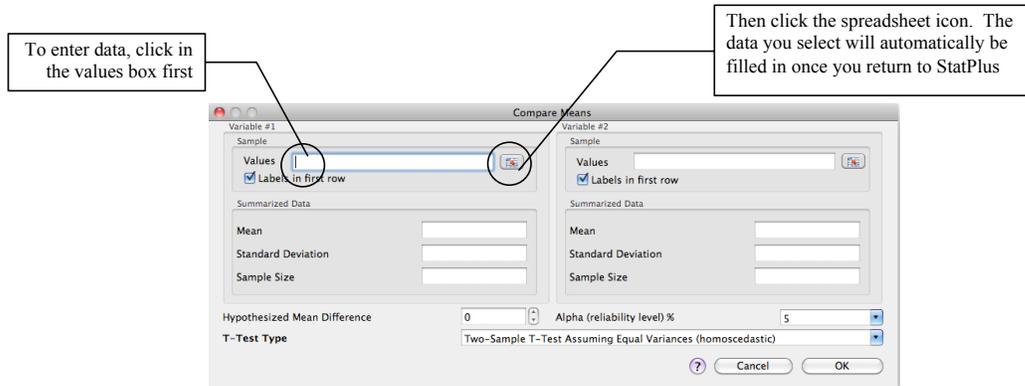
If you don't have both sets of data (if you are comparing your measurements to a standard reported in the literature only as mean plus or minus standard deviation), you have to do more work. Ask your instructor for guidance.

Calculating T With StatPlus.

If StatPlus is in the dock, open it by clicking on it.

If it isn't in the dock, open it from the applications folder. Doing so will put an icon for it  in the dock, and put its main menu bar (StatPlus Spreadsheet Statistics Data Charts Help) at the top of the screen.

Choose Basic Statistics and Tables from the Statistics menu, and Comparing Means (t-test) from that menu. That should pull up a dialog box that looks like the one below.



To enter your data, click in the Values box, then click on the little spreadsheet icon , which should take you to your spreadsheet if it is open, or to a new Excel file. From that new blank file, you can open your data file. Select the data for the first variable, return to StatPlus, and repeat for the second variable.

Check the settings:

If you included labels in the selection make sure StatPlus knows there are labels in the first row.

Leave the summarized data blank.

Set the hypothesized mean difference to zero, (because the null hypothesis is that the two collections of data were chosen at random from the same population) and

Set the alpha (the maximum probability at which you will reject the null hypothesis) to 5% (= 0.05) by convention.

For T-Test Type, choose the two-sample test assuming unequal variances (heteroscedastic). You are not comparing paired data, and your two sets of values almost certainly do not have the same variance (standard deviation squared).

When you click OK, it chugs for awhile, then opens a new file called StatPlusMacResults.xlt, with something like this on a tab called Comparing Means:

Comparing Means [t-test assuming unequal variances (heteroscedastic)]

<i>Descriptive Statistics</i>			
<i>VAR</i>	<i>Sample size</i>	<i>Mean</i>	<i>Variance</i>
	9	6.77778	1.94444
	8	9.375	1.125
<i>Summary</i>			
<i>Degrees Of Freedom</i>	15	<i>Hypothesized Mean Difference</i>	0.E+0
<i>Test Statistics</i>	4.34884	<i>Pooled Variance</i>	1.56204
<i>Two-tailed distribution</i>			
<i>p-level</i>	0.00057	<i>t Critical Value (5%)</i>	2.13145
<i>One-tailed distribution</i>			
<i>p-level</i>	0.00029	<i>t Critical Value (5%)</i>	1.75305
<i>G-criterion</i>			
<i>Test Statistics</i>	#N/A	<i>p-level</i>	#N/A
<i>Critical Value (5%)</i>	#N/A		
<i>Pagurova criterion</i>			
<i>Test Statistics</i>	4.34884	<i>p-level</i>	0.99938
<i>Ratio of variances parameter</i>	0.60573	<i>Critical Value (5%)</i>	0.0255

In this output table,

VAR: Probably short for *variable*. This should have the labels from your two columns of data. If not, type them in before you forget which comparison this is.

Sample Size: Sample size tells you the number of values in each group. Use it to do a reality check: do the sample sizes reported by StatPlus match the number of values in each of your categories? If not, you should repeat your data selection.

Mean: Is this the same as what you had Excel calculate for you on your spreadsheet? (Another reality check.)

Variance is standard deviation squared. It is another measure of the spread of data around the mean.

Degrees of freedom is the total number of data points minus 2 (it's the number of data points that are free to vary before the remaining ones are set).

The **Hypothesized Mean Difference** should be 0 (given as 0.E+0 in Excel-style scientific notation).

Test Statistics is your t value.

The Pooled Variance is an estimate of variance that assumes the true variance of the two samples is the same. You can ignore it.

The p-level (the probability of getting results this different from each other if the two sets of values were really drawn from the same pool) is given twice, once for a one-tailed distribution, and once for a two-tailed distribution.

In general, you should use the p value for the **two-tail** test, since this tests for both positive and negative differences between the means. The one-tail test gives you greater power to

detect a difference in only one direction, but increases the possibility of getting a false indication of significant difference.

The **t Critical value (5%)** is the t associated with that probability. If your t is greater than the critical value, your p-value is less than 0.05. Remember, the bigger the difference between the two groups, the higher the t and the lower the probability of getting a t that big by random factors alone.

It is OK for our purposes to ignore the **G** and **Pagurova criteria**.

This output table is static. If you change anything about your data, you must re-do the analysis.

Excel reports the values with up to nine decimal places. ***This in no way obligates you to do the same!*** Round the values to the number of significant digits of your least precise measurement. You can do this by formatting the cells in Excel, or correcting the number when you paste a result into Word. Using extra significant digits makes false claims about your ability to make fine distinctions, e.g., reporting a weight as 4.98730958 grams means you can tell the difference between a specimen weighing 4.98730958 g and one weighing 4.98730959 g.

Multiple comparisons – a caution

What if you have more than two sets of data to compare (e.g., two experimental treatments and one control)? It is tempting to do multiple t-tests to find out which means are different from each other. However, this should only be undertaken with great caution, because the more comparisons you make, the more likely you are to find high t-values by chance alone. Remember, the p-value is a calculation of the likelihood of getting a t as high as yours even if the values in both groups were actually drawn from the same population (*i.e.*, by chance), so if you did 100 comparisons, you can predict that chance alone would produce about 5 “significant” results. See Adjusting the critical α , on p. 26.

Analysis of Variance

One improvement on multiple comparisons is to begin with an analysis of variance (ANOVA) on all the data first, to find out whether the variation between groups differs from the variation within each group. ANOVA tests the null hypothesis that there are no differences between the groups, that in fact, all the data are pulled from what is effectively a single population.

ANOVA uses the F-statistic, which is the ratio between the mean square (a measure of variation) between groups and the mean square within groups. The null hypothesis is that the variation between groups is no bigger than the variation within groups, or that $F \leq 1$. As with the t-test, a large F statistic occurs rarely when the groups are the same, so the associated p-value is a measure of how likely it is to get an F as large as yours by chance alone.

Use the one-way ANOVA in StatPlus, assuming you have only one variable (e.g., “treatment”), which consists of different conditions (e.g., “control”, “treatment 1”, “treatment 2”, etc.). Set your data up in columns, with labels in the first row.

Select one-way ANOVA from the StatPlus Statistics menu, click in the “variables” box, click on the spreadsheet icon, find your datafile, and select your data including the labels. Let it run. It will produce a spreadsheet with a table like this:

Analysis of Variance (One-Way)

Summary					
<i>Groups</i>	<i>Sample size</i>	<i>Sum</i>	<i>Mean</i>	<i>Variance</i>	
<i>control</i>	7	20.	2.85714	1.80952	
<i>treatment 1</i>	7	4.	0.57143	0.28571	
<i>treatment 2</i>	7	9.	1.28571	0.57143	

ANOVA						
<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p-level</i>	<i>F crit</i>
Between Groups	19.14286	2	9.57143	10.76786	0.00084	4.90007
Within Groups	16.	18	0.88889			
<i>Total</i>	35.14286	20				

In this output table:

Groups should accurately label your sets of data. If not, repeat your data selection.

Sample size should accurately reflect the number of data points in each group. If not, repeat, with correct selection of data from your spreadsheet.

Sum should be what you get if you add up all the values in one group. The **Mean** should be the average of the data in the group, and **Variance** is standard deviation squared.

SS stands for the sum of (deviation) squares. It should remind you of the standard deviation. Squaring the difference between the individual values and the mean or between the group means and the overall mean gets rid of the negative values and magnifies the impact of a large deviation. SS_{between} is the sum of the squares of the differences between each group mean and the total mean, times the number of scores in the group, and SS_{within} is the sum of the squares of the differences between every raw score and its sample mean.

$$SS_{\text{between}} = \sum (\bar{x}_i - \bar{x}_{\text{total}})^2 N$$

where

$$\begin{aligned} \bar{x}_i &= \text{the mean of group } i \\ \bar{x}_{\text{total}} &= \text{the mean of all raw scores combined} \\ N &= \text{the number of scores in group } i \end{aligned}$$

$$SS_{\text{within}} = \sum (x_i - \bar{x}_i)^2$$

where

$$\begin{aligned} x_i &= \text{raw score } i \\ \bar{x}_i &= \text{the mean of } i\text{'s group} \end{aligned}$$

Df, or degrees of freedom, is the number of values that are free to vary and still give you the same statistic. Within groups, this should be 1 less than the number of groups. Between groups, this should be the total number of samples minus the number of groups.

MS stands for mean square, or variance, and it is the ratio between SS and df. This yields something like an average, and corrects for the number of scores.

F is the ratio between MS_{between} and MS_{within} ($= \frac{MS_{\text{between}}}{MS_{\text{within}}}$). The null hypothesis is that the variation between groups is no bigger than the variation within groups, or that $F \leq 1$. A larger **F** indicates greater variance between groups than within each group.

The **p-level** is the probability of getting an **F** as big as yours by chance alone, even if all the data come from the same population.

The **F crit** is the **F** associated with the p-value (probably 0.05) you specified when you set up the test. It is the smallest **F** that would let you conclude your groups differ statistically from each other.

If you get a statistically significant result (i.e., a p less than 0.05) with ANOVA, you can test the treatment groups in pairs by the t-test. Doing so, however, increases the chance of finding spurious “significance.” You can guard against that in various ways. The two simplest are to adjust the α (that is, the maximum p-value you will use to reject the null hypothesis of no difference between the groups), or caution your reader that you have performed multiple comparisons.

Adjusting the critical α

The probability of getting a large t rises with the number of t-tests you perform. Therefore you should tighten your decision criterion a little. The easiest way is to correct the critical α by dividing your usual α (0.05 for most purposes) by the number of pairwise comparisons you made. This protects you very well against claiming statistical significance where none exists, but not so well against claiming no difference when there is one.

For small numbers of comparisons, this is nearly identical to the next simplest correction, calculated as $1-(1-\alpha)^{1/n}$. As the number of comparisons grows, this correction stops the α from slipping into infinitesimal territory. This helps prevent you from being unable to detect any differences at all.

You can be very confident of any difference that passes either of these corrected α tests.

There are other ways to guard against spurious significance that are beyond the scope of this chapter, but which a statistician would be delighted to teach you.

Reporting Your Statistical Results.

Use the criteria above, that is, a significance level of $p \leq 0.05$, to decide whether the values in two sets of data can be called statistically different from one another.

Remember that the p value only tells you how likely it is to get a t as big as yours if the two samples really were picked from the same population, so resist the urge to say you’ve proved chance did or did not cause the difference you observe. Here are a few conventional expressions of such a result:

The mean [whatever was measured] of Group 1 (5 ± 2 , $n = 10$) is statistically lower than that of Group 2 (7 ± 2 , $n = 10$) by the t-test ($t = 2.236$, $p = 0.049$).

Group 2 had significantly higher [whatever was measured] by the t-test ($t = 2.236$, $p = 0.049$).

There was no statistically significant difference between Groups 3 and 4 by the t-test ($t = 0.745$, $p = 0.475$).

The Chi Square Test for Counted Data

One test for statistical significance applicable to many experiments that count data in categories (e.g., number of cells in particular phases of mitosis) is the Chi (pronounced like *sky* without the *s*) squared (χ^2) test. It tests for whether the values are distributed in the categories as predicted (expected) by chance.

$$\chi^2 = \sum_{i=1}^m \frac{(o_i - e_i)^2}{e_i}$$

χ^2 is based on the difference between a series of observed values (o_1, o_2, \dots, o_m) and expected values (e_1, e_2, \dots, e_m). Bigger deviations from the expected yield a bigger χ^2 .

The differences between observed and expected are squared for two reasons. The first is that this makes all the terms positive so the negative ones don't cancel out the others. The second reason is that this accentuates large differences and minimizes small ones.

The squared difference is divided by the expected value to obtain something like the percent deviation, thereby weighting the individual $\frac{(o_i - e_i)^2}{e_i}$ terms so that the relative deviation rather than the absolute deviation is used in calculating the statistic.

The individual $\frac{(o_i - e_i)^2}{e_i}$ terms are summed to give an overall indicator of the difference between observed and expected across all the categories.

χ^2 lets you determine the probability that the deviation from your prediction was due to chance alone.

Example:

For the entering class in 2012, UMass accepted 11,918 out of 18,006 female applicants, and 9552 out of 16320 male applicants*. Is this distribution according to chance, or is there some non-random factor at work? Imagine the admissions process taking place by the admissions officials putting all the applications into a giant bag, and pulling out 21,470 (11918 + 9552) without looking. How likely would they be to pull out 11,918 applications from females and 9552 from males?

It helps to look at a 2-way table:

gender	accepted	rejected	total applications
female	11,918	6088	18,006
male	9,552	6768	16,320
totals	21,470	12,856	34,326

What would we expect from the grab-bag admissions technique? The applicant pool was 52.5% female (18,006/34,326), so the population of admitted students should also be 52%

* UMass Amherst At a Glance 2012-2013, http://www.umass.edu/oapa/publications/glance/FS_gla_01.pdf, accessed 8/26/13

female. Similarly, 62.5% of all applicants were accepted, so we expect 62.5% of the female applicants to be accepted. Here's the two-way table again, with the expected values calculated:

gender	accepted	rejected	total applications
female	$\frac{18006}{34326} \times 21470 = 11,262.3$	$\frac{18006}{34326} \times 12856 = 6743.7$	18,006
male	$\frac{16320}{34326} \times 21470 = 10,207.7$	$\frac{16,320}{34326} \times 12856 = 16,320$	16,320
totals	21,470	12,856	34,326

$$\chi^2 = \frac{(11918 - 11262.3)^2}{11262.3} + \frac{(9552 - 10277.7)^2}{10277.7} + \frac{(6088 - 6743.7)^2}{6743.7} + \frac{(6768 - 6112.3)^2}{6112.3} = 1.5 \times 10^{-48}$$

This particular calculation gives us a minute χ^2 , because the observed values are actually quite close to the expected ones.

The calculated χ^2 by itself has very little meaning. The significance (literally) comes from calculating the probability of getting a χ^2 at least this big by random assignment of individuals to categories. Imagine the admissions committee doing the grab-bag technique over and over and over. Most of the time, a random assignment procedure will produce results similar to the prediction.

Every once in a while, however, the person pulling applications from the bag to accept will grab mostly applications from women, or mostly from men. This would produce a much larger χ^2 than we calculate from the actual data.

In our actual UMass admissions example, almost any result other than the reported one would give a χ^2 bigger than ours, so the probability of getting results at least this different from the expected is quite high. (In fact, Excel reports it as 1.)

The one remaining element needed in order to calculate p from χ^2 is the degrees of freedom, which is the number of categories you have to count to be able to fill in the rest by subtraction. This assumes you know all the subtotals. The degrees of freedom = m-1, or one less than the number of categories.

You can get Excel to do the heaviest of the lifting for you, that is calculating χ^2 from arrays of observed and expected values, and calculating p from the χ^2 . However, you do have to get Excel to calculate the expected values for each category. Here's how I did the calculations for the application example:

I typed in the raw data from the UMass website, then used formulas to calculate the subtotals and the grand total:

	A	B	C	D
1	gender	accepted	rejected	total applications
2	female	11918	6088	18006
3	male	9552	6768	16320
4	totals	21470	12856	34326

Then I made another table for the expected values, and typed in the appropriate formulas to have Excel do the calculations for me.

	A	B	C	D	E
1	gender	accepted	rejected	total applications	
2	female	11918	6088	18006	
3	male	9552	6768	16320	
4	totals	21470	12856	34326	
5					
6					
7	expected	11262.3	6743.7		
8		10207.7	6112.3		
9					

I used the formula builder to calculate χ^2 with the function CHISQ.TEST. You can see why it matters how you array your observed and expected values. Their relative positions must match.

CHISQ.TEST =CHISQ.TEST(B2:C3,B7:C8)

CHISQ.TEST(actual_range, expected_range)

CHISQ.TEST

Description
Returns the test for independence: the value from the chi-squared distribution for the

More help on this function

Arguments

CHISQ.TEST

actual_range	B2:C3	{11918,6088}
expected_range	B7:C8	{11262.274}

Result: 1.50116E-48

Finally, I calculated p with the function CHISQ.DIST.RT. (It only examines the right side of the χ^2 distribution, because that's where the larger χ^2 values are, and we are only interested in the probability of getting a χ^2 value as high as or higher than our calculated value – but that's probably more than you wanted to know.)

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D
1	gender	accepted	rejected	total applications
2	female	11918	6088	18006
3	male	9552	6768	16320
4	totals	21470	12856	34326
5				
6				
7	expected	11262.3	6743.7	
8		10207.7	6112.3	
9				
10				
11	X2	1.5E-48		
12	p=	CHISQ.DIST.RT(B11,1)		
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				

The Formula Builder pane shows the following information for CHISQ.DIST.RT:

- Description:** Returns the right-tailed probability of the chi-squared distribution.
- SYNTAX:** CHISQ.DIST.RT(x,deg_freedom)
- Arguments:**
 - x: B11 (1.501E-48)
 - deg_freedom: 1 (1)
- Result:** 1

Remember, when you have finished these calculations, you do not have *proof* that chance did or did not cause your results; you simply know how likely it is to get results like yours if chance were the only factor. If results as skewed as yours would be unlikely to occur by chance alone (less than 5% of the time), then you can reject the hypothesis that your results are randomly assigned to categories. In the case of the UMass admissions data, we can conclude that there is no apparent gender bias in acceptance rates. (See page 26 for a further discussion of the meaning and reporting of statistical results.)

The Q-test for the Outlying Data Point

Sometimes you collect data by repeated measurements that seem to cluster around a certain value, except for one odd data point. You want to be honest; you really did get that measurement, and you don't want to cheat. However, it does seem to be from another universe. For example, suppose these are the readings from your experiment:

25.1 21.2 27.5 22.7 23.8 26.3 40.6 22.9

That 40.6 not only looks out of place, but it has an effect on the mean and standard deviation of your data set:

	mean	standard deviation
including maximum value	26.1	6.3
excluding maximum value	24.0	2.1

It may also have an effect on statistical significance when you compare this distribution to another one by a t-test.

Is it a mistake? Is it noise? Did your recording device hiccup? What is the likelihood of getting an outlier this far from the next value when you are picking from a population with these characteristics?

Chemists (and other scientists) often test for outliers with the Q-test. This test calculates the ratio between the putative outlier's distance from its nearest neighbor and the range of values:

Notice that as the distance between the potential outlier and its nearest neighbor increases, so does Q.

$$Q = \frac{|\text{potential outlier} - \text{nearest neighbor}|}{\text{maximum value} - \text{minimum value}}$$

For our hypothetical data listed above,

$$Q = \frac{|\text{potential outlier} - \text{nearest neighbor}|}{\text{maximum value} - \text{minimum value}} = \frac{40.6 - 27.5}{40.6 - 21.2} = 0.675$$

The significance test consists of comparing your calculated Q to the theoretical Q that is expected to occur 5% of the time if you were sampling from a population with this mean and standard deviation. The null hypothesis is that all your data points come from the same pool. If your Q is as big as or bigger than the critical Q, then you can reject that null hypothesis with 95% confidence, and safely exclude the one odd data point.

Table 1, on page 32, lists critical values of Q at the 5% significance level. Notice that the smaller the number of data points, the larger Q must be for you to reject a data point. For a set of 8 data points, according to the table, Q would be greater than 0.526 just by luck of the draw 5% of the time or less. Since our experimental Q (0.675) is larger than 0.526, we may safely omit the 40.6 value from our data set, because such a large value is unlikely to occur by chance alone. We can conclude that that data point measures something other than what all the others are measuring. To maintain honesty, we can report something like this: "One very large (or small) value was rejected from the data set by the Q-test with 95% confidence."

You may only reject one data point from a data set by this method.

Excel Tricks to help you calculate Q

Sort

Select your data, then choose Sort from the Data menu. This makes visual inspection for outliers easier. *Note: if there are identifiers adjacent to the data values, be sure to include them in the sort, or you will lose the connection between the labels and the values!*

MAX

From the f_x menu, under statistical. This returns the maximum value from a list of values. The result goes in the cell where you typed the equation. You can specify the data array either by typing the range of cells where your data are found, or by clicking and dragging over them.

=MAX(data array)

MIN

This function works just like MAX, but tells you the minimum value in a series.

=MIN(data array)

COUNT

This tells you how many numerical values there are in an array. (If you use COUNTA, it will return the number of non-empty cells, rather than the number of data points.)

=COUNT(data array).

LARGE

This function tells you the kth largest value in a series. Set $k = 2$ to find the second largest value. Set $k = n-1$ to find the next-to-last value in a set of n data points.

=LARGE(data array, k)

SMALL

This function tells you the kth smallest value in a series. Set $k = 2$ to find the second smallest value.

=SMALL(data array, k)

Table 1. Critical values of Q at the 95% confidence ($\alpha = 0.05$) level, for data sets up to $n = 30$

number of data points[*]	Q_{0.05}
3	0.970
4	0.829
5	0.710
6	0.625
7	0.568
8	0.526
9	0.493
10	0.466
11	0.444
12	0.425
13	0.410
14	0.396
15	0.384
16	0.374
17	0.365
18	0.356
19	0.349
20	0.342
21	0.337
22	0.331
23	0.326
24	0.321
25	0.317
26	0.312
27	0.308
28	0.305
29	0.301
30	0.298

* Table from David B. Rorabacher, 1991. Statistical treatment for rejection of deviant values: critical values of Dixon's "Q" parameter and related subrange ratios at the 95% confidence level. *Anal. Chem.*, 63 (2): 139–146

<http://pubs.acs.org/doi/abs/10.1021/ac00002a010>, accessed 9/10/09.

DNA, Amino Acids, and the Genetic Code

DNA

DNA is a polymer of the four nucleotides illustrated in Figure 6. Each nucleotide consists of a phosphate, a sugar (deoxyribose), and one of four nitrogenous bases.

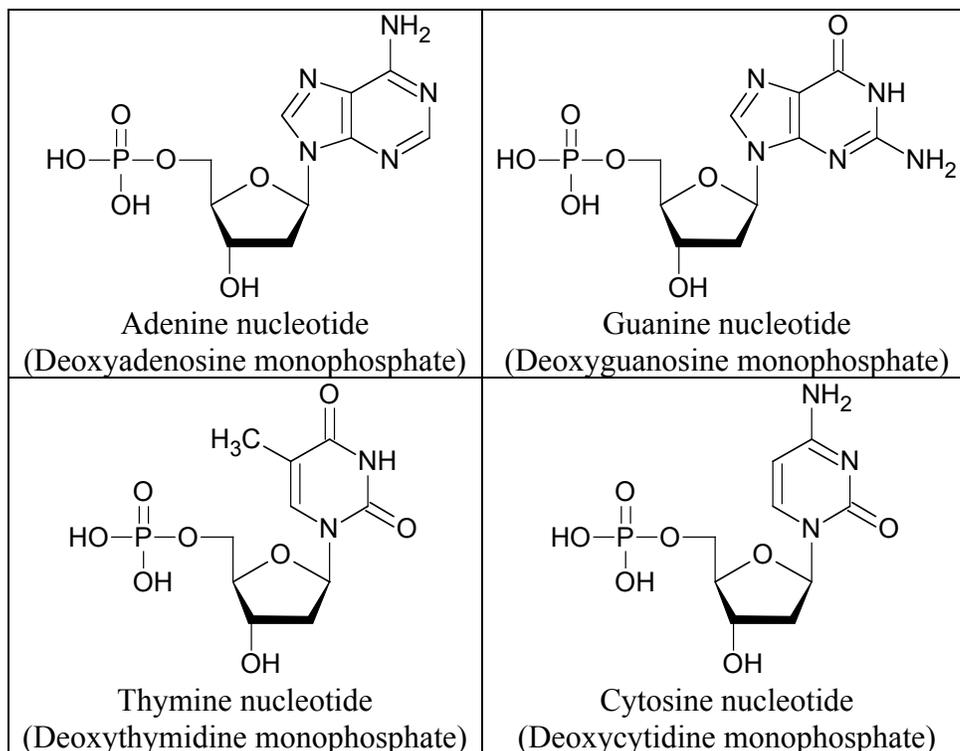


Figure 6. The four nucleotides that make up DNA

The phosphate of one nucleotide is covalently bound to the hydroxyl group of the sugar on the next, producing a backbone of alternating sugars and phosphates, as shown in **Figure 7**.

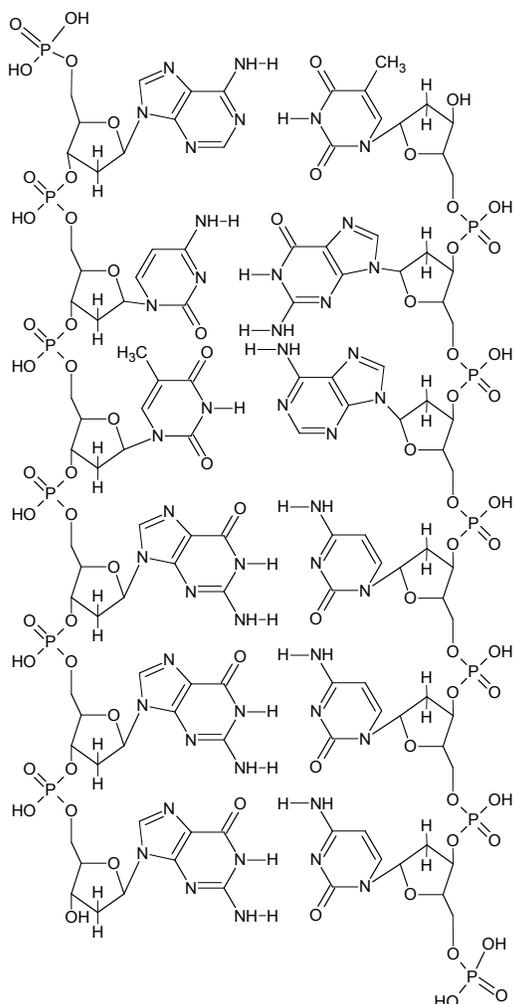


Figure 7. A segment of DNA. The left strand (reading top to bottom) is 5'-ACTGGG-3' . Its complement on the right is 5'-CCCAGT-3'.

DNA is double stranded (almost always in nature, and always in this course, except for your primers).

The two DNA strands are complementary because of base pairing. (Adenine forms two hydrogen bonds with thymine, and cytosine forms three with guanine.)

They are also anti-parallel. One runs from 5' to 3', the other, from 3' to 5'

The phosphate is on the 5' end, the hydroxyl on the 3' end.

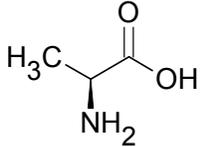
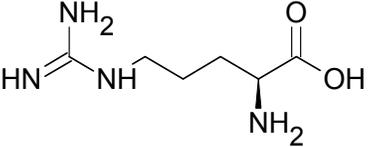
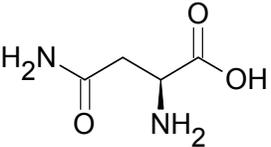
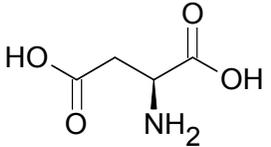
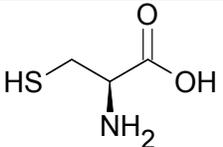
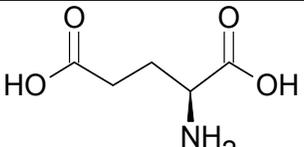
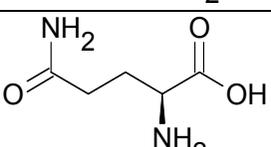
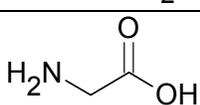
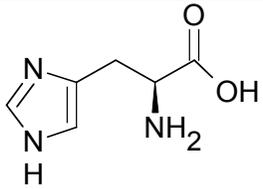
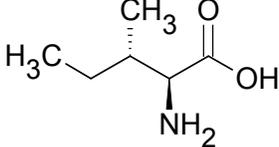
The sense (or coding) strand of the DNA contains the same sequence as the mRNA.

The antisense (or non-coding) strand is complementary to the sense strand and is the template for the mRNA.

Figure 7 shows the sequence ACTGGG. Note the following conventions:

- The letter of the base (A, C, T, G) stands for the whole nucleotide
- The sequence is written from 5' (on the left) to 3' (on the right)
- The 5' and 3' ends are typically not marked.
- The complementary strand is not written down.

The Amino Acids

Amino Acid	1-letter code	3-letter code	Structural Formula
Alanine ^o	A	Ala	
Arginine ^c	R	Arg	
Asparagine ^p	N	Asn	
Aspartic Acid ^c	D	Asp	
Cysteine ^o	C	Cys	
Glutamic Acid ^c	E	Glu	
Glutamine ^p	Q	Gln	
Glycine ^o	G	Gly	
Histidine ^c	H	His	
Isoleucine ^{lh}	I	Ile	

c = charged, p = polar, lh = large, hydrophobic, o = other

Amino Acid	1-letter code	3-letter code	Structural Formula
Leucine ^{lh}	L	Leu	
Lysine ^c	K	Lys	
Methionine ^{lh}	M	Met	
Phenylalanine ^{lh}	F	Phe	
Proline ^o	P	Pro	
Serine ^p	S	Ser	
Threonine ^p	T	Thr	
Tryptophan ^{lh}	W	Trp	
Tyrosine ^{lh}	Y	Tyr	
Valine ^{lh}	V	Val	

c = charged, p = polar, lh = large, hydrophobic, o = other

The Genetic Code

The sequence of nucleotides in the sense strand of DNA specifies the sequence of amino acids in a protein. After the anti-sense strand is transcribed into mRNA, the sequence is translated into protein according to the following rules:

First nucleotide	Second nucleotide				Third nucleotide
	U	C	A	G	
U	F	S	Y	C	U
	F	S	Y	C	C
	L	S	stop	stop	A
	L	S	stop	W	G
C	L	P	H	R	U
	L	P	H	R	C
	L	P	Q	R	A
	L	P	Q	R	G
A	I	T	N	S	U
	I	T	N	S	C
	I	T	K	R	A
	(start) M	T	K	R	G
G	V	A	D	G	U
	V	A	D	G	C
	V	A	E	G	A
	V	A	E	G	G

DNA Molecular Weight Standard

Many times in this course you will need to know the length of a fragment of DNA that you have amplified by PCR or digested with restriction enzymes, or produced by reverse transcription. This is done by loading the DNA in an agarose gel, applying a voltage, and letting the DNA migrate through the gel. Shorter fragments move faster than long ones, and you can use the distance traveled during the experiment as an indicator of fragment length. In order to calibrate DNA length and distance traveled, you compare the distance your fragment traveled with the distances traveled by a series of standard fragments of known length. The DNA standard used in this course is shown in Figure 8.

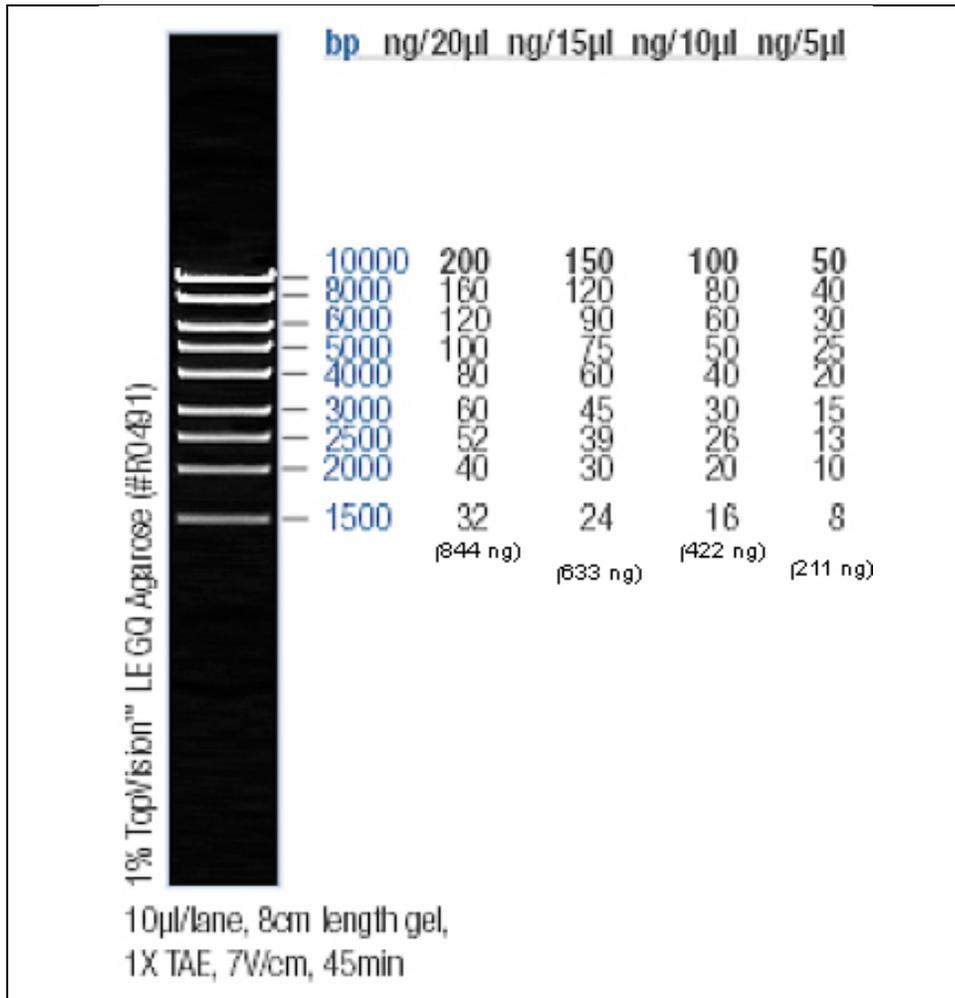


Figure 8. Fermentas #SM0393 MassRuler High Range DNA Ladder

Metric Prefixes

(adapted from Greg Pearce, chemformula bbcmb@sigmaxi.org,
<http://www.unc.edu/~rowlett/units/prefixes.html>)

In your introductory chemistry and biology labs, you probably used only a few of the most common metric prefixes, such as milli-, centi-, and kilo-. Various textbooks and lab manuals contain longer lists of prefixes, but not many contain a complete list. There is no point in memorizing all of them, but it is nice to be able to look them up when you encounter them as you read the scientific literature. The smaller prefixes, such as nano-, pico- and femto-, are becoming increasingly common as analytical chemistry and biotechnology develop more sensitive methods. The larger prefixes, such as mega-, giga-, and tera-, are becoming more common in computing as engineers find ways to pack more and more memory into tiny devices. Table 2 presents prefixes ranging from 10^{24} to 10^{-24} . To help you visualize the effect of these prefixes, the column labeled “a sense of scale” gives some examples of the magnitudes represented.

Table 2. Metric Prefixes

Prefix	Abbreviation (note upper and lower case)	Meaning	synonym	A sense of scale (approximate)
yotta-	(Y-)	10^{24}	1 septillion	Mass of water in Pacific Ocean ~1 Yg Energy emitted by sun per second ~400 YJ Volume of earth ~1 YL Mass of earth ~6000 Yg
zetta-	(Z-)	10^{21}	1 sextillion	Radius of Milky Way galaxy ~1 Zm Volume of Pacific Ocean ~1 ZL Annual world energy production ~0.4 ZJ
exa-	(E-)	10^{18}	1 quintillion	Age of universe (12 billion yr) ~0.4 Es
peta-	(P-)	10^{15}	1 quadrillion	1 light-year ~9.5 Pm
tera-	(T-)	10^{12}	1 trillion	Sun-to-Jupiter distance ~0.8 Tm
giga-	(G-)	10^9	1 billion	Human life expectancy ~3 Gs 1 light-second ~0.3 Gs
mega-	(M-)	10^6	1 million	Two weeks ~1.2 Ms
kilo-	(k-)	10^3	1 thousand	
hecto-	(h-)	10^2	1 hundred	
deka-	(da-)	10	1 ten	
deci-	(d-)	10^{-1}	1 tenth	
centi-	(c-)	10^{-2}	1 hundredth	
milli-	(m-)	10^{-3}	1 thousandth	
micro-	(μ -)	10^{-6}	1 millionth	Diameter of human ovum ~1 μ Volume of mosquito blood meal ~2 μ L Volume of wood frog egg ~3 μ L
nano-	(n-)	10^{-9}	1 billionth	Radius of chlorine atom ~0.1 nm
pico-	(p-)	10^{-12}	1 trillionth	Mass of bacterial cell ~1 pg
femto-	(f-)	10^{-15}	1 quadrillionth	Radius of proton ~1 fm Volume of <i>E. coli</i> cell ~1 fL Volume of red blood cell ~100 fL
atto-	(a-)	10^{-18}	1 quintillionth	Time for light to cross an atom ~1 as Bond energy of C=C ~1 aJ
zepto-	(z-)	10^{-21}	1 sextillionth	600 atoms or molecules ~1 zmol
yocto-	(y-)	10^{-24}	1 septillionth	Mass of proton or neutron ~1.7 yg

GENERAL LAB TECHNIQUES

Introduction to micro-pipetting

Goal for this lab:

Master micropipetting technique.

Introduction

One of the most important skills you will need in this course is your ability to use a micropipettor. Micropipettors are used to make accurate measurements of extremely small volumes—from one milliliter down to one microliter (1 mL to 1 μ L). Most of what we do in molecular biology involves manipulating volumes of liquid in this range. If you learn to do it accurately now, your experiments will go much more smoothly later.

Micropipettors

These are precision scientific instruments, and must be treated with respect. The pipettor is used to draw liquid up into a cheap disposable tip. The three pipettors you will use take up and deliver liquids in the volume range from \sim 0.5 μ L to 1.0 mL. Your instructor will show you how to use this device. Read and follow these guidelines to maintain the accuracy and precision of your pipettors.

Rotate the volume adjustor to the desired setting. Note the change in plunger length as the volume changes. Be sure to properly locate the decimal point when reading the volume setting. (Your instructors will demonstrate.)

You have three sizes of pipets in this lab: LTS20s, which can measure between 1 μ L and 20 μ L; LTS200s, which can measure between 21 μ L and 200 μ L; and LTS1000s, which can measure between 200 μ L and 1000 μ L (1 mL).

There are three numbers on the display of each pipettor. Look at the top of the pipet to see which one you are holding, then look at the display. The numbers represent volumes as shown below. The color change represents crossing the decimal place or changing units.

L 20		L 200		L 1000	
1	10 μ L	1	100 μ L	1	1000 μ L = 1.0 mL
0	1 μ L	0	10 μ L	0	100 μ L
0	0.1 μ L = 100 nL	0	1 μ L	0	1 μ L

Firmly seat a proper-sized tip on the end of the micropipettor. The tips boxes are color-coded to match the label on the plunger.

When withdrawing or expelling fluid, always hold the tube firmly between your thumb and forefinger, keeping it nearly at eye level to observe the change in the fluid level in the pipet tip. Do not pipet with the tube in the test tube rack or have another person hold the tube while you are pipetting.

Hold the tube in your hand during each manipulation. Open the top of the tube by flipping up the tab with your thumb. During manipulations, grasp the tube body (rather than the lid), to provide greater control and to avoid contamination of the mouth of the tube.

For best control, grasp the micropipettor in your palm and wrap your fingers around the barrel; work the plunger (piston) with the thumb. Hold the micropipettor almost vertical when filling it.

Notice the friction “stops” on the two-position plunger with your thumb. Depressing to the first stop measures the desired volume. Depressing to the second stop introduces an additional volume of air to blow out any solution remaining in the tip.

To withdraw the sample from a reagent tube:

1. Depress the plunger to **first** stop and hold it in this position. Dip the tip into the solution to be pipetted, and draw fluid into the tip by gradually releasing the plunger. Be sure that the tip remains in the solution while you are releasing the plunger.
2. Slide the pipet tip out along the inside of the reagent tube to dislodge any excess droplets adhering to the outside of the tip. To avoid future pipetting errors, learn to recognize the approximate levels to which particular volumes fill the pipet tip.
3. If you notice air space at the end of the tip or air bubbles within the sample in the tip, carefully expel the sample back into its supply tube. Collect the sample at the bottom of the tube by pulsing it in a microcentrifuge.

To expel the sample into a reaction tube:

1. Touch the tip of the pipet to the inside wall near the bottom of the reaction tube into which the sample will be emptied. This creates a capillary effect that helps draw fluid out of the tip.
2. Slowly depress the plunger to the first stop to expel the sample. Depress to second stop to blow out the last bit of fluid. Hold the plunger in the depressed position.
3. Slide the pipet out of the reagent tube with the measurement plunger depressed, to avoid sucking any liquid back into the tip.

Use the ejector button (located at the back and different from the plunger) to remove the tip into a waste container.

Important pipettor don'ts:

- Never rotate the volume adjustor beyond the upper or lower range of the pipet.
- Never use the micropipettor without the tip in place; this could ruin the piston. Pipettors use disposable plastic tips. Every molecular biology lab circulates its own version of the story of the not-too-bright grad student who did not use a tip. Do not be this student!
- Never invert or lay the micropipettor down with a filled tip; fluid will run back into the piston.
- Never let the plunger snap back after withdrawing or expelling fluid; smooth motions are the key to success.
- Never immerse the barrel of the micropipettor in fluid. Only the disposable tip touches the liquid.
- Never reuse a tip. Tips are pretty cheap (about \$0.59 per rack). The risk of cross contaminating your solutions is too great to get tricky with tips. Just use a new one every time unless there is no possibility of cross contamination—like if you are pipetting the same solution into multiple empty tubes.

Measurement Repeatability

An essential property of good science is that an experiment gives the same results even in different hands. Repeated measurements of the same thing should give the same value, no matter who makes the measurement.

Our first job, therefore, is to make sure our measurements are reliable. We will be checking both inter- and intra-observer reliability, as well as the reliability of our instruments.

Figure out what each of these volumes of water should weigh, and which pipet should be used to dispense that volume:

μL	mL	weight (g)	weight (mg)	pipet
1000 =		=		
100 =		=		
10 =		=		

Choose one of these volumes for pipet calibration.

Put a weigh boat on the balance pan, and zero the balance.

Have one member of your group dispense the appropriate volume of water onto the pan. Record your result.

Re-zero the balance and repeat 10 times.

Repeat the exercise with the other member(s) of the group.

Data Analysis

Examine your data by eye. Does anything stand out? Any strange values? Next week we will learn a statistical test to see whether you can confidently omit one outlying value. See p. 31 for a sneak peak.

Using the built-in functions in Excel, calculate the mean and standard deviation of each of your sets of repeated measurements.

Describe any differences between measurements made by different people, or with different instruments. Make a few repeat measurements as necessary to distinguish between technique and instrument. (Hint: the biggest source of discrepancy is likely to be either using the second stop on the pipet plunger or letting go of it suddenly during filling. Think about which technical error would give you a measure higher than expected, and which, lower.)

Dilutions

It is often very important to know the precise concentration of some chemical you are using in your experiment. Various units of concentration are used in biology and chemistry:

Molarity (M) moles/Liter

mg/mL g/L

% w/vol g/100 mL (because 1 mL of water weighs 1 g)

% vol/vol mL/100 mL

Sometimes you can start from scratch, that is, weigh out the substance and dissolve it in the appropriate amount of solvent (usually water for biological applications).

What if I need 10 mL of 5M NaCl (MW = 58)? You can calculate how much NaCl to dissolve in 10 mL water this way (notice how all the units but g cancel out):

$$\frac{5\text{mol}}{\text{L}} \times \frac{58\text{g}}{\text{mol}} \times 10\text{mL} \times \frac{1\text{L}}{1000\text{mL}} = 2.9\text{g}$$

What if I need 1 mL of 5mM NaCl?

$$\frac{5\text{mmol}}{\text{L}} \times \frac{1\text{mol}}{1000\text{mmol}} \times \frac{58\text{g}}{\text{mol}} \times 1\text{mL} \times \frac{1\text{L}}{1000\text{mL}} = 0.00029\text{g}$$

Unfortunately, 0.00029 g is an almost imaginary amount of sodium chloride. Even if you made 10 times what you needed, you'd have to weigh out 0.0029 g, and most of our prep room balances only report three decimal places. You could make a liter of 10mM NaCl with 0.29 g, because sodium chloride is cheap to buy and legal to pour down the drain, but many reagents are much harder to come by and dispose of, so we need other ways of making solutions.

If you already had your 10 mL of 5 M NaCl, you could make your 1 mL of 5 mM NaCl by dilution, that is, by taking a small volume of your 5 M NaCl and diluting it in water. To calculate how, we use the lab preparer's best friend:

$$c_1v_1 = c_2v_2$$

c_1 = the concentration of the initial solution used to make the more dilute solution

v_1 = the initial small volume of the first solution used to make the dilute solution

c_2 = the concentration of the second, more dilute solution

v_2 = the final volume of the second solution after the dilution is carried out

So we need to figure out what volume (v_1) of the 5 M (c_1) solution to use to make 1 mL (v_2) of a 10 mM NaCl (c_2) solution.

$$v_1 = \frac{c_2v_2}{c_1} = \frac{5\text{mM} \times 1\text{mL}}{5\text{M}} \times \frac{1\text{M}}{1000\text{mM}} = 0.001\text{mL}$$

Another way to look at this is to calculate the dilution factor, that is, the ratio between the initial and final concentrations. Diluting a 5M solution to a 5mM solution is a 1000-fold dilution:

$$\frac{c_2}{c_1} = \frac{5mM}{5M} \times \frac{1M}{1000mM} = \frac{1}{1000}$$

Therefore, you need 1 part stock solution to make 1000 parts of your final solution.

$$v_1 = \frac{c_2}{c_1} \times v_2 = \frac{1}{1000} \times 1mL = 0.001mL$$

Fortunately, I have a micropipettor that can deliver 0.001mL (=1 μ L). So I can make the dilute solution by mixing 0.001mL of my concentrated solution with 0.999 mL of water.

What if I don't really trust* my pipettor down in the single μ L range?

Worse, what if my concentrated sodium chloride stock solution is 1M, rather than 5M?

The initial volume would be 0.0002 mL (=0.2 μ L), and I *really* don't have a pipettor I trust to deliver that small a volume.

I can solve the problem of a v_1 too small to deliver by using a *series* of dilutions, each transferring a volume I *can* accurately deliver, to achieve my intended volume and concentration. For this procedure, I use a modification of $c_1v_1=c_2v_2$:

$$c_1v_t = \left(\frac{c_1}{d}\right)(v_f + v_t)$$

c_1 = the concentration of the initial solution used to make the next dilution in the series

v_t = the transfer volume

d = the dilution factor (*e.g.*, 2 for a 2-fold dilution series, where the concentration of each solution in the series is half that of the previous one)

$c_2 = \frac{c_1}{d}$ = the concentration of the second solution in the series

v_f = the desired final volume of each solution

$v_f + v_t$ = the initial volume of a diluted solution, until the transfer volume is removed to make the next dilution in the series

This procedure is illustrated in Figure 9.

* When I say I don't trust the pipettor, I mean that I worry about tiny errors in delivery representing a large percentage of the volume I'm transferring.

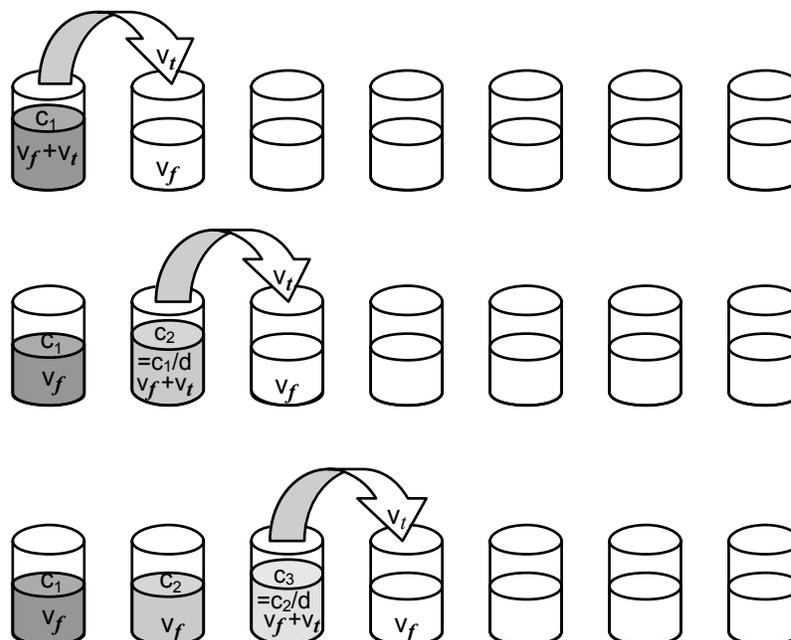


Figure 9. A serial dilution. A small volume (v_t) of solution is transferred to a vessel containing a volume (v_f) of solvent. After these are mixed, v_t of the second solution is transferred to the third vessel, also containing v_f of solvent. The concentration of each solution is $1/d$ the concentration of the previous one in the series.

So if I didn't think I could trust a $1\ \mu\text{L}$ transfer volume, I might try a 10-fold serial dilution ($d = 10$) (both because calculations with 10 are easy, and because I want the last dilution in my series to be one one-thousandth of the initial concentration). Suppose that the $1\ \text{mL}$ volume in the example above was more than enough for my experiment. An easy 10-fold dilution series could be done with a $0.1\ \text{mL}$ transfer volume, as follows:

$$\begin{aligned} \text{Starting with the general equation:} & \quad c_1 v_t = \left(\frac{c_1}{d}\right)(v_f + v_t) \\ \text{Substituting the initial concentration, the dilution} & \quad 1\text{M} \times 0.1\text{mL} = \left(\frac{1\text{M}}{10}\right)(v_f + 0.1\text{mL}) \\ \text{factor, and the transfer volume:} & \\ \text{Solving for } v_f: & \quad 1\text{M} \times 0.1\text{mL} \times \left(\frac{10}{1\text{M}}\right) - 0.1\text{mL} = v_f = 0.9\text{mL} \end{aligned}$$

So by starting with $1\ \text{mL}$ ($v_f + v_t$) of my 1M (c_1) solution, and transferring $0.1\ \text{mL}$ (v_t) of it to $0.9\ \text{mL}$ (v_f) of solvent gives me $1\ \text{mL}$ ($v_f + v_t$) of 0.1M solution.

Repeating this process will give me a series of solutions, each one-tenth the concentration of the one before it.

What if I don't know what concentration range of a particular substance is suitable for use in my experiment or in my measuring device? You need to test a wide range of concentrations. Serial dilution is the best way to achieve such a range. With a 2-fold dilution series of 10, you can take a solution down to less than a 1000^{th} of its original value:

$$\frac{1}{2^{10}} = \frac{1}{1024} = 0.000977$$

With a 10-step 10-fold dilution series, you can get down to 10^{-10} (less than a billionth!) of the original concentration.

What if the manufacturer gives me exactly 1 mg of the chemical? You can take advantage of the manufacturer's precision by dissolving the chemical, still in its original package to prevent loss of material, in a small, known volume of suitable solvent (which the package insert should identify for you), thereby making a stock solution of known concentration (*e.g.*, 1 mg/mL). From that, you can make whatever dilute concentrations you need.

DNA EXTRACTION AND QUANTIFICATION

Lab 1: DNA Extraction

Goal for this lab:

Extract DNA from various leafy vegetables.

Introduction

In this lab, we will extract DNA from common salad ingredients. The DNA we obtain today can be used for many different experiments, so this is an important activity!

In any DNA purification (or extraction) protocol, there are three important considerations:

1. We must break up the tissues and cells to release the DNA.
2. We must prevent the DNA from being degraded by cellular enzymes during the purification.
3. We must get rid of cellular components that are not desired—in this case, proteins, lipids, carbohydrates (cell walls), and RNA.

Break up tissues and cells.

We will cause tissue and cell disruption in three ways. The first is mechanical grinding. We will use a mortar and pestle to grind frozen leaf tissue into a fine powder. Freezing (in liquid nitrogen) makes the leaf tissue brittle and easy to grind. This breaks up the extracellular matrix in which the cells reside and allows better access of our chemicals to the cells themselves. The second method of disruption is chemical. The DNA extraction buffer that we use contains a high concentration of a detergent, sodium dodecyl sulfate (SDS). You will recall that detergents have both hydrophobic and hydrophilic moieties, and thus will bind both lipid and water. This detergent will solubilize the lipid bilayers in the cells, so that the plasma membrane and also the internal membranes are dissolved. In addition, we use the metal chelating compound ethylene diamine tetra-acetic acid (EDTA), which tightly binds calcium (Ca^{++}) and magnesium (Mg^{++}) effectively locking it away from cellular components. In the absence of Ca^{++} and Mg^{++} , proteins denature and membranes lose their integrity. The last disruption is high temperature. This encourages the detergent and EDTA to work efficiently and hastens cellular disintegration. This high temperature step is critical for efficient cellular disruption and release of DNA.

Prevent DNA degradation.

We prevent DNA degradation by inhibiting the enzymes (deoxyribonucleases, or 'DNases') that degrade DNA. Conveniently, DNases require Mg^{++} as a co-factor, so, by chelating the Mg^{++} with EDTA, we inhibit the activity of DNases. Prior to the addition of EDTA, we control DNase activity using low temperature (freezing in liquid nitrogen). If we were to break up the cells at room temperature, cellular DNases released during grinding would immediately start to degrade the DNA. Thus, it is imperative to get the EDTA containing DNA extraction buffer well mixed with the ground tissue before the temperature is raised.

Get rid of undesirable molecules.

We will use two methods to accomplish this. Both rely on differential solubility of specific macromolecules in particular solutions. Proteins and complex carbohydrates (such as those found in cell walls) are insoluble in solutions containing high concentrations of potassium

acetate (KOAc). By adding KOAc, we force the proteins and carbohydrates to precipitate (become solid and fall out of solution) while the nucleic acids will remain in solution. Nucleic acids are insoluble in solutions containing high levels of sodium plus the alcohols ethanol (EtOH) or isopropanol (ISOP), while lipids are quite soluble in these conditions. Proteins and carbohydrates are somewhat soluble in alcohol. By adding alcohols to a salty solution, we can force the nucleic acids to precipitate, while leaving the lipids, and the remaining carbohydrates and proteins, in solution. The main problem with alcohol precipitation of nucleic acids is that the sodium precipitates along with the nucleic acids. This is undesirable, and so we always rinse nucleic acid pellets with 70% EtOH (containing no salts) to remove the sodium from the pellets prior to dissolving the nucleic acids in the next solvent.

Materials:

Reagent	conc		pH	also known as
DEB	100 mM	NaCl	8.0	Sodium chloride
(DNA Extraction Buffer)	50 mM	Tris		trishydroxymethylaminomethane
	25 mM	EDTA		Ethylene diamine tetra-acetic acid
	1 %	SDS		Sodium dodecyl sulfate
	10 mM	BME		β -mercaptoethanol
T ₁₀ E ₅	10 mM	Tris	8.0	
	5 mM	EDTA		
T ₁₀ E ₁	10 mM	Tris	8.0	
	1 mM	EDTA		
EtOH	95 %		-	Ethanol
EtOH	70 %		-	Ethanol
ISOP	100 %		-	Isopropanol
KOAc	5 M			Potassium acetate (CH ₃ COOK)
NaOAc	3 M		5.2	Sodium acetate (CH ₃ COONa)
LN ₂				Liquid nitrogen

Mortar and pestle

Miracloth and a funnel

Paper and tape

A choice of edible leaves

Method:

Before you begin, read through the directions and get everything out that you will need so you don't have to fumble for it in the middle of some critical step in the extraction procedure.

SAFETY: LN₂ is very very cold, -320° F (-196° C), and will severely damage tissue that it contacts! Use extreme caution while handling to avoid freezing injuries.

1. Pre-cool your mortar and pestle by pouring some liquid nitrogen (LN₂) into the mortar and letting the pestle sit in it. Allow the LN₂ to evaporate.

2. Now, add some more LN₂, and ~0.25-0.5 g of leaves (about 1/3 spinach leaf for example). **Do not let the tissue thaw** once it is frozen. Grind slowly at first, so you won't slop the LN₂ and leaves everywhere. Then, when the LN₂ has mostly evaporated (but while the tissue is still frozen) grind hard to get the tissue into a fine powder. The more you grind, the more you'll get. You can carefully add additional LN₂ to keep your tissue frozen if necessary.
3. Transfer the powder into a 15 mL ROUND BOTTOMED plastic tube which contains 6 mL of DNA extraction buffer (DEB). (A paper cone made from a sheet of printer paper will make this much easier.) **The amount of tissue used should fit easily on top of the extraction buffer in your tube. Do not overload the tubes!** Overloading with too much tissue will result in poor quality DNA.
4. **Use a spatula to shove the powder down into the tube and mix well**, then snap the cover down to the second stop and **shake vigorously** to mix. Wait (mixing occasionally) until the mixture has completely thawed before going on to the next step.
5. When your sample is ready, transfer the tube to a 65°C water bath for 10 minutes.
6. Place the tube on ICE and add 2.5 mL 5M KOAc. Mix well by inverting the tube repeatedly. Incubate ON ICE for 20 minutes.
7. Centrifuge 5000 rpm 10 minutes at 4°C.
8. Remove your tube from the centrifuge and notice the very large pellet consisting of tissue debris and precipitated protein. REMEMBER, you are trying to GET RID of the pellet in this step!
9. Transfer the supernatant (the liquid on top of the pellet) through Miracloth filter into a fresh round-bottomed 15 mL tube containing 6 mL of 100% isopropanol. Mix well by inverting the tube gently but repeatedly. **Precipitating nucleic acids may be clearly visible at this point.**
10. Spin in the centrifuge at 5000 rpm for 10 minutes at 4°C.
11. Remove your tube from the centrifuge. REMEMBER, this time, YOU WANT YOUR PELLETT—the pellet is the nucleic acids—including your DNA. Pour off the supernate.
12. Rinse your pellet carefully with 2-3 mL 70% EtOH. Aspirate as much liquid as humanly possible, then let the pellet air dry for a few minutes. Getting rid of the EtOH is essential before proceeding to the next step.
13. Add 450 µL T₁₀E₅ to your pellet. It is important to get your DNA pellets to dissolve completely into the T₁₀E₅. We still use a high concentration of EDTA at this step, because not all the proteins are gone yet, so nucleases could still attack our DNA. Transfer the liquid into a labeled 1.5 mL microtube.
14. Add 50 µL 3M NaOAc pH = 5.2. Mix. Add 1 mL 95% EtOH. Mix gently by inversion until phases are completely mixed. Fluffy whitish clouds of DNA should almost fill the tube. If you do not see DNA (or can't lift it out of the tube in the next step), consult an instructor. You may have to spin it down and rinse it with ethanol, and let it air dry before going on.
15. Use a 1 mL pipette tip like a spatula to lift out the nucleic acid fluff and transfer it into a clean, dry, labeled 1.5 mL microtube. Add 1 mL of 70% EtOH to rinse. If it cannot be scooped, spin it down, rinse with ethanol and air dry before going on.

16. Use a pipette tip to transfer the nucleic acid fluff into a clean dry labeled 1.5 mL microtube.
17. Spin briefly in the microfuge—15 seconds is plenty of time. Remove as much EtOH as humanly possible. If necessary, you may try drying the sides of the tube with a sterile swab.
18. Gently dissolve* the pellets in 100 μ L T₁₀E₁. We use a low concentration of EDTA now, because our nucleic acid preparation is fairly pure—protein contamination is minimal now. However, we keep some EDTA around ‘just in case’. If we had too much EDTA in our final solution, we could never do subsequent enzymatic manipulation of our DNA. Let the pellets dissolve COMPLETELY-- leaving LABELED tubes **in the refrigerator overnight if necessary**. Store at -20°C once it is **completely** dissolved.
19. If time permits, measure the concentration of your DNA by the Nanodrop, as explained beginning on page 59.
20. Label your tube carefully with your group initials, the date, and gDNA (genomic DNA). Put it in a rack in a freezer box with your group name on it in label tape.
21. Write in your lab notebook what you did, either “exactly as in manual”, or “as in manual, except that ...” Give the date, and what label you put on the tube. Later on in the semester, when you will need this DNA again, all this will help you remember (or figure out) what’s in the tube you just put in the freezer.

Clean-up:

Alcohols are safe to go down the drain *only when they are less than 10%*. *Run plenty of water while discarding alcohols so that this dilution will be achieved.*

Save your reagents. If you have to make DNA again later, you’ll be glad to have them. You will use your T₁₀E₁ again.

Put the tubes that contained any β -mercaptoethanol in the marked bucket in the hood.

Put plant debris and used potting soil in the bucket labeled “Compost”.

Dispose of glass pipettes in the glass trash box.

Set your mortar and pestle to soak as directed by your instructor.

* If it doesn’t go into solution, add more T₁₀E₁ and heat to 65°C for a few minutes. Do not “rack” more than once.

Lab 2: DNA Quantification

Goal for this lab:

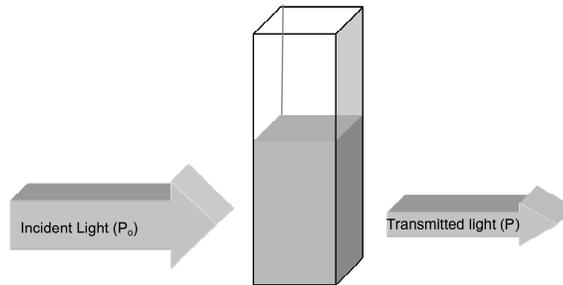
Quantify plant DNA by two methods.

Introduction:

Reliable measurement of DNA concentration is important for many applications in molecular biology. DNA quantification is generally performed by spectrophotometric measurement of the absorption at 260 nm, or by agarose gel analysis. In this lab, you will compare the two methods for quantifying the genomic DNA that you prepared during our last meeting.

DNA quantification by molecular absorbance spectroscopy.

DNA concentration can be determined by measuring the absorbance at 260 nm (A_{260}) in a spectrophotometer using a UV-transparent cuvette (usually quartz). In molecular absorbance spectroscopy, a beam of ultraviolet or visible light (P_o) is directed through a sample. Some of the light may be transmitted through the sample (P). Light that was not transmitted through the sample was absorbed. Transmittance (T) is defined as the ratio of P/P_o . Absorbance (A) is defined as $-\log(T)$. The light that was absorbed is what we care about: the absorbance at 260 nm, or A_{260} .



Unfortunately, spectrophotometric measurement does not differentiate between the two nucleic acids commonly found in a cell (DNA and RNA). The OD₂₆₀ measurement thus tells you something about the total amount of nucleic acid in a preparation, but can't let you measure the concentration of DNA in a preparation that also contains RNA. RNA contamination of a DNA preparation can lead to overestimation of DNA concentration, if spectroscopy is the only method used to determine concentration.

DNA quantification by agarose gel electrophoresis

Quantification of DNA can also be done using agarose gel analysis. This method is only effective with small amounts of DNA, since high concentrations of DNA cannot be accurately measured this way. As little as 5 ng DNA can be detected by agarose gel electrophoresis with SYBR-Safe staining. In this method, the DNA sample is run on an agarose gel alongside known amounts of DNA standards. The amount of sample DNA loaded can be estimated by comparison of the band intensity with the standards using a scanner or imaging system.

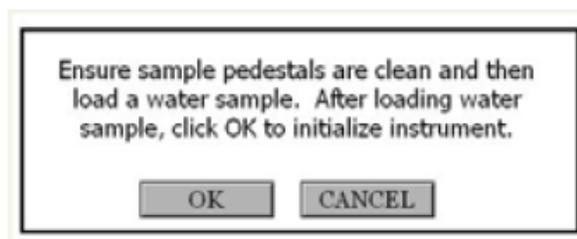
Quantification using agarose gel electrophoresis is also extremely useful when DNA samples are contaminated with RNA, since RNA is separated from DNA in agarose gels by virtue of the difference in length (size) of an average genomic DNA fragment versus the average size of a mRNA, tRNA or rRNA molecule. RNA is small: tRNA is only ~100 nucleotides (nt) long, and mRNA averages ~1.5 kilobases (kb), depending on the size of the encoded protein.

Genomic DNA is theoretically the size of a whole chromosome (1000s of kb), although breakage during handling causes the size of an extracted DNA sample to be smaller (~34-50 kb). By analyzing DNA that has been *physically* separated from RNA on agarose gels, the concentration of DNA can be measured in the absence of contaminating RNA.

Method 1: Spectrophotometric quantitation via Nanodrop^{*}

Module Startup

When the software starts, you should see this message:



For best results, ensure measurement pedestal surfaces are clean, load a water sample onto the lower measurement pedestal and then click 'OK'. The message "Initializing Spectrometer- please wait" will appear. When this message disappears, the instrument will be ready for use. All data taken will automatically be logged in the appropriate archive file.

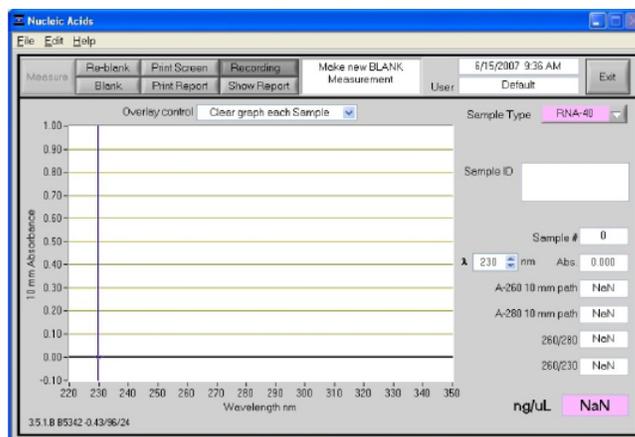
Measure (F1)

Each time a software module is opened (initiated), the Measure button is inactive as noted by its "grayed-out" appearance. A blank must first be measured before the Measure button will become active.

The Measure button is used to initiate the measurement sequence for all samples (non-blanks). It is actuated by depressing the F1 key or clicking the Measure button. The entire measurement cycle takes approximately 10 seconds.

Blank (F3)

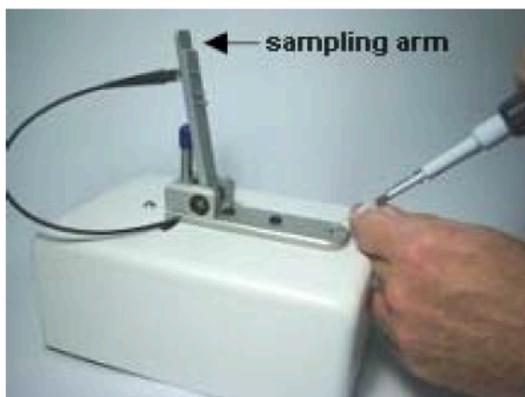
Before making a sample measurement, a blank must be measured and stored (see "Blanking and Absorbance Calculations" on page 62 for more details on absorbance calculations). After making an initial blank measurement, a straight line will appear on the screen; subsequent blanks will clear any sample spectrum and display a straight line, as shown in the following image:



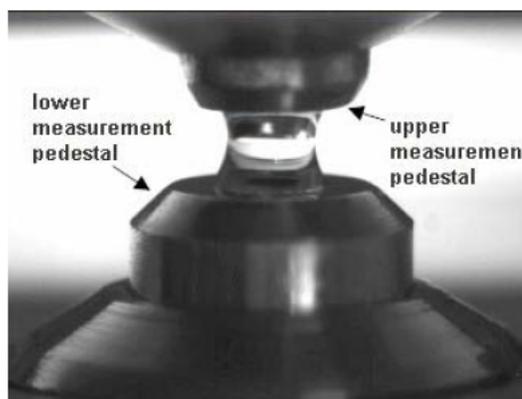
^{*} Information and instructions taken from Thermo Scientific's NanoDrop 1000 Spectrophotometer V3.8 User's Manual, <http://www.nanodrop.com/library/nd-1000-v3.7-users-manual-8.5x11.pdf>, accessed 1/14/13.

For the most consistent results, it is best to begin any measurement session with a blanking cycle. This will assure the user that the instrument is working properly and that the pedestal is clean. Follow the steps below to perform a blanking cycle:

1. With the sampling arm open, load a 1 μL blank sample (the buffer, solvent, or carrier liquid used with your samples) onto the lower measurement pedestal. Make sure there are no bubbles!



2. Close the sampling arm. The sample column is automatically drawn between the upper and lower measurement pedestals. Click the “Blank” button (F3)



3. When the measurement is complete, open the sampling arm and wipe the sample from both the upper and lower pedestals using a soft laboratory wipe. Simple wiping prevents sample carryover in successive measurements for samples varying by more than 1000-fold in concentration. See the Nanodrop website (www.nanodrop.com) for performance data on sample carryover.

Wiping the sample from both the upper and lower pedestals (as shown to the right) upon completion of each sample measurement is usually sufficient to prevent sample carryover and avoid residue buildup. Although generally not necessary, 2 μL water aliquots can be used to clean the measurement surfaces after particularly high concentration samples to ensure no residual sample is retained on either pedestal.



4. Analyze an aliquot of the blanking solution as though it were a sample. This is done using the “Measure” button (F1). The result should be a spectrum with a relatively flat baseline. Wipe the blank from both measurement and pedestal surfaces and repeat the process until the spectrum is flat.
5. Analyze 1 μL of your sample with the “Measure” button. Do a second measurement to be sure, following the steps above.

Blanking and Absorbance Calculations

When the NanoDrop 1000 Spectrophotometer is “blanked”, a spectrum is taken of a reference material (blank) and stored in memory as an array of light intensities by wavelength. When a measurement of a sample is taken, the intensity of light that has transmitted through the sample is recorded. The sample intensities along with the blank.

intensities are used to calculate the sample absorbance according to the following equation:

$$\text{Absorbance} = -\log \frac{\text{Intensity sample}}{\text{Intensity blank}}$$

Thus, the measured light intensity of both the sample and of the blank are required to calculate the absorbance at a given wavelength.

Concentration Calculation (Beer’s Law)

General

The Beer-Lambert equation is used to correlate the calculated absorbance with concentration:

$$A = E \times b \times c$$

Where

A is the absorbance represented in absorbance units (A),

E is the wavelength-dependent molar absorptivity coefficient (or extinction coefficient) with units of liter/mol-cm,

b is the path length in cm, and

c is the analyte concentration in moles/liter or molarity (M).

Nucleic Acids

For nucleic acid quantification, the Beer-Lambert equation is modified to use an extinction coefficient with units of ngcm/mL . Using this extinction coefficient gives a manipulated equation:

$$c = A \times e \times b$$

Where

c is the nucleic acid concentration in $\text{ng}\mu\text{L}$

A is the absorbance in AU

e is the wavelength-dependent extinction coefficient in $\text{ng}\times\text{cm}\mu\text{L}$

b is the path length in cm

The generally accepted extinction coefficients for nucleic acids are:

Double-stranded DNA: 50 ng×cm μ L

Single-stranded DNA: 33 ng×cm μ L

RNA: 40 ng×cm μ L

For the NanoDrop 1000 Spectrophotometer, path lengths of 1.0 mm and 0.2 mm are used compared to a standard spectrophotometer using a 10.0 mm path. Thus, the NanoDrop 1000 Spectrophotometer is capable of measuring samples that are 50 times more concentrated than can be measured in a standard spectrophotometer.

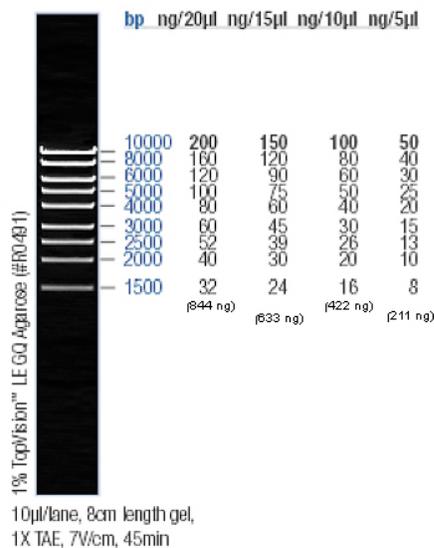
Note: Absorbance data shown in archive files are represented as displayed on the software screen. For Nucleic Acid, Protein A280 and Proteins and Labels modules, data are normalized to a 1.0 cm (10.0 mm) path. For MicroArray, UVVis, Protein BCA, Protein Bradford, Protein Lowry and Cell Culture modules the data are normalized to a 0.1 cm (1.0 mm) path. For high absorbance UV-Vis samples, data are normalized to a 0.1 mm path.

Method 2: Quantification using agarose gels.

Materials:

Reagent	conc		pH	also known as
DB	10	mM	Tris	7.5 Dilution Buffer
1× TAE	40	mM	Tris	8.0 Running Buffer
	20	mM	Acetic acid	(dispense it yourself from the carboy at the back of the room)
	1	mM	EDTA	
Agarose gel	0.9	%	Agarose	
			SYBR-safe	DNA binding dye
RNase	0.9	%	Agarose	
Agarose gel			SYBR-Safe	
	1.0	μ g/mL	RNase	
Loading dye	30	%	glycerol	
	0.3	%	Bromphenol blue	
	0.3	%	Xylene cyanol	

Fermentas #SM0393 MassRuler High Range DNA Ladder (See below.)



SAFETY: *SYBR safe is a weak mutagen and a possible carcinogen, so must be handled carefully. WEAR GLOVES when handling gels and electrophoresis buffers to protect your skin. Discard your gloves when you've stopped handling your gel. AVOID DRIPS so that other people don't accidentally come into contact with SYBR-safe. The best way to avoid drips is to use a Kimwipe to wipe off your gel tray before moving around with it.*

Using agarose gel electrophoresis, you can simultaneously assess both the **yield** and **quality** of your DNA isolations by visualizing the DNA on 0.9% agarose gels containing SYBR-Safe. SYBR-safe is a dye that can intercalate between the bases of DNA in a manner that is essentially proportional to the amount of nucleic acid present. Because there is some double strandedness in most RNA preparations, SYBR-safe will also stain RNA. When exposed to ultraviolet light (250-350 nm), SYBR-safe fluoresces yellow, which is then visualized by photography. The DNA concentration from your samples can be estimated by comparing the SYBR-safe fluorescence intensity to the intensities observed for standards of known DNA concentration run on the same gel.

Running a gel

(Your instructor will demonstrate setting up the gel apparatus.) Prepare two 0.9% agarose gels. ***The gel contains SYBR-Safe, so wear gloves.*** Get two 50-mL tubes of 0.9% agarose from the 65°C water bath. Make sure that one tube contains RNase, and the other does not. Set up the gel molds with a comb in place to form the wells. Use label tape to mark which one contains RNase.

Pour your RNase-free gel first, to prevent contaminating your gloves, etc. with RNase and spreading it to other things in your workspace.

Get approximately 250 mL of running buffer (1× TAE) from the carboy for each gel.

When your gels have set, transfer the gel ***in its little plastic tray*** to the electrophoresis apparatus, and add running buffer. Avoid cross-contaminating RNase from one gel apparatus to the other.

DNA is negatively charged, and so will migrate toward the positive electrode. Make sure you have the wells at the correct (NEGATIVE) end. (An easy way to remember is that gels run in alphabetical order: from black to red.) Remove the comb very gently and carefully, pulling straight up.

While your gel is setting up, you can prepare your samples.

Dilute your DNA

Make four dilutions of the DNA sample you prepared at our last meeting^{*}: 1/20, 1/50, and 1/100. Make the dilutions as follows: Place the requisite volume of dilution buffer into a labeled 1.5 mL microtube, add the appropriate volume of DNA stock solution, mix, then spin the liquid to the bottom of the tube.

^{*} If you had to dilute your sample to get the DNA into solution, and your volume is >500 µL, you may need much less dilution: 1/2, 1/5, 1/10. Check with your instructor.

Dilution	DNA stock solution	Dilution buffer
1/5	40 μL	160 μL
1/20	10 μL	190 μL
1/50	4 μL	196 μL
1/100	2 μL	198 μL

Save these samples on ice. Do not discard. You will use them again later today.

Remember, you are running two gels, so prepare two of each of your samples. Into labeled 1.5 mL microtubes, place 5 μL of your diluted sample (1/5, 1/20, 1/50 or 1/100). To each, add 1 μL loading dye. This dye serves two purposes: it contains glycerol, which is heavy, and will make your solution sink under the electrophoresis buffer in the gel wells. It also contains two dyes (xylene cyanol (XC) and bromophenol blue (BPB)), which migrate at different speeds in the gel, and allow you to track your gel's progress visually while it runs.

Load your gels. You will include two lanes of 'standards': one lane with 2.5 μL (105.5 ng) of MassRuler—the DNA standard of known molecular weights and masses—and a second lane with 5 μL (211 ng) of the same standard. (See previous page.) Then load each of your diluted samples into one of the four adjacent lanes. Record the order in which the samples were loaded in your notebook.

Run the gel at 100 volts for 45 minutes.

Photographing your gel

Take the tray with your gel out of the electrophoresis apparatus and place the tray on a paper towel to catch excess buffer.

Take your gel (on its tray) over to the gel doc system.

Place the gel (without the tray) on the UV box, and close the door.

TAKE OFF YOUR GLOVES—NO GLOVES ON COMPUTERS, EVER.

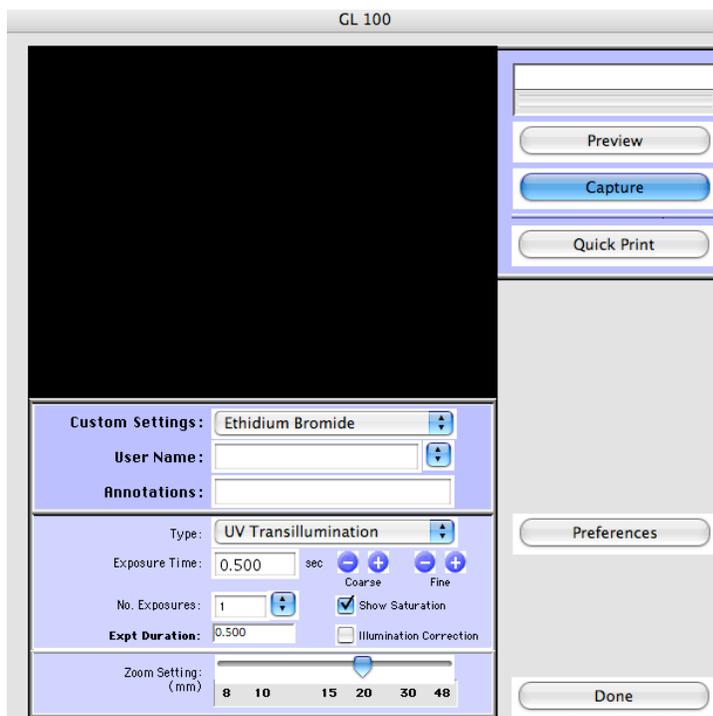
Open the Kodak software (the big red K on a yellow background in the dock at the bottom of the screen).

Back in the Kodak software, go to the "File" menu, and "Select Digital Camera". In the dialog box that follows, select the GL100/GL200 camera, and click OK.

Look in the extreme upper right corner of the screen. You should see a button saying 'Capture GL 100'. HOWEVER, more likely, you will see a button saying, 'Capture GL2200'. If you do, go back to the file menu, select digital camera again, and select GL100/GL200 camera, and click OK. This time, you will be asked whether you want a GL100 (you do) or a GL200. Choose the GL100.

Look in the extreme upper right corner again. Now it should correctly say, 'Capture GL100.' Click the button.

Now you see a dialog box for the camera (See below).



Choose 'Preview' to view your gel.

If the gel is not visible, make sure that the UV light is on. If the gel is still not visible, use the 'IOXperts Camera Control' program (in the dock, to the right of the Kodak icon—see below) to adjust the gain on the camera.



When your image is to your liking, click on the 'Capture' button (in the Kodak MI software). Your image will open.

Save your image to the desktop. Remember to save the image you take to your USB drive and/or to your UDrive space. Do note that files saved from the Kodak software will be '.bip' files, which can be read ONLY by the Kodak software. This is enough for right now, but later you want to be sure to save the image in other formats for the purposes of presentation in lab reports and wiki pages.

Analyzing the image of your gel

Estimating the size of DNA fragments using semi-log paper

DNA moves through the matrix of the agarose gel in response to the electric field generated by in the gel apparatus. Longer pieces of DNA move more slowly, and are found nearer the wells, than shorter ones. In order to figure out the sizes (that is the length in base pairs) of the DNA fragments in our samples, we compare the distance they travel to distances traveled by fragments of known size, that is, our *standards*.

Figure 10 illustrates the band pattern of a DNA standard sold by Fermentas Life Sciences. The wells would have been at the top of this picture, and the smallest fragments (100 bp) have migrated farthest during the 45 minute run.

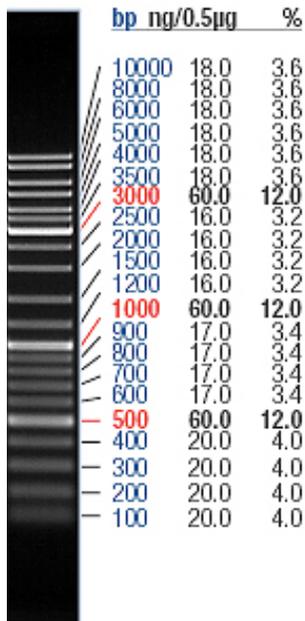


Figure 10. Fermentas GeneRuler™ DNA Ladder SM0331 (from <http://www.fermentas.com/catalog/electrophoresis/generulers.htm>, accessed 7/25/08).

1µg/lane, 8cm length gel,
AE, 7V/cm, 45min

Slight differences in agarose concentration, voltage, or running time result in different distances traveled by DNA of a given length in different replicates of the same experiment. Because of these variations, it is necessary to run a set of standards with every gel, and use their migration distances to create a standard curve, which can then be used to estimate the length of experimental (unknown) DNA fragments running on the same gel.

To create a standard curve from the standard bands, we must measure the distance traveled by each band. The easiest way to do this is to enlarge the image, print it, and put a ruler to it. Measuring the distance from the top of the image to the leading edge of each of the bands produced the values in Table 3.

Table 3. Migration distance of Fermentas SM0331 standards

fragment length (base pairs)	migration distance (cm from top _* to leading edge of band)
10000	5.4
8000	5.7
6000	6.3
5000	6.6
4000	7.3
3500	7.6
3000	8.1
2500	8.6
2000	9.3
1500	10.6
1200	11.5
1000	12.3
900	12.7
800	13.2
700	13.8
600	14.4
500	15.1
400	15.8
300	16.8
200	17.7
100	18.6

Plotting these values produces the curve shown in Figure 11. The relationship is distinctly non-linear. This non-linearity makes it hard to distinguish the lengths of the fragments that travel the farthest, because the slope gets shallower and shallower as the distance gets longer.

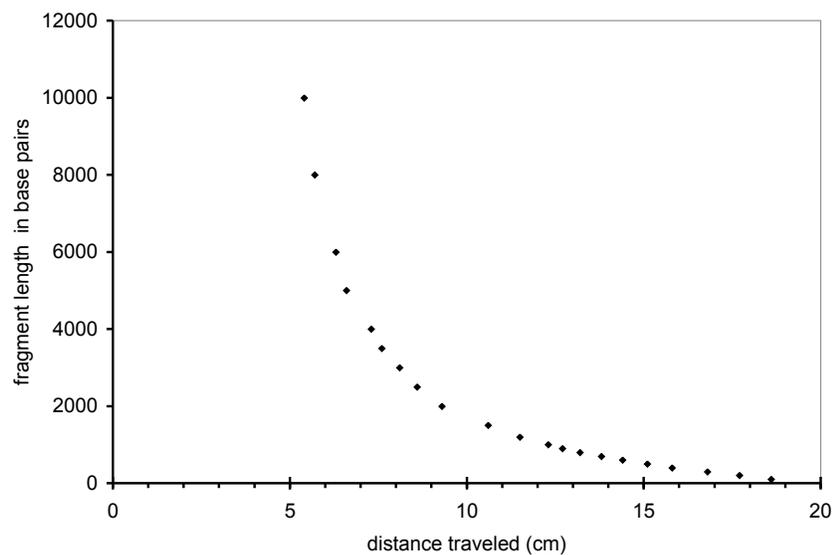


Figure 11. Relationship between distance traveled by DNA fragments and their length in base pairs.

* Note that since we are really interested only in the distance traveled by our sample bands *relative to* the distance traveled by the standard bands, we could have measured the distance from any particular starting point. It is not necessary to measure from the wells themselves.

One simple way to straighten out many non-linear curves (that is, logarithmic ones) is to replot the points using the log of the sizes instead of the actual sizes. Doing so produces the curve illustrated in Figure 12, which is much more nearly linear. (Yours may not look quite so nice, since this graph was made with values taken from the best photograph the manufacturer could produce of the best-looking gel ever run in their facility.) To figure out the length of your unknown DNA fragment, all you have to do is measure its distance from the same starting point you used to measure the standards, find that distance on the X axis of your graph, draw a vertical line from that point up to intersect the regression line, and a horizontal one over to the Y axis, which will give you ... the \log_{10} of its length. Let's see – if $x = \log_b y$, then $b^x = y$. Ugh. Isn't there a way to read the actual length of the fragment off a graph?

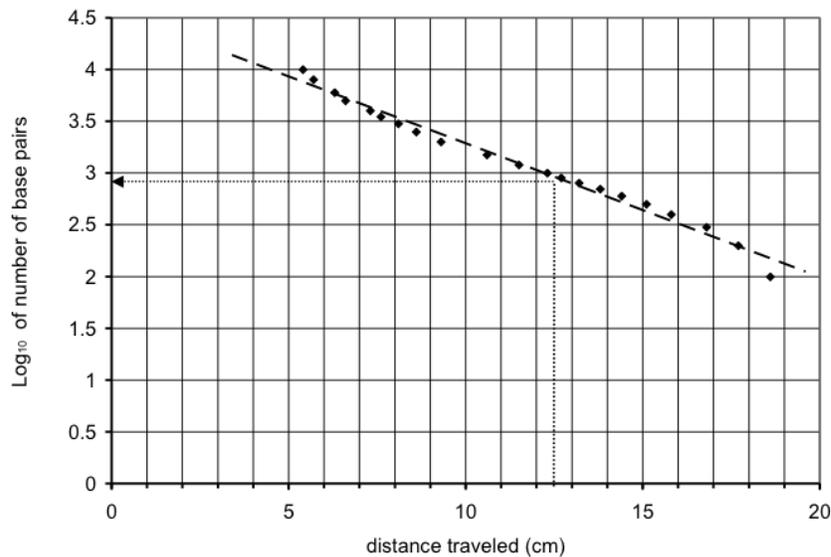


Figure 12. Relationship between distance traveled and \log_{10} of the number of base pairs.

One way to achieve the goals you're probably listing right now (do no calculations, use a straight line, and read the number of base pairs right off the graph) is to plot your data on semi-log paper, which calculates the log for you via the spacing of the lines. Figure 13 shows what this type of graph paper looks like. To get a straight line plot of your points like that in Figure 12, you can plot distance traveled vs. number of base pairs.

This graph paper differs from regular graph paper in that the major divisions of the Y axis increase by a factor of 10, not 1 (1, 10, 100, 1000, etc., rather than 1, 2, 3, 4). There is no zero on this axis; it is always possible to divide by 10 to get the next major division down (0.1, 0.01, 0.001, etc.). Each of these major divisions is called a cycle. Figure 13 is an example of 3-cycle semi-log paper.

The minor divisions increase by the amount of the value of the major division. The first cycle on this paper would be divided as follows: 1, 2, 3, 4, 5, 6, 7, 8, 9, and the second cycle would be: 10, 20, 30, 40, 50, 60, 70, 80, 90. These divisions are further subdivided, with more subdivisions for the lower values in each cycle.

To use this paper, you must assign powers of 10 to the Y axis that let you plot all your values. Since the fragments in the GeneRuler range from 100 to 10,000 base pairs, the major

divisions can be assigned to be 100, 1000, and 10,000. Then plotting the migration distance on the X axis and the base pair length on the Y generates a graph like the one in Figure 14.

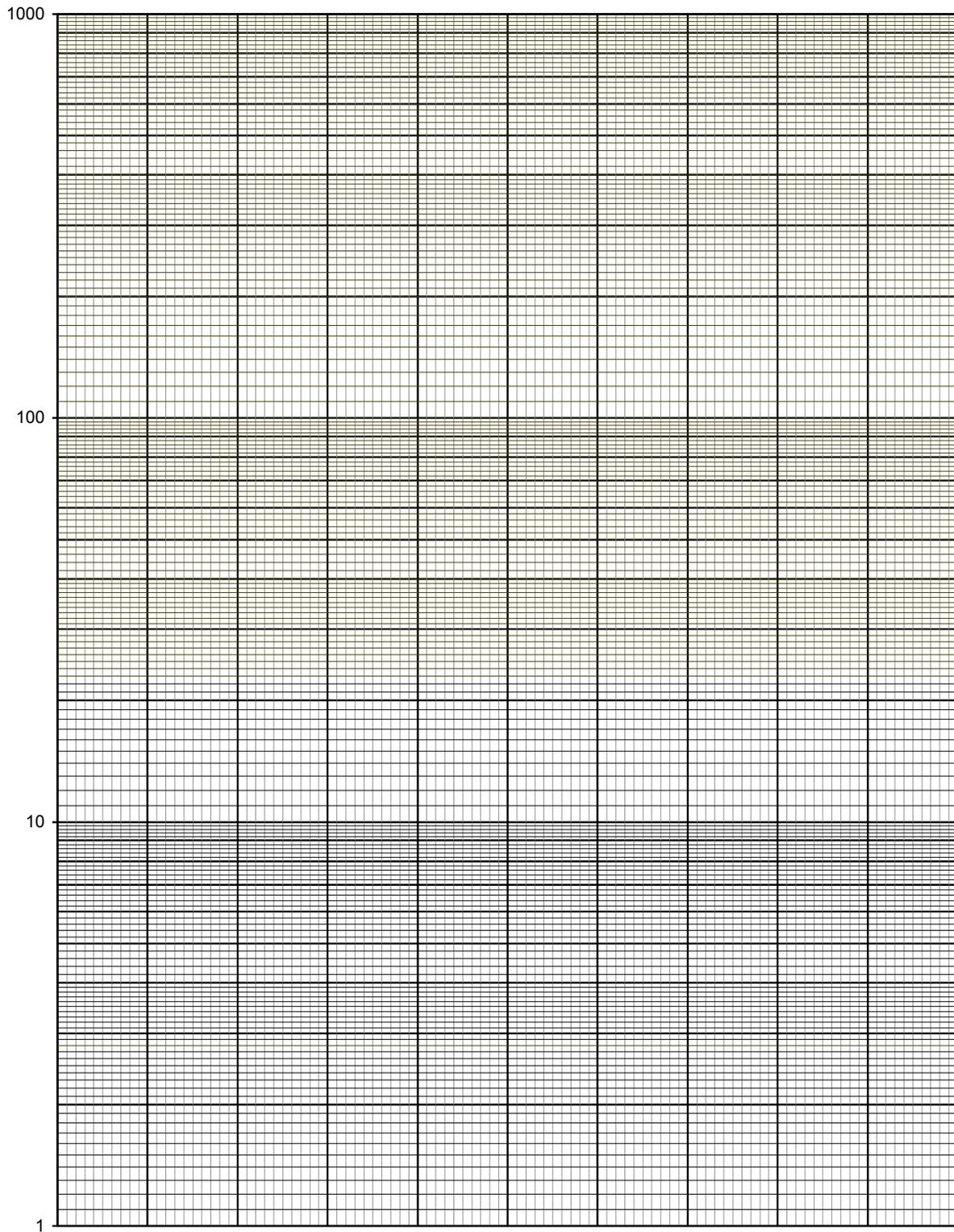


Figure 13. Semi-log graph paper. Plotting values on the Y-axis automatically takes their log.

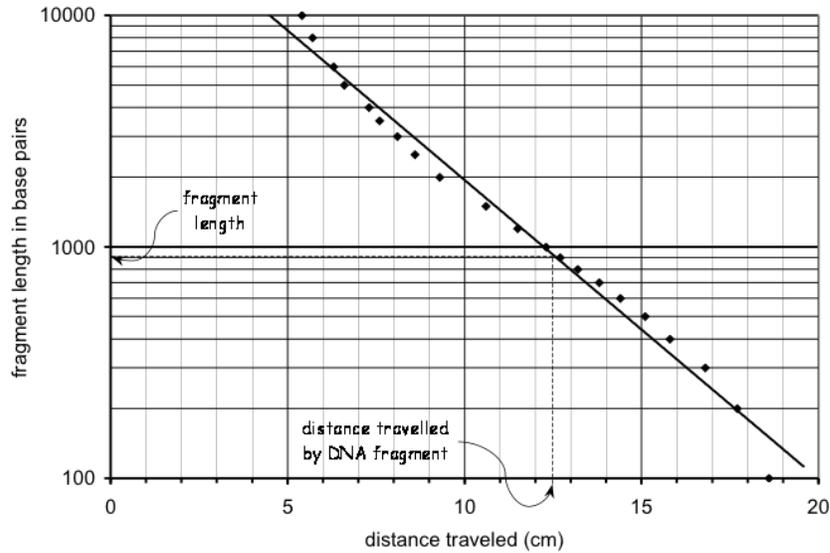


Figure 14. Same data as Figure 11, plotted on semi-log paper. A band that traveled 12.5 cm (see vertical arrow) would be about 900 base pairs long (see horizontal arrow).

The data plotted this way give us a pretty good straight line, which we can fit to the points by eye with a straightedge. The straight line, in turn, gives us a way to estimate the length of the DNA fragments from their migration distance. For example, if one of the samples produced a band that traveled 12.5 cm in this gel, we would draw a straight line from the 12.5 cm mark on the X axis to the linear regression line. Then we can draw a horizontal line over to the Y axis and read the base pair length from it.

There is a bit of an art to creating a standard curve. Notice that the points do not fit exactly on a line, and the slope seems to vary slightly from region to region on the graph. This will lead you to underestimate the length of DNA in certain bands (e.g., those that travel very short distances) and overestimate others (e.g., the bands that travel farthest). How you fit your line to your points depends on where your data points will lie on the graph. As illustrated in **Figure 15**, there are several better fitting short lines you may wish to use, depending on the distance traveled by the band you are interested in evaluating.

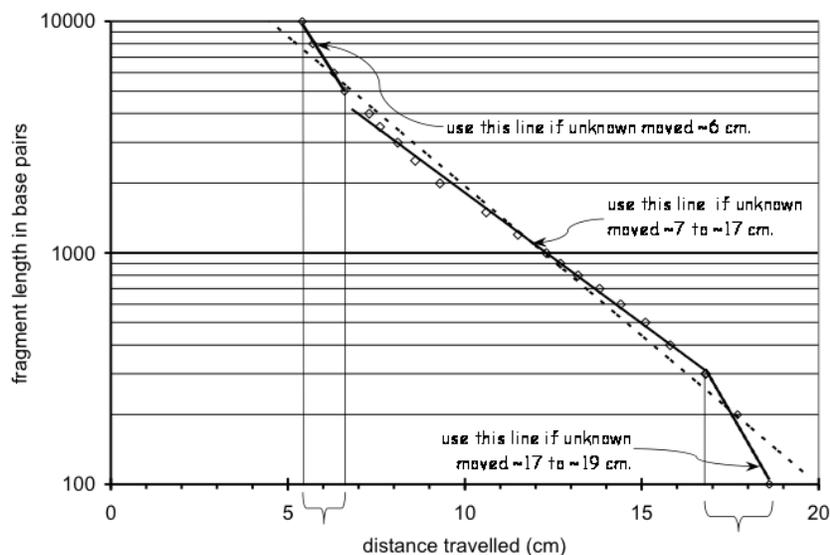


Figure 15. Same as Figure 14, showing how several short lines (solid lines) fit the data better locally than the single best fit line (dashed line).

If possible, analysis should be performed on the desktops in 364.

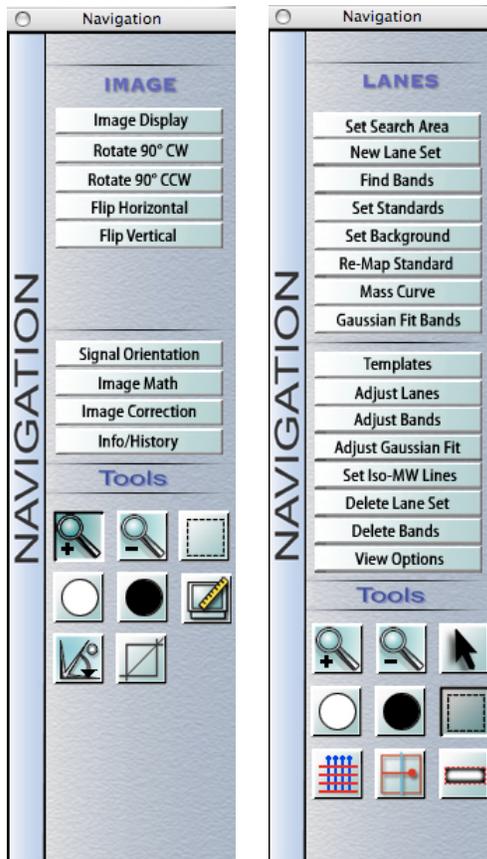
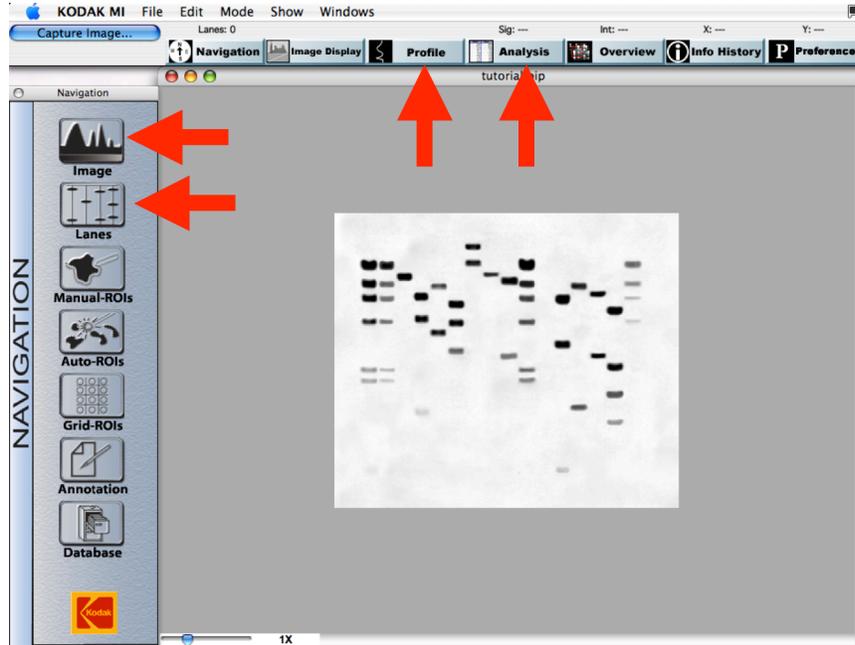
The DNA quality can be assessed by examining the proportion of fluorescence that is associated with a discrete band of high molecular weight DNA relative to a range of smaller (<20 kbp) fragments that appear as a **smear** that runs the length of the gel lane. That tight high-molecular weight band at the top is your intact genomic DNA. Any smear you may see below it is degraded DNA. ***Make an assessment of the quality of your DNA preps based on these criteria.*** Be sure to consult with an instructor if you are not sure.

In the gel that did not contain RNase, you may also see RNA, which will appear far down the gel, and will either look like a big bright smear, or as two bright bands close together. Of course, in the gel you ran that contained RNase, no RNA will be observed. Remember, the purpose of running the gel was to separate the RNA from the DNA, so that you could analyze the DNA by itself. Thus, you should not include RNA blobs or bands in your analysis (see below).

Back at your own computer, you will analyze the images of your gels. Go to UDrive or use your USB drive to get the file that is the image you took of your gel. Then, get the Kodak software running on your computer. In the File Menu, Open your image. The screen you see will look like the image on the next page. I have highlighted the buttons you will use with arrows.

Before you go on, print the picture of your gel, and place the image(s) in your lab notebook. Be sure to label each lane appropriately. Next, ***export*** the image from the Kodak software as a JPEG ('.jpg') file and save the file to your USB drive or UDrive space. JPEG files can be read by many different software packages, and can be used for creating figures for your lab reports and wiki pages. Images of gels for display should not have the lane markings you use in the analysis below. At this point you will have a hard copy of the image of your gel (in your notebook) and will have two files on the computer for each gel image: a bip file used by the Kodak software, and a .jpg file that you will use to make figures later. Make note of the filenames you used in your lab notebook.

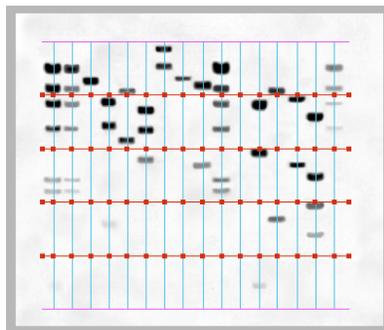
Note on using canned software: The best way to learn new software is to play. Don't be afraid to click stuff and see what happens. If you get lost doing this, just quit and start over. As long as you do not save (just say no when prompted, unless you are sure...), you really cannot make anything bad happen. There is a walkthrough below, and your instructors can help if you have problems.



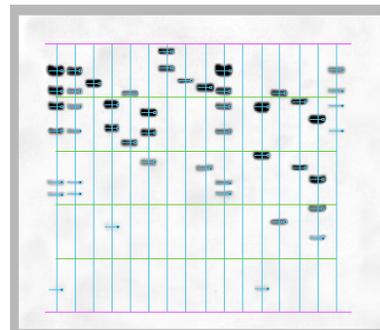
First, you may need to fix the orientation of the picture, so that the lanes are at the top. To do this, click the 'Image' button in the navigation bar on the left. Now you will be looking at the 'Image' subsection, as shown at far left. Use the flip or rotate commands until your image is oriented correctly. Then, to get back to the 'Navigation' tool bar, click on, "Navigation" written horizontally at the far left.

Now, you must "show" the software the location of the lanes of your gel. To do this, click the 'Lanes' button in the navigation bar. Now you will be looking at the 'Lanes' subsection, as shown at near left. Use your cursor to select a rectangle that surrounds the active (i.e., the lanes you used) parts of your gel image. Your rectangle must be 'tight' to the lanes you used—do not include extra space on the sides, top or bottom of the lanes. Then click 'New Lane Set' in the lanes navigation window. You'll see that the software has made a vertical line through each of your lanes, and has made some horizontal lines as well, to help your visual orientation. See below, left:

Take a hard look at the lanes that the computer has identified. Does each lane have a vertical line down the center? Is each vertical line centered well on the lane? If not, start over. If you are having difficulties, using the 'Adjust Bands' button, you can try setting band width to 80%. Your instructors can help you if you experience difficulties after a few tries.

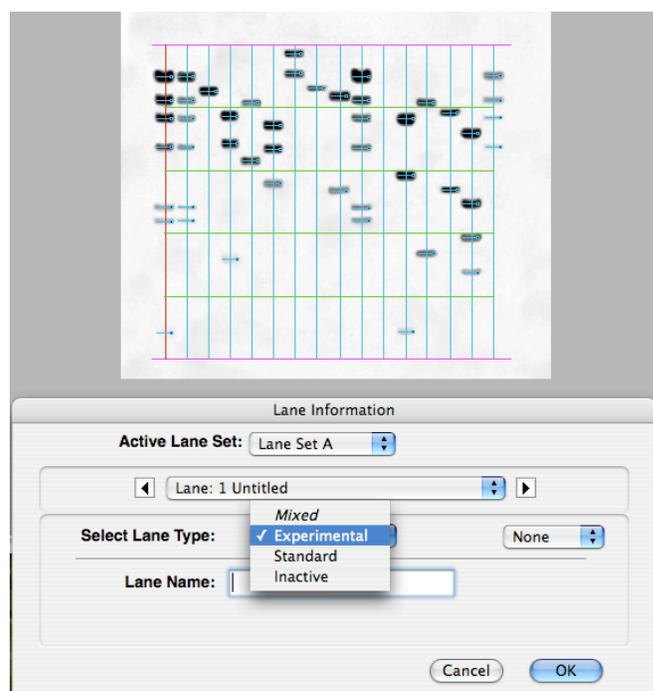


Next, you must allow the software to identify the bands on your gels, and fix any mistakes the software may have made. To do this, click the 'Find Bands' button. The software will automatically figure out how wide and how bright your bands are and will identify them as on the right:



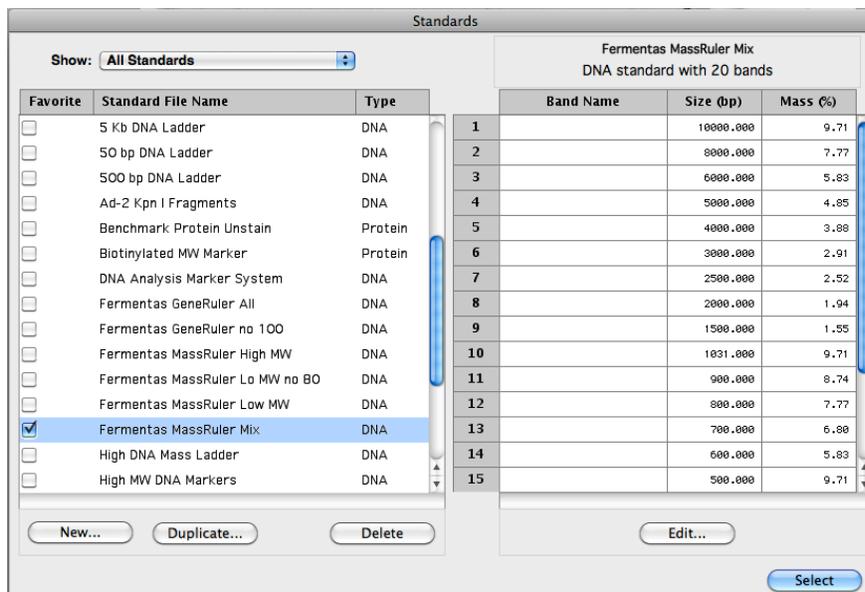
Take a good look to make sure that all your bands have been identified correctly. Select and delete any 'extra' bands that the software found. If there was RNA in the gel, the software probably marked it as a band or multiple bands. Get rid of these: you are not trying to quantify RNA, and, in fact, you cannot use this method, because you do not have a set of RNA standards to use. Add in any bands that you see but the computer missed. Look carefully at both lanes of standards – did the software identify the correct number of bands (9)? If not, correct this problem before proceeding.

Now we need to "show" the computer which lane contains the standards, and give it some information about the standards we used. To do this, use your cursor to click on the vertical line going through your first standards lane: the one in which you loaded 5 μ L of standards. This should highlight the line, which will change color. In the 'Lanes' bar at left, click "Set Standards". This will open up a dialog box, as shown below:



In the dialog box, click ‘Select Lane Type’ to bring up the drop down menu shown in the picture above. The lane you are defining is a lane of standards, so choose this. (You can also choose to mark each of your experimental lanes, and to mark lanes as inactive, if you have left any blank lanes.)

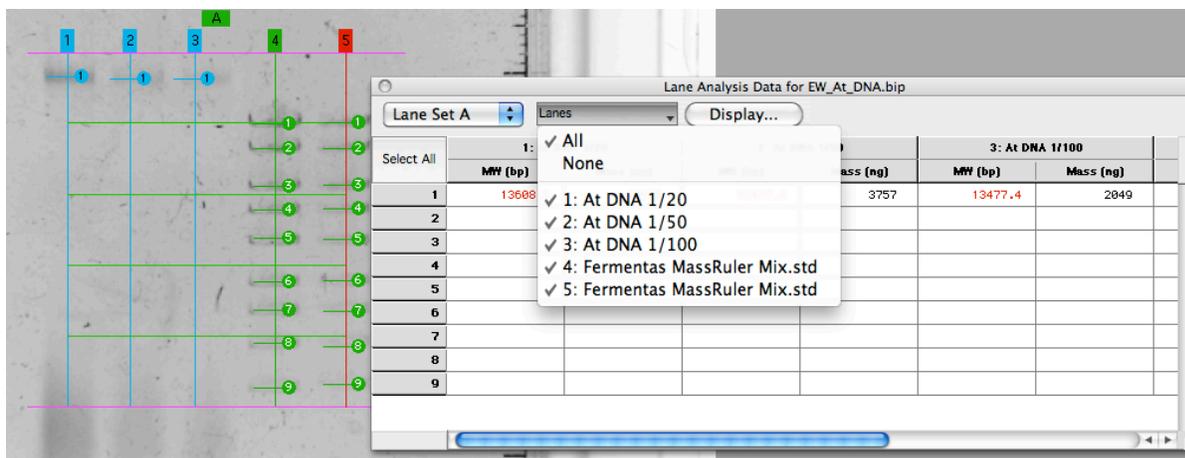
You need to tell the software two things about your standards: What they are, and how much of them you have loaded. When you told the computer that it was looking at a lane of standards, you got an option to identify the standards. That will open a dialog box:



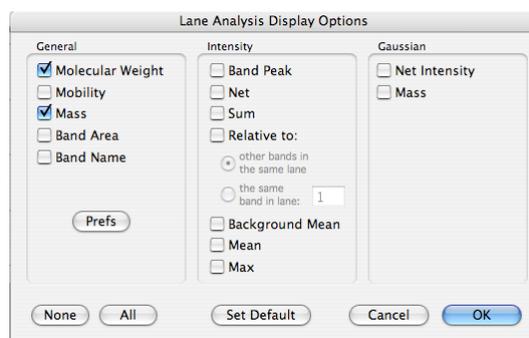
We are using MassRuler™ DNA Ladder Mix, Fermentas #SM0393. (See on page 12.) Find this choice in the list, and select it. On the right, you will see the size and mass (as a percentage) of each band. Click ‘Select’. Back in the ‘Lane Information’ dialog box, you must specify how much standard you have loaded. Since we loaded 2.5 μ L of standard, we will have 257.5 ng total mass in the lane. The software is looking for 4 digits, so enter your amount as 2575 (if measuring in ng) or as 0.2575 (if measuring in μ g). Click OK to get out of the ‘Lane Information’ dialog. Now go through the same process for your second standards lane, remembering that there is twice as much mass (515 ng) in the lane.

With the information you have just provided, the software is now ready to calculate, based on brightness, the mass in each experimental lane—and that, of course, is the information you are looking for!

In order to see the calculations, we need to click the ‘Analysis’ box at the top of the main window. That will open up a new window showing the calculations. You only need to tell the software which lanes you want to see – click on “Lanes” and select “All”, and what information you want to see. (See picture below.)



Right now, we only want mass, so click on ‘Display’ to get ‘Lane Analysis Display Options’ as shown below:



De-select ‘Molecular Weight’. (This is totally inaccurate, since the genomic DNA bands are far larger than any of the standards. We’ll use this feature later in the semester.) Make sure that ‘Mass’ is selected. Click OK. Voilà! A table will appear showing the mass of the band(s) in each lane. Print it and/or copy the information into your notebook. ***You should always double check that the masses predicted for the bands in the standards are what you expected them to be.*** If not, there was a problem, and you need to go back and fix it!

Remember to export an image without lane and band markers to use in your Wiki.

Calculating the concentration and yield of DNA

Using the new set of measurements you just made, recalculate the concentration of your stock DNA preparation and your total yield of DNA. Notice that you will need to think through your calculation—we will not be giving you a formula. You have plenty of information to make the calculation: you know how many ng of DNA was in each lane, you know what volume you loaded into each lane, and you know how much each sample was diluted prior to loading. ***Check your calculations with an instructor.***

Settle on a final concentration and yield for your DNA preparation

You now have a series of numbers, which may or may not be in good agreement with each other. It is up to you to make a well reasoned conclusion about the best/most accurate concentration measurement, so that you can use this for all future calculations. Before you leave, ***check your reasoning with an instructor. Be sure that you have settled on your final concentration and yield before you leave today.*** You will be using these numbers again throughout the semester, so it is important to get them settled correctly now.

Clean-up

Pour the used running buffer into the specially marked waste container.

Put your used gel, as well as the 50 mL tube that held the agarose in the specially marked containers in the hood.

The rest of your trash, unless you think it is particularly contaminated with ethidium bromide (e.g., a paper towel that held the gel), can go in the regular trash containers.

Rinse the living daylighters out of all your gel equipment: casting trays, electrophoresis apparatus, etc. Are you willing to touch it barehanded? Is it free of all traces of RNase? If not, rinse it some more! Let everything air dry.

Rinse your Erlenmeyer flasks thoroughly, first in tap water, then once with distilled water from the big carboy. Hang them up to air dry.

See your instructor about cleaning up any spills.

Unused running buffer can go down the drain.

Save your DNA standards and your loading dye.

Discard your diluted DNA, but save your genomic DNA!

Lab write up:

Please be sure to address the following in your notebook:

- What differences did you notice between the gel with RNase and the one without RNase? (Explain possible reasons that account for the differences.)
- Did the two DNA quantification methods (spectrophotometric and gel-based) give you the same results? If not, why not? Which do you consider to be more accurate and why?
- Did all three dilutions of your genomic DNA result in the same estimate of DNA concentration of the stock solution? If not, why not?
- Assess the quality of your genomic DNA preparation.
- Report both the yield (total amount in μg) and the concentration (in $\mu\text{g}/\mu\text{L}$) of DNA that you purified.
- If you had problems with your DNA purification, where do you reckon the problems arose? Be specific – “experimental error” explains nothing.

Review and follow the conventions given on page 14 regarding the description of DNA sequences.