BIOL20: The Dynamic Genome in the Neil A. Campbell Science Learning Laboratory

This is a brand new facility and Biology 20 is the first college class to work here! The NACSLL has a bioinformatics lab and a wet lab, so students can learn skills used in research labs on campus. NACSLL was designed for courses that use cutting edge technologies, so professors can, for example, offer courses modeled after research in their laboratory and give students the opportunity to experience research. The first part of the course will introduce techniques and core biological concepts. These concepts will provide a solid background in genetic information transfer and genome biology that will be useful in future classes and in understanding research projects. The Laboratory techniques taught, including computer skills, while limited in number, are used routinely in many research labs on campus. Additionally, the first course segment will focus on experimental design and interpretation of results.

During the second course segment students apply the tools and concepts learned in the first part of the course to complete a research project. This project will focus on analyzing the genomes of strains of rice currently being investigated in Dr. Wessler's laboratory.

Today we will do the following:

1. Take preliminary quizzes and surveys
2. Introduce ourselves
3. Log onto the computers and use PubMed
4. Learn basic lab skills
    a. Lab safety
    b. Dilutions
    c. Pipetting exercises
    d. Lab notebook recordkeeping

Computers in the NACSLL

The NACSLL is equipped with twenty-eight 17-inch MacBook Pros that students and instructors use in the bioinformatics lab. Each computer is numbered and is stored in the lockable recharging cabinet. Students will pick a computer today and use it for the whole quarter. Computers are labeled with a number on the bottom. Select a computer and record the number here: _____.

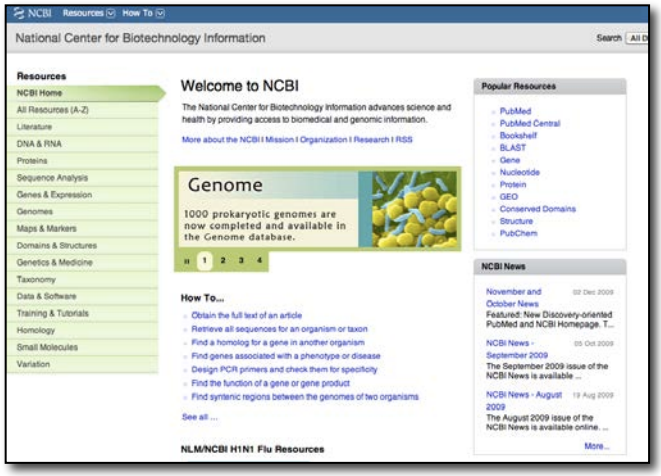
Log on information: Select the Student account and enter "1234" as the password. Please do not change the password and do not store any sensitive material on this computer. You can customize the computer such as choosing a different desktop picture

if you wish. For storing and sharing files, it is a good idea to sign up for a free Dropbox account (https://www.dropbox.com/).

The class will use the Safari Browser for web work. The browser is already set up with bookmarks in the Bookmark Bar for websites used in the course. The first bookmark is "BIOL20 Fall 2011 Home" and links to the class webpage. This page will be the source for information in the class including assignments, syllabus, and data. This course will not use iLearn. (https://ulb-103-199.ucr.edu/Classes/2011-Fall/index.html)

Introduction to the NCBI website and PubMed

Biological sequence data and journal articles are collected, indexed, and made available by the National Center for Biotechnology Information (NCBI). NCBI is a unit of the National Library of Medicine (NLM) at the National Institutes of Health (NIH). Because it is a part of the NIH, the collections of sequence data and journal articles are available free to anyone at http://www.ncbi.nlm.nih.gov/. This is what the NCBI home page looks like….



NCBI also provides tools for searching and downloading the databases it maintains through the web portal NCBI Entrez. While the search tool for the literature database is PubMed, sequence data is searched though a number of tools collectively called Blast. PubMed indexes thousands of biological journals going back as far as 1950. It also contains thousands of full-length articles in PDF format available for free download in a collection called PubMed Central. Blast searches on DNA sequence databases that are often referred to as GenBank. There are three public repositories for sequence data: NCBI, DDBJ (Japan), and EMBL (Europe). All share data on a nightly basis. Although the

file formats and search tools may differ between the three repositories, they are essentially redundant.

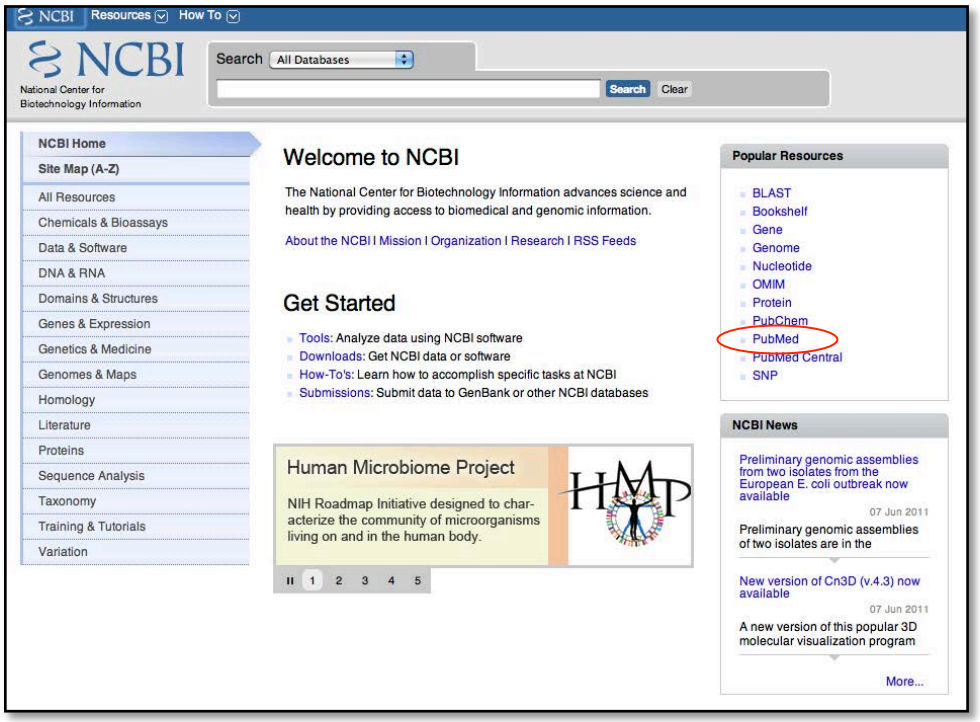
Spend some time browsing much of the NCBI Entrez portal. This is an amazing resource. There are plenty of help tutorials, free full-length (slightly dated) textbooks, and a lot of interesting information.

PubMed

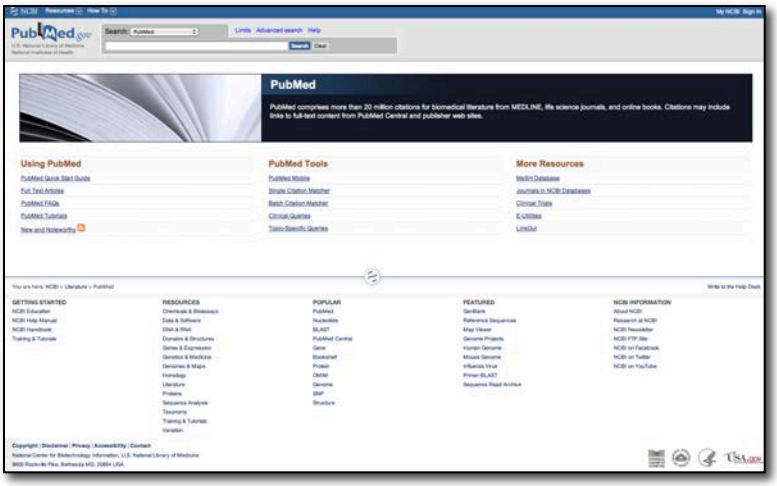
Literature searches about a biological problem are very easy. PubMed makes the index available on its website with no access limitations. PubMed (and Blast) can be used from any computer and with any Internet connection and access through the University network is not required. However, to access many full-length articles and download PDF copies using UCR's network is necessary.

Steps for a PubMed search:

1. Open NCBI in a web browser by going to the NCBI home page and click on PubMed located in the list on the right.



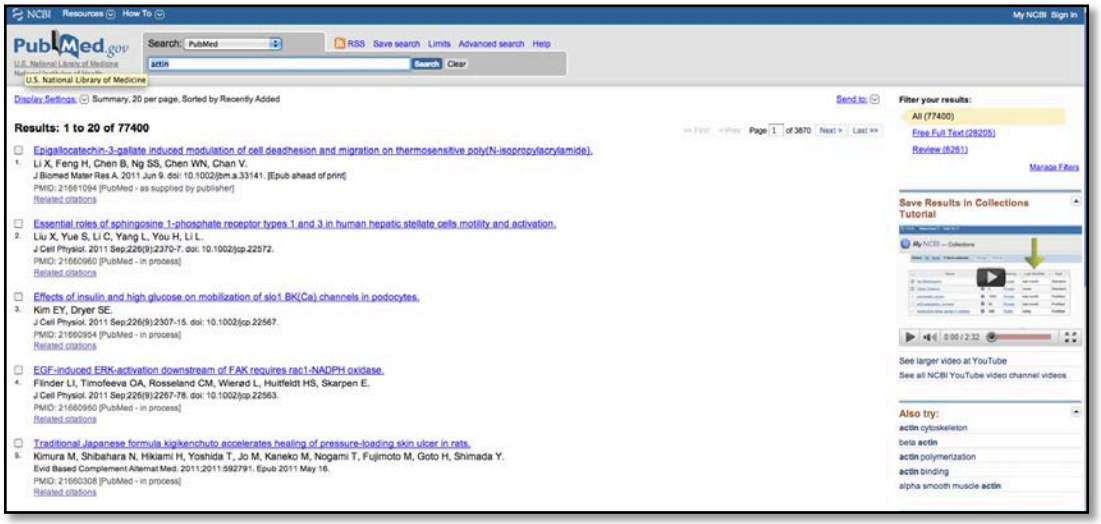
This is the PubMed homepage…



2. To search for a topic, think about the topic to identify keywords, author last names, journal titles, publication year, or institution that can be used as search terms.

For the first search enter 'actin' and click 'Go.' The result of the search is shown below.

4. The details of the results list like the one shown below will be discussed in class.



Basic Lab Skills

Everyone who works in a biology laboratory must know how to make accurate dilutions – whether of chemicals or microbes. For example, concentrated stock solutions must be diluted prior to adding it to a reaction mixture in a test tube. In this activity you will learn about (i) units of measurement, (ii) dilutions and (iii) how to use the micropipettes.

Activity

Part 1 – Dilutions and Micropipetting

In the molecular biology lab we use volumes in the range of microliters (µl) to liters.

Liter, l = 1
Milliliter, ml = 0.001 l or $1x10^{-3}$ l
Microliter, µl 0.000001 l or $1x10^{-6}$ l or $1x10^{-3}$ ml

Solutions are often supplied by companies as concentrates (also called stock solutions) that have to be diluted with water before use. Stock solutions are given names that refer to just how concentrated they are. For example a stock solution that is 10-fold concentrated is known as a 10X solution. A 10X solution needs to be diluted 10-fold to a 1X solution prior to use. This activity will show you how it is done.

To calculate a dilution you use the formula $c_s v_s = c_f v_f$ rearranged in this way:

$$\frac{\text{Final concentration } (c_f)}{\text{Stock concentration } (c_s)} \times \text{Final volume } (v_f) = \text{Volume of stock } (v_s)$$

If you have a 10X solution and need 100 µl of a 1X solution, how would you use the formula above? An instructor will lead you through the calculation in class.

The food coloring is a 10X solution and you need to dilute it before using. Calculate the dilutions for the dyes and fill in columns 3 and 4 of Table 1.

Table 1. Dilutions for 10X food coloring stocks

Tube Label	Amount of dye (µl)	Amount of dH_2O (µl)	Total volume (µl)
Red			500
Blue			1000
Green			750
Yellow			870

Micropipettes measure microliter (µl) volumes. Remember that 1 milliliter (ml) = 1000 µl. You have four pipettes that measure different volume ranges. The pipette is named based on the largest volume it can measure with the letter 'P' added:

P2	0.5 µl – 2 µl
P20	2 µl – 20 µl
P200	20 µl – 200 µl
P1000	200 µl – 1000 µl

Your instructors will demonstrate how to use the micropipettes.

1. Label 4 tubes Red, Blue, Green, and Yellow.
2. Make dilutions according to your calculations in Table 1 by mixing the dye and water volumes indicated.
3. Label 4 new tubes with 5—8.
4. Using Table 2 transfer the correct amount of the diluted food coloring into the new tubes to create new colors.

Table 2. Create new colors using diluted food coloring.

Tube number	Green (µl)	Yellow (µl)	Red (µl)	Blue (µl)	Total volume (µl)
5	0	300	60	0	360
6	0	0	30	16	46
7	27	0	0	120	147
8	27.5	50	57	12.5	147

5. Mix all tubes well.
6. Compare your results to the standards provided by your instructors.

Part 2 – Pipetting Consistency

To measure your pipetting consistency you will pipette the same volume several times onto a balance and weigh the volume of liquid each time. Water has a density of 1 g/ml. If you pipette 1000 µl of water, how much should it weigh?

1. Place a "weigh boat" on a balance and "zero" or "tare" the balance.
2. Set a P1000 to 500 µl.
3. Pipette 500 µl of water on the weigh boat. Record the mass.
4. Repeat step 3 four times.
5. Select a new volume greater than 200 µl and take five measurements recording the weights each time.
6. How consistent are your repeated pipettes?

Extension Questions

Start these in class. They will be completed as homework using Google Docs.

1. 6X Loading Dye is a stock solution that will be used often. If the final volume of a solution is 12 µl calculate the amount of 6X Loading Dye and water you would use to dilute the Loading Dye to 1X. Show your work.
2. In Part 3 of class water was repeatedly pipetted onto a balance. When repeated measures are taken it is useful to know the range of the measurements. One way to report the range is to use the standard deviation (SD) of the measurements. Follow the steps below to calculate the SD for the measurements made in Part 3 for the 500 µl volume.
    a. Calculate the average of the measurements. The average is calculated by adding up all of the measurements and dividing by the number of measurements.

        Example weights: 0.1, 0.2, 0.4, 0.3, 0.1 (ml)

        $$\text{Average} = \frac{0.1+0.2+0.4+0.3+0.1}{5} = \frac{1.1}{5} = 0.22$$

    b. Subtract each of the measurements from the average to obtain the deviations.

        Deviations

        0.1 – 0.22 = -0.12
        0.2 – 0.22 = -0.02
        0.4 – 0.22 = 0.18
        0.3 – 0.22 = 0.08
        0.1 – 0.22 = -0.12

    c. Square the deviations. This is done to turn negative values into positive ones.

        $-0.12^2 = 0.014$
        $-0.2^2 = 0.0004$
        $0.18^2 = 0.0324$
        $0.08^2 = 0.0064$
        $-0.12^2 = 0.014$

    d. Sum the squares of the deviations.

        0.014+0.0004+0.0324+0.0064+0.014 = 0.0672

    e. Divide the sum of the squares by one less than the number of measurements. This is the variance.

        Variance = 0.0672/(5-1) = 0.0168

    f. Take the square root. This is the standard deviation.

        √0.0168 = 0.130

        Accuracy is 0.2 ± 0.130 ml or 200 ± 130 µl

    g. Comment on your pipetting accuracy.
3. Calculate the standard deviation for the 200 µl volume.

Learning Goals

Students will:

1. Be able to calculate dilutions.
2. Be able to properly dilute common reagents.
3. Be able to use a micropipette to accurately measure liquids.
4. Be able to calculate the standard deviation of a set of numbers.
5. Be able to use the standard deviation to discuss the variation in a set of numbers.
6. Be able to use PubMed to search scientific journals.

2: Genetic Information Transfer

The information to make proteins is stored in the genes of an organism's genome. The process of going from genes to proteins is called genetic information transfer and is the topic for discussion on Monday.

1. View the Khan Lecture "DNA: An Introduction to DNA" http://www.khanacademy.org/v/dna?p=Biology
2. View the animation from the Dolan DNA Learning Center on the structure of the macromolecules involved in information flow. http://dynamicgene.dnalc.org/index.html

Click on "Structure" at the bottom and step through the animation.

Answer these questions before class.

1. Briefly define the following terms:
    a. Replication
    b. Transcription
    c. Translation
    d. DNA
    e. RNA
    f. tRNA
    g. Splicing
    h. Ribosome
    i. DNA Polymerase
    j. RNA Polymerase
    k. Spliceosome
2. Describe in words and diagrams information transfer.

Learning Goals

Students will be able to:

1. Define the following terms
    a. Replication
    b. Transcription
    c. Translation
    d. DNA
    e. RNA
    f. tRNA
    g. Splicing
    h. Ribosome
    i. DNA Polymerase
    j. RNA Polymerase
    k. Spliceosome
2. Relate the terms in #1 to each other to diagram and describe the transfer of genetic information in a cell.

3: Restriction Digestion and Gel Electrophoresis

Background

Before it can be studied in the lab, a gene must be isolated from the genome and then amplified into large quantities. For example, a scientist may want to amplify the actin gene (the gene that codes for the actin protein) to determine its exon/intron structure (as you will do). There are two common methods for amplifying a gene of interest. One requires that the gene be inserted into a bacterium where it is replicated as the bacterium divides and replicates its own chromosome (Figure 1a). The second method, called the Polymerase Chain Reaction (PCR) uses the DNA replication machinery in a test tube to make many copies of the gene of interest (Figure 1b). We will discuss PCR in detail during the next class. For now we will focus on using bacteria to amplify DNA.

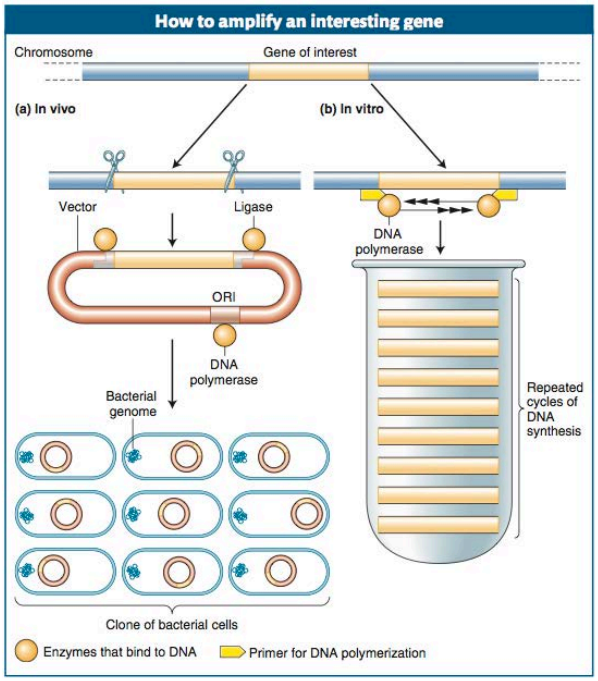


Figure 1. Two methods of isolating and amplifying a gene are (a) *in vivo*, by tricking the replication machinery of a bacterium into amplifying the plasmid containing the gene, and (b) *in vitro*, in the test tube using the polymerase-chain-reaction technique. Both methods employ the basic principles of molecular biology: the ability of specific proteins (yellow) to bind to DNA and the ability of complementary single-stranded nucleic acid segments to hybridize together (the primer used in the test- tube method).

To amplify a gene using bacteria, the gene is cut out of the genome using restriction enzymes. These enzymes bind to DNA at specific sequences and cleave the DNA backbone. For example in Figure 2 the restriction enzyme *EcoRI* binds to the sequence GAATTC and cleaves the backbone between the G and the A. The fragment of DNA that contains the gene of interest is then inserted into a DNA molecule called a plasmid. Plasmids are small, circular DNA molecules found in

some bacteria. The plasmid is first cut with *EcoRI* and then the linear plasmid is combined in the same test tube with the gene fragment. Because the gene fragment and the plasmid have complementary ends (called sticky ends), they will stick together and the enzyme ligase will connect the DNA backbone creating a circular recombinant molecule. This animation demonstrates cutting and ligating DNA: http://www.dnalc.org/view/15476-Genetic-engineering-inserting-new-DNA-into-a-plasmid-vector-3D-animation-with-with-basic-narration.html.

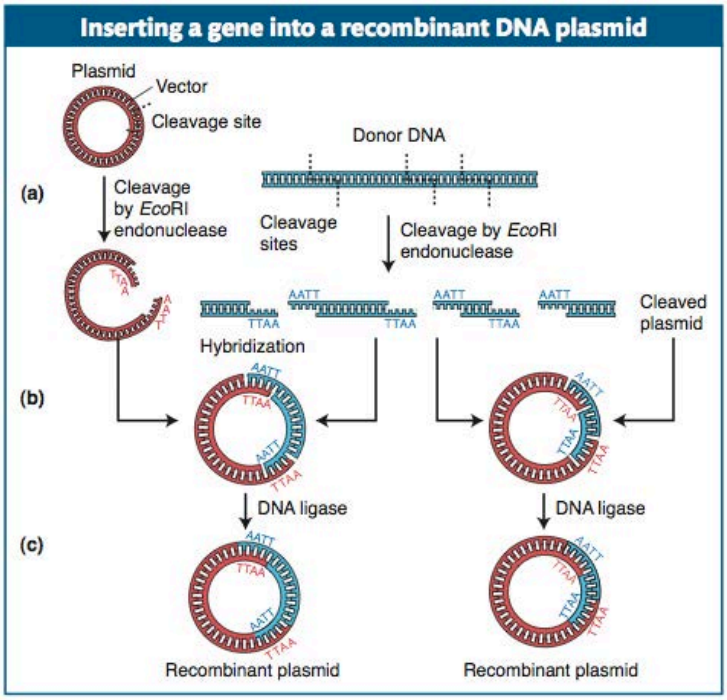


Figure 2. Method for generating a collection of recombinant DNA plasmids containing genes derived from restriction enzyme digestion of donor DNA.

The plasmid containing the new gene is inserted into the model bacteria *E. coli* using a technique called transformation. You will transform bacteria with plasmid DNA at a later time. Inside the bacterium, the plasmid is replicated as the cell divides, making many copies of the gene (Clone of Bacteria cells in Figure 1). The plasmid can be extracted, isolated and stored in a test tube in the refrigerator for future use.

In this activity you will digest plasmid DNA with the restriction enzyme *PvuI* that cuts sequence the CGATCG between the T and the C. The plasmid will be cut one or more times creating linear DNA fragments. After the plasmid is completely cut, you will visualize the resulting DNA using agarose gel electrophoresis. This technique separates DNA fragments based on size in an electrical field. The carbohydrate agarose, obtained from seaweed, is melted in a solution that conducts electricity. As the agarose solidifies a sieve like network of holes is created. DNA will move through the gel when an electrical current is applied at a rate that is inversely proportional to its size. Watch the "Gel Electrophoresis" (http://www.dnalc.org/resources/animations/gelelectrophoresis.html) animation from the Dolan DNA Learning Center. One lane (called a slot in the animation) is used for a DNA size reference called a "ladder." A copy of the reference is provided at the end of

this section. Cut it out and tape it onto the inside of the front cover of your lab notebook. The ladder will be used to determine the sizes of the DNA fragments and the overall size of the plasmid.

Logistics/overview

-Each student will be given one plasmid to digest and then run on 2 lanes that will include the experimental (with enzyme) and the control (without enzyme).
-There will be 4 students/bench and each will have a different plasmid.
-Each bench will run one gel that will include 8 samples plus 2 DNA ladder lanes.
-All students at the bench will take a gel picture and analyze the data independently and then in a group.

Activity

There are 4 parts to this activity:

1. Set-up restriction digest.
2. Pour a gel.
3. Load, run and photograph gel
4. Analyze results and determine which plasmid you have.

Part 1 – Restriction digest

You will set up two digests:

A. Plasmid plus enzyme.
B. Plasmid no enzyme.

The "no enzyme" reaction is the negative control. The reaction with enzyme should have a very different pattern from the negative control if the enzyme successfully digests the DNA.

Materials:
Tubes
Pipettes and tips
Plasmid
Enzyme mix
No Enzyme mix
Lab marker
Water bath at 37˚C
Ice block

Cautions:

Keep tubes on the ice block until the 37˚C incubation step!

Method:

1. Label 1 tube 'A' and 1 tube 'B,' and put your initials on both tubes. Place in ice block.
   Tube A: Plasmid + Enzyme mix
   Tube B: Plasmid + no enzyme mix
2. Add 19.0 µl of *PvuI* enzyme mix to Tube A.
3. Add 19.0 µl of No enzyme mix to Tube B.
4. Add 6.0 µl of plasmid to both Tube A and Tube B.
5. Mix both tubes well.
6. Spin tubes in centrifuge to collect all liquid at the bottom. The Instructors will demonstrate how to use the centrifuge.
7. Place your tubes in the 37˚C water bath for at least 15 minutes. Record the time in your notebook. The enzyme catalyzes optimally at 37˚C.

Part 2 – Pour Gel

Gels will be used through out this class. Today the instructor will demonstrate how to set up the gel rig and how to melt agarose. Each bench will need one gel.

Gels are used in different percentages depending on the size of the DNA fragments that need to be separated. Today we will pour a 1% gel that contains 0.5 grams of agarose in 50 ml of TAE buffer. To pour a gel:

Materials:
Gel rig (Base, tray, comb, and lid)
Power supply
Agarose
Tris-Acetate EDTA Buffer (TAE)
Ethidium Bromide (EtBr)

Caution:
Ethidium bromide intercalates between the bases in a DNA helix and can cause mutations. You must wear gloves when working with EtBr. Use only pipettes labeled EtBr for pipetting ethidium bromide.

Method:
1. Assemble the gel rig.
2. Weigh out 0.5 gram of agarose and place in a 200 ml flask.
3. Add 50 ml of TAE buffer to the flask. Swirl to distribute the agarose.
4. Microwave the flask on HIGH for 30-45 secs. Using the Hot Hands swirl the flask. **CAUTION: The liquid will be very hot!**
5. Microwave the flask on HIGH for 1 min. Swirl the flask to make sure that all of the agarose is melted. If you have "floaters", microwave for an additional 30 seconds.
6. Put on gloves and add 5.0 µl of Ethidium Bromide to the gel.

7. Pour the agarose into the gel tray where it will solidify in about 15 minutes.

Part 3 – Load, Run, Photograph Gel

Materials:
6X Loading Dye

Method:
Your instructors will demonstrate how to prepare the gel for loading and how to load the gel.

1. Take Tubes A and B out of the water bath. Add 5 µl of 6X Loading Dye to each tube. Mix.
2. Wearing gloves, remove the gel tray from the gel rig and turn 90 degrees. Put tray back in rig. Pour TAE Buffer in rig until gel is just covered.
3. Load the DNA Ladder into the first well of the gel. Load the other samples. Write down what you are loading in each well.
4. Put lid on gel rig carefully. Plug leads into the power supply. The leads are black for negative and red for positive and DNA will migrate towards the positive pole. Remember, "Run to the red." An Instructor will show you how to adjust the voltage to 170 V and set the timer for 20 minutes.
5. When the gel is done, photograph it using the GeneFlash. An instructor will show you how this is done.

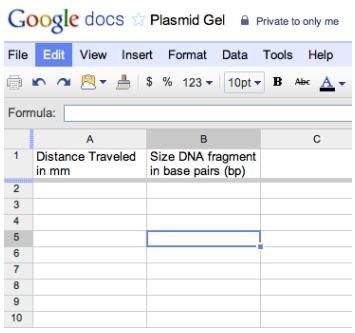
Part 4 – Data Analysis

DNA fragments migrate through an agarose gel at a rate that is proportional to the size of the fragment. The DNA Ladder that was placed in the first well is the reference. The size of the bands is known (see image at the end of this section) and we can measure the distance each band moves. This information is used to estimate the size of the DNA bands by using graph paper or computer spreadsheets. Google spreadsheets will be used in this class to estimate bands sizes on agarose gels.

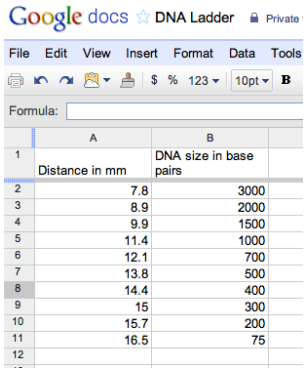
An Instructor will demonstrate how to take gel measurements and set up the spreadsheet.

Method:

1. Open a Google Spreadsheet and save it as "Plasmid gel."
2. Label the first two cells (A1, A2) as shown in the screen shot.



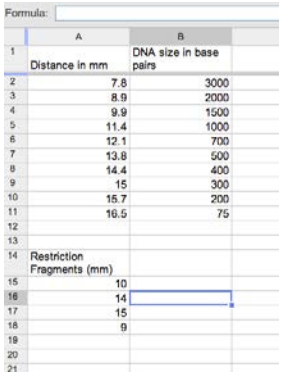
3. Tape the gel photograph into your notebook.
4. Using a ruler draw a line across the gel wells. This is your baseline.
5. Measure the distance between the baseline and each band in the ladder from 200 bp to 4,000 bp. Record the distance and DNA band size in the spreadsheet.



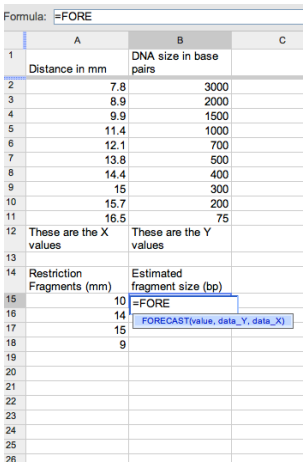
This is the reference for this gel. Steps 1-5 must be repeated for each new gel.

Add graphing just for an example.

6. Measure the distance between the baseline and a restriction fragment. Use the standard curve to estimate the length of the fragment. Record on the spreadsheet.



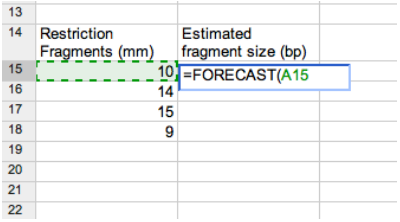
7. Double click the cell next to the first band size start typing “=FORECAST()”.



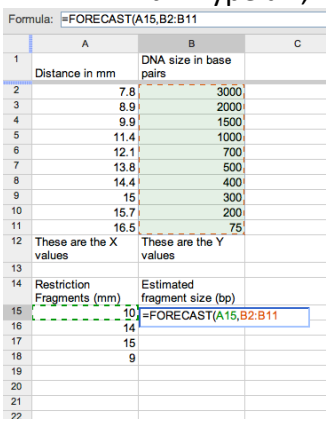
The function called FORECAST calculates the linear regression equation (y=mx+b) and will calculate the size (y) based on the distance travelled (x). FORECAST takes three pieces of data:

a. Value = the distance the band of interest travelled
b. data_y = the y data, in this case the sizes of the DNA ladder bands.
c. data_x = the x data, in this case the distances measured for the DNA ladder bands.

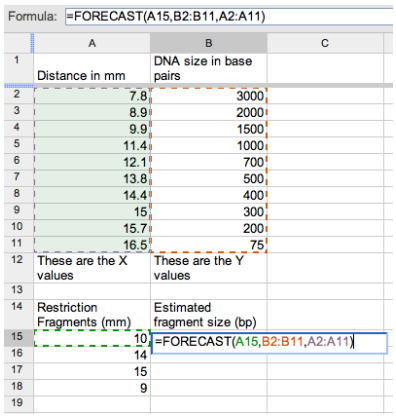
8. Between the parentheses:
   a. Click the distance travelled in the cell to the left. This places a holder for the value in the cell.



   b. Type a "," and then highlight the sizes of the DNA ladder.



   c. Type a "," and then highlight the distances for the DNA ladder.



9. In the formula type "$" in front of the letter and number in the data_y and data_x values. This one will read A15, $B$2:$B$11, $A$2:$A$11. This holds the data values constant.
10. Press Enter to complete the formula and the value for the band size is shown.
11. Click on the little blue square on the cell highlight and drag down to fill in the other band sizes. When copying the formula from one cell to another, the first value is replaced with the next distance value.

Plasmid and Restriction Fragment Lengths.

Plasmid	Total Length (bp)	Fragment Lengths (bp)
pBR322	4361	4361
pUC19	2686	895, 1791
pRS413	4970	1916, 3054
pRS415	6021	1916, 4105

Extension Questions

1. A 1% gel contains 0.5 g of agarose in 50 ml of TAE buffer. How many grams of agarose is needed for a 1.5% gel in 50 ml of TAE? Show your work.
2. Complete documenting your gel results using the spreadsheet. Determine the sizes of all of the plasmids on the gel. Use the sizes predicted in Part 2 to determine which plasmid is which.
3. Why are the sizes determined on the gel not exactly the same as the actual sizes?

Learning Goals

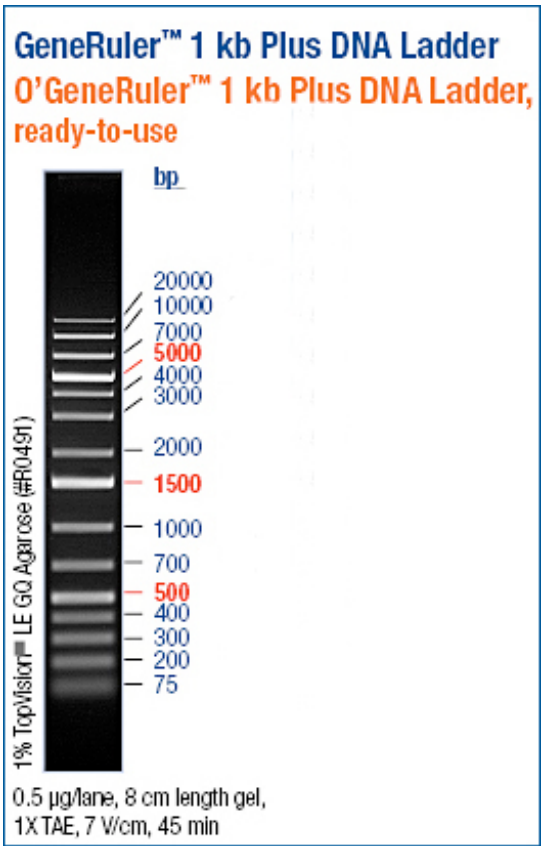
Students will:

1. Be able to define plasmid and restriction enzyme.
2. Understand that restriction enzymes cut DNA into smaller fragments.
3. Be able to set up a restriction digest and understand the purpose of the no enzyme control reaction.

4. Understand how DNA fragments are resolved on agarose gels.
5. Be able to analyze the results of gel electrophoresis to determine DNA fragment length.

DNA Ladder Reference

Cut this image out and tape it onto the inside front cover of lab notebook.



4: Genome Annotation

The human genome contains 2.5 billion base pairs (bp). Coincidentally the maize genome is approximately the same size. Genome annotation is the process of giving meaning to those billions of Gs, As, Ts, and Cs. A combination of experimental evidence and computer predictions is used to annotate genomes. In this class we will use a computer tool called the DNA Subway for genome annotation. The Polymerase Chain Reaction (PCR) will be used to generate experimental data for the computer annotations.

Today we will:

1. Review how introns are experimentally determined.
2. Review the process of PCR.
3. Use PCR to amplify the actin gene.
4. Use the DNA Subway to annotate a portion of a genome.

Background

Genes can be predicted by computer programs, and those predictions need to be verified by experimental evidence. A scientist has to go into the lab and do experiments to obtain this evidence. In the lab today an experiment will be started to gather evidence for the gene structure of Actin. To do this two DNA samples will be isolated: genomic DNA and DNA made using the mRNA extracted from the plant. PCR will be used to amplify the Actin gene from these two DNA samples and the results will be visualized on a gel. In the next class the sequence of the isolated regions will be compared using bioinformatics tools to determine the gene structure.

PCR

PCR is a powerful technique used to isolate and amplify specific regions of DNA. The Actin gene that is the focus of this activity is about 2 kb and is one of over 20,000 genes in the genome. PCR is a remarkable technique because a scientist can use it to isolate a single gene from the entire genome. Figure 1 describes the steps in PCR.

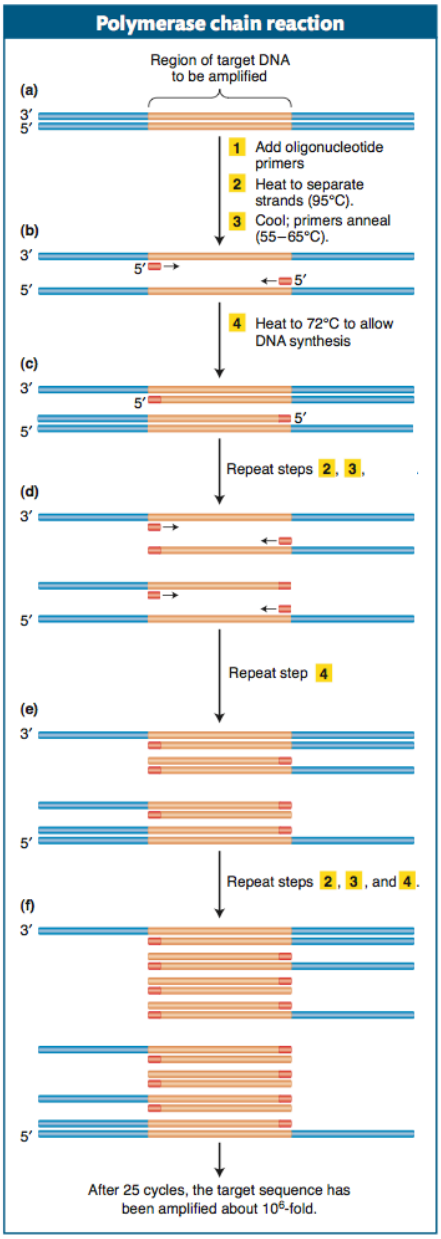


Figure 1. The polymerase chain reaction specifically copies a target DNA sequence. (a) Double-stranded DNA containing the target sequence. (b) Two primers have sequences complementing primer-binding sites at the 3' ends of the target gene on the two strands. The strands are separated by heating (called denaturation); then cooled to allow the two primers to base pair to the primer-binding sites (called annealing). Together, the primers thus flank the targeted sequence. (c) After the temperature is raised, Taq polymerase then synthesizes the first set of complementary strands in the reaction (called extension). These first two strands are of varying length, because they do not have a common stop signal. They extend beyond the ends of the target sequence as delineated by the primer-binding sites. (d) The two duplexes are heated again, exposing four binding sites. After cooling, the two primers again bind to their respective strands at the 3' ends of the target region. (e) After the temperature is raised, Taq polymerase synthesizes four complementary strands. Although the template strands at this stage are variable in length, two of the four strands just synthesized from them are precisely the length of the target sequence desired. This precise length is achieved because each of these strands begins at the primer-binding site, at one end of the target sequence, and proceeds until it runs out of template, at the other end of the sequence. (f) The process is repeated for many cycles, each time creating more double-stranded DNA molecules identical with the target sequence.

Watch these two videos about PCR from the Dolan DNA Learning Center :
3D Animation: http://www.dnalc.org/view/15475-The-cycles-of-the-polymerase-chain-reaction-PCR-3D-animation-with-audio.html
2D Animation: http://www.dnalc.org/view/15924-Making-many-copies-of-DNA.html. Click on "Animation" in the Flash window and step through.

cDNA is synthesized from mRNA

In order to determine gene structure one could find the location of exons by comparing the sequence of spliced mRNA to the sequence of genomic DNA. Unfortunately this simple approach is not possible because it is very difficult to directly determine the sequence of RNA molecules. Instead, scientists first synthesize a DNA copy of an mRNA using the enzyme reverse transcriptase (originally isolated from viruses similar to HIV/AIDS) and then sequence the DNA copy. Reverse transcriptase is a DNA polymerase that uses RNA instead of DNA as template. Like other DNA polymerases, reverse transcriptase requires a short DNA primer. To prime mRNAs specifically, a special primer (of ~20 T's, called an oligo-dT primer) is added to the total RNA isolated from the cell to bind the poly(A) tail of mRNA (Figure 2). The DNA copy synthesized by reverse transcriptase is called complementary DNA (cDNA). The cDNA is made double stranded using a DNA polymerase. Unlike RNA, cDNA can serve as a template for PCR.

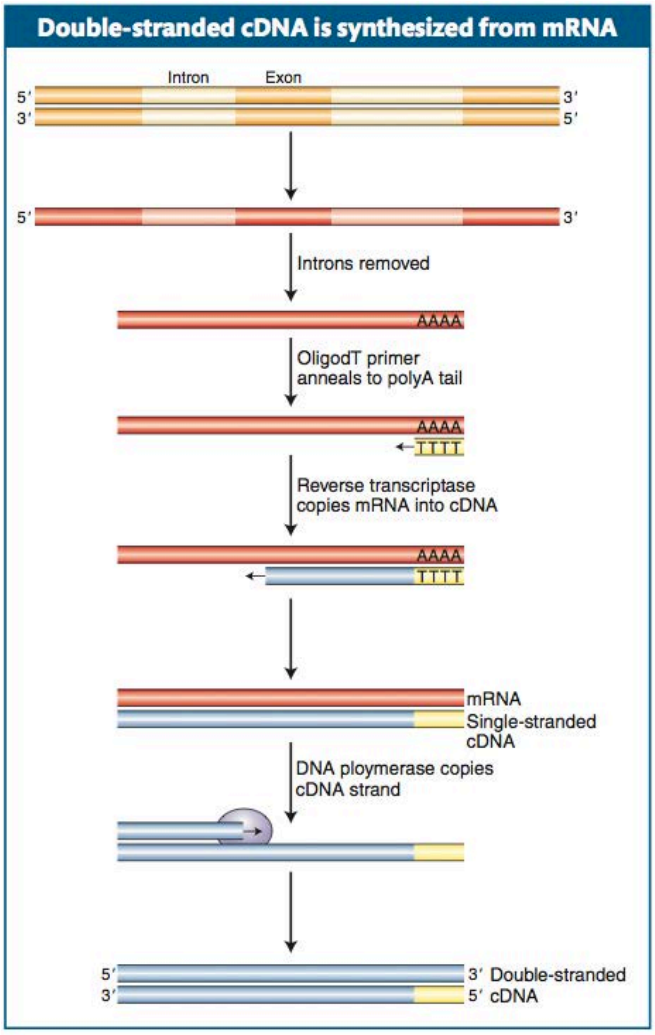


Figure 2. The formation of cDNA for the Actin gene. The Actin gene (with its introns) is transcribed into pre-mRNA. The introns are removed in the cell by splicing, and A residues are added to the 3' end to form polyadenylated mRNA. In the laboratory, mRNAs are isolated from cells and a short oligo(dT) primer is hybridized to the poly(A) tail of all mRNAs to prime synthesis of complementary DNA from the RNA template by reverse transcriptase. When the mRNA strand has been degraded (by treatment with NaOH or RNAseH), addition of a second primer (step not shown) permits initiation by DNA polymerase and completes the synthesis of double-stranded cDNA.

PCR products (called amplicons) from genomic DNA and cDNA can be visualized by agarose gel electrophoresis and, if desired, the resulting bands can be isolated and

mailed to a company for sequencing. Exon/intron boundaries are identified by comparing the cDNA sequence to the genomic sequence.

Activity 1. Gene structure determination

Part 1 –Set up PCR

PCR will be used to isolate the Actin gene from genomic DNA and cDNA extracted from Arabidopsis, rice and maize. Each student will select one plant. After PCR is complete the results will be visualized by gel electrophoresis. The expected sizes of the amplicons for the genomic DNA from each plant are listed below.

Plant source of DNA	**Genomic DNA amplicon size (bp)**
Arabidopsis	871
Maize	921
Rice	1007

In addition to setting up PCR with cDNA and genomic DNA samples, control PCRs need to be set up. Control reactions are as important as the experimental samples because if the experiment does not work, the controls should provide reasons why.

Recall that the restriction digestion experiment included reactions with and without enzyme. The no-enzyme reaction was actually a negative control where nothing is expected to happen to the plasmid DNA if there is no enzyme. If the negative control does not come out as expected then you need to make new reagents and start all over. We will discuss both positive and negative controls in class.

Setting up a reaction usually requires mixing different reagents most of which are added to all the samples (including the controls). These reagents are mixed together into a master mix that is then distributed to the individual sample tubes. The variable reagent is then added to a subset of the sample tubes. All PCR reactions contain enzyme, dNTPs, buffer and water. The variable reagents are the primers and/or the DNA template. Today, the primers will be the same in all reactions and the DNA will vary. To further simplify setting up PCR, companies provide concentrated 2x Taq enzyme mixes that contain enzyme, dNTPs, and buffer. This enzyme mix is diluted during the preparation of the master mix to 1x. We will discuss how to calculate the volumes for the master mix in lab.

Materials:

Large 1.5 ml tubes
Small 0.2 ml tubes

2x Taq enzyme mix, contains:
- Taq enzyme
- dNTPs
- buffer
- gel loading dye (that's why its green!)

Template DNA
Primers (choose the primers that match the plant

Method:

1. Label a 1.5 ml tube as "MM" for master mix.
2. Label 0.2 ml strip tubes with A, B and C. Write your initials on one of the tubes. Instructors will demonstrate labeling tubes. A, B and C correspond to these reactions:
    A. Genomic DNA
    B. cDNA
    C. No DNA Negative control
3. Mix the following the MM Tube in the order listed:

Reagent	1 reaction (µl)	4 reactions (µl)
Water	6.5	26.0
2X Taq Mix	12.5	50.0
Primer F	0.5	2.0
Primer R	0.5	2.0
Total	20.0	80.0

4. Mix the contents of the tube well. Spin briefly to get the contents to the bottom.
5. Place 20 µl of Master Mix in Tubes A, B and C in the strip. There should be leftover Master Mix.
6. To the strip tubes add:
    A. Add 5 µl of genomic DNA
    B. Add 5 µl of cDNA
    C. Add 5 µl of water

    The tubes will now have a total volume of 25 µl.
7. Gently tap each tube in the strip to mix.
8. Place the strip of tubes in the thermocycler.
9. Discard the MM tube. Keep unused 2X Taq Enzyme Mix in your freezer box for later use.

Part 2 – Pour Gel

The PCR products will be visualized on a gel and analyzed. There will be one gel per group. Today we will use a 1.5% gel. Calculate the amount of agarose needed for a 100 ml gel. Don't forget the ethidium bromide. Each bench will run one gel. Place two combs in the gel tray as demonstrated by an instructor.

Part 3 – Run Gel (After DNA Subway work.)

1. Load 12 µl of each reaction on the gel.

2. Load the DNA Ladder in the first and last lanes.
3. Run at 150V for 30 minutes.
4. Photograph gel.
5. Analyze gel and discuss the results with your bench group.

Genome Annotation using the DNA Subway.

Background

In lab one gene out of thousands was targeted for study. This is a powerful way to carefully annotate a gene, but impractical for annotating all of the genes in a genome. To do that many computer programs and a lot of time are required. To simplify the process the DNA Subway was developed. The "subway map" contains three major steps in genome annotation.

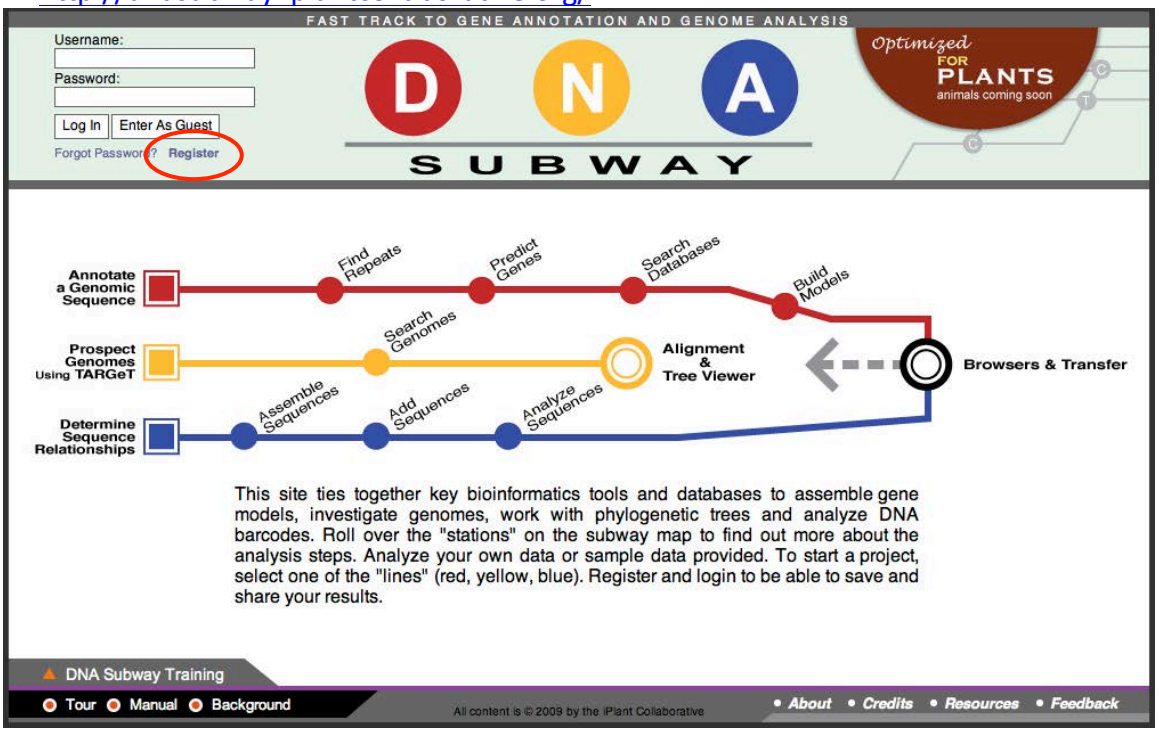
1. **Find Repeats** – The genome is largely repetitive DNA. For example the same nucleotide may be repeated many times or a triplet combination such as CAG may be repeated hundreds of times. There are also mobile pieces of DNA called transposable elements (abbreviated TEs) that exist in many copies. TEs will be discussed in more detail during a later class. To find repeats in the genome, a computer program called RepeatMasker compares the genome to a large library of repeat sequences. When one is found, RepeatMasker replaces each bp in the repeat with the letter 'N'. The sequence is now said to be masked which means that other programs annotating the same sequence will ignore these regions. Masking a genome can hide 50% or more of the sequence.
2. **Predict Genes** – Gene prediction involves finding the sequences in the genome that encode proteins or functional RNAs (like tRNA or rRNA). This turns out to be a difficult task that often involves reaching a consensus from the results of several different programs. The DNA Subway uses the programs Augustus, FGenesH and Snap to predict protein coding genes. It uses tRNA Scan to predict tRNA genes.
3. **Search Databases** – DNA sequences generated in laboratories all over the world are collected in large databases – like GenBank (at the NCBI website). Anyone who wants to find out something about a DNA sequence (called the query) can search GenBank using a tool called BLAST. BLAST quickly searches millions, even billions of sequences to find those that are very similar to the input sequence (the query). There are two types of BLAST searches used by the DNA Subway:
    a. BlastN – Searches a database of DNA sequences.
    b. BlastX – Searches a database of protein sequences.

Activity

Today you will use the DNA Subway to annotate a region of a plant genome that contains the gene Actin 1. Actin is found in all eukaryotes and is a major component of the cytoskeleton. The protein sequence of actin is very similar in many species of plants and animals (it is said to be conserved). We will use it to learn how computer programs predict genes from DNA sequence and how scientists provide experimental evidence to support the predictions. DNA sequences from three species of plants (*Arabidopsis thaliana, Oryza sativa and Zea mays*) are available on the website. Choose one to annotate.

DNA Subway Instructions

1. Create an account on the DNA Subway by clicking on "Register." *Do not use your UCR password for this account.*
   http://dnasubway.iplantcollaborative.org/



   Record your Username and Password here:

   Username:_____

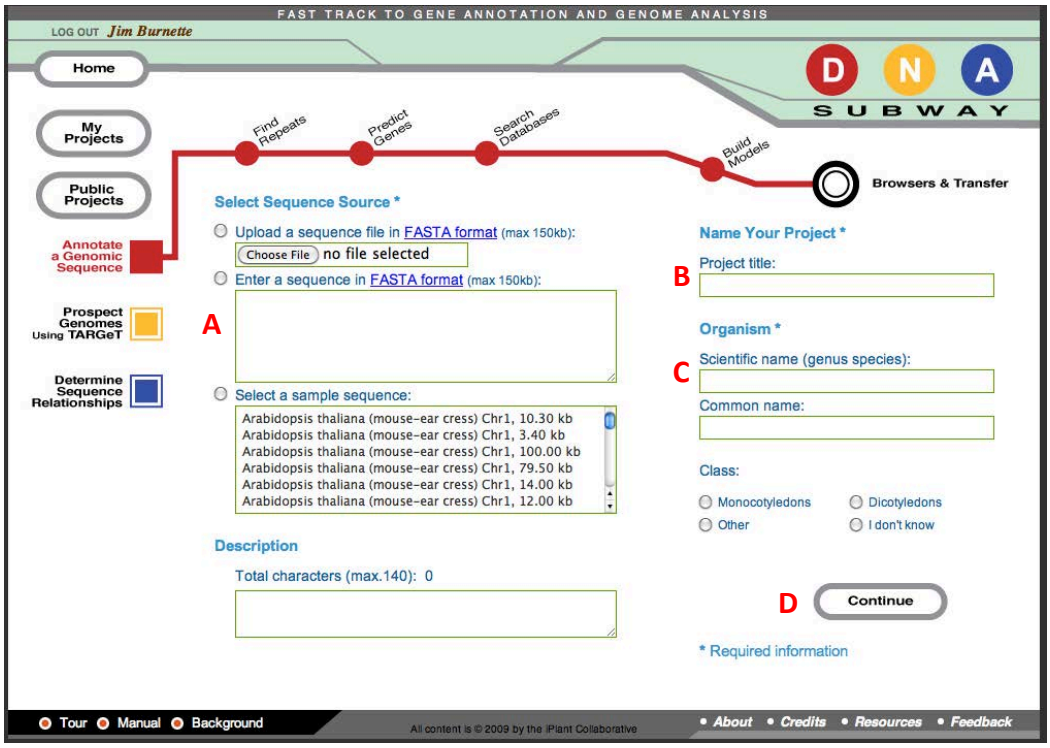
   Password:_____

2. Select the Red station "Annotate a Genomic Sequence."

3. Open a new browser tab and open the link (http://dynamicgenome.ucr.edu/courses/schedule/) for the DNA sequence from your plant.
4. The first step in using the subway is to create a project by entering the DNA sequence to be annotated. The remaining fields document the source of the DNA for future reference. Do the following to get started:
   A. Copy and paste the sequence into the textbox labeled “Enter a sequence”.
   B. Provide a project of your choice in the “Project title field”.
   C. Fill in the organism’s scientific name, common name and class.

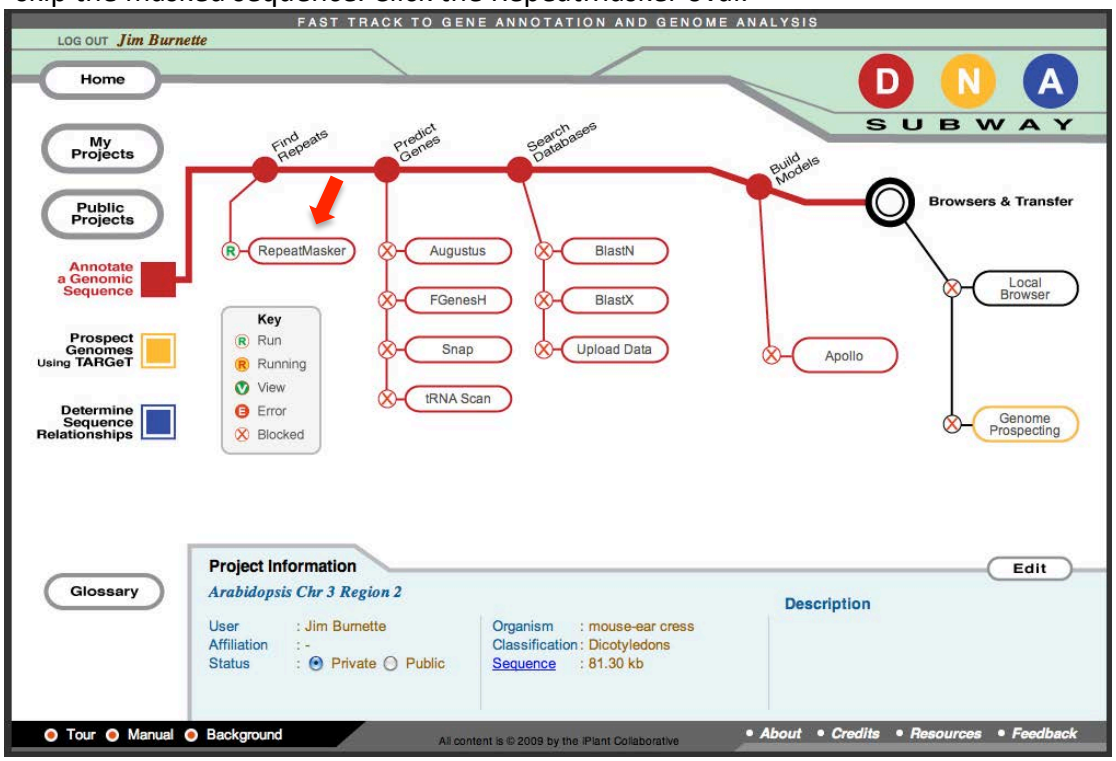
Scientific Name	Common Name	Class
Arabidopsis thaliana	mouse-ear cress	Dicotyledons
Citrus clementina	Clementine	Dicotyledons
Oryza sativa	Rice	Monocotyledons
Zea mays	Maize	Monocotyledons

   D. Click Continue.

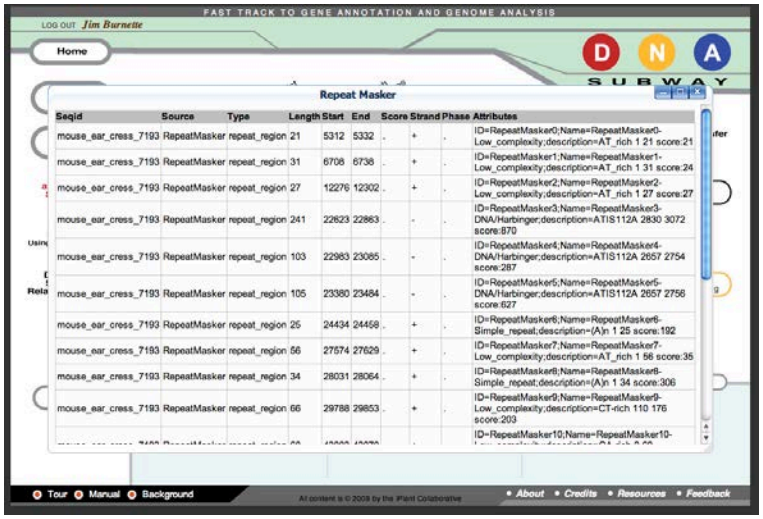


5. Now you are ready to ride the Subway. You will see that there is a green ‘R’ (for run) by the RepeatMasker stop indicating that it is ready to run. All of the other stops have red ‘Xs’ by them. They cannot be used until RepeatMasker is finished its analysis because the repeats “confuse” gene prediction software and slow

them down. RepeatMasker will find the repeat regions in the sequence and mask them. The gene prediction programs used in the next stop on the line will skip the masked sequence. Click the RepeatMasker oval.

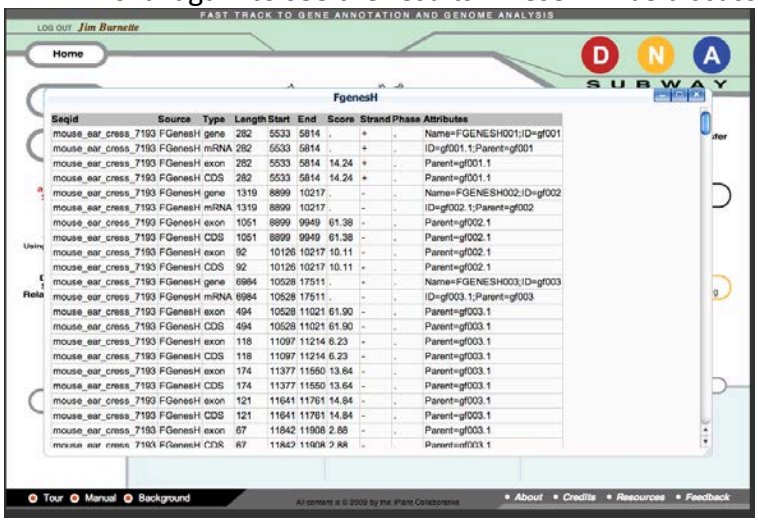


6. When RepeatMasker finishes a ‘V’ will appear. You can now click the RepeatMasker oval again to see the results. We will discuss these in class.

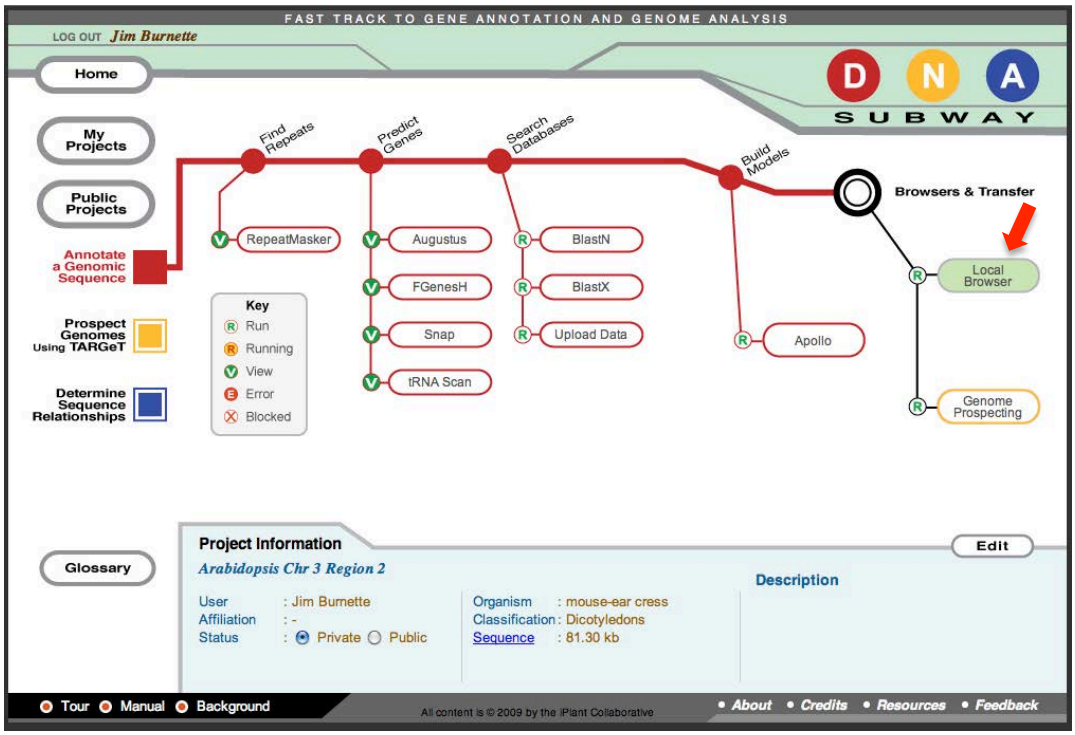


7. Now that the sequence has been masked, gene prediction programs such as FGenesH and Augustus can be used to find genes in the sequence. Click the 4

ovals under “Predict Genes” to start the gene prediction programs. Click on the oval again to see the results. These will be discussed in class.

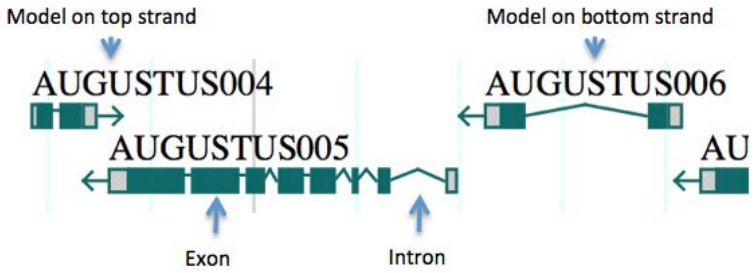


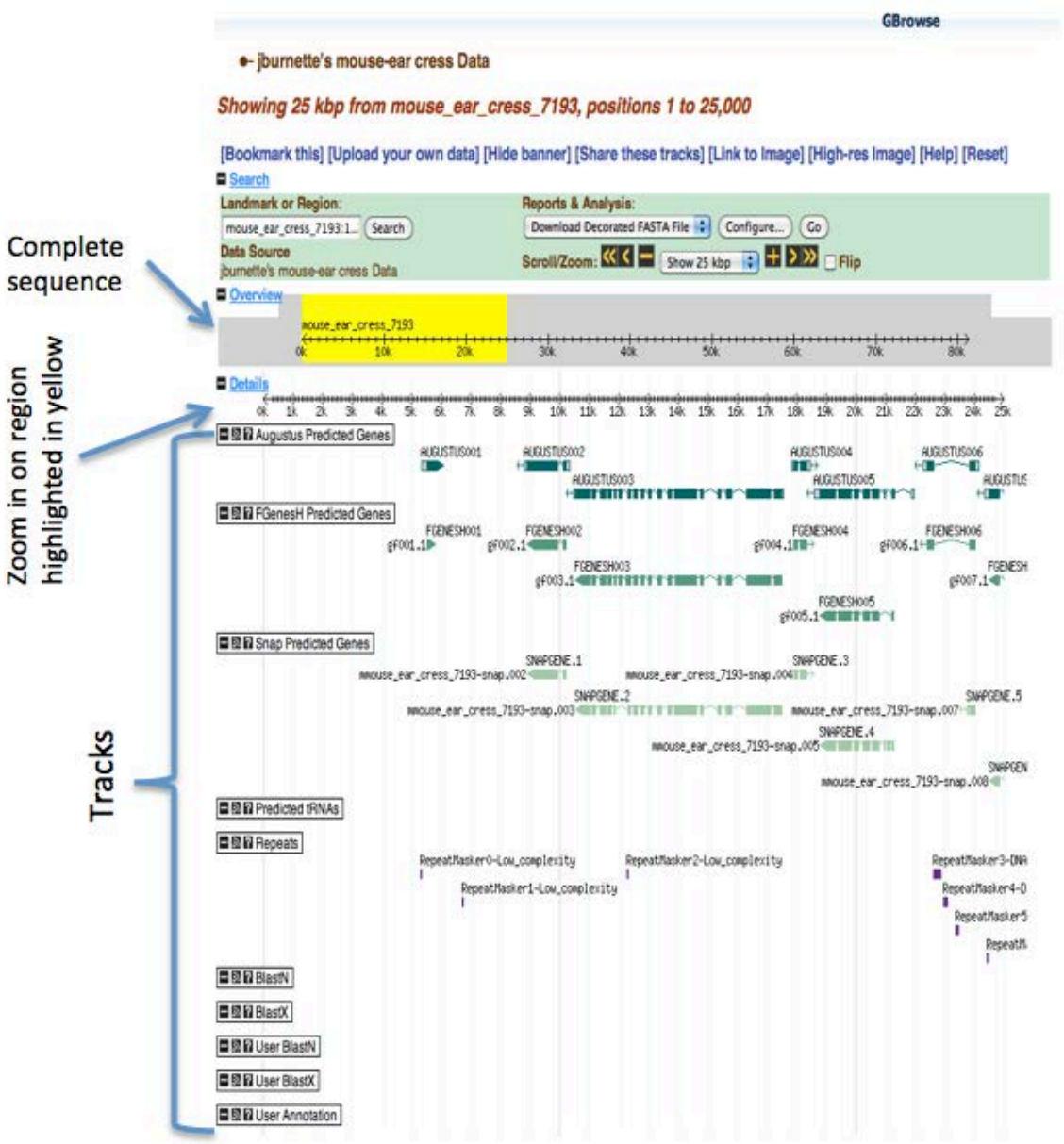
8. View the results in a genome browser by clicking on the oval labeled “Local Browser” to pop-up the genome browser called GBrowse. Be patient, this can take several seconds to load.



9. The output of RepeatMasker and the gene prediction programs is a long list of text that is hard to read. Genome browsers convert this text information into a graphic that is easier to interpret. Information is organized into tracks. We will spend some time exploring the tracks of the genome browser.

A. The line in the Overview represents the complete sequence. The region highlighted in yellow is shown in detail in the tracks below. Use the controls in the green region to zoom in or out of the sequence or to change the region shown in detail.
B. Tracks in the details display the results of the gene prediction programs and of RepeatMasker. Gene predictions are shown as green exons joined by lines representing introns. The right-most or left-most exon will end in an arrow. Remember DNA is double stranded and either strand can encode a gene. If the exon arrow points to the right then the gene is on the top strand. If the exon points to the left, the gene is on the bottom strand.





10. Although gene prediction programs can provide good gene models, they are only hypotheses because they are not supported by experimental evidence. Experimental support for computer models comes from several resources that can be found at the NCBI databases including experimentally derived DNA and protein sequences. Programs such as BlastN and BlastX are used to quickly search large sequence databases. Click on the BlastN and BlastX ovals to search the databases. After the searches end go back to the Local Browser to view the results. Do all of the gene models have supporting evidence?
11. Compare the genomic region you annotated with the regions annotated by your neighbors. How to they compare? Do some have more or less gene predictions, more or less repeats?

Extension Questions

Answer these questions in your lab notebook.

1. What are the sizes of the PCR products? Use a Google spreadsheet to answer this.
2. Compare the genomic DNA amplicon and the cDNA amplicon from your samples. Generate a hypothesis that explains the observed size difference.
3. How would test this hypothesis?
4. Compare the sizes of the cDNA bands obtained by other students for your plant. What is the standard deviation of the sizes? Why is there this variation? How about the genomic band?

Learning Goals

Students will:

1. Be able to describe how PCR and primers isolate a specific region of DNA.
2. Be able describe how PCR primers link bioinformatics to experiments.
3. Be able to describe the steps in a PCR cycle.
4. Be able to describe exponential amplification.
5. Be able to set up a controlled PCR experiment.
6. Be able to describe a negative control and why it is used in PCR.
7. Be able to analyze an agarose gel of the PCR results.
8. Be able to discuss the steps of genome annotation.
9. Be able to use a genome browser to learn about regions of a genome.

5: DNA Sequence Analysis

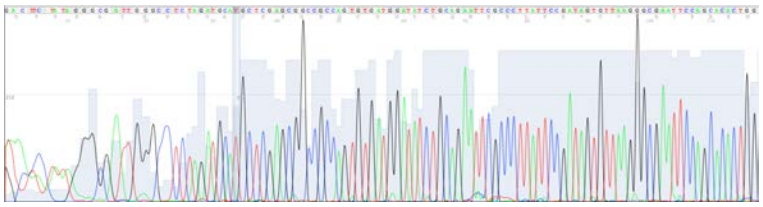
In the previous activity you amplified the Actin gene from genomic DNA and cDNA. The amplicons could be cut out of the gel, purified to remove the agarose and the DNA sequenced. This was done previously and the amplicons were sequenced by the facility at UCR. In this activity the sequencing results will be analyzed to determine the gene structure of Actin. The sequences of the Actin genes from the different species will also be compared.

Activity

Use the following steps to analyze the DNA sequences:
Analyze the DNA samples using the following steps:

Part 1 – Download sequence and check quality.
Part 2 – BLAST the query sequence against GenBank database.
Part 3 – Compare the cDNA and genomic DNA from the same species using Blast2Sequences.
Part 4 – Compare all the genomic sequences.

Part 1 – Check Sequence Quality



The raw output from a DNA sequencer is a called a chromatogram (shown above). High quality sequence will have tall peaks for each base. There will also be a few "N" calls. "N" calls mean that the software could not determine the base at that position. The first 10-20 bases are usually low quality (as seen in the figure above) and the end of the sequence is usually of low quality. Low quality sequence is trimmed off before further analysis.

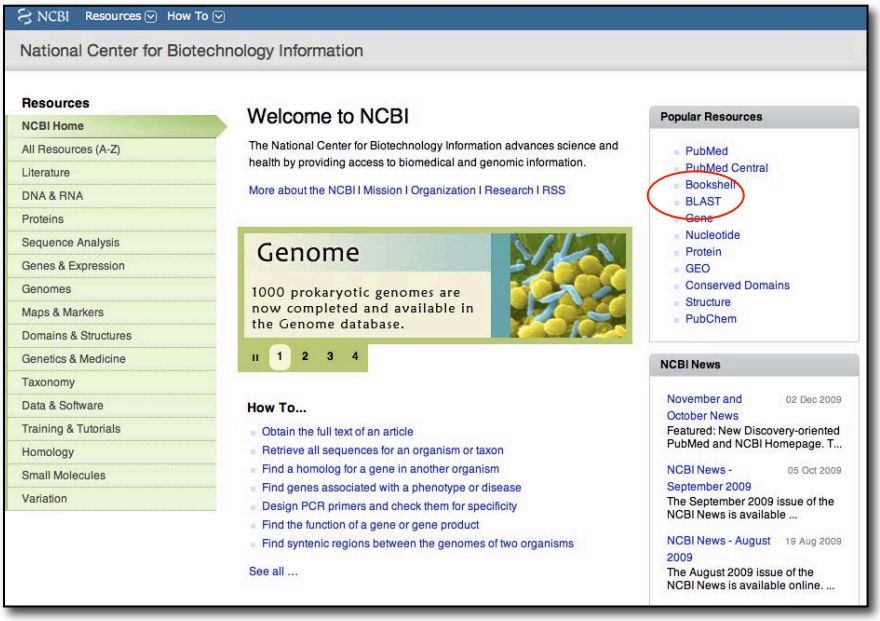
The sequence traces for the actin experiment can be found on the Data page of the course website.

Part 2 – Blast sequence against GenBank

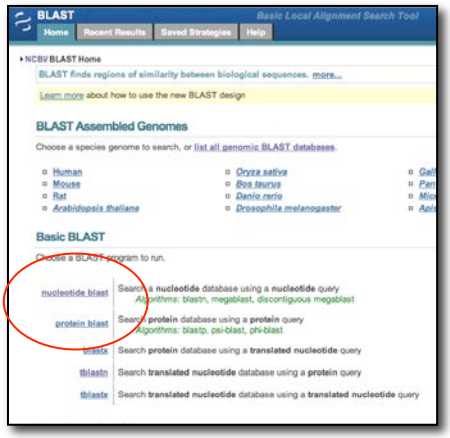
As mentioned in a prior section, GenBank is a database of billions of DNA sequences deposited by scientists from around the world. GenBank is hosted by the National Center for Biotechnology Information (NCBI) one of the institutes of the National Institutes of Health (NIH). A computer program called BLAST is used to search the billions of DNA sequences in GenBank. To do a BLAST search, a query sequence is

entered and the computer program finds similar sequences based on statistical scores. One statistic used is the e-value where a low e-value (usually less than 0.1) indicates a strong match between the query and the "subject or hit". The e-value is the probability of that same match occurring in random DNA sequence. BLAST is very fast. It can search the billions of sequences for matches in seconds. Today we will perform a nucleotide blast (written as blastn) where a DNA sequence (the query) is used to search a DNA database. If Actin 1 is the query, blastn will find related sequences.

Access Blast by clicking on the BLAST link on the NCBI home page (http://www.ncbi.nlm.nih.gov/).



Click on the BLAST to go to the BLAST home page. Bookmark this page because it will be used frequently:



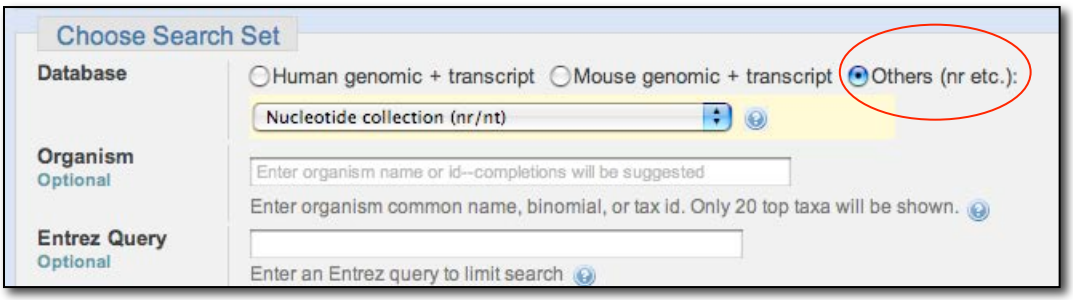
There are five different versions of BLAST. The different versions make it possible to use nucleotide and protein queries to search nucleotide or protein databases and are summarized in the screenshot above. For today use BLASTN because a nucleotide query will be used to search a nucleotide database.

Nucleotide Blast (BLASTN): The most straightforward search using a nucleotide sequence (the query) and "BLAST" it against a nucleotide database (the subject) and can reveal

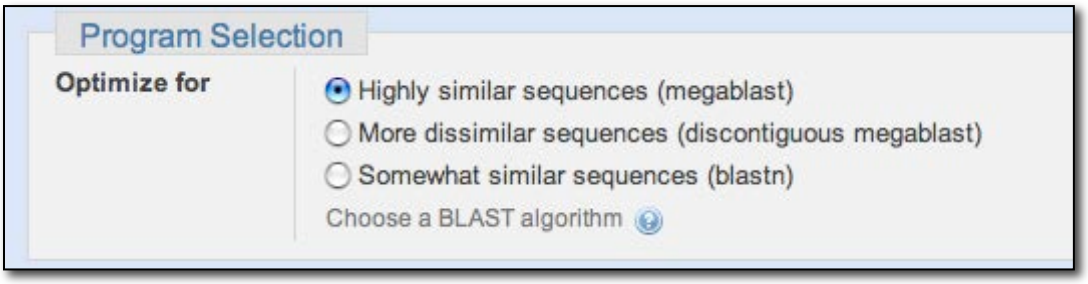
a. whether there are prior publications that reported information about this sequence
b. where the sequence is located in the genome (more on location in class).
c. whether the sequence is found in genomes of closely related organisms.
d. whether the sequence codes for an RNA and/or a protein and what, if anything is known about the protein.

1. Select 'nucleotide blast.' Cut and paste the Actin 1 sequence from the course website in the Query text window (Enter accession number….).

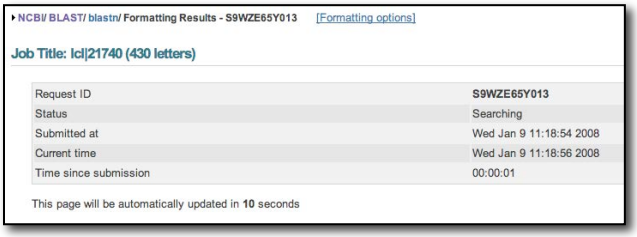
2. Under "Choose Search Set" select "Others" and the drop down list changes to "Nucleotide Collection (nr/nt)." This is the complete non-redundant nucleotide database.



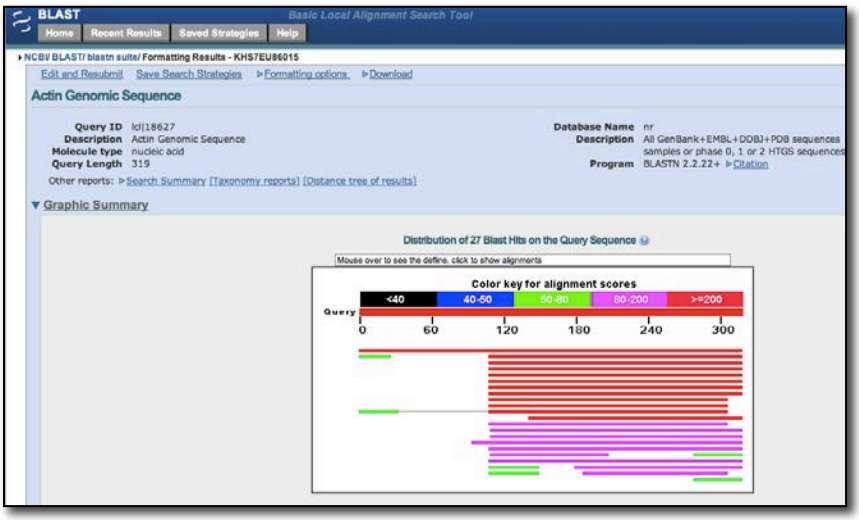
3. There are three versions of BLASTN. Choose megablast (default).



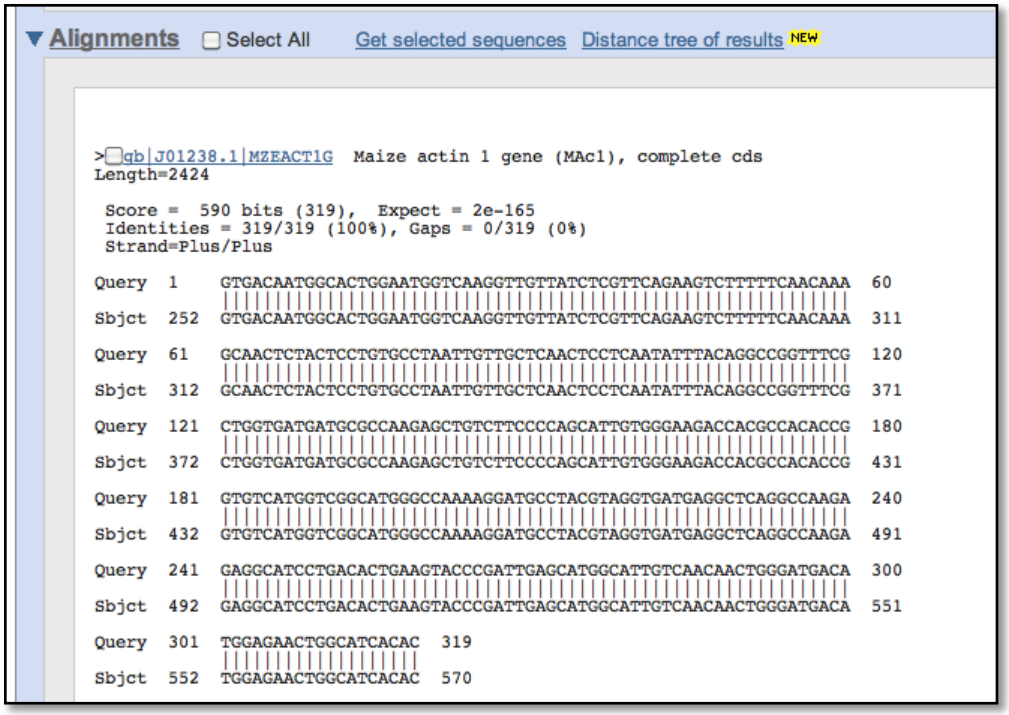
4. Select the "Blast" button and the queue page will appear.



5. The queue page will automatically refresh until the BLASTN is finished. The BLASTN results will be discussed in class.



6. Details of the Alignment (to be discussed in class)



Alignments Select All Get selected sequences Distance tree of results NEW

>gb|J01238.1|MZEACT1G Maize actin 1 gene (MAc1), complete cds
Length=2424

Score = 590 bits (319), Expect = 2e-165
Identities = 319/319 (100%), Gaps = 0/319 (0%)
Strand=Plus/Plus

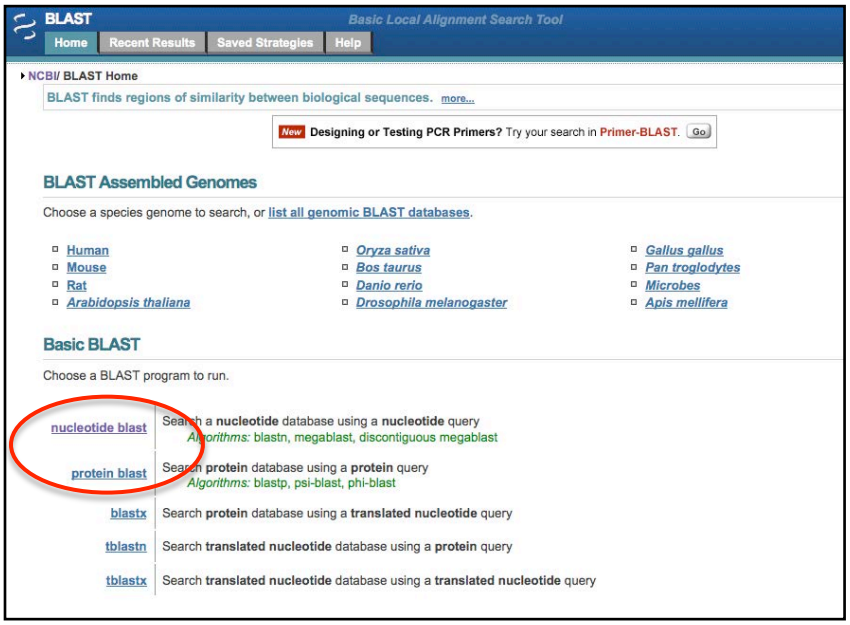
```
Query  1    GTGACAATGGCACTGGAATGGTCAAGGTTGTTATCTCGTTCAGAAGTCTTTTTCAACAAA  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  252  GTGACAATGGCACTGGAATGGTCAAGGTTGTTATCTCGTTCAGAAGTCTTTTTCAACAAA  311

Query  61   GCAACTCTACTCCTGTGCCTAATTGTTGCTCAACTCCTCAATATTTACAGGCCGGTTTCG  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  312  GCAACTCTACTCCTGTGCCTAATTGTTGCTCAACTCCTCAATATTTACAGGCCGGTTTCG  371

Query  121  CTGGTGATGATGCGCCAAGAGCTGTCTTCCCCAGCATTGTGGGAAGACCACGCCACACCG  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  372  CTGGTGATGATGCGCCAAGAGCTGTCTTCCCCAGCATTGTGGGAAGACCACGCCACACCG  431

Query  181  GTGTCATGGTCGGCATGGGCCAAAAGGATGCCTACGTAGGTGATGAGGCTCAGGCCAAGA  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  432  GTGTCATGGTCGGCATGGGCCAAAAGGATGCCTACGTAGGTGATGAGGCTCAGGCCAAGA  491

Query  241  GAGGCATCCTGACACTGAAGTACCCGATTGAGCATGGCATTGTCAACAACTGGGATGACA  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  492  GAGGCATCCTGACACTGAAGTACCCGATTGAGCATGGCATTGTCAACAACTGGGATGACA  551

Query  301  TGGAGAACTGGCATCACAC  319
            ||||||||||||||||||
Sbjct  552  TGGAGAACTGGCATCACAC  570
```

7. What is the Accession number and gene name of the best hit to the query sequence?
8. Repeat with the cDNA sequence. What is the best hit?

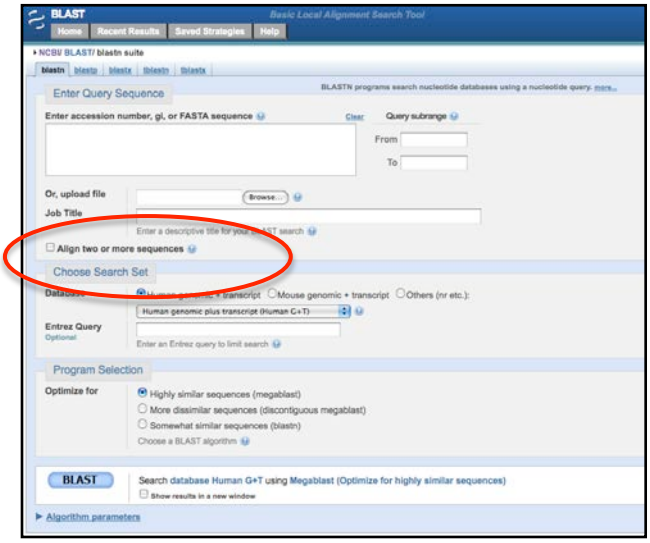
Part 3 -- Compare the cDNA to the genomic DNA sequences.

The cDNA and genomic DNA sequences need to be aligned to determine where they are similar and where they differ. While this could be done by hand (also called manually) for short PCR sequences, it would be time consuming. Fortunately, a modified version of blastn called Blast2Sequences exists that performs alignments quickly and accurately. In this case the query (Genomic Actin 1 DNA) will be compared to a single subject (cDNA) instead of to a database.

1. Open a web browser, go to the Blast Website (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi) and click on the 'nucleotide blast' link.



2. Check the 'Align two or more sequences' checkbox.

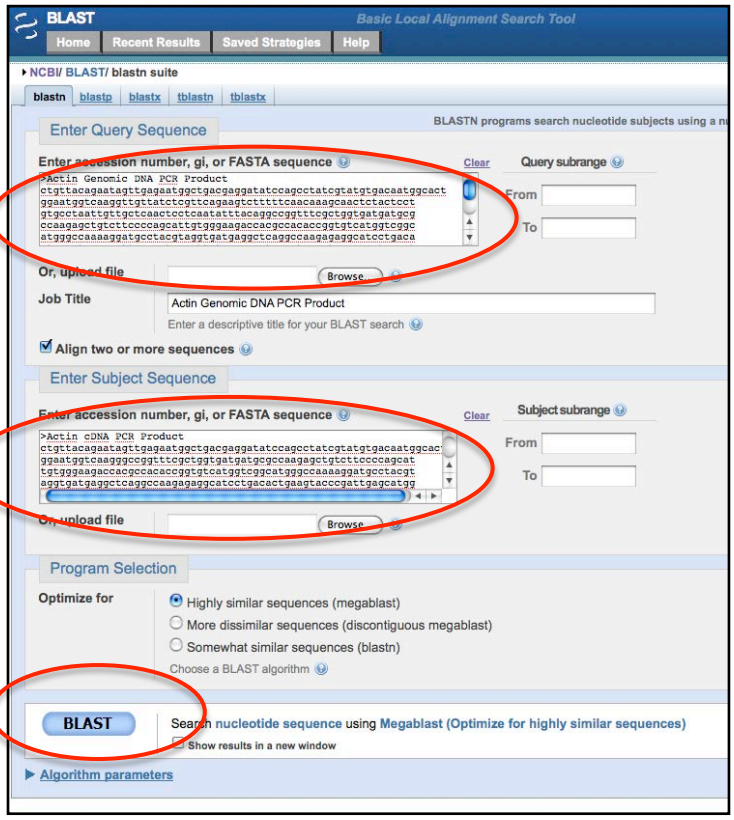


3. A) Enter the 'Actin Genomic DNA sequence' in the Query (top) textbox and B) the Subject (bottom) 'Actin cDNA sequence' in the bottom text box. C) Click 'Blast.'

A.

B.

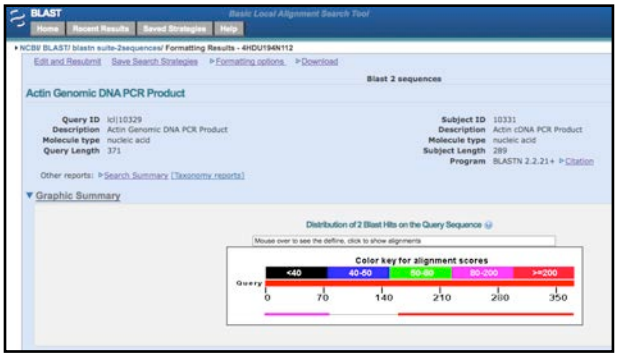
C.



4. The results.

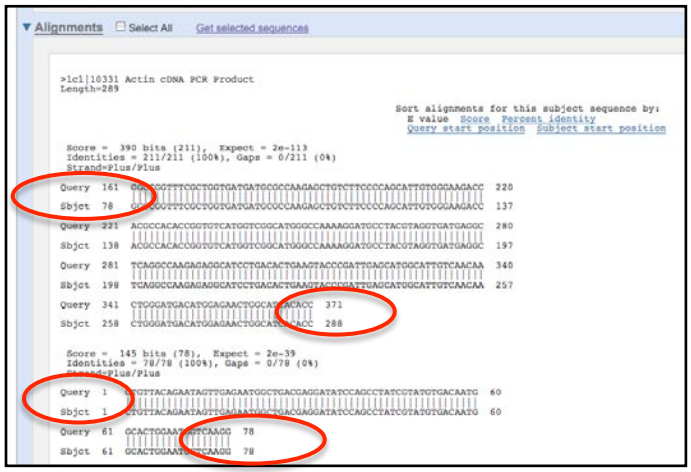
The results are presented diagrammatically in the top half of the page with the query shown as a red, thick rectangle. Any similarity between the query and the subject is shown as thin rectangles below the query. The color of the rectangle indicates the hit score; the higher the score the better the hit.

In this case there are two hits between the query (genomic DNA) and the subject (cDNA). Why are there two hits and what does the gap represent?



At the bottom of the page are alignments of the two hits nucleotide by nucleotide. When bases at the same position are the same a vertical line is placed between them.

Alignments are ordered by score from highest to lowest. Notice that the second alignment starts with 1 in both query (genomic DNA) and subject (cDNA) while the first alignment starts at nucleotide 161 in the query and 78 in the subject.



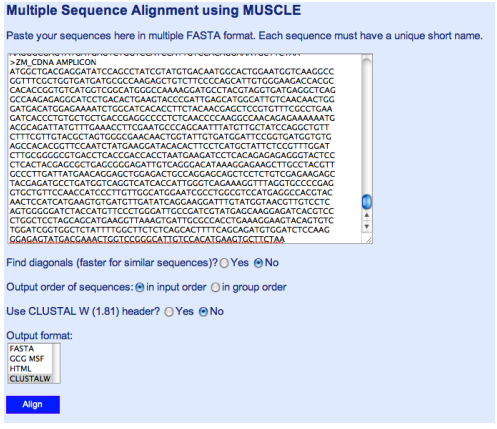
5. Draw the gene structure for the Actin sequence. Indicate the start and stop locations of the exons.

Part 4 -- Compare all of the DNA sequences to each other using multiple sequence alignment.

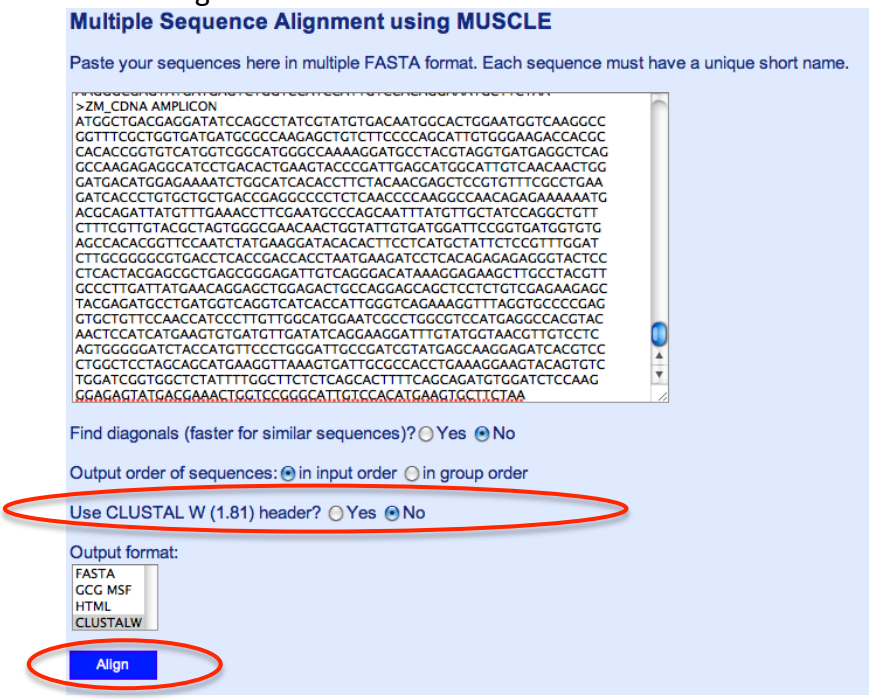
Although Blast2Sequences works well when comparing two sequences, it is often necessary to compare many sequences. This is done with a different tool called a Multiple Sequence Alignment (MSA). There are several programs that will do a MSA and the one we will use is MUSCLE. After the alignment is generated the results are viewed with a program called Jalview (Java alignment viewer).

1. Click MSA: Muscle Bookmark (the complete URL is on the course website).

2. In a separate browser tab open the class website and go to the Data page. Copy and paste the actin sequences into the text field on the MUSCLE page.



3. Click the “in input order” radio button for “Output order of sequences ”. Click Align.



4. On the results page click the “Start Jalview” button to open the alignment. We will discuss how to interpret the alignment in class.

Extension Questions

1. Why is there a large "gap" in the cDNA sequence? Is this an actual gap in the sequence? Draw a diagram to show the relationship between the cDNA and genomic DNA.
2. What would be the impact, if any, of the sequence polymorphisms in the introns (seen in the MSA) on *Actin 1* expression?
3. Similarly, what would be the impact, if any, of sequence polymorphism in the exons of the actin genes of the different species?
4. Translate the first 10 amino acids using the codon table found here for one cDNA: http://www.neb.com/nebecomm/tech_reference/general_data/genetic_code.asp.
5. The class website has the protein sequences (minus the first 10 amino acids) for Actin in single letter amino acid code. Add the first 10 amino acids (question 4) to the actin protein sequence and use it as the query for a protein blast (blastp). What is the accession number of the top hit?
6. Use MUSLCE to do an alignment for all of the protein sequences found on the webpage. Comment on the observed similarities and differences. Refer to this webpage to decode the single letter amino acid code if necessary: http://www.neb.com/nebecomm/tech_reference/general_data/genetic_code.asp.

Learning Goals

Students will be able to:

1. Use NCBI Blast to query a DNA database.
2. Describe the results of a BLAST query.
3. Use BLAST and related tools to annotate an intron.
4. Discuss why the cDNA sequence will show the locations of introns.
5. Describe why there are differences in sequence among different species.
6. Use multiple sequence alignment tools to analyze several sequences at once.

6: Genome Variation

Background

The DNA sequence of genomes is dynamic. Sequence variation was apparent when the actin genes of *Arabidopsis*, maize and rice were compared. Today we will investigate sequence differences among individuals of the same species. If two individuals have sequence differences at the same place in the genome it is called a polymorphism (poly=many, morph=form).

A difference of one nucleotide is called a single nucleotide polymorphism (SNP, pronounced "snip"). The mutation that results in sickle cell anemia is due to a SNP: the normal allele has a glutamic acid codon GAG whereas the mutant allele has a valine GTG codon at exactly the same position in the protein. This is an example of a nonsynonymous mutation and, in this case, it is detrimental to protein function. In contrast, many SNPs do not change the protein sequence (like TTT and TTC which both code for phenylalanine). These are called synonymous mutations. Many synonymous mutations occur at the third position of a codon. Look at a table of the genetic code and figure out why this is the case. Where else could SNPs occur without affecting the amino acid sequence?

Another type of polymorphism is the insertion/deletion (indel) and occurs when one or more nucleotide of DNA is present in one individual and absent in another. The indel highlighted in yellow could be an insertion of three bases in Individual A or as a deletion in Individual B. SNPs and indels are the results of errors during the DNA replication process.

```
Individual A: GGCCTACTCGGTACT
Individual B: TCCG---TGAGCCATGA
```

A third type of variation, which is the focus of the Wessler laboratory, is due to the insertion of mobile pieces of DNA called transposable elements (TEs). TEs have a characteristic structure shown in Figure 1. The autonomous element contains the gene encoding transposase enzyme (TPase) necessary for movement of both the autonomous and nonautonomous elements. The nonautonomous element lacks a functional TPase gene and depends on the autonomous element for movement.

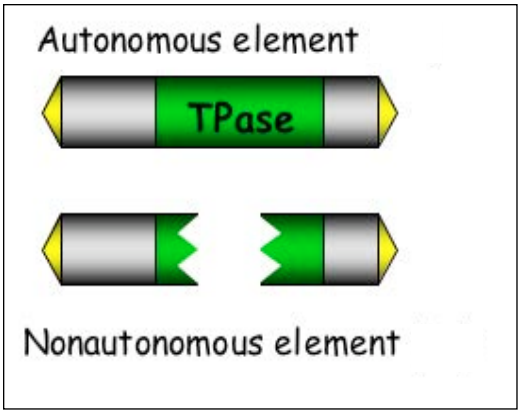


Figure 1: Molecular structure of autonomous and nonautonomous elements.

The yellow arrows at the element ends represent the terminal inverted repeats (TIRs) (Figure 1 above and 2 below). Sequence 4 (below) is the inverse of sequence 1 and likewise 3 is the inverse of 2. TIRs are the binding sites of transposase and it is the action of transposase that results in insertion polymorphism. When a TE moves, polymorphism is generated at the site that it left and at the new insertion site.

Figure 2. The DNA sequences at the ends of a TE.

One of the big surprises of the genomics era is that TEs account for over 50% of the genomic DNA of many species. Most TEs in a genome are inactive meaning that they can no longer move. Active TEs have often evolved mechanisms to avoid harming the organism. They insert between genes, into other TEs or into introns. What do you think happens to an organism when a TE inserts into an exon of a gene that encodes an important protein?

Activity

This activity will investigate TE insertion polymorphism in the maize (corn) genome by comparing the genome of different strains. Maize is a good species for TE research because it has lots of active TEs that contribute to genome variation. Collections of maize strains are available that capture this variation. Different strains (isolated by breeders and seed companies) are adapted to the diverse climates where maize grows - from cool northern states to hot southern states; from high plateaus to prairies at sea level. To study TE insertion polymorphism we will locate TE insertions in the reference B73 strain of maize. PCR will be used to isolate and analyze the DNA surrounding the TE insertions from DNA extracted from over 20 maize strains to determine if the TE is present in all strains.

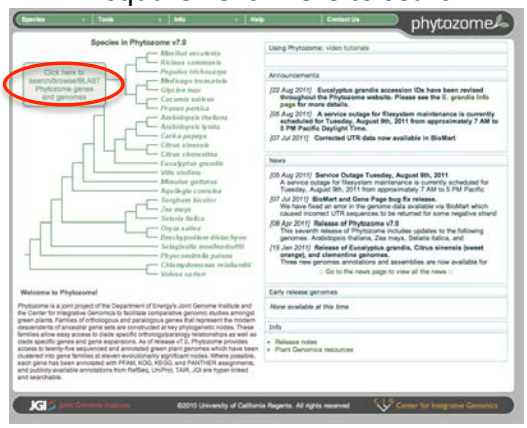
Part 1 – Finding TE insertions in B73

The sequence of a transposable element will be used as a BLAST query in the B73 maize genome browser. From the hundreds of hits TE insertions in introns will be chosen using the following criteria:

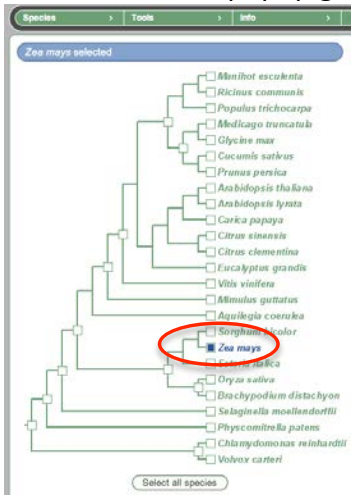
- ✓ Only 1 TE insertion in the intron.
- ✓ Intron less than 1 kb.
- ✓ EST support for the exons flanking the intron.

The Maize Genome Browser at the Phytozome website will be used to annotate TE insertions. In lab, primers to the flanking exons will be used to amplify the intron region from B73 and many other strains of maize.

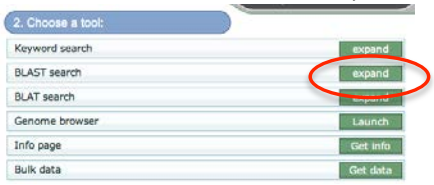
1. Open a web browser to Phytozome (www.phytozome.org) and click on the gray square "Click here to search…".



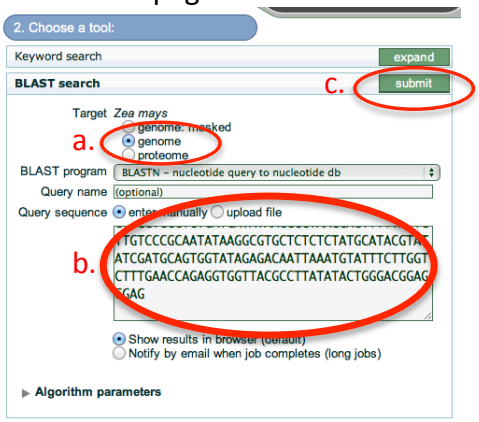
2. Close the popup gray box and select *Zea mays* on the tree.



3. In the Choose a tool, click "Expand" next to the BLAST Search option.

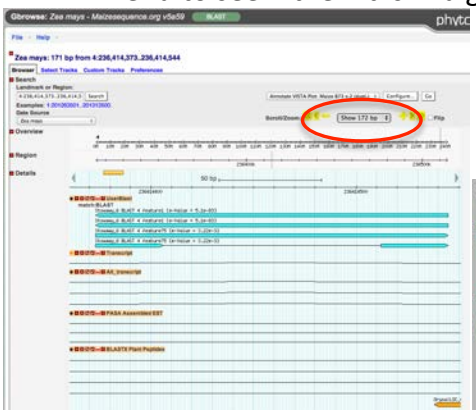


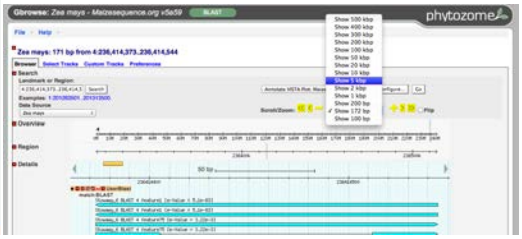
4. In the box, a) select “Genome” as the target, b) copy and paste a TE query in the Query text box and c) click “Submit”. The queries can be found on the course web page.



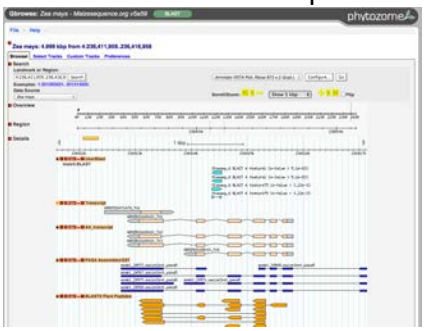
5. In the results window click “Switch Views”.

6. In the new results window, click on a good Blast hit. A browser window will open zoomed in on the blast result. Zoom out to 5 kb by clicking on the “Show n bp” menu to see if the hit is in a gene.





7. Using the information in the browser, does the hit meet the criteria stated above? If so keep the screen and be prepared to discuss the genomic region.



Now that intronic TE insertions have been found we need to design a PCR experiment to amplify the region out of genomic DNA samples. We will do a think-pair-share to design the experiment.

Part 2 – PCR Set-up

The PCR protocol for this activity is similar to the one used to amplify. There will be DNA samples from 28 different strains of maize and each student will choose three. Design a PCR experiment using the one done last Thursday as a model. What other reactions should be included? What reagents are needed? Write the protocol in the lab notebook.

Part 3 – Gel Electrophoresis.

Follow the protocol from page 3.12 and pour one gel with two combs per bench. Gels will be stored at 4˚C until the next class. During the next class load DNA ladders in the first and last lanes and 15 µl of each sample.

Part 4 – DNA Sequence Analysis

After gels are run during the next class the DNA sequences of the PCR amplicons will be placed on the website. What questions can be addressed using this sequence? What bioinformatics tools will be needed to address these questions?

Extension Questions

1. Analyze the bands on the gel. What do you learn about the TE insertion?
2. Using the DNA sequence available on the course website and bioinformatics tools determine why some bands are larger and some are smaller.
3. Create a multiple sequence alignment that includes 5 different strains. Label the TE insertion and identify other polymorphisms. What consequences if any could the polymorphisms have on the gene? On the Strain?

Learning Goals

Students will:

1. Be able to describe three types of genome variation-SNPs, indels and TEs
2. Be able to describe the source of the variation.
3. Be able to design and set up a PCR experiment to amplify across and intron.
4. Be able to identify variation though comparative DNA using BLAST and multiple sequence alignment programs.

7: Bacterial Transformation

Last week a plasmid was digested with the restriction enzyme *PvuI* and the results were visualized on a gel. Bacteria were used to generate the large quantities of plasmid needed for that experiment. Plasmids are introduced into bacteria through a process called transformation. In the experiment that follows bacteria will be transformed with a plasmid that contains the gene for Green Fluorescence Protein (GFP). A successful transformation will result in bacteria that glow green under blue of ultraviolet light. In the next experiment, GFP will be used to detect TE activity.

Background

GFP

GFP comes from the jellyfish *Aequorea victoria* and fluoresces green when exposed to blue light. Researchers have found GFP extremely useful for an important reason: visualizing the presence of the gene doesn't require sacrificing the tissue to be studied. That is, GFP can be visualized in living organisms by using fluorescent-imaging microscopy. The importance of the GFP reporter gene to modern science was evident when the 3 scientists responsible for its discovery and adaptation to the lab were awarded the 2008 Nobel Prize in Chemistry. You can learn more about this at this site: http://nobelprize.org/nobel_prizes/chemistry/laureates/2008/press.html.

GFP was cloned from a jellyfish. Usually when a gene is cloned for expression in a different organism the introns cannot be present. In order to clone an intronless version of the GFP gene did the scientists clone the genomic DNA or cDNA? Remember that introns are not found in bacteria. Another consideration is how will the gene be turned on? The promoter of a gene is the switch that a cell uses to turn a gene on or off. Promoters are very specific to individual species. The scientists had to add a bacterial promoter so that the gene will be expressed in bacteria. In the next experiment the GFP gene will be expressed in a plant called Arabidopsis and a different promoter has to be used for expression in plants.

Transformation

Transformation is the process of introducing DNA into a cell. In this lab *E. coli* will be transformed with a plasmid. Transformation is very inefficient and only a tiny percent of cells will contain the plasmid (less than 1 per million). To recover only the cells that take up a plasmid, called transformants, antibiotic selection is used. The plasmid contains a gene for resistance to antibiotics in the growth medium. Transformants will divide while all other cells will die. A common antibiotic is ampicillin and the plasmid shown in Figure 1 has the ampicillin resistance gene (AMP).

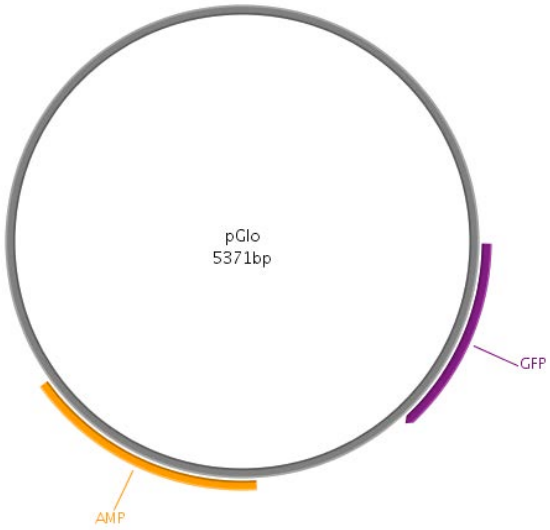


Figure 1. Plasmid map of pGlo.

Plasmids are mixed with bacterial cells that have been treated with calcium chloride to make them competent to take up DNA. The mixture is incubated at high temperature for a brief period and during this heat shock the plasmid is taken up by the cells. After the cells recover they are spread (plated) on agar petri plates that has nutrients, ampicillin, and agar (impure agarose). Transformants resistant to the ampicillin will divide many times during to form a colony on the surface of the plate.

Activity

In this activity *E. coli* will be transformed with pGLO to see the result of introducing a foreign gene into an organism. A group of two will set up one transformation reaction containing the plasmid and one that does not (the negative control). Each transformation will be spread on two petri plates: one with ampicillin and one without. What will happen on each plate? If pGLO were expressed in *E. coli* what would the result be?

Part 1 – Transformation

Materials:

Competent *E. coli* cells (purple cap tube)
pGLO (1ng/µl)
2 LB plates minus ampicillin – rich media for bacteria (1 black stripe on the edge)
2 LB plates plus ampicillin (1 black stripe and 1 red stripe)
Liquid LB Broth
42˚C bath
37˚C bath
Sterile beads

Cautions:

- Lab strains of *E. coli* do not cause disease. Regardless they are treated as hazardous material so follow instructions on discarding wastes.

- Keep the competent cells on ice.
- Do not centrifuge competent cells. The cells will settle to the bottom of the tube. Gently tap the tube with your finger to resuspend.

Method:

1. Label one tube "+DNA" and one tube "- DNA".
2. Add 20 µl of competent cells to each tube.
3. Add 5 µl of plasmid to the "+DNA" tube.
4. Mix the tubes by gently tapping.
5. Place on ice for 10 minutes.
6. After 10 minutes on ice, place tubes in the 42˚C water bath for 30 secs.
7. Add 500 µl (0.5 ml) of LB broth to each tube. Tap gently. Place the tubes at 37˚ for 20 minutes.
8. Label the plates as follows.
    a. Label the bottom of one of each type of plate with the date, your initials and "+ DNA".
    b. Pipette 100 µl of the "+DNA" transformation onto each plate.
    c. Label the bottom of one of each type of plate with the date, your initials and "Minus DNA".
    d. Pipette 100 µl of the "-DNA" transformation onto each plate.
9. Pour 3-4 glass beads onto the plates. Shake gently to spread the liquid and cells evenly across the surface of the media. Empty the beads into the Used Bead Beaker.
10. Tape all plates, 7 in total, in a stack using masking tape. Place the plates lid down in the incubator. The plates will incubate overnight at 37˚C degrees (12-16 hours) and then be placed at 4˚C to stop growth. The results will be analyzed during the next lab period.

Part 2—Analysis (completed during next class)

The plates from the previous lab period will be analyzed to determine if the transformation was successful. The transformation efficiency will be calculated and the results of the class will be compared.

Transformation efficiency

To analyze the transformation experiment the transformation efficiency, which is a measure of the number of transformants per input plasmid, will be calculated.

1. Remove the plates from the incubator. Count the number of colonies and on each transformation plate. Record the colony counts in your notebook in a table similar to the one below. If a plate has too many colonies to count, TMTC is recorded.
2. View each of the plates using UV light. Turn the plate lid down and remove the lid. Hold the agar surface above the UV light. Record the results in the table.

Colonies should have formed on both LB-Amp plates but not on both LB+Amp plates. Why? Why do some colonies "glow" green?

Plate	Transformation minus plasmid	Transformation plus plasmid	GFP Fluorescence
LB-Amp			
LB+Amp			

3. Calculate the transformation efficiency expressed as transformants/ng.

$$\text{Transformation Efficiency} = \frac{\#\text{ colonies}}{5\text{ ng of plasmid}} \times \frac{500\ \mu\text{l transformation}}{100\ \mu\text{l plated}}$$

ng of plasmid = 5 µl of plasmid used X 1 ng/µl = 5 ng.

Viable counts are expressed in colony forming units (cfu)/ng.

4. Good transformation efficiency is around $1x10^6$ cfu/ng. How does your efficiency compare? How does this compare with your bench mates' results? What are some possible reasons for this variation?

Learning Goals

Students will be able to:

1. Describe bacterial transformation and its uses.
2. Discuss the need for antibiotic selection.
3. Discuss how each plate contributed to the interpretation of the experiment:
    a. +DNA LB+AMP
    b. +DNA LB-AMP
    c. -DNA LB+AMP
    d. -DNA LB-AMP
4. Explain why the bacteria are "competent."
5. Define GFP and why it is such a useful protein.
6. Calculate simple transformation efficiency and explain what the number means.

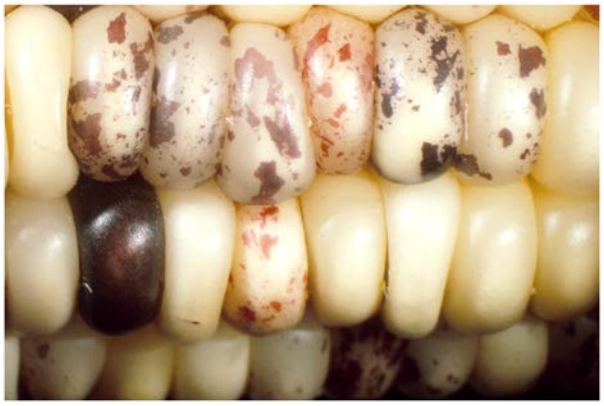
8: Discovery of transposable elements and impact on genomes

Background

The Discovery of Transposable elements

It all began more than 60 years ago with a far-sighted scientist named Barbara McClintock who was studying the kernels of what we informally call "Indian corn"--those ears with richly colored kernels that we associate with Thanksgiving and that scientists call maize.

Maize and corn are the same species. Maize is a grass that is taxonomically related to other familiar cereal grasses like barley, rice, wheat and sorghum. By the 1920s, researchers had found that maize kernels were ideal for genetic analysis because heritable traits such as kernel color and shape are so easy to visualize. The results of early studies on maize led to an understanding of chromosome behavior during meiosis and mitosis. As a result, by McClintock's time, maize was one of two model genetic organisms - the other being *Drosophila melanogaster* (the fruit fly).



As early as the 1920's it was known that maize had 10 chromosomes [this is the haploid number (n) - maize, is a diploid (2n) with 2 sets of 10 chromosomes]. In addition to being a superb geneticist, McClintock was one of the best cytologists in the world and her specialty was looking at whole chromosomes. Maize was ideal for this analysis because it has a large genome (recall - 2500 Mb) and its chromosomes were easily visualized using a light microscope. The first thing of note that McClintock did as a scientist was to distinguish each of the 10 maize chromosomes of maize. This was the first time anyone was able to demonstrate that the chromosomes (of any organism) were distinct and recognizable as individuals.

In the course of her studies of various maize strains, she noticed the phenotype shown below in **Figure 1a.** This phenotype is characteristic of chromosome breakage. While chromosome breakage is commonly observed in maize, it had not previously been observed at a single site (locus) in one chromosome. In one particular strain chromosome 9 broke frequently and at one specific place or *locus*. After considerable study, she found that the breakage was caused by the presence in the genome of two genetic factors. One she called *Ds* (for *Dissociation* -it caused the chromosome to "dissociate"), and it was located at the site of the break. But another genetic factor was needed to activate the breakage. McClintock called this one *Ac* (for *Activator*). Because she could not genetically map the position of Ac in the genome she hypothesized that it was capable of moving around (transposing). For example, Ac could move from chromosome 1 to chromosome 3.

As she followed the descendants of this strain, she identified rare kernels with fascinating phenotypes. One such phenotype was a colorless kernel containing pigmented spots. This is summarized in **Figure 1b**.

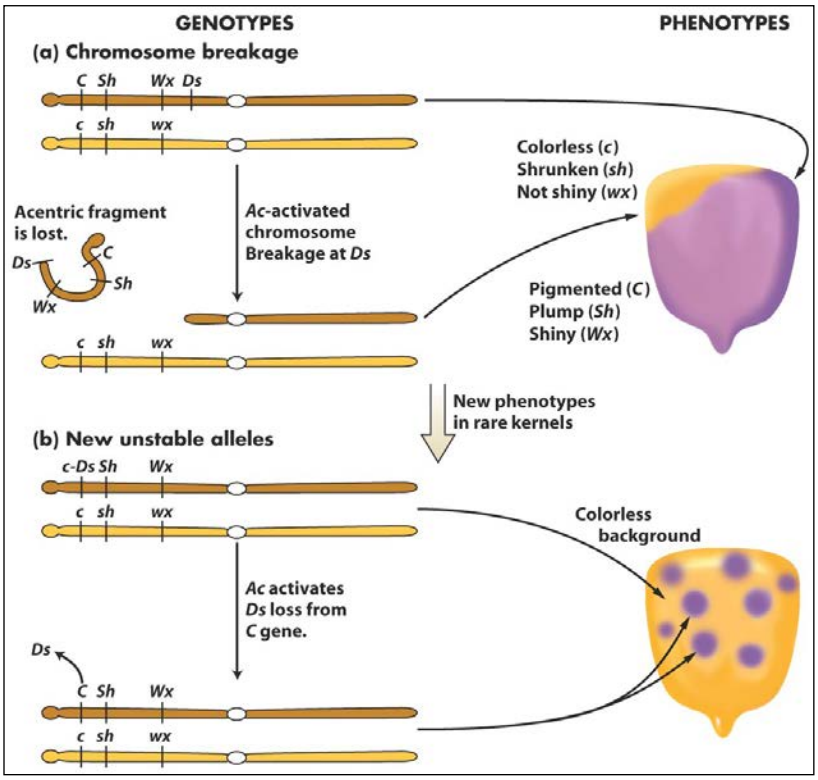


Figure 1. New phenotypes in corn are produced through the movement of the *Ds* transposable element on chromosome 9. (a) A chromosome fragment is lost through breakage at the *Ds* locus. Recessive alleles on the homologous chromosome are expressed, producing the colorless sector in the kernel. (b) Insertion of *Ds* in the *C* gene (top) creates colorless corn kernel cells. Excision of *Ds* from the *C* gene through the action of *Ac* in cells and their mitotic descendants allows color to be expressed again, producing the spotted phenotype.

What she soon knew conclusively was this: *The TEs that she was studying were inserting into the normal genes of maize and were causing mutations. What she had discovered was a different type of mutation - one that was caused by a transposable element and one that was reversible. This contrasts with other mutations that you have learned*

about like base pair changes and insertions/deletions that are essentially irreversible. Her logic is summarized in the figure below. Furthermore, she provided the following explanation for what was going on with the spotted kernels:

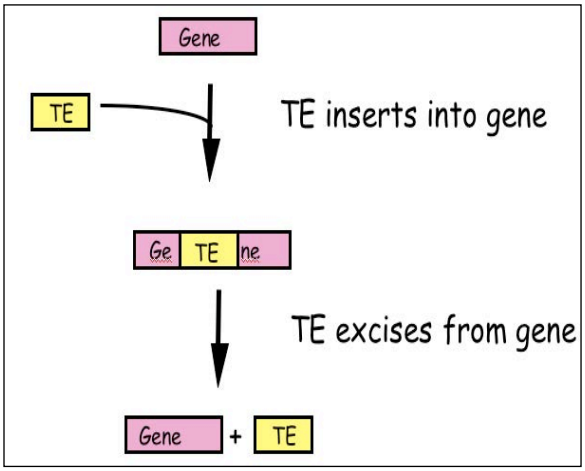
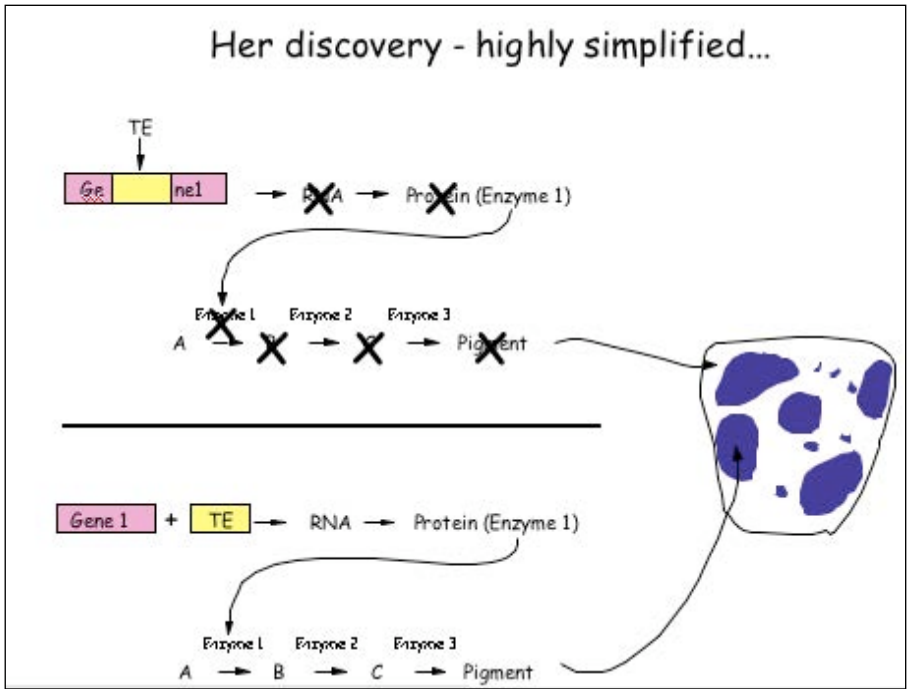


Figure 2: McClintock hypothesized that TEs were a source of "reversible" mutation. Their ability to transpose allowed them to excise from mutant genes leading to phenotypic reversion.



What DNA transposable elements look like to the geneticist (Ac, Ds)

Barbara McClintock discovered the TEs Ac and Ds when she figured out that they were responsible for the spotted kernel phenotypes. She was a geneticist - and their main experimental tool is the genetic cross.

Here are some of the properties of Ac/Ds that McClintock figured out through observation of kernel phenotypes and by performing carefully designed crosses:
(1) Ac and Ds could insert into a variety of genes - e.g. those involved in pigment production, starch biosynthesis, and early embryo development, to name but a few.
(2) Ac and Ds were normal residents of the corn genome - they were not, for example, introduced into the genome by a virus.
(3) Ds could not move without Ac in the genome, whereas Ac could move itself or Ds. Thus, Ac was called an autonomous element while Ds was called a non-autonomous element.

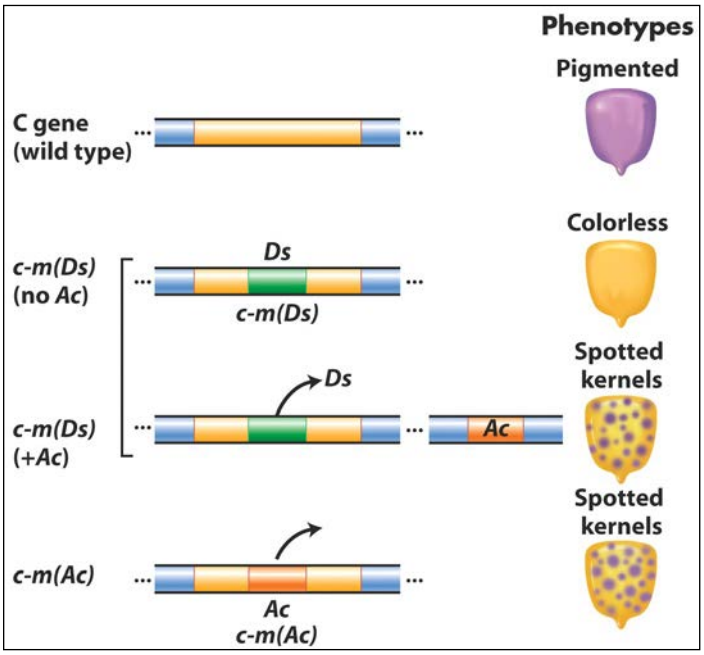


Figure 3 Summary of the main effects of transposable elements in corn. *Ac* and *Ds* are used as examples, acting on the *C* gene controlling pigment.
In maize (but not many other organisms), normal alleles are capitalized and mutant alleles are written in lower case. In addition, McClintock designated alleles caused by the insertion of a TE as "mutable", m for short [e.g. c-m(Ds) or c-m(Ac)].

TEs are in all organisms: After her initial results were reported in the late 1940's, the scientific community thought that TEs were oddities and possibly restricted to maize and perhaps to a few other domesticated plant species. However, this proved not to be the case as in subsequent years TEs were discovered in the genomes of virtually all organisms from bacteria to plants to human. It is for this reason that McClintock was awarded the Nobel Prize in Medicine or Physiology in 1983, almost 40 years after her discovery. We will return to Barbara McClintock often during this course.

What transposable elements look like to the molecular biologist (Ac,Ds):

With the advent of molecular cloning biologists were able to isolate and sequence gene-sized fragments of DNA from the genomes of plants and animals. Figure 4 is a simplified figure showing what Ac and Ds look like at the DNA level.

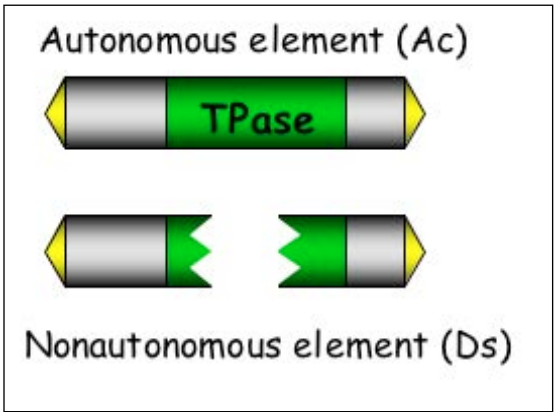


Figure 4: Molecular structure of Ac and Ds.
Ac: Tpase is the gene encoding the transposase enzyme which is necessary for movement of both Ac and Ds.
Ds: Ds requires Ac for movement because it is a defective version of Ac where the Tpase gene has been deleted.
Yellow arrows at the ends are the terminal inverted repeats - this site where transposase binds and cuts the element out of the surrounding genomic DNA.

Ac contains a single gene - that encodes the transposase. Figure 4 shows how this protein catalyzes the movement of Ac and Ds.

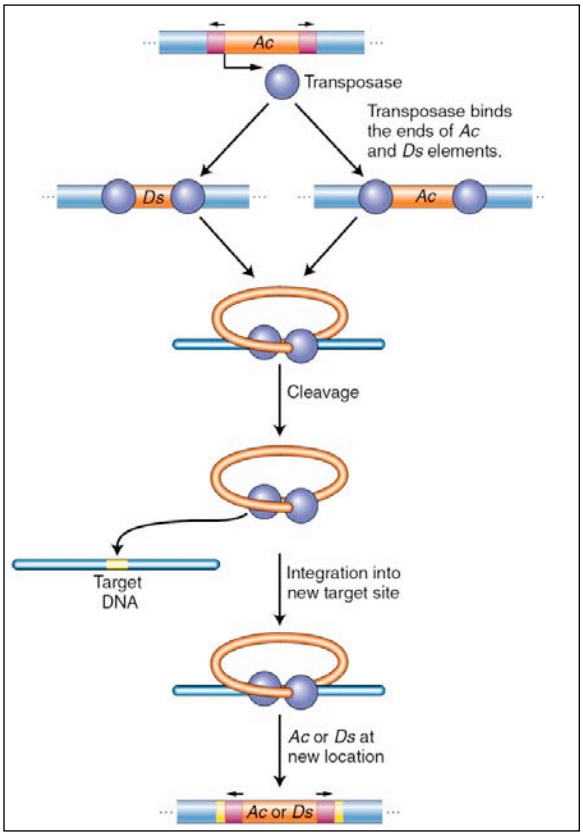


Figure 5 Activator transposase catalyzes excision and integration. The maize *Ac* element encodes a transposase that binds its own ends or those of a *Ds* element, excising the element, cleaving the target site, and allowing the element to insert elsewhere in the genome.

Like many other proteins, the transposase protein can multi-task. First, it is a DNA binding protein that is able to bind specifically to the ends of the Ac element. The protein also binds to the ends of Ds as it is identical to the Ac ends. Such "sequence-specific binding" is mediated by precise contacts between the amino acids of part of the

transposase (called the binding domain) and the precise nucleotide sequences at the Ac (and Ds) ends. Second, it is an enzyme. Once bound, the two transposase molecules form a dimer (via protein-protein interactions) and another region of the transposase (called the catalytic domain) cuts the element out of the surrounding genomic DNA. The two transposase proteins bound to the TE then cuts the chromosome at another site (the target) in the host genome and the TE inserts.

Finally, for now at least, there is one other feature of TEs that needs to be introduced. This is the target site duplication (TSD) that is created during insertion of virtually all TEs. How it is generated is shown below in Figure 6.

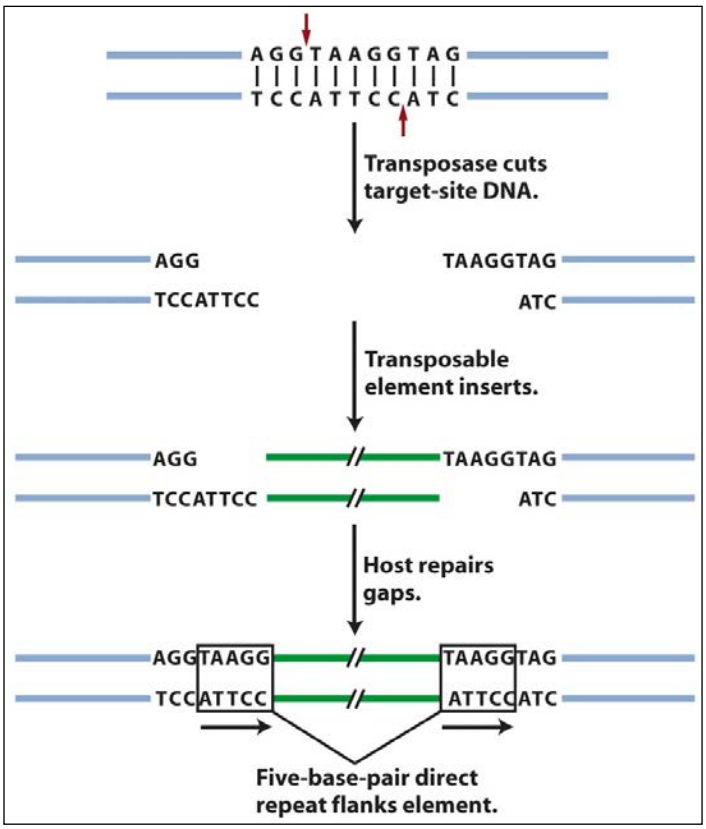


Figure 6: An inserted element is flanked by a short repeat. A short sequence of DNA is duplicated at the transposon insertion site. The recipient DNA is cleaved at staggered sites (a 5-bp staggered cut is shown), leading to the production of two copies of the five-base-pair sequence flanking the inserted element. This is called a target site duplication (TSD).

What transposable elements look like to the bioinformaticist

As you know, Human Genome Project ushered in the genomics era which is characterized by the availability of increasing amounts of genomic sequence from a variety of plant and animal species [animals - including human, fruit fly (*Drosophila*), earthworm, dog, mouse, rat, chimp; plants - including *Arabidopsis thaliana*, rice, maize (corn) cottonwood (a tree)]. For now, it is sufficient to know that TEs make up the vast majority of the DNA sequence databases and recognizing TEs in genomic sequence is usually the first step in the modern analysis of TEs.

The elements you will be analyzing in experiment 2 are the Ping and mPing (for miniature Ping) elements - which were first identified by computational analysis of the rice genomic sequence (see page 14 below for how this was done). The figure below shows that Ping is the larger coding element like Ac. Unlike Ac Ping contains 2 genes (ORF1 + Tpase).

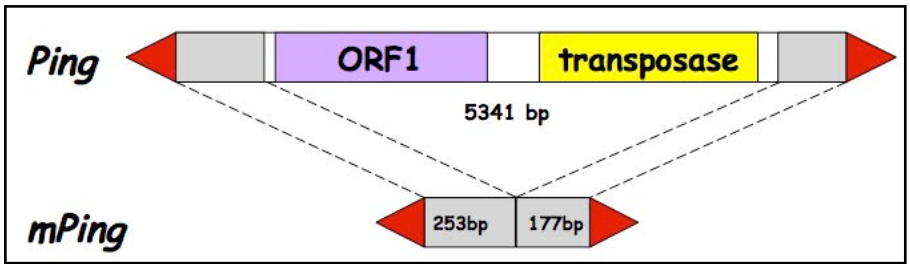


Figure 7: Ping encodes two genes: the transpoase (TPase) and ORF1 (open reading frame) (function unknown at this time). The red arrows are the terminal inverted repeats (14bp). mPing is 253bp + 177bp long.

As you can see, like Ds which is derived from Ac by a large deletion, mPing is derived from Ping by a large deletion. Our hypothesis is that Ping encodes a protein that binds to the ends of mPing and catalyzes its transposition.

So, this should be a snap, right? Let's just study an mPing element that is inserted into a rice gene and monitor its movement in the same way as McClintock did with spotted kernels (rice grains in this case). Well, unfortunately, we can't do that - because - like most TEs in the genome, mPing is not inserted into a rice gene - but rather - it is inserted between rice genes!

Digression - how can organisms survive with so many TEs? Where are TEs located in the genome?

At this point we need to go up to 30,000 feet in order to understand a larger concept: the connections between TEs, evolution and natural selection. In short, the distribution of TEs in most genomes is due to the action of natural selection —the foundation for all modern biology.

These are two kinds of selection that will need to understand:

*Negative selection
*Positive selection

It is important first to know something about natural selection itself. Here's a slightly edited version of its definition in Wikipedia: ". . . In the context of evolution, certain traits or alleles of a species may be subject to selection. Under selection, individuals with advantageous or `adaptive' traits tend to be more successful than their peers reproductively—meaning they contribute more offspring to the succeeding generation

than others do. When these traits have a genetic basis, selection can increase the prevalence of those traits, because offspring will inherit those traits from their parents."

Negative selection is the elimination of a deleterious trait from the population by natural selection. It is also called "purifying selection." In the context of TEs, insertions into genes are deleterious and, as such, are eliminated from the population. The word *elimination* in this case means that an individual with the TE insertion will either not be viable or will not be able to reproduce.

Positive selection occurs when a certain allele has a greater fitness than others, resulting in an increase in frequency of that allele. This process can continue until the entire population shares the fitter phenotype, then the allele is said to be "fixed" in the population. An example of this is a TE insertion that affects a gene in some positive way that makes the organism more adaptive in a particular environment. Such a change would be incredibly rare, though, because there are thousands of genes in a genome where a TE can insert and most insertions in a gene are harmful. Think of a population where the climate has changed and become much drier. Increasing the expression of one particular gene in the genome might increase drought tolerance and allow an organism with such a "mutation" to survive. For a TE to insert into just that gene, in the right place so that it increases the expression of the gene, is extremely unlikely. However, when we think of probabilities it is important to keep in mind that there are lots of TEs in a genome, that there can be many individuals in a population and finally - evolution occurs over very long time periods - that's why it's called evolution, not revolution!

So what does all this have to do with transposable elements?

Transposable elements can insert into all regions of the genome - in genes and between genes. However, if we look at an entire genome, we usually find most of the TEs between genes and in noncoding regions of a gene (e.g like introns). This is because insertions into genes have fallen victim to negative selection. In contrast TEs between genes remain for generations, hundreds of generations, because they are not harmful. Rather, they are usually neutral and may even be beneficial.

Most of the TEs in the genome are INACTIVE

This leads to a second point you need to remember: The vast majority of transposable elements in a genome are inactive (they can't move anymore). TEs can be inactivated in one of at least two ways—through mutation or through what is called "epigenetic silencing."

The mutation part is easier to understand. All DNA is susceptible to mutation - usually base pair changes or deletions. This happens (very rarely) when there is an error during

replication and the wrong base is inserted - for example a G is put opposite T (instead of an A). This change could alter an amino acid in a protein. Mutation can also happen by "free radicals" - chemicals that accumulate in our cells and can damage our DNA. Finally, mutagens in our environment - like UV light or cigarette smoke - can damage our DNA.

There are dramatically different consequences of a mutation in a gene vs. in a TE. Stately simply, mutation in a gene is usually eliminated from the population by natural selection (negative selection), whereas mutation in a TE will be neutral and, as such, will persist in the population. Thus, unlike genes, TEs will accumulate mutations and become inactive. (NOTE - TEs AND GENES SUSTAIN MUTATIONS AT THE SAME FREQUENCY. HOWEVER, IF YOU STUDY AN ORGANISMS GENOME, MOST OF THE GENES WILL BE ACTIVE WHILE MOST OF THE TEs WILL HAVE SUSTAINED INACTIVATING MUTATIONS)

Epigenetic regulation of TEs

Eukaryotic chromosomes exist in the nucleus as chromatin - an equal mixture of DNA and protein. The basic unit of chromatin is the nucleosome - about 180bp of DNA wound around a core of histone protein (shown as a ball in figure 8 below). Chromatin can be loosely organized (open chromatin, called euchromatin) or highly condensed (where nucleosomes are tightly packed, called heterochromatin) as seen in Figure 8. In most plant genomes such as maize, transposable elements are frequently clustered and associated with condensed chromatin. Genes in condensed chromatin cannot be expressed and are inactive. This is the fate of the vast majority of the TEs in a genome.

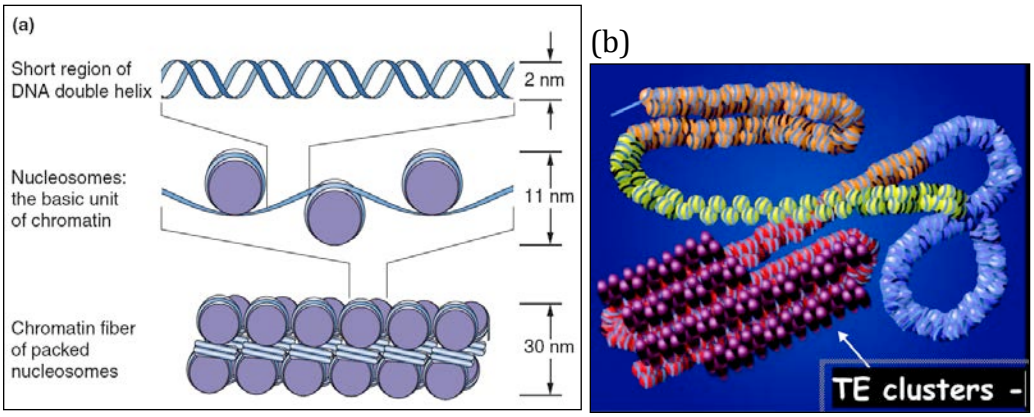


Figure 8. The nucleosome in decondensed and condensed chromatin.
(b) Chromatin structure varies along the length of a chromosome. The least-condensed chromatin (euchromatin) is shown in yellow, regions of intermediate condensation are inorange and blue, and heterochromatin coated with special proteins (purple) is in red.

Learning Goals

Student will be able to:

1. Define a transposable element and be able to draw the basic DNA structure.
2. Discuss the relationship of Ac and Ds elements.
3. Discuss how the organisms survive with TEs in their genome.

9: TE Superfamilies

It turns out that the rice genome can contain up to 1000 copies of mPing and from zero up to seven Ping elements (this depends on the rice strain). Ping and all of the mPing elements in the rice genome make up a TE family.

The genomes of plants and animals contain many different families of transposable elements. This concept is central to understanding what genomes are made of.

What is a TE family?

We have already been introduced to two TE families. One family (from maize) contains the Ac and Ds elements while the second family (from rice) contains Ping and mPing elements.

In functional terms, a TE family contains all the elements that can be mobilized by a particular transposase. A TE family usually contains autonomous elements (e.g. Ac, Ping) and nonautonomous elements (e.g. Ds, mPing) elements. When we analyze the DNA sequence of entire genomes we often find that a family contains several elements including one or more autonomous elements and many copies of nonautonomous elements (the maize genome has over 50 copies of Ds and, as mentioned above, some rice genomes have up to 1000 copies of mPing).

The transposase encoded by the Ac element can mobilize both Ac and Ds elements. If there is no Ac element in the genome, all of the Ds elements will be "stuck" where they are - they will not able to move elsewhere in the genome because there is no transposase to catalyze their movement. The same is true for Ping and mPing in rice – mPing will be stuck in place if Ping is not in the genome.

A very important feature of TE families is that each family is independent. In practical terms this means that the Ac transposase cannot mobilize Ping or mPing elements and, similarly, the Ping transposase cannot mobilize Ac or Ds elements. Or, as shown in **Figure 1**, the transposase from family A cannot move the elements in family B. The reason for this is simple. A transposase works by first binding to a specific DNA sequence near the ends of the element called the Terminal Inverted Repeat or TIR. The Ac transposase first binds to a specific sequence of nucleotides that is only near the ends of Ac and Ds elements while the Ping transposase binds to a specific sequence that is only near the ends of Ping and mPing elements. (Recall that in addition to catalyzing chemical reactions, proteins can also bind to DNA. Transposases are proteins that do both: bind to DNA and then catalyze the transposition reaction.)

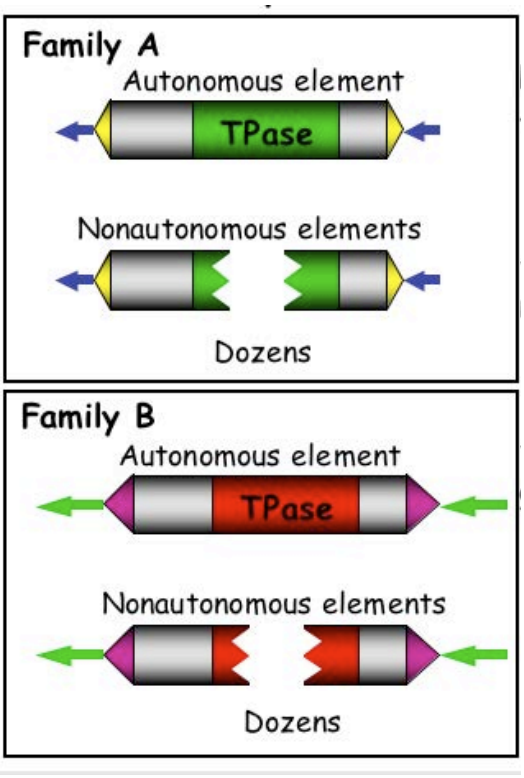


Figure 1: A TE family contains autonomous elements and all the nonautonomous elements in the genome that its transposase can move. Genomes have many TE families that are independent. This is because the transposase of Family A, for example, cannot bind to the ends of the elements from Family B and vice versa.

What is a TE superfamily?

After Barbara McClintock discovered Ac and Ds (in the 1940's) she then discovered a second TE family, which she called Spm (for Suppressor-mutator). The autonomous element in this family is called Spm and the nonautonomous element is called dSpm (for defective-Spm). Thus, Spm-dSpm is another TE family.

McClintock's discoveries resulted from genetic analyses of corn plants. After the discovery of TEs in maize, researchers working with other model organisms, including *Antirrhinum majus* (a.k.a. snapdragon) *Drosophila melanogaster* (a.k.a. the fruit fly) *and Caenorhabditis elegans* (a.k.a. the worm) also identified TEs through genetic studies. In the 1980's when it became possible to isolate specific genes, researchers isolated McClintock's Ac, Ds, Spm and dSpm elements and the elements from snapdragon (called Tam 1,2,3 etc), the fly (called P-elements, mariner elements and others) and the worm (called Tc1, 2 and 3 elements).

When the DNA sequences of these elements were determined and compared (by computer analysis), researchers were surprised to find that the transposases encoded by some of the elements from different species, even from different kingdoms (animal vs. plant), were similar. For example, the amino acid sequence of the transposase from the maize Ac element was similar to the amino acid sequences of the transposases of Tam3 from snapdragon and the P element from the fly, while the transposases of the mariner (fly) and Tc1 (worm) elements were similar.

These similar transposases were subsequently organized into superfamilies. Fortunately, after all of the sequencing of genomes and comparisons of TEs, there are now known to be fewer than 10 superfamilies of transposases. Some superfamily names and elements and some members include: hAT (includes Ac, Tam3, P elements), CACTA (includes Spm, Tam1), PIF/Harbinger, Mutator and Mariner. The distribution of some of the superfamilies across the tree of life is summarized in **Figure 2**.

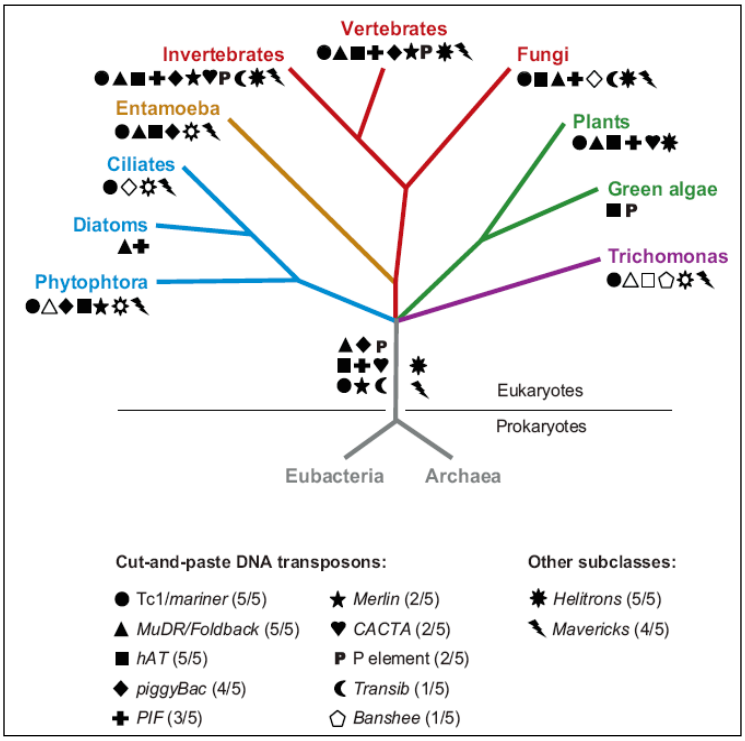


Figure 2. Distribution of the major groups of DNA transposons across the eukaryotic tree of life. The tree depicts 4 of the 5 "supergroups" of eukaryotes where DNA transposons have been detected. The occurrence of each TE superfamily is denoted by a different symbol. (*Feschotte ·Pritham* Annu. Rev. Genet. 2007.41:331-68).

How many families and superfamilies can an organism have in its genome?

In short, many. First, members of most superfamilies are present in all plant genomes including maize, rice and *Arabidopsis*, and are also present in most animal genomes (**Figure 2**). For example, the rice genome has Mariner, PIF/Harbinger (Ping), hAT (Ac/Ds), CACTA and Mutator elements. In addition, each superfamily usually contains many families in one genome.

Structural features shared by superfamily members:

Before you can study a TE superfamily, we need to look closely at the structural features of transposable elements in more detail because these features are usually shared within a superfamily.

Figure 3: Structural features of transposable elements that are shared by superfamily members. TSD = target site duplication, TIR = terminal inverted repeat, Tpase = transposase gene that is present in all autonomous elements, ORF1 – a second gene that is only encoded by members of the Ping/PIF/Harbinger superfamily.

The terminal inverted repeat (TIR)

In the figure above, the sequence of the blue triangles is shown. Look closely and you will see that the sequence of the right TIR is the reverse-complement of the left TIR. These sequences help define a TE family because they are bound by TPase produced by a family member. While all members of a TE family have identical or near identical TIRs, the TIRs of superfamily members (elements from different species) are usually similar but not identical. In addition, the length of the TIR can vary. For example, the length of the Ping TIR is 15bp while the length of the Ac TIR is 11bp.

Target site duplication

The target site duplication (TSD) is a direct repeat sequence that flanks the TIR. It is generated during the insertion of virtually all TEs into genomic DNA. How it is formed is shown below.

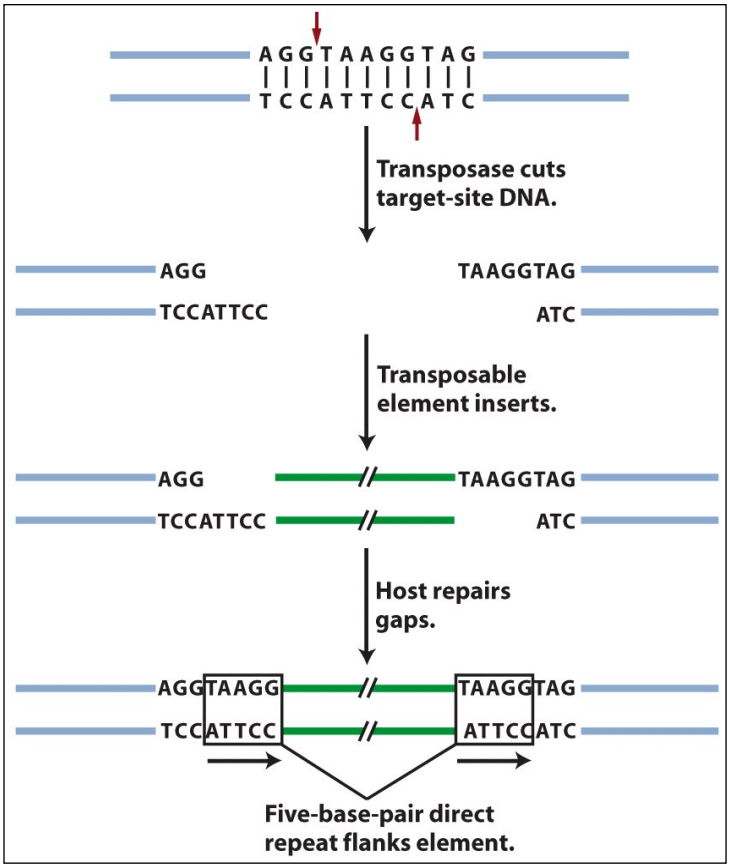


Figure 4: An inserted element is flanked by a short repeat. A short sequence of DNA is duplicated at the transposon insertion site. The recipient DNA is cleaved at staggered sites (a 5-bp staggered cut is shown), leading to the production of two copies of the five-base-pair sequence flanking the inserted element.

The length of the TSD, but usually not the sequence, is a common feature of a TE superfamily. For example, members of the hAT superfamily (Ac, Tam etc) all have an 8bp TSD, while members of the Mutator family have a 9bp TSD. Ping has a 3 bp TSD that is almost always TAA or TTA.

The Transposase (Tpase) gene

The sequence of the TPase is also characteristic of a superfamily. In fact, the tpase sequence (or part of it) is THE feature used to define superfamilies. Later in the course you will use the sequence of the TPase and the sequence of the TIR to find Ping family members in rice and in other plant species.

Using computational analysis to find all elements related to Ping in rice and other genomes

You can identify all TEs related to Ping in rice because the whole genome of rice has been sequenced. To do this one performs a Blast search using either the DNA sequence of the whole element or the protein sequence of the TPase. Using the whole element as query would retrieve only very related elements. To explore the diversity of the superfamily (in the rice genome or in other sequenced genomes) you would use the amino acid sequence of the TPase protein or part of the sequence. The Blast results in

either case would be numerous and determining relatedness of the elements is impossible from a Blast output. To analyze the relationships between large numbers of related DNA sequences we use phylogenetic trees. These trees are similar to species trees.

Earlier BLAST was used to identify related sequences using a single query of a large database. The result is many 'hits' that must be compared to each other in order to determine which sequences are most closely related. This process is called multiple sequence alignment and there are several computer programs for this task. Once you have a multiple alignment a different software program is used to infer a phylogenetic tree. The process of generating a tree can be time consuming and tedious. Luckily, Yujun Han (a former graduate student in the Wessler lab) streamlined this process by creating a single web-based bioinformatics pipeline called TARGeT. Today TARGeT will be used to explore TE families in the rice genome.

Phylogenetics

Understanding how to interpret phylogenetic trees can be very useful in understanding the relationships of DNA or protein sequences as well as species. Here we have a graphical representation of a phylogenetic tree. Notice the terms and what they refer to.

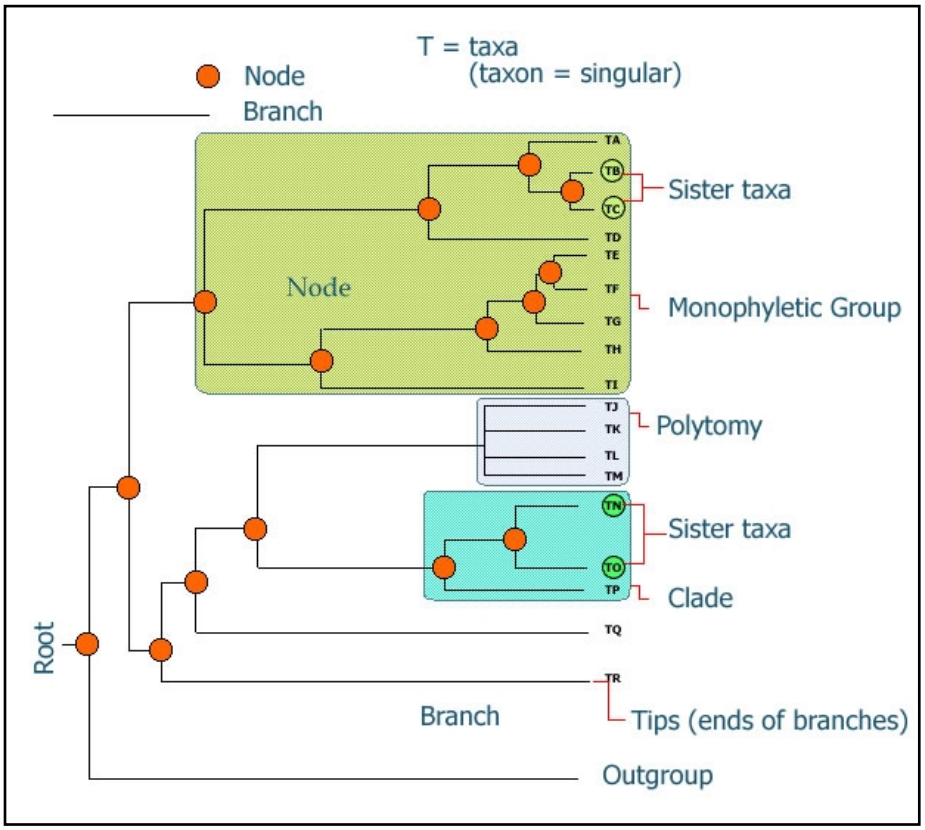


Figure 1: This figure shows a graphical representation of a phylogeny. The important features of the phylogenetic tree are labeled.

Before you can interpret a tree you need to understand some terminology. The tips of a tree represent the sampled individuals. These units are called taxa (taxon = singular). We use the term taxa to refer to any level of organization or any named group of organisms. A taxon can represent all individuals in a defined species, a single individual, or a specific amino acid or nucleotide sequence. In a species tree these represent the living organisms that were sampled to reconstruct the phylogeny. In **Figure 1** our species are labeled taxa A-R. Other types of trees can be made using specific genes or gene families, or in our case these input taxa represent the DNA sequences that are obtained from a database. The individual members of the tree are placed on horizontal lines called branches and branches intersect to form nodes. A node represents the common ancestor (in this case the last common sequence) shared by all members that branch from that node. In Figure 1 the nodes are labeled with orange dots. Ancestral node sequences are inferred based on the extant (existing) sequences. These nodes represent what the last common ancestor of that group may have 'looked like' to the best of our knowledge. We can infer the sequence of the nodes using the information we have from the tips. The ancestral sequence is a best guess based on the available data.

When looking at a tree we are able to visualize the relatedness of the individual members that make it up. Individuals that are placed next to each other on the tree (they are connected by only one node) are called sister taxa. In our case the sister taxa on our transposon trees represent the sequences with the highest level of sequence identity (they are most similar to each other).

All members that arise from the same node are said to be in a clade (also called lineages). If all members of a group occur in the same clade the group is said to be monophyletic. If all members of a defined group are not included in a single clade the group is considered either polyphyletic or paraphyletic. **Figure 2** shows us the distinction between these two states. In the polyphyletic situation all members of the group do not share a most recent common ancestor. In the paraphyletic case some but not all of the descendants from a most common recent ancestor are included in the group.

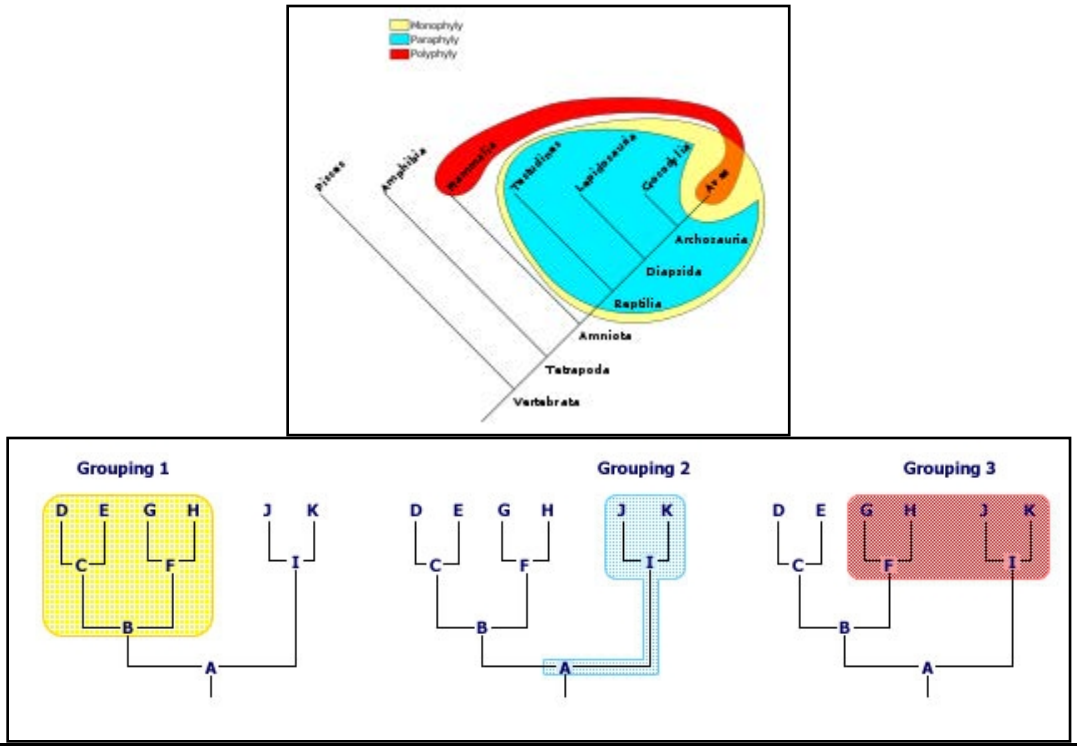


Figure 2: Monophyletic groups are highlighted in yellow, paraphyletic groups are highlighted in blue, and polyphyletic groups are highlighted in red. The tree of the vertebrates gives us an example of a monophyletic group, the sauropsids, a paraphyletic group, the reptiles, and a polyphyletic group, the warm-blooded animals.

In some trees the actual length of the branches connecting the sequences (or species) represent the number of base pair changes over time. So, long branches represent many changes while short branches represent few changes. Branches where more then one tip emerges from one node are called polytomies. If we find polytomies in our transposable element trees, we can assume that the elements placed in the polytomy were very recently active, as all of the sequences are virtually (or are exactly) identical. This example is presented below in **Figure 3**. In the case of the Ponga, Pongb, Pongc, Pongd, and Ponge clade the branches were too short to draw because these are virtually identical copies in the rice genome, and thus they are represented as a polytomy. We can use this information to design new experiments.

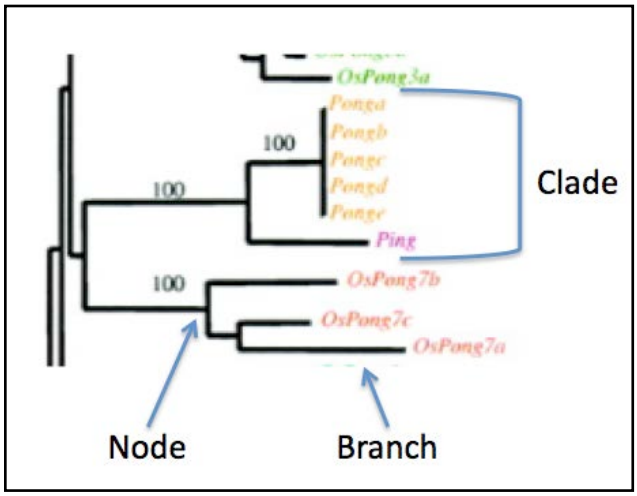


Figure 3. A magnified view of a tree includes the elements most closely related to the Ping element in the rice genome.

How are phylogenetic trees constructed?

Generally, trees are constructed by identifying shared derived characters, also known as synapomorphies. These characters can be, morphological (e.g., beak dimensions or the presence of a hinged jaw), developmental (e.g., presence of a developmental stage such as gastrulation), or DNA or amino acid sequences. In our case we will use the transposase amino acid sequence as the basis for our comparisons.

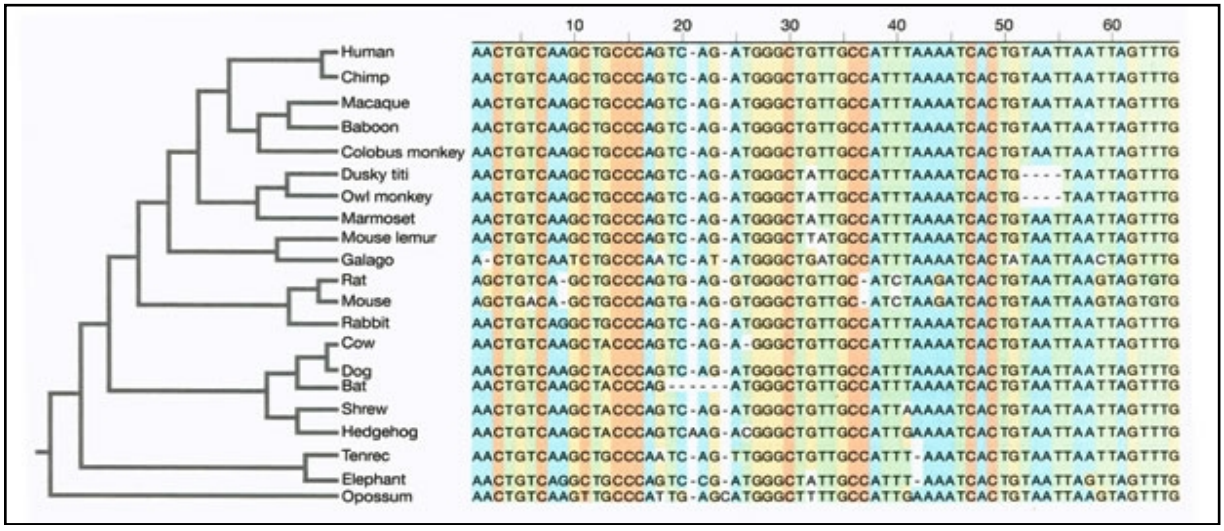


Figure 4: Sequence alignments show the relationships between the mammals. The nucleotide sequences alignment visually demonstrates nucleotide similarities and differences while also showing the presence of gaps in the sequences. The level of similarity between the sequences is used to reconstruct the phylogeny.

When reconstructing a phylogeny we first collect our data and assign similarity between the individuals based on how many characters differ between them. In our example we would place the specific sequence for each individual into a table, with each individuals

sequence in a separate row. The nucleotides (or amino acids) in these rows are then aligned with one another such that each position in the alignment is counted as a character. Any deletions, insertions or base pair differences between the individual's sequences are highlighted by the alignment. See **Figure 4** above for an example of this process. Once the alignments are complete, all pair wise comparisons of the sequences are made. What this means is that each sequence is used as a starting point and is compared to all other sequences present in the alignment. The differences between the sequences (point mutations, deletions and insertions) are noted and the cumulative numbers of changes between sequences are used to generate a value describing how similar the sequences are to one another. From these comparisons a distance matrix is constructed. The more similar a sequence is to another sequence the lower the distance. A sequence compared to itself would have a distance of 0, as all the characters (nucleotides or amino acids) are the same. The more differences we see between any particular comparisons, the higher the distance value. Once the distance matrix is generated based on all of the pair wise comparisons a tree can be drawn.

Although this sounds simple, it is not. If we look at 2 taxa there is 1 possible tree, 3 taxa there are 3 possible trees, 4 taxa there are 15 possible trees, 5 taxa there are 105 possible trees. Once we get up to even the modest number of 10 taxa there are 34,459,425 possible trees. If we want to look at 20 taxa there are 8,200,794,532,637,891,559,375 ($8x10^{20}$) possible trees. Even with the best computers available we cannot efficiently investigate and evaluate the likelihood of all possible trees for any reasonable data set (more then 15 individuals/sequences in the study). Our distance matrix from our multiple alignments can rule out many of these possible trees as impossible given the data, but there are still many trees that 'fit' the data.

In order to pick the best tree, programs use complex algorithms to find the tree(s) that require the fewest number of changes to explain each step in the tree. A tree with the least number of steps or changes needed to explain the relationships between the taxa is the most parsimonious tree. Parsimony simply defined means 'less is better'. In other words the path that requires the fewest changes is the most likely answer. There are many different approaches to generate the best tree.

When looking at a tree you will notice that there are numbers on the branches, these represent what we call bootstrap values. This is a confidence level indicator of how probable that clade is based on the data available. If a clade has a bootstrap value of 100 we can be very confident that this relationship is accurately pictured in the tree. If the bootstrap value is 60 we have less confidence in this portion of the tree. The bootstrap value is analogous to a p-value or confidence interval in statistics or the e-value of BLAST.

Bootstrap values are generated as follows. Let's say that we have 100 individuals in our data set. We first use all the samples to generate the best fit tree. Once we have the best fit tree we take a sub-set of the original data set, 50 individuals, and re-run the

program generating a new tree. The new tree generated from the smaller sub-sample is compared to the original tree generated from all the data. The original tree is evaluated by counting the number of times the same groupings are generated in the sub-sampled data sets. If the same relationships are seen again and again then we have more confidence in their biological reality. A value of 100 indicates that the clade was generated every time the data was sampled. A value of 60 indicates that that particular clade was found 60% of the time that the data was sampled. With a bootstrap value of 60, we say that this clade would not be "well supported". This process of sub-sampling is done over and over again using a different random set of 50 individuals each time. Typically 100 to 200 bootstrap replicates are used to estimate tree reliability. The more often the same clades are constructed using different subsamples, the higher the bootstrap value, and the more confident we are that the relationships are represented accurately. As you might imagine, generating so many trees is an enormous task that would not be possible without computers.

Phylogenetic Analysis of TE families in Rice

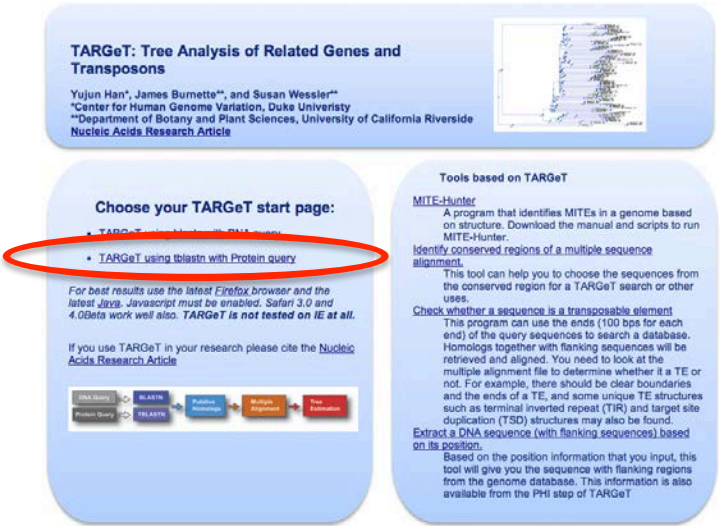
Phylogenetic analysis requires many steps to obtain related sequences and translate the relationships into a visual diagram of a tree. The steps are:

1. Related sequences are obtained often using BLAST searches.
2. The sequences are extracted from BLAST output.
3. A multiple sequence alignment is done on the sequences.
4. The tree is inferred based on the multiple sequence alignment.

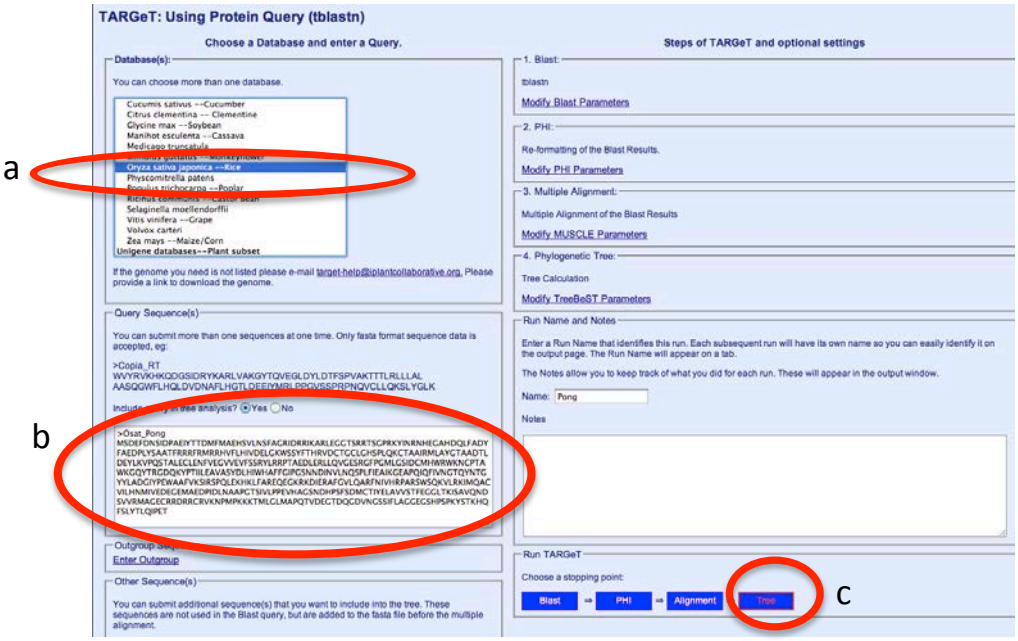
To streamline and speed up this process TARGeT was developed. What used to take several class periods due to a lot of tedious reformatting of data can be done in minutes.

In the past a nucleotide query was used against a nucleotide database. Today a protein query will be used against a translated nucleotide database (called a tblastn). Based on the previous discussion of TE sequence evolution, why is it better to use the protein sequence as the query rather than the nucleotides?

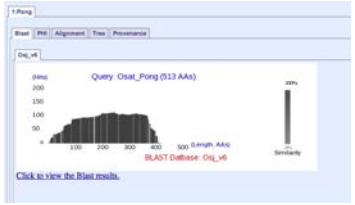
1. Open a web browser and open the page for today in one tab and TARGeT in another tab.
2. Choose "TARGeT using tblastn with a protein query.



3. Transposase sequences for different TE superfamilies are available on the course website.
    a. Choose Oryza sativa japonica—Rice for the database.
    b. Choose one of these sequences and paste it in the query window of TARGeT.
    c. Click the “Tree” button.



4. The results of the TARGeT search will be discussed in class.



Learning Goals

Students will be able to:

1. Define TE family and superfamily and identify the difference between the two terms.
2. Discuss the discovery of mPing and Ping using bioinformatics.
3. Explain that the bioinformatics results are a hypothesis.
4. Use TARGeT to identify TE families in a plant.
5. Define phylogenetics and “tree terms:”
    a. Clade.
    b. Branch.
    c. Node.
    d. Taxon/Taxa.
    e. Monophyletic.
    f. Polytomy.
6. Recognize recent TE movement using a phylogenetic tree.

10: Microscope Activity

This activity will introduce the compound microscope using transmitted white light. Using dry mounts and wet mounts various specimens will be observed. Microscopy is integral to many research areas especially cell biology. During the next lab period Arabidopsis leaves will be observed using the dissecting scope and fluorescence.

Background

The Compound Light Microscope

Structural elements.

1. **Binocular eyepiece tube.** This assembly holds the ocular lenses. It is designed to allow adjustment of the distance between the oculars and the focus of one of them. The eyepiece tube is held in position by a setscrew (not shown). **This screw should remain tightened at all times** and not used for changing the position of the eyepiece tube.

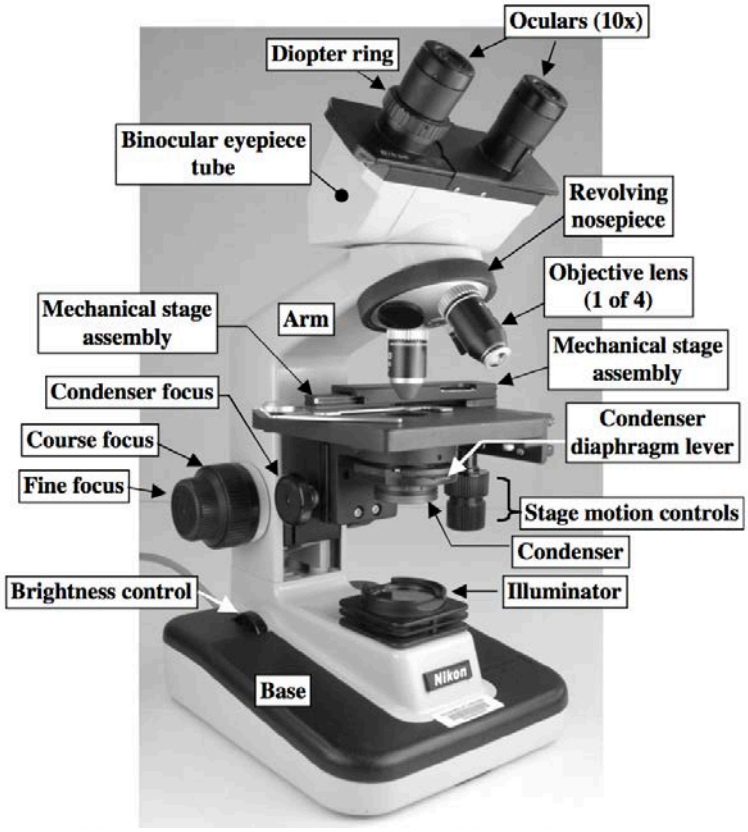


Figure 1. Parts of the compound light microscope; The *Nikon Alphaphot.*

2. **Revolving Nose Piece.** The objective lenses are threaded into this structure. The nosepiece rotates in a manner that allows objectives of different power to be "clicked" into position in the optical path. There are grooves on the edge of the nose piece that allow it to be rotated easily. **Note: do not change objective by pushing or pulling on the objective – this could loosen the objectives and result in scope or slide damage.**

3. **Arm** and **Base.** These are structural elements that support the optical system. The light source is built into the base. Off to the side of the base are the combined on/off

switch and rheostat, which controls brightness.

4. **The stage and associated parts.**

- Stage. The stage is the plane surface on which slides are placed for viewing.

- Focus controls. These controls raise and lower the stage and thus vary the distance between the specimen and the objective lens (this distance is referred to as the **working distance** for that objective). The specimen is focused in this manner. The larger knob that moves the stage greatly when rotated is the coarse focus. The smaller knob that moves the stage only slightly is the fine focus. **The coarse focus should only be used when working with low power objectives. Further, the stage should never be moved upward with the coarse focus control when looking through the microscope - doing so could damage the objective and/or slide. Such damage can be prevented by watching, from the side of the stage, to see that the slide will not touch objective when the stage is raised with the coarse focus.**

- Condenser mounting bracket. This bracket holds the condenser assembly and can be moved up and down with the condenser focus control.

- Mechanical stage assembly. This mechanism functions to precisely move the slide from side to side and/or front to back. It is has two parts: 1. **Movable slide clamp.** This device is spring-loaded and "grasps" the slide **by its edge.** 2. **Stage motion controls.** These are 2 knobs (located one on top of the other) that project downward from the stage - one moves the slide from side to side and the other from front to back

The fine focus control has a limited range of movement. In practice, this means that there are situations when fine focus control cannot be used to sharpen the image adequately. To correct this, rotate the 10x objective into position and then rotate the fine focus knob to the middle of its range of travel. Next, **very carefully** use the coarse focus to sharpen the image (this will require only the slightest movement of the coarse focus knob). Once this is done, the fine focus control should be usable.

The Optical Components.

The microscope has four optical components: 1) the light source, 2) the condenser lens, 3) the objective lenses, and 4) the ocular lenses.

1. **The light source.** Light from the sub-stage bulb is diffused by a frosted glass surface located above the bulb. Light intensity is controlled by the power/brightness knob. The amount of illumination necessary for adequate viewing is dictated by the material to be examined and the magnification used. **It must be kept in mind that excessive**

brightness will result in a substantial loss of image contrast.

2. **The condenser.** The function of the condenser lens is to focus the maximum amount of light from the light source upon the specimen. Because this lens "condenses" a broad, dim circle of light into a smaller and brighter spot, it needs to be focused properly for best results. For our microscopes, the condenser is properly focused when the condenser assembly is near the uppermost range of adjustment. The function of the condenser can be demonstrated as follows:

a. Plug in the microscope and turn on the illumination to an intermediate level.

b. Move the condenser to its uppermost position with the condenser focus control and open the condenser diaphragm completely with its control lever.

c. With the lowest power objective in position, lay flat a small (2" x 2" or so) piece of paper over the hole in the center of the stage. When looking directly at the paper (not through the scope), you should see a bright spot of light projected upon the paper.

d. Observe what happens to the spot of light when the condenser is lowered and then raised again.

e. At this time value of having the condenser properly focused should be clear, that is, positioned near the top of range of movement. NOTE: At one level of focus, the clear background of the microscopic field will have a grainy and colorful appearance. This "grain" can be eliminated by slightly lowering the condenser with the condenser focus control.

While a properly focused condenser contributes to a high quality image, it can also cause a potential problem when viewing living specimens. This problem is that the small spot of illumination created by the condenser is also hot in addition to being bright. This heat may harm living specimens. To avoid such problems, make the required observations quickly, or lower the light intensity with the brightness control knob **(do not lower the condenser).**

The condenser diaphragm regulates the diameter of the light beam passing through the condenser. As the condenser diaphragm is closed, specimen contrast increases. This phenomenon can be utilized to great advantage when observing specimens that have low inherent contrast such as living cells. Be advised that microscopic resolution *(see next section)* is lost as the condenser diaphragm is closed.

3. **The objective lenses.** The function of the objective lens is to accomplish the first stage of magnification in the compound microscope. These lenses are threaded into the revolving nosepiece. NOTE: Remember that rotating the nosepiece should be done with

the use of the “grooved” surface on the edge of the revolving nosepiece and **not by pushing or tugging on the objectives themselves.** Doing so could loosen the objectives and eliminate the parfocality of the microscope – this could result in scope or slide damage.

A further word of caution about the objective lenses concerns **lens cleaning** - improper cleaning of the objectives can result in a scratched lens. This kind of damage is prevented by keeping in mind that all exposed lens surfaces should be cleaned **with lens paper only. DO NOT USE** KLEENEX, LAB TISSUE, OR HANDKERCHIEFS! A TA will demonstrate the best way to clean the lenses of the microscope.

At this point note the markings on the three or four objectives on the nosepiece. These are shown in Figure 2. Two points need to be made here:

First, the magnification listed gives the magnification achieved only with the objective lenses. Since the ocular lens, to be discussed later, magnifies the image formed by the objective lens ten times, the value for each objective lens magnification times the 10x magnification contributed by the ocular gives the total magnification for the combined (compound) lens system.

Second, numerical aperture (NA) is a measure of the light gathering capacity of the objective lens. The greater the value of NA, the greater the resolving power of the lens.

Figure 2. Objectives of the Nikon Alphaphot. The upper large numbers indicate the magnifying power of the objective; 4 - 100x. The number to the side or just below the mag. value indicates the numerical aperture of the objective; 0.1-1.25. The numbers at the bottom give some other information about the objective.

Some optical principles involving the objective lens system.

The **resolving power** of a lens system is defined as the minimum distance that two points can be spatially separated and still be seen as two individual points. The resolving power of the unaided human eye is 0.1 mm (100 µm). That is, if two points are spaced closer than 100 µm, the human eye sees them as a single point. The maximum resolving power of the light microscope is about 0.2 µm. Since many cellular organelles are 1.0 µm or less in length, they are at the limit of the resolving power of the light microscope.

Depth of field for a given objective is a term that relates to the thickness of the microscopic field that is in usable focus at a particular time. This definition may be obscure at the moment, but the following exercise will help illustrate it. Before beginning, please refer to the content of Section II in the SUMMARY OF MICROSCOPE

USE PROCEDURES above. Use the techniques described there for this and all subsequent microscopy exercises.

a. Obtain a slide with three overlapping threads of different color.

b. Using the protocol referred to above, obtain a low power (40x) image of the area where the threads overlap and then determine which of the three threads is on top, in the middle, and on the bottom.

c. Now focus as best as possible on the middle thread. Are the top and bottom threads clearly visible?

d. While still focused on the middle thread, bring the medium power objective into position and sharpen the focus on the middle thread. Are the top and bottom thread clearly visible?

e. Repeat step "d" using the high power objective. Are the top and bottom threads still visible? Depth of field decreases as magnification increases. This relationship can be used by the microscopists in two ways: First, lenses with a thicker depth of field should be used when either locating smaller specimens or when studying the spatial relationships of larger structures where high resolution is not important. Lenses useful here are low power lenses. Second, lenses with a thinner depth of field should be used when to study the spatial relationships of smaller structures where high resolution is more important. High power lenses with higher numerical apertures are most useful here.

Working Distance of the objectives lenses. As mentioned previously, working distance for a focused objective is the distance between the tip of that objective and the specimen being viewed. Working distance varies greatly between the objectives supplied with our microscopes. To observe this variability, do the following.

a. Starting with the microscope configuration achieved in part "c" of the Parfocality exercise (40x objective focused upon the letter "e") observe the working distance of the 40x objective.

b. Rotate the 10x objective into position, focus, and note the working distance.

c. Finish by rotating the 4x objective into position, focus, and note the working distance. It should be clear that working distance decreases with an increase in the power of the objective. While an awareness of the working distance for the different objectives has no direct influence upon the quality of the image, it does have two practical uses. First, the microscopist must always be mindful of the short working distances of high power objectives and the potential for unintended

contact between the objective and the slide. This potential problem is averted by **always beginning observations on a new slide with lower power objectives.** Second, the long working distance of the low power (4x) objective offers plenty of room for exchanging slides and **eliminates the need for lowering the stage when a new slide is selected.**

4. The Ocular Lenses. The ocular lenses use the real image formed by the objective lens as its object and produces the virtual image that the viewer sees (please see Figure. 3). The reason that the virtual image has this name is because it has the "virtual" appearance of being present in a plane located below the microscope. Once again, this type of microscope is called "compound" because the image that the viewer sees is formed by two lenses in optical series.

Observations relevant to the virtual image of the compound microscope:

• Obtain a focused image of the letter "e" at low power. What direction does the image move when the slide is moved toward you? Away from you? To the right? To the left? Is the image upright or inverted?

The presence of two ocular lenses (one for each eye) provides an advantage to the microscopist as well as a disadvantage. The advantage is that, with a binocular viewing system, the microscopist has greater depth perception of the selected field of view. The disadvantage is that because of individual differences in vision, the oculars need to be adjusted to accommodate different users. These adjustments are made as follows:

- Adjustment of inter-ocular distance (I.O.D.). While looking through the microscope, the eyepieces are moved apart or together to accommodate the varying I.O.D.'s of different viewers.

- Adjusting focus of the ocular lenses. Note that one of the oculars has a fixed focus (right side) while the other is adjustable with the diopter ring (left side). This permits the viewer to compensate for differences in the vision of each eye. Making this adjustment: With the left eye closed and the right eye open, focus the microscope with the lower focus controls until the image is sharp. With the right eye closed and the left eye open focus with the diopter ring until a sharp the image is obtained. The microscope is now adjusted to compensate for the differences between eyes. Some people have difficulty using binocular microscopes that is not related to either of the above adjustments. One explanation for this problem is that the microscope has been damaged by a negligent previous user thus making binocular viewing impossible. Another common problem here is that the viewer is holding his/her eyes too close to the oculars. The solution is to back your eyes off slightly from the oculars and redo the two adjustments described above. If you suffer from near- or far-sightedness, you should use the microscope without your glasses; it corrects for faulted vision.

The microscope cannot correct for astigmatism.

Activity

PRACTICAL MICROSCOPE USE AND THE CONCEPT OF IMAGE CONTRAST

In this section, we will look at some plant and animal cells with the goals of practicing microscopy skills and introducing another important microscopy concept; **contrast.**

Contrast.

What can be seen in the light microscope is limited by the size of the object. Another limiting factor in microscopy concerns image contrast. Contrast relates to the ability of the viewer to distinguish an object from its background. For example, it is easier to see dark spots on a white suit than it is to see dark spots on a dark suit. Problems of inadequate contrast are common in microscopy. Two methods to increase contrast will be used in this lab: partial closure of the condenser diaphragm and differential staining.

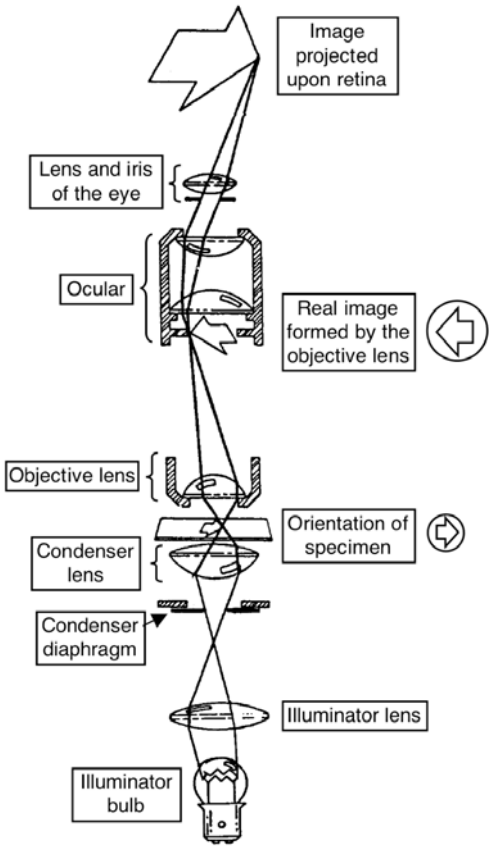


Figure 3 An image forming ray path in a Compound Light Microscope

1. Closure of the condenser diaphragm. For each of the specimens that will be provided, first observe fresh, untreated tissue. It is difficult to see much detail when the condenser diaphragm is wide open. However, if the diaphragm is slowly closed while observing the specimen, more detail becomes visible. This improvement is realized only up to a certain point, after which contrast is not improved and resolving power is lost. Examine the effect of manipulating the condenser diaphragm for each of the available tissues and then make a sketch of the best images in your lab notebook. **NOTE: The condenser must be properly focused to optimize the use of the condenser diaphragm to increase contrast.**

2. Differential staining. Objects seen in the compound microscope are discerned as a result of their interaction with light. Biological material frequently absorbs only a small amount of light and thus has poor contrast. Differential staining involves the use of stains (dyes) which bind to various components of the specimen to increase contrast. Stains usually bind specifically to particular substances in cells or tissues with a unique chemical make-up. Today, three different stains with

three different binding affinities will be used: Acetocarmine. A natural stain that comes from an insect. This stain binds to chromatin and stains it red; thus it is useful for identifying nuclei. I2KI. This stain is a mix of iodine (I2) and potassium iodide (KI). I2KI works through a chemical reaction between the iodine and the starch granules of green plants. The resulting starch-iodine complex has a deep purple/blue color. Toluidine Blue. Unlike the preceding two stains, toluidine blue is a "metachromatic" stain when applied to biological material at an acid pH. This means that when the stain binds to different cellular components, the color produced may be different and dependent upon the chemical composition of that component. For each of the specified tissues, draw images of the specimens reacted with these stains and comparing these images with those drawn from the unstained material. While doing this, think about what is uniquely colored with the different stains.

The Tissues. – each lab table will do one of the following tissues and report their findings to the rest of the class.

Onion epidermis. A TA will demonstrate how to obtain a small strip of onion epidermis that is one cell in thickness. Once you can do this with ease, proceed as follows:

1. Make a wet mount of a small epidermal strip as described in the section below entitled the "Making of wet mounts." Use tap water as the mounting liquid and float the strip on the drop with the torn surface against the water. Don't forget to add a cover slip.

2. Place the slide on the microscope stage. At low magnification, find an area with a minimal number of bubbles. Work up to a 400x image.

3. Manipulate the condenser diaphragm until an image with the best contrast has been obtained.

4. Sketch this image in your lab notebook.

5. Perfuse a drop of acetocarmine under the cover slip as described in the section below. The best staining will occur near the edge of the tissue.

6. Repeat steps 2 - 4 above.

Potato slice. Here make a wet mount of a **thin** section (slice) of potato tissue. The TA will give tips about how to make a section that is thin enough without hurting yourself.

1. Make a wet mount of a small thin section of potato.

2. Proceed as described in steps 2 - 6 for the onion epidermis with the exception of

substituting I2KI for the acetocarmine used in step 5. The thinnest edge of the tissue will be the best place for making observations.

Cheek epidermal cells. The epidermal cells lining the inside of your cheeks will be observed. Use the flat end of a toothpick to scrape the inside of your cheek to remove some of these cells. Gentle pressure is all that is needed. Proceed as follows:

1. Make a wet mount of cheek cells using a drop of saliva as the mounting liquid.

2. Proceed as described in steps 2 - 6 for the above tissues with the exception of substituting 0.1% toluidine blue for the stain. Be careful not to withdraw liquid from under the cover slip too quickly - the cells might sweep out from under the cover slip.

This concludes the practical part of this laboratory session. Turn off the microscope illuminator and clean the low power objective and oculars if necessary.

When finished, write a short summary (a sentence or two should be sufficient) of the results for each of the four contrast enhancing techniques (that is one for the condenser manipulation and one each for the three stains). Then use the results to compare and contrast the structure and composition of the cells that were examined.

**** * Two Useful Microscope Techniques**

Making of wet mounts:

A wet mount is a common form of specimen preparation for light microscopy that is used for the observation of a wide range of specimens. They are prepared as follows:

1. Place a single drop of water or other mounting liquid on the center of a microscope slide.

2. Place the sample upon the surface of or into the mounting liquid as specified. Grasp the opposite edges of a single cover slip between the thumb and forefinger. Be careful here, coverslips are quite thin and easily broken.

   Touch the outer free edge of the cover slip to the slide surface at a point close to the drop. While holding the edge of the cover slip against the slide and slightly angled from the slide surface, move the slide toward the drop until they touch - the meniscus of the drop should “grab” the cover slip.

3. Slowly lower the cover slip onto the drop and release it from your grasp as the cover slip approaches the slide.

Perfusion of stains under a cover slip: An easy way to apply aqueous stains to a slide

that has been “cover-slipped” is as follows:

1. Apply a full drop of the stain to the slide surface at the edge of the cover slip - be careful not to get the stain on the top of the cover slip.

2. Place the edge of an absorbent material, such as a small piece of a paper towel, against the edge of the cover slip opposite to that which the stain was applied. This should pull some of the stain under the cover slip.

3. Blot off any excess stain. ***SUMMARY OF MICROSCOPE USE PROCEDURES:***

I. Obtaining the desired image with your microscope:

1. Starting checklist:

a. Plug in the ‘scope, turn on the sub-stage illuminator, and raise the brightness to an intermediate level.

b. Check to see that the low power (4x) objective is in position. If it is not, use the edge of the revolving nosepiece to move it into position.

c. Check to see that the condenser assembly is close to its uppermost possible position. If not, use the condenser focus control to move the condenser assembly upward.

d. Check to see that the condenser diaphragm is not completely closed.

2. With a slide on the stage of the microscope, open the movable slide clamp by pushing it to the left and position the slide so that it can be held by the clamp. Remember that the slide clamp holds the slide by its edge.

3. Using the stage motion controls, move the slide so the object to be observed is positioned directly over the condenser. (If *you cannot see the specimen, focus on the edge of the cover slip – this will give an adequate starting focus.)*

4. While looking at the microscope stage from the side, carefully move the stage to its uppermost position with the coarse focus control.

5. While looking through the microscope, slowly lower the stage with the coarse focus control until an image of the specimen appears. Sharpen the focus with the fine focus control.

6. If higher magnification is required, center an important feature of the specimen in the

field of view and sharpen the focus using the fine focus control. Then rotate the revolving nose piece so that the next strongest objective clicks into position. Some slight focusing will be necessary here to produce a sharp image of the specimen.

7. If still higher magnification is required, repeat step 6.

II. Changing slides:

1. Rotate the low power (4x) objective into position. Because of the long working distance of this objective, **THERE IS NO NEED TO LOWER THE STAGE WHEN CHANGING SLIDES.**

2. Remove the slide from the grasp of the slide clamp.

3. Insert the new slide into the slide clamp and move the specimen into the light path. If the focus was not changed, the new specimen should be in close enough focus for beginning observations.

III. Concluding the microscopy session:

1. Rotate the low power (4x) objective into position and remove the slide from the grasp of the slide clamp.

2. Using the stage motion controls, center the stage mechanism over the stage.

3. Clean the ocular lenses and the 40x objective if necessary.

4. Turn off the illuminator.

Learning Goals

1. Be able to name the microscope parts shown in the labeled diagrams.

2. Be able to describe what constitutes improper use of (1) the coarse focus control, (2) the revolving nose piece, and (3) the movable stage clamp and to list at least one major consequence of improper use of each of these parts.

3. Be able to routinely utilize all of the microscope use procedures described here.

4. Be able to safely clean the ocular and higher power objective lenses.

5. Be able to adjust the inter-ocular distance and the oculars (diopter ring) for

comfortable viewing.

6. Be able to describe how the objective and ocular lenses interact to form the magnified image that the viewer sees.

7. Be able to describe the relationship between resolving power and numerical aperture.

8. Be able to describe what the condenser lens does and how an improperly focused condenser can influence both image quality and the usefulness of the condenser diaphragm.

9. Be able to define the following terms: resolving power, depth of field, working distance, parfocal (parfocality), contrast, differential staining.

10. Be able to describe how the following microscope parameters change as one moves from the lowest power objective to the highest power objective: (1) resolving power, (2) depth of field, and (3) working distance.

11. Be able to describe the value of binocular viewing to the microscopist.

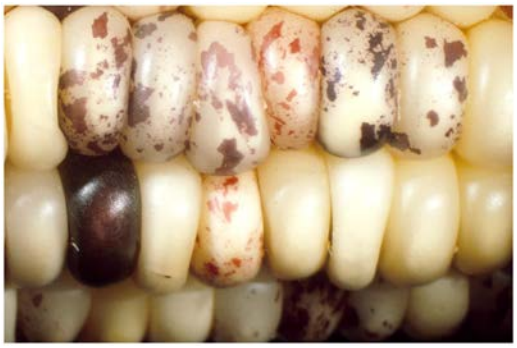
12. Be able to give a positive and a negative aspect of using the condenser diaphragm to increase contrast.

13. Be able to give a positive and a negative aspect of using differential staining to increase contrast.

11: Discovery of an active rice TE

A visual assay for the movement of rice TEs in *Arabidopsis thaliana*

We need to create an experimental system that mimics the one used by McClintock with TEs inserted into pigment genes and expressed in the kernel. For this experiment, we need to use a visual assay to test for movement of the rice mPing element in Arabidopsis.



Creating a visual phenotype: You know what a reporter is—someone who goes out, gathers facts, brings back information, and turns it into ordered and accessible information. Just so, scientists use so-called reporter genes to attach to another gene of interest in cell culture, animals, or plants. Certain genes are chosen as reporters because the characteristics they confer on organisms expressing them are easily identified and measured. Most reporter genes are enzymes that make a fluorescent or colored product or are fluorescent products themselves. Among the latter kind is one that is central to your work this quarter, called Green Fluorescent Protein or GFP, the same gene that was on the plasmid pGLO used in the transformation experiment.

In our experiments, the GFP reporter gene will substitute for the maize pigment gene. The mPing element has been engineered into the GFP gene so that it cannot produce fluorescent protein. If mPing excises the GFP gene will be able to function again.

Arabidopsis thaliana

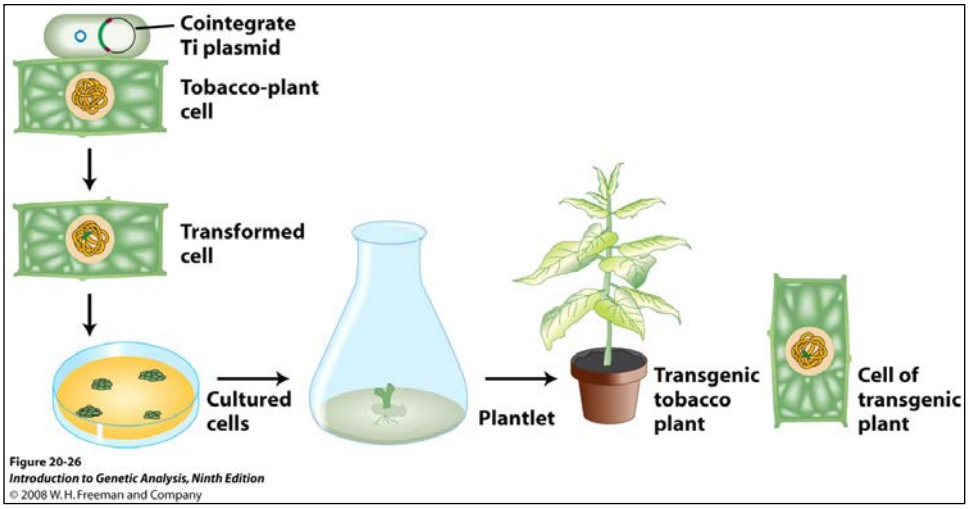
Model organisms include *E.coli*, yeast (*Saccharomyces cerevisiae*), *Drosophila melanogaster*, *Caenorhabditis elegans* (a.k.a. the worm), mouse (*Mus musculus*), and *Arabidopsis thaliana*. Like the other model organisms, *A.thaliana* is easily transformed by foreign DNA and is small and has a relatively short generation time (~6 weeks). This small flowering plant is a genus in the family *Brassicaceae*. It is related to cabbage and mustard. *A. thaliana* is one of the model organisms used for studying plant biology and the first plant to have its entire genome sequenced (~125 Mb, about the same as *Drosophila*).

Agrobacterium tumefaciens: introducing foreign DNA into plants

A crown gall tumor. Infection by the bacterium *Agrobacterium tumefaciens* leads to the production of galls by many of plant species.

In 1977, two groups independently reported that crown gall is due to the transfer of a piece of DNA from *Agrobacterium* into plant cells plants (Mary Dell Chilton, a postdoctoral associate at the University of Washington, and two other researchers working in Germany named Marc Van Montagu and Jeff Schell). This resulted in the development of methods to alter *Agrobacterium* into an efficient delivery system for gene engineering in plants. In short, *Agrobacterium* contains a plasmid (the Ti-plasmid) that contains a fragment of DNA (called T-DNA). Proteins encoded by the Ti-plasmid facilitate the transfer of the T-DNA into plant cells and ultimately, insertion into plant chromosomes. As such, the Ti-plasmid and its T-DNA is an ideal vehicle for genetic engineering. This is done by cloning a desired gene sequence into the T-DNA that will be inserted into the host DNA by *Agrobacterium*.

As shown in the figure below, foreign DNA is inserted in the lab into the T-DNA (shown as the green DNA in the "cointegrate Ti plasmid below), which is then transformed into Agrobacterium, which is then used to infect cultured tobacco cells. The Ti plasmid moved from the bacterial cell to the plant nucleus where it integrated into a plant chromosome. Tobacco cells can be easily grown into "transgenic" plants where all cells contained the engineered T-DNA.



Schematic of how Agrobacterium has been exploited to deliver foreign DNA into plant chromosomes.

The foreign DNA inserted into the T-DNA included both a gene of interest and a "selectable" marker, in this case, an antibiotic resistance gene. This is necessary because the procedure for transferring a foreign DNA into a plant via Agrobacterium-mediated transformation is very inefficient. By using media/agar containing the antibiotic, only the cultured cells with the T-DNA in their chromosomes will be resistant to the antibiotic and able to grow.

Back to Ping and mPing: how they were discovered

Isolating Ping and mPing from rice: Geneticists had never isolated an active TE from rice like the Ac and Ds elements discovered by Barbara McClintock in maize. The logic used to isolate the first active rice TEs, Ping and mPing, is described.

Rice (*Oryza sativa*) has the smallest genome of all cereal grasses at 450 million base pairs (Mb). By contrast, the maize genome is almost six times larger at 2500 Mb. About 40 percent of the rice genome comprises repetitive DNA and most of this is derived from TEs. As discussed above, most of the TEs in a genome are inactive due to mutation. Because the full genome sequence for rice is known, members of the Wessler lab were able to use a computational approach to identify TEs that were potentially active based solely on their sequence characteristics.

To find an active TE in rice, researchers compared the publicly available genome sequence of rice to itself. This sounds confusing, but here is what it means: Scientists first used computers to compare the genome sequence of *Oryza sativa* (domesticated rice) to itself and identified several sequences that were repeated (called families or repeats). The repeat families were then analyzed (by computer again) to identify families that contained identical or almost identical sequences. The researchers reasoned that actively moving TEs should be represented by several identical or nearly identical copies in a genome. The reason for this is that when an element moves, an identical copy inserts elsewhere in the genome. Over millions of years these originally identical copies accumulate mutations (more on this later) and start looking different.

By analyzing the genome this way, the researchers found a 430bp sequence with 50 nearly identical copies scattered across the 12 rice chromosomes. They named it "*mPing*" for "*miniature Ping*."

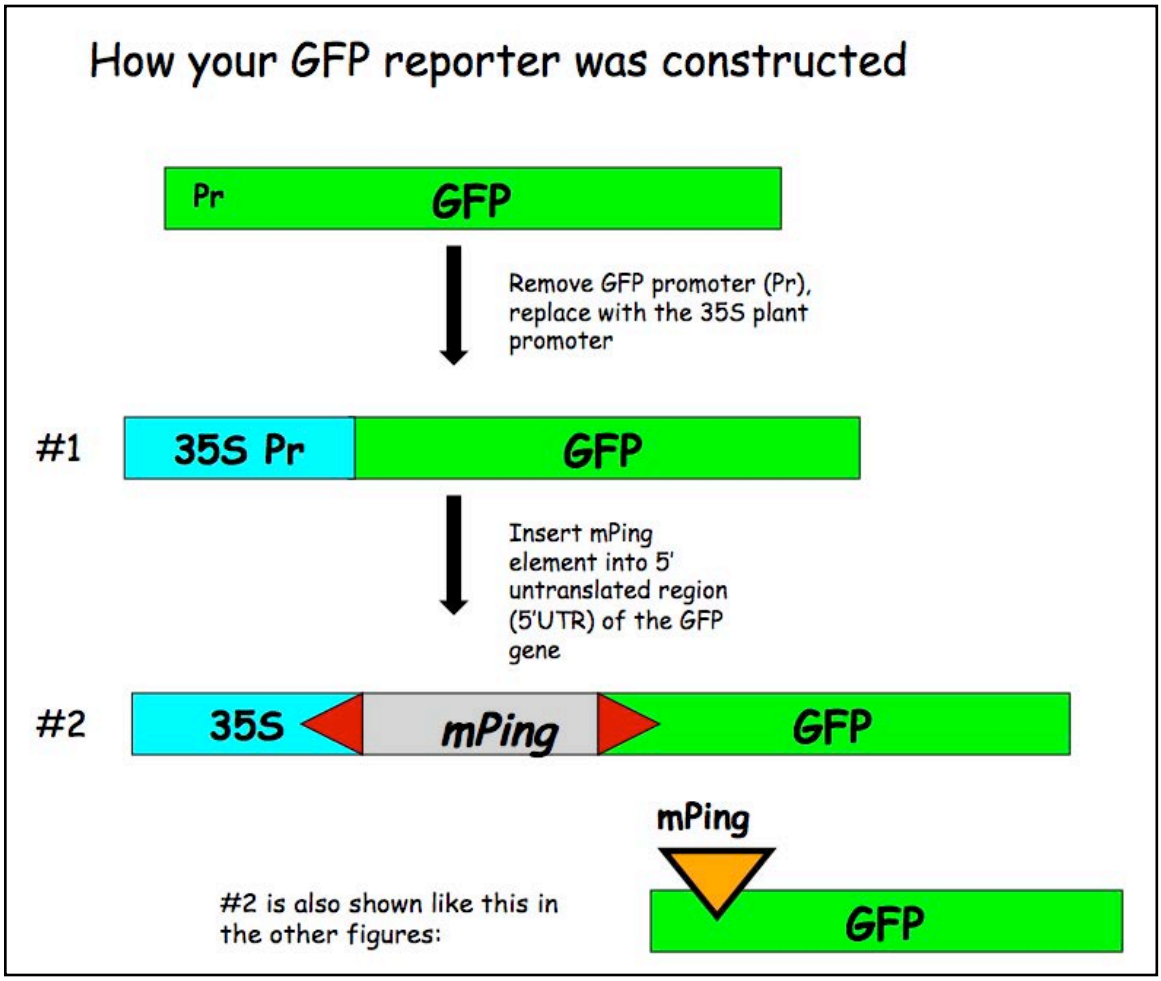
A note here about the precision of words that scientists use to describe experimental results. In this case, the researchers called *mPing* a "candidate" for an active transposon" and not simply an "active transposon." The reason is that computational analysis usually identifies sequences that must be tested further by experiments. In other words, finding identical copies of a TE in a genome is not sufficient evidence to conclude that mPing is in fact an active element. In Experiment 1 you will test whether *mPing* is actually able to move - right before your eyes.

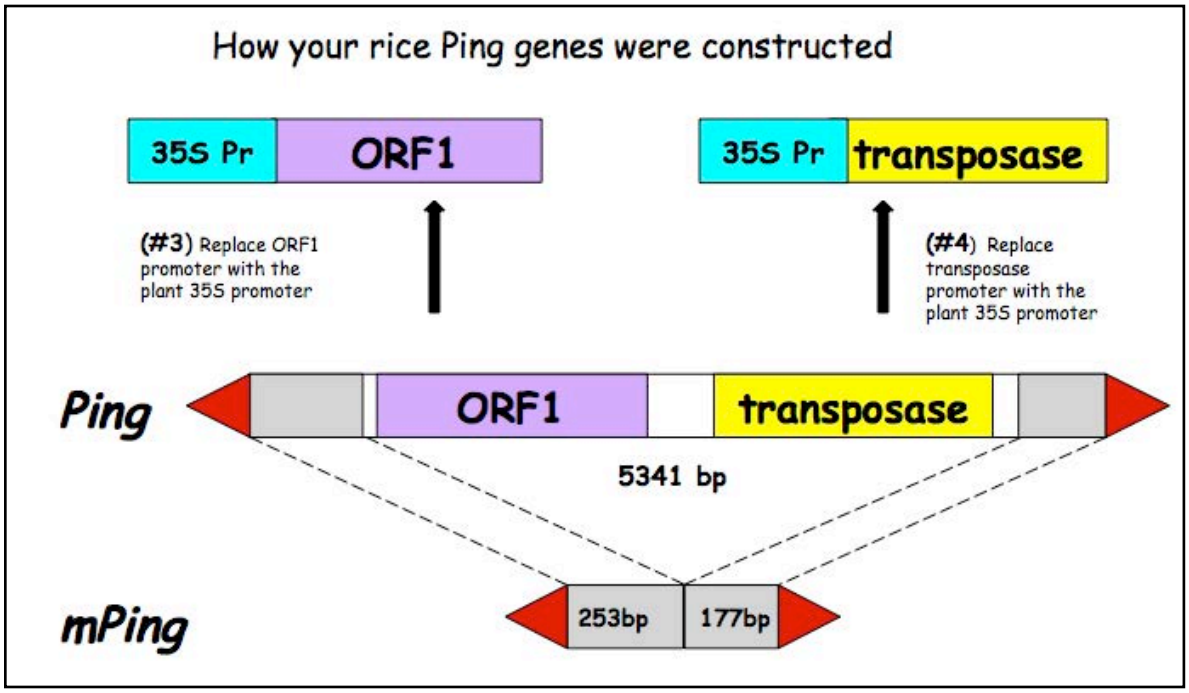
It was puzzling to understand how *mPing* could transpose because it is very small and does not code for any proteins and is thus unable to move on its own. The researchers reasoned that there must be a protein-encoding transposon in the rice genome that encodes the transposase necessary to enable itself and other related elements to move. To find this coding element, the researchers searched the rice genomic sequence for longer related elements. They found a candidate TE which they called *Ping* - that had the same ends as *mPing* but was much longer (~5000 bp) and contained two ORFs. One encodes the transposase gene and the second (called ORF1) is of unknown function (see Figure 7, page 7).

The purpose of this experiment will be to test whether any or all of the proteins encoded by *Ping* can mobilize the *mPing* element. In other words, can either ORF1 or the transposase or both mobilize the *mPing* element?

Design of the experiment and controls

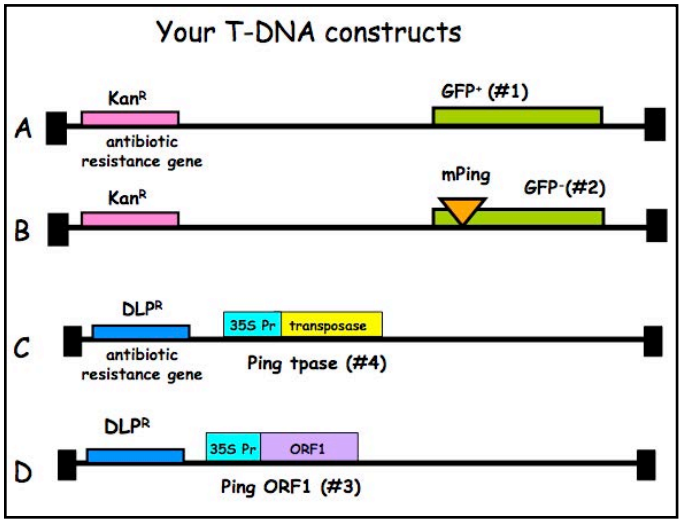
Now up until this point *Ping* and *mPing* were considered active TE candidates, - as there was no evidence that these TEs were actually capable of moving around nor was there evidence that *Ping* produced a proteins that could catalyze the movement of *mPing*. Experimental evidence was necessary to move these elements from candidates to bona fide active TEs. To address these questions, transgenic Arabidopsis plants were generated by engineering T-DNA in the test tube and using *Agrobacterium tumefaciens* to deliver the following constructs into *Arabidopsis* plants. These are described in detail below.


How your GFP reporter was constructed
Pr
GFP
Remove GFP promoter (Pr), replace with the 35S plant promoter
#1
35S Pr
GFP
Insert mPing element into 5' untranslated region (5'UTR) of the GFP gene
#2
35S
mPing
GFP
#2 is also shown like this in the other figures:
mPing
GFP



In this experiment we are testing whether the *Ping* encoded protein(s) can catalyze the transposition of *mPing*. So, there are actually three questions we will be attempting to answer:

--Can ORF1 protein by itself excise (move) *mPing*?
--Can Tpase protein by itself move *mPing*?
--Can both proteins work together to move *mPing*?
To address these questions you will analyze mPing excision in transgenic *Arabidopsis* plants containing <u>one or two</u> of the following T-DNA constructs:



The transgenic *Arabidopsis* plants used in this experiment contain one or two of the 4 T-DNA insertions in their genome.

(A) Plants containing this T-DNA in their genome are the positive control. These plants should be green under UV light because the GFP protein is produced (designated GFP^+).

(B) Plants containing this T-DNA in their genome are the negative control. These plants should be red under UV light because there is no GFP protein (designated GFP^-). Note that the red color is due to chlorophyll fluorescence.

(C) Plants with this T-DNA are part of your experimental unknown.

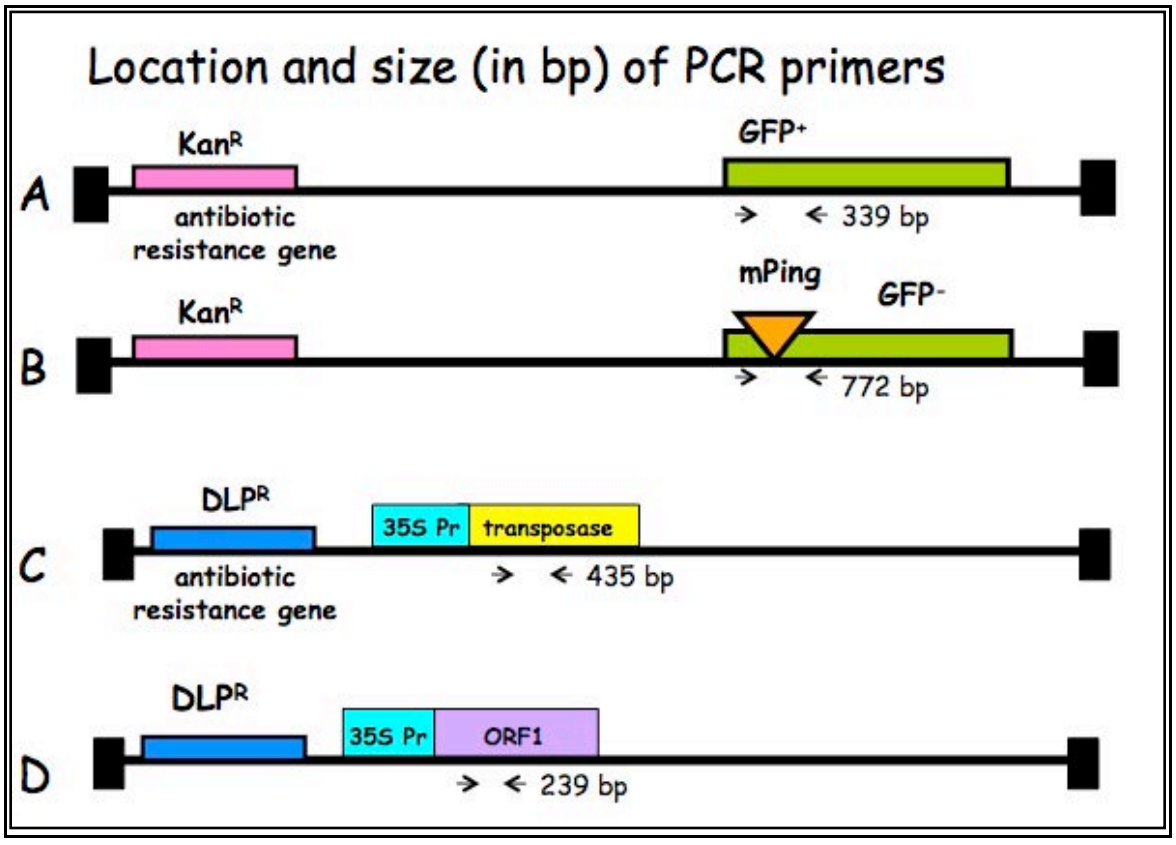
(D) Plants with this T-DNA are also part of your experimental unknown.

(E) Not shown – NO T-DNA at all. This is the wild type control.

Note that A and B have the same antibiotic resistance gene and C and D share a different one.

A closer look at the regions that will be amplified in the experiment

The regions to be amplified by PCR in this experiment are shown below as arrows to indicate the PCR primers and the direction of DNA synthesis.
Once you have grown your *Arabidopsis* seedlings, you're ready to isolate leaf DNA and to do PCR.



Learning Goals

Students will be able to:

1. Define and give an example a reporter gene.
2. Describe how the reporter assay works in the *mPing* experiment.
3. Discuss the controls in the *mPing* experiment.
4. Describe why some leaves have GFP spots.
5. Design a PCR experiment and interpret the results.
6. Describe why the two *Ping* genes were tested independently.
7. Contrast Arabidopsis transformation with bacterial transformation.

12: Detecting TE Excision—Day 1

During this lab period the Arabidopsis seedlings will be observed using fluorescence microscopy. Using blue to ultra-violet light, GFP will fluoresce green if it is being expressed in a seedling. After observing the seedling, they will be ground up and DNA extracted to genotype the seedlings using PCR.

Background

Fluorescence Microscopy

The compound microscopes used in the previous lab had a light source below the specimen that illuminate the whole specimen. In contrast, fluorescence microscopy uses an intense light source to excite a fluorescent dye or protein such as GFP from above. For example if blue light (Figure 1) is used to illuminate a specimen that has GFP expression, the GFP will absorb the blue photons and emit green ones. The green light is then transmitted through the objective and to the viewer. The remainder of the field of view is usually dark. The light source emits all colors of light. Glass filters are used to select the desired wavelength (excitation filter). The dichroic mirror and the emission filter are used to block the excitation light and allow the emitted light to pass to the oculars. This lab uses a dissection scope instead of a compound microscope because resolution at the cellular level is not needed. The dissection scope used in this lab will also excite chlorophyll in the leaves which will fluoresce red. The emission filter will allow the red light to pass so the GFP will appear as green patches on a red background.

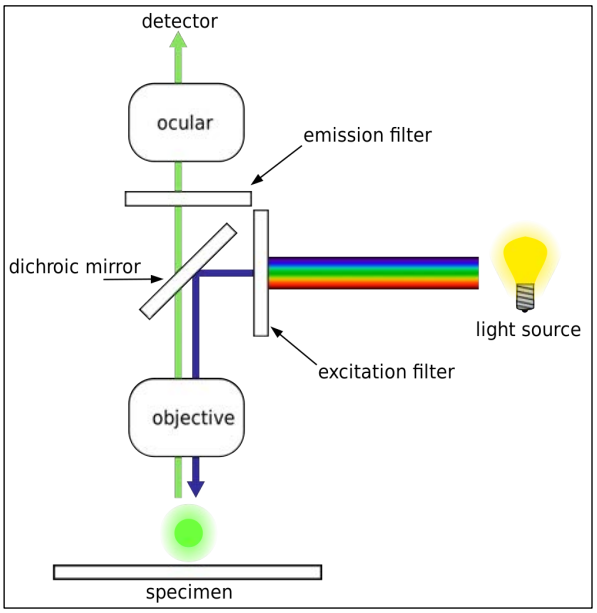


Figure 1. Schematic for fluorescence microscopy optics. Source: http://en.wikipedia.org/wiki/Fluorescence_microscope

Activity

First the plants will be observed for phenotype. There are six plants and each will be looked at using the fluorescence scope. Then the plants will be ground for DNA will be extraction. The introduced genes will be amplified using primers to the three introduced

genes to determine the genotype of each plant. What genes are required to excise mPing can be determined using the phenotype and genotype information.

Viewing GFP Expression

Two weeks ago seeds for 6 genotypes of Arabidopsis were planted. The 6 genotypes are:

1. Wild Type
2. GFP
3. mPing Reporter (GFP with mPing)
4. mPing Reporter + ORF 1
5. mPing Reporter + Transposase
6. mPing Reporter + Transposase + ORF1

An instructor will demonstrate how to use the scope. Record observations in the lab notebook. Images will be made available on the course website.

Genotyping the plants

The seedlings must be genotyped to determine which genes are present in the GFP expressing plants. This is done by extracting the genomic DNA from each line. PCR with primers against the three transgenes (GFP, ORF1, and Transposase) will be done to determine what genes are present.

DNA Extraction

The DNA extraction will be started today, but possibly not finished. A stopping point will be announced in class.

Each group of two will extract DNA from the six lines (each partner does three extractions).

Materials list:

Extraction Buffer
10% SDS (sodium dodecyl sulfate)
5M KOAc (Potassium Acetate)
3 pieces of Miracloth (filter cloth)
100% Isopropanol
70% Ethanol
Sterile water
Freezer block
Liquid nitrogen
65˚C heating block
Sterile 1.5 ml tubes (2 for each prep)
Sterile applicators for grinding
Mortars

Method:

1. Label 2 tubes for each plant. Set one set of tubes aside.

2. Harvest 5-6 seedlings and place in a tube. Ladle liquid nitrogen in a mortar and dip the end of the tube in it. Grind the frozen tissue with a sterile stick. Remove the tube from the liquid nitrogen. If the tissue freezes just wait until it thaws. *Freezing the tissue makes it easier to break open the cell walls.*

3. Add 1 ml of Extraction Buffer, and grind some more in the buffer. *The Extraction buffer contains EDTA a chemical that removes magnesium ions from solution and prevents DNases from destroying the DNA. The buffer keeps the pH slightly basic to further block enzyme function.*

4. Add 120 µl of 10% SDS. Mix by inverting. *SDS is a detergent that denatures and binds to the hydrophobic portion of protein and lipids. SDS is the main detergent in shampoo where it does the same thing to clean hair.*

   Prepare all samples to this step. Keep them on ice until all are ready for step 5.

5. Place tubes at 65˚C for 20 minutes. Pour one 1.5% gel per group.

6. Add 300 µl 5M KOAc. Mix well by inverting several times (important!), then place on ice 5 minutes. KOAc causes the SDS to precipitate along with the proteins and lipids. The solution will turn cloudy.

7. Spin for 5 minutes at top speed in a centrifuge.

8. Squirt about 700 µl of the supernatant through miracloth funnel into the second tube. The instructors will demonstrate how to make a funnel out of the miracloth. This step removes solids that did not pellet during the centrifugation.

9. Add 600 µl of isopropanol. Mix the contents thoroughly by inverting. *The isopropanol (rubbing alcohol) causes a change in the hydrophobicity of the solution and the DNA precipitates. DNA precipitate may or may not be visible at this point; don't worry if you don't see much. Potential stopping point.*

10. Spin for 5 minutes at top speed. *Look for a small glassy pellet in the bottom of the tube that may be visible. The instructors will help you.*

11. Carefully pipette off supernatant. Use a P200 set to 200 µl to remove most of the solution in several steps. Then use a P20 set to 20 µl to remove the last drops.

12. Add 500 ul of 70% ethanol and flick until the pellet comes off the bottom. *The ethanol removes residual isopropanol and salts from the pellet. Potential stopping point.*

13. Spin 3 minutes.

14. Pipette off the ethanol with a P-1000. Use a P200 set to 200 µl to remove most of the solution in several steps. Then use a P20 set to 20 µl to remove the last drops. Make sure the pellet stays in the tube!

15. Leave the tube open on the bench and air dry for 5-10 minutes to allow the remaining ethanol to evaporate.

16. Resuspend the DNA in 50 µl TE. Let sit at RT for about 30 minutes, then mix by pipetting. Depending on amount of starting material may need to be diluted for PCR.

Learning Goals

Students will be able to:

1. Compare and contrast bright field microscopy with fluorescence microscopy.
2. Interpret the images from the microscopy in relation to TE activity.
3. Extract DNA from plants.

13: Detecting TE Activity—Day 2

Today the DNA extraction will be completed. The concentration and purity of the DNA will be determined using spectroscopy. PCR using primers against GFP, ORF1, and Transposase will be used to amplify those genes from the DNA samples. The PCR results will be visualized on an agarose gel.
Today

1. Finish DNA Extraction
2. Check the DNA quality
3. Determine the DNA concentration

Activity

Part 1 – DNA Extraction

Complete the DNA extraction from the previous lab period, if necessary.

Part 2 – Determine DNA quality

One way to check the quality of the DNA is to run it on a gel. The pattern on the gel will tell you if the DNA is in large pieces or degraded. What do you predict the banding pattern will look like if the DNA is high quality and in large pieces? If it is degraded?

Part 3 – Determine DNA concentration

DNA absorbs UV light at a specific wavelength and the amount of absorption is proportional to the concentration of DNA in the solution. Proteins absorb UV light at different wavelengths and this can be used to determine concentration as well as protein contamination in DNA samples. This activity will demonstrate the relationship of UV absorbance to DNA and protein concentration and purity. The concentration and purity of the Arabidopsis DNA samples will be determined. This information is critical to decide how much DNA to use for PCR.

Absorption Spectra and Beer's Law

Materials:

DNA Standard
Protein Standard
Water

Method:

Work in groups of 4 to complete this part. An instructor will demonstrate how to use the Nanodrop Spectrophotometer.

1. Open the UV-Vis setting on the NanoDrop software.

2. Blank the NanoDrop using 2 µl of water.

3. Clean the pedestal with a lab wipe and blank it with 2 µl of water.

4. Clean the pedestal and measure the absorbance of 2 µl of the DNA Standard.

5. The software will display a graph with the wavelength of light on the x-axis and the Absorbance on the y-axis. The vertical black line is the "Abs. 1 marker" and you can slide it left or right to select the wavelength and see the measured absorbance. Place the black line at the point on the vertical axis where the absorbance is the highest. This is lambda-max for DNA. Record the lambda-max and the absorbance in your notebook.

6. Record the absorbance at 280 nm.

7. Beer's Law describes the relationship between sample concentration and absorbance. Beer's Law is A=**ε**cl where:
    a. A is the absorbance.
    b. **ε** is the molar extinction coefficient and is equal to the absorbance of a 1.0 M solution.
    c. c is the concentration in moles per liter (M).
    d. l is the path length in cm. For the NanoDrop, check the graph carefully to determine what path length was used.

8. The concentration of the DNA standard is XXXXX M. Use Beer's Law to calculate the molar extinction coefficient, **ε** for DNA. Ask an instructor to check your work and results.

9. Clean the pedestal with a lab wipe. Take a reading for the Protein Standard. There should be two prominent peaks. Record both lambda values and the absorbance values in your notebook. Protein is a common contaminate in DNA preparations.

10. The concentration of the protein standard is XXXXXM. Calculate **ε** for protein using the 280 value.

11. A measure of purity of a DNA sample is the 260/280 ratio. Calculate this ratio by dividing the absorbance recorded for 260 nm by the absorbance recorded for 280 nm. Pure DNA samples should have a ratio of 1.8.

DNA Concentration and Purity using NanoDrop the simple way

In the steps above, you used the absorption value at 260 nm to determine the DNA concentration. This version uses a modified form of Beer's law since it is known that a DNA solution with A=1.0 will have a concentration of 50 ng/µl. For routine use, the

NanoDrop software will scan a DNA sample and calculate the concentration for you along with some ratios that indicate the purity of the sample.

1. Select the "DNA" option in the NanoDrop software pane.
2. Calibrate and Blank the NanoDrop if necessary.
3. Take readings for your DNA samples. Record the concentration and the 260/280 ratio for each sample. Earlier you found that one lambda peak for protein is 280 nm. The 260/280 ratio measures the amount of DNA to protein in the sample. For Pure DNA the ratio 1.8. Anything between 1.7 and 2.0 is good. A ratio of 260/230 nm is also reported. Several chemicals including phenol, guanidine, carbohydrates and protein absorb at 230 nm. The ratio should be between 1.8 and 2.0. The DNA extraction protocol in this lab does not use phenol or guanidine so a bad 260/230 ratio could mean protein or carbohydrate contamination.

Part 4--Amplify genomic DNA using PCR

Three pairs of primers are needed: one pair for GPF, one pair for ORF1, and one pair for Tpase. The primers for ORF1 and Tpase will be mixed and used in a single PCR tube. This is called duplex PCR. A separate PCR will be used for the GFP primers.

For each group, one person will set up the GFP PCRs and the other will set up the ORF1+Tpase reactions. Each person will have six DNA samples to analyze. An additional negative control is needed making seven reactions in total for each primer set. What is this additional control? Design the experiment in your notebook. The volume of DNA used may change depending on the concentration and an instructor will determine how much DNA sample to use. An instructor will check it before you continue.

Pour a 1.5% gel for each group and use two combs per gel. The gel will be run on Monday, October 24.

After everyone is done, your samples will be cycled with the following conditions:

1 cycle for:	initial denaturation	94°C	3 min
30 cycles for:	denaturation	94°C	30 sec
	annealing	58°C	30 sec
	extension	72°C	1 min
1 cycle for:	final extension:	72°C	10 minutes

Sequences of the Primers used and sizes of the amplicons:

GFP Primers (772 and/or 339 bp amplimers)

GFP-R 5'- AGA CGT TCC CAA CCA CGT CTT CAA AGC -3'
GFP-F 5'- CCT CTC CAC TGA CAG AAA ATT TGT GC -3'

ORF 1 Primers (239 bp amplimer)

ORF1-FOR 5'- CAC TGG TCA AGG TTG AAG TCA GCG ATC TCT G -3'
ORF1-REV 5'- CAG CAT CCA TTT CGC TCT TGT CTT TCT CTG -3'

Tpase Primers (435 bp amplimer)

TPase-For 5'- GGT ATG TTC GGT AGC ATT GAC TGT ATG CAT TGG C -3'
TPase-REV 5'- GAA TCG ACG TTG TAG AAC ACC AAA TGC TCT CTC -3'

Learning Goals

Students will be able to:

1. Use a Nanodrop spectrophotometer to measure DNA concentration and purity.
2. Describe the usefulness of the molar extinction coefficient.
3. Define lambda max of absorption spectra.
4. Discuss why the 260/280 ratio is useful measure of DNA purity using the absorption spectra for DNA to defend their answer.
5. Understand duplex PCR.
6. Use genotype and phenotype data to draw conclusions from the results.

Appendix 1: The Dynamic Genome Lab Notebook Guidelines

The lab notebook is a written record of the experiments performed and the data collected in lab. Use the following guidelines for keeping the lab notebook.

The lab notebook will be graded before class and you will not be allowed to start an experiment if the notebook is not complete.

Ink should be used for writing in the lab notebook. Write larger and neater than normal to make it easy to read the notebook while conducting and experiment.

Guidelines

Date

The notebook is maintained in chronological order. The date should be recorded using the following format, Day, Date and Year (e.g., Friday, December 18, 2009).

Purpose

This is a short statement of 2-3 sentences about the experiment and why it is being done.

Materials

A list of materials and incubation temperatures required for the experiment.

Protocols

Write out a protocol completely the first time it is used. The protocols are provided in the course book with many details. In the lab notebook, write the "bare bones" instructions. For example the eleventh step of DNA extraction has this instruction:

> 11) Add 500 ul of 70% ethanol and flick until the pellet comes off the bottom (for best washing results). Spin 3 min, then pipette off the ethanol with a P-1000. Suck off the rest of the ethanol with a P-20 pipette. Make sure the pellet stays in the tube! Let air dry in hood for about 5 minutes with the caps open.

You should write in your notebook:

> II. Add 500 ul of 70% Ethanol. Spin 3 min. Pipette off ethanol. Dry for 5 mins.

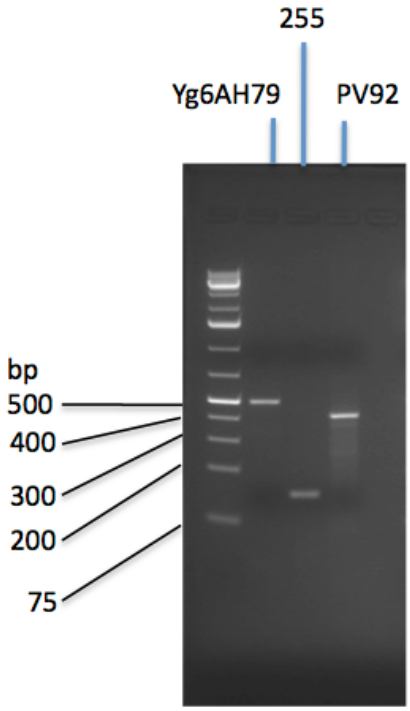
Place a "✓" next to the step as it is completed. This makes it easy to keep track of what steps have been completed. If incubation is necessary, record the starting time. Also indicate modifications done to the protocol while at the bench.

The next time you do the same protocol you can write:

PCR was set-up as on Wednesday, June 21. The following primers and DNA samples were used.

Data

As data is collected such as gel photographs and Nanodrop readings enter them into your notebook. Gels should be labeled so that it is easy to see what is in each lane. Relevant sizes of markers should also be labeled for easy reference. An example of a well-labeled gel is shown below:



Appendix 2. Using Google Docs

Google Docs is an online web app similar to Microsoft Word. It is free and available using any computer with a web browser. What's even better is that all files created using Google Docs are stored in the cloud and providing access anywhere in the world. We will use Google Docs in class for homework assignments and papers.

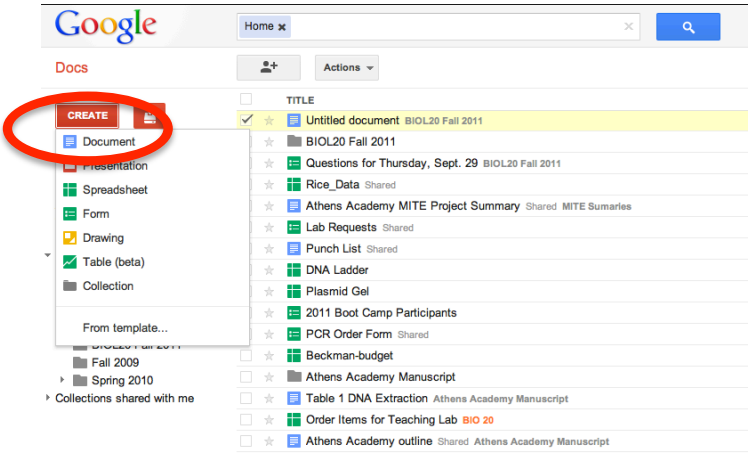
Using Google Docs

Create Account

Signing up for Google Docs is free and easy. Go to docs.google.com create a new account. It is a bad idea to use the same password for Google Docs and your UCR account. You do not need a Gmail account to create a Google Docs account.

Create a Doc Document

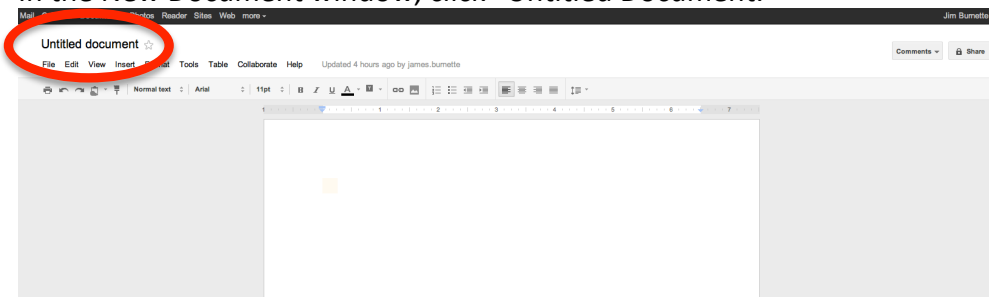
Click the "Create" button and select "Document".



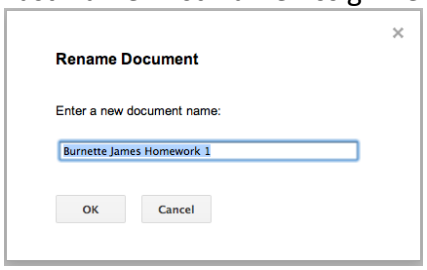
Google Docs also has spreadsheets and presentation which will be used during the quarter also.

Naming the Document

In the New Document window, click "Untitled Document."



In class you will be required to use the following naming scheme:
Last Name First Name Assignment Name, e.g., Burnette James Homework 1



Entering Text

Use Google Doc just like any other word processor. Even the same keyboard shortcuts do what you expect. Remember that on a Mac the "command" key is used in place of the "ctrl" key.

Uploading an image

Click the image icon to upload an image into a Google Doc.

Screen Shots

In class it is often necessary to take a screen shot, that is capture a small area on the screen as an image.

On a Mac

Compose the image and press the "command" + "shift" + 4 at the same time. The Cursor will turn into a crosshair. Click and drag the cursor around the area to be included in the shot. Release the mouse. An image will be placed on the desktop called "Screen Shot" with the date and time. This can be uploaded to a Google Doc for documentation.

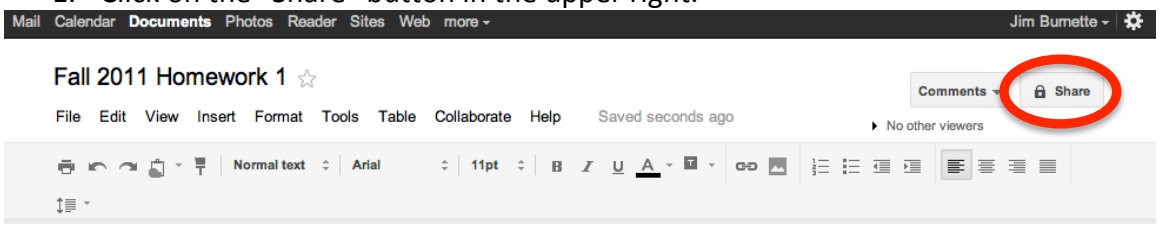
On a PC

For instructions go to: http://take-a-screenshot.org/.

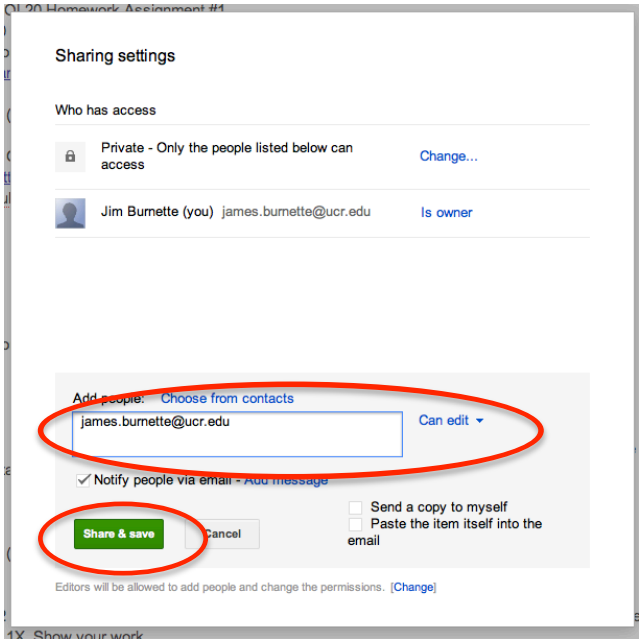
Sharing a Document

Once an assignment is complete the document must be shared with the grader.

1. Click on the "Share" button in the upper right.



2. Enter the e-mail address of the grader. The grader will be listed in the instructions.



3. Click “Share & save”.