

PBIO/BIOL2240L: The Dynamic Genome
Freshman College Summer Experience
Summer 2010

Table of Contents

Syllabus

Grading Policy

Lab Notebook Guidelines

Chapter 1: Background on Information Transfer and Experiment 1: Actin gene annotation		1-57
A. Background	1-14	
B. NCBI and PubMed	15-19	
B. Experiment 1: Actin gene	20-34	
C. Genomic and cDNA sequence analysis	35-47	
D. Using Genome Browsers	49-57	
Chapter 2: The Discovery of TEs and Experiment 2: mPing		59-80
A. Background	49-71	
B. Experimental Design	72-76	
C. Protocol for Experiment 2	77-80	
Chapter 3: Introduction to Genome Browsers		81-87
Chapter 4: Project: Intron polymorphism in maize		89-106
A. DNA Subway	89-99	
B. Primer Design	100-105	
C. Experimental Design	106	



Syllabus Summer 2010



Syllabus

Grading

Data

	Lecture Topic	Lab Work	Due dates	Downloads
Fri, July 2	<ul style="list-style-type: none"> • Assessment • Course Overview • Mechanics 	Pipettors		Start CURE Survey Class Survey
Tues, July 6	Info transfer	DNA Extraction		Exp. 1
Weds, July 7	Intro PCR	PCR, pour gel		
Thurs, July 8	<ul style="list-style-type: none"> • Review • Analyze results 	Gel of PCR		
Fri, July 9	<ul style="list-style-type: none"> • Quiz • Discussion Making of Fittest 			

Mon, July 12	mPing background	Observe Plants		mPing Handout Slides
Tues, July 13	Review	Extract DNA		
Weds, July 14	PubMed	PCR		PubMed Handout
Thurs, July 15	blantn Analyze results	Gel of PCR		
Fri, July 16	Quiz/Mid-Term Projects			
Mon, July 19	Maize Browser			
Tues, July 20	Primer Design Finish Annotation			Browser and Primer Design Handout
Weds	Greenhouse Tour			

Weds, July 21	Sequencing Facility Tour			
Thurs, July 22		PCR, pour gel		
Fri, July 23	Quiz	Gels, more PCR		
Mon, July 26		Lab		
Tues, July 27		Lab		Homework 4
Weds, July 28		Lab		
Thurs, July 29	Course Review	Lab		
Fri, July 30	Final	<ul style="list-style-type: none"> • Surveys: • Post-Survey • CURE • Department FCE <ul style="list-style-type: none"> ◦ 2 FCE paper evaluations 		

BIO/PBIO 2240L The Dynamic Genome
Summer 2009

Dr Susan Wessler and Dr Jim Burnette
Ryan McCarthy, TA

Course website: <http://www.dynamicgenome.org/classes/fcse10/>

User: dynamicgenome

Password: tesjump

	Dr. Susan Wessler	Dr. Jim Burnette	Ryan McCarthy
Office	Plant Sciences 4510	Plant Sciences 1506	Plant Sciences 3507
Phone	706-542-1870	706-542-4581	706-542-5622
Hours	By appointment	By appointment	By appointment
E-mail	sue@plantbio.uga.edu	jburnette@plantbio.uga.edu	rmccarthy@plantbio.uga.edu

Attendance: We require 100% attendance and class participation. Any missed lab will be difficult to make up. If you know you will be absent for any class, make arrangements in advance with the instructor. Discuss unplanned absences immediately upon returning to class. **If you have a fever, do not come to class. Call Dr. Burnette and go to the Health Center.**

Class participation is a major part of this course. You are expected to be prepared for each day, participate in all discussions, and ask a lot of questions. Twenty percent of your grade is based on class participation.

Restrict cell phone/texting/earphone use and personal web browsing/e-mail to breaks. Cell phones should not be on your desk or lab bench at any other time. Do not use class time to work on assignments for other classes. Do not listen to music with earphones during lab work. We will provide a stereo for the whole lab.

In the computer lab, place your backpacks in the cubby-holes.

For your safety, you must wear closed toe shoes (no flip-flops or sandals). Long shorts are permitted. Long hair should be tied back away from the face for all labs. Eating is permitted in the computer lab (room 1503A) but not the wet lab room 1606.

The syllabus and other handouts can be found on the website link above.

Assignment due dates

Assignment	Date
Notebook Checks	Tuesday, July 7, 2009
Homework	Thursday, July 9, 2009
Quiz I	Friday, July 10 2009
Homework	Thursday, July 16, 2009
Mid-Term	Friday, July 17, 2009
Homework	Thursday, July 23, 2009
Quiz II	Friday, July 24, 2009
Homework	Thursday, July 30, 2009
Final	Thursday, July 29, 2009
Final Presentation	Friday, July 30, 2009

The first Notebook check will be on July 7. Spot checks will be done for the rest of the semester.

Quizzes: 30 min. or less and will cover material recently presented in class.

Mid-term and Final will be cumulative. Format will be discussed in class.

The final presentation will be a class poster worked on in class.

Homework assignments will be extensions of in-class lectures and exercises to prepare for the quizzes and tests.

Notebook check	10%
Quiz average	30%
Homework average	15%
Mid-Term	10%
Final Presentation	15%
Participation	<u>20%</u>
	100%

Grading Percentages: The final grade will be calculated using the following break down. Letter grades will be assigned using the standard plus/minus system: A=95-100, A-=90-94, B+=86-89, B=83-85, B-=80-82, C+=76-79, C=73-75, C-=70-72, D=60-69, F<60.

The Dynamic Genome Lab Notebook Guidelines

The lab notebook is a written record of the experiments you perform and the data you collect. You should plan your experiments in the notebook before you do them, record changes made to the protocol as you conduct the experiment, enter and document data, and then provide a summary of the results and next steps. When you are asked where is a particular gel image or how many microliters did you use your response better be, "It's in my lab notebook." Also remember, if it is not in your notebook it never happened!

General outline:

Date

The notebook is maintained in chronological order with the date being entered before starting a day's work. You should use the following format, Day, Date, Year (e.g., Friday, December 18, 2009).

Purpose

This is a short statement to remind you what you are doing. It does not need to be in complete sentences. (e.g., PCR using MITE 54 primers on the B73 Genome to ...)

Materials

A list of materials and incubation temperatures required for the experiment. This is to remind you to collect what you need and turn on incubators before starting experiments.

Protocols

The first time you do a protocol write it out completely. The protocols are provided in the course book with many details. You should write the "bare bones" in your lab notebook. The course book protocol will say:

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:

*11. Add 500 ul of 70% Ethanol. Spin 3 min. Pipette off ethanol.
Dry.*

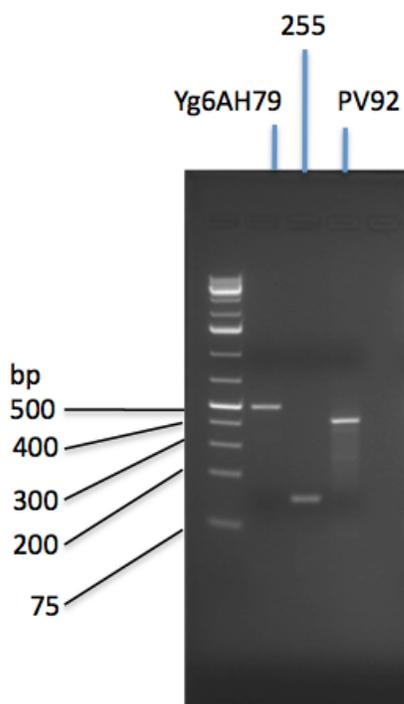
As you complete a step of the protocol or add a reagent to a tube, place a "✓" next to the step or reagent. That way you know what you have completed. '

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

DNA was extracted from B73 using the protocol on Tuesday, January 5, 2010 with the following modifications:

Data

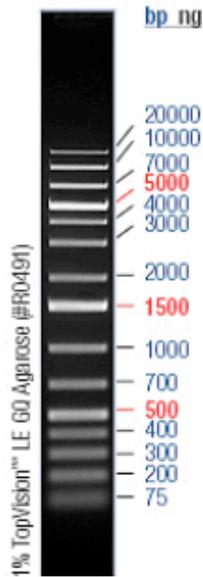
As you collect data such as gel photographs and Nanodrop readings you must enter them into your notebook. Do this as you get the data. Do not allow gels to stack up because you will forget what they are for. Gels should be labeled so that it is easy to see what is in each lane. Relevant sizes of markers should also be labeled for easy reference. An example of a well-labeled gel is shown below:



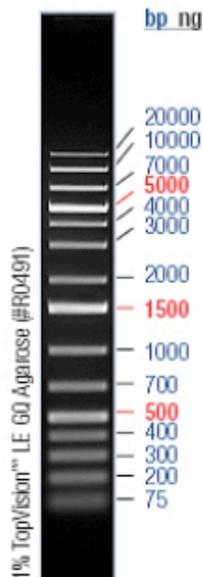
A well-documented experiment will help you tremendously when studying and writing reports. If you continue in science or medicine a complete and up-to-date notebook will make writing a paper much easier.

DNA “ladder” or “marker” used for gel electrophoresis. You should cut out one and tape it to the inside cover of your lab notebook for easy reference.

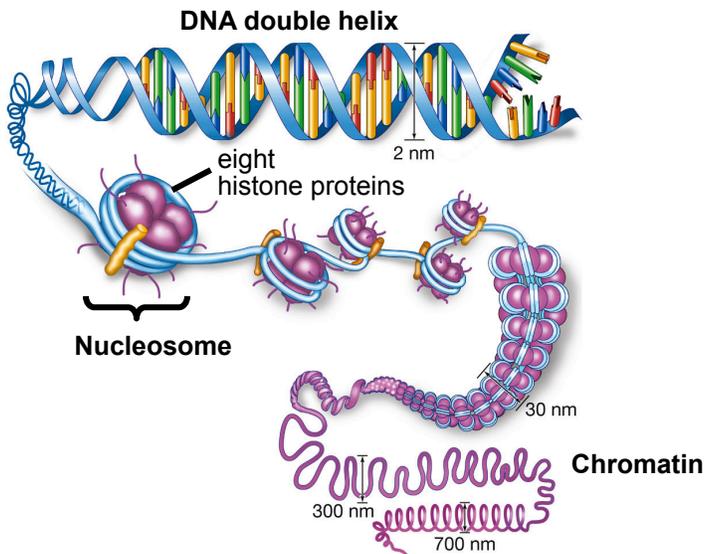
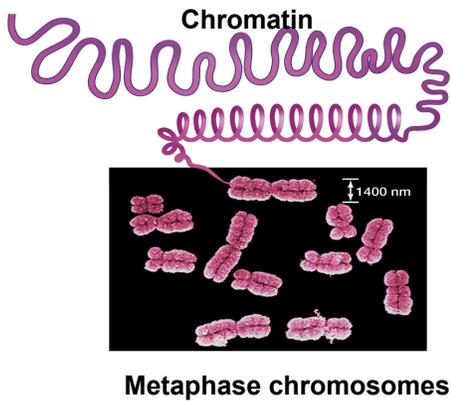
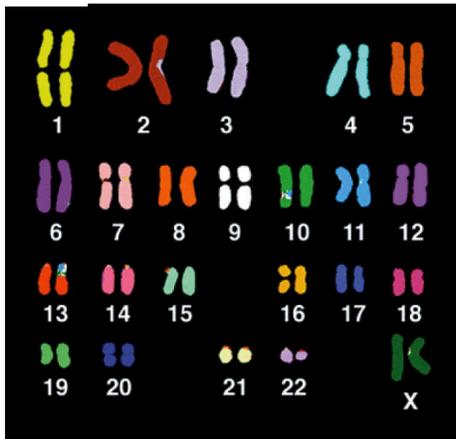
bp	ng/7μl
20000	28
10000	28
7000	28
5000	105
4000	28
3000	28
2000	28
1500	112
1000	25
700	25
500	105
400	34
300	34
200	34
75	34



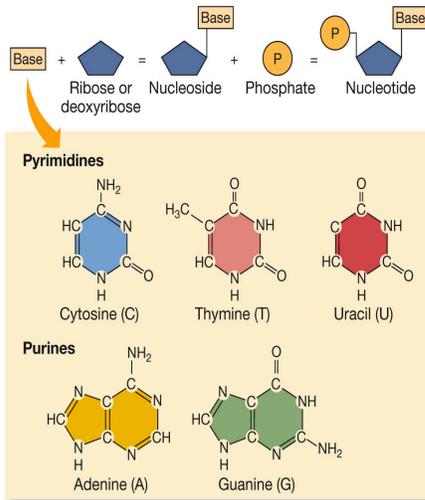
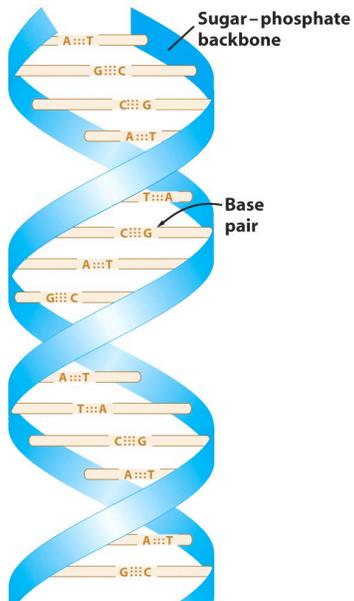
bp	ng/7μl
20000	28
10000	28
7000	28
5000	105
4000	28
3000	28
2000	28
1500	112
1000	25
700	25
500	105
400	34
300	34
200	34
75	34



What is the Genome?



How is the DNA Double Helix Organized?

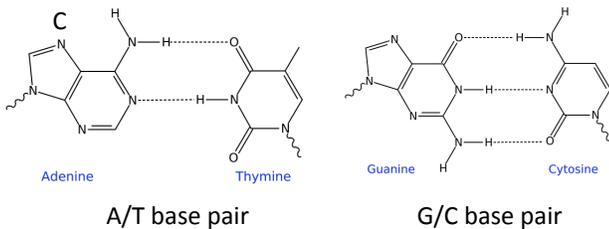


LIFE 8e, 1

A 
 A single strand of DNA. The backbone is in blue, and the bases are colored.

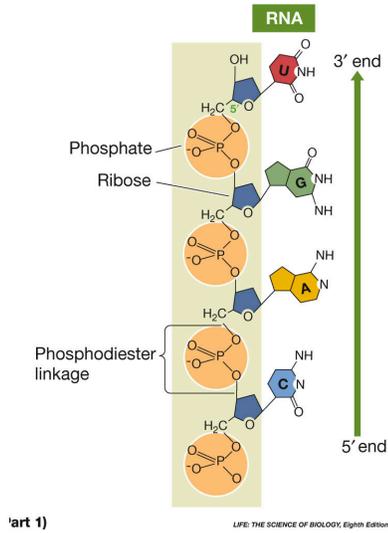
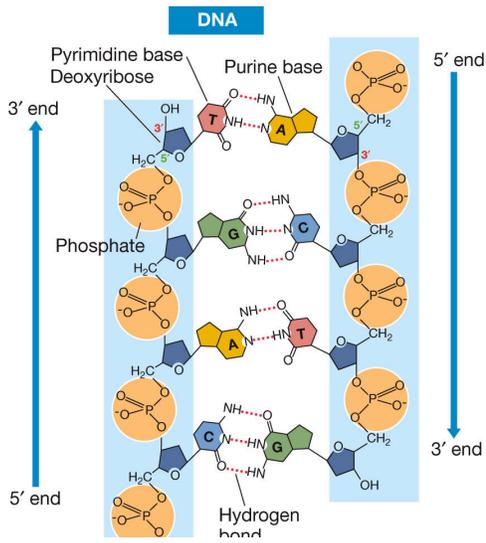
B 
 Base Pairs

A double strand of DNA. The bottom strand is the complement of the top strand.



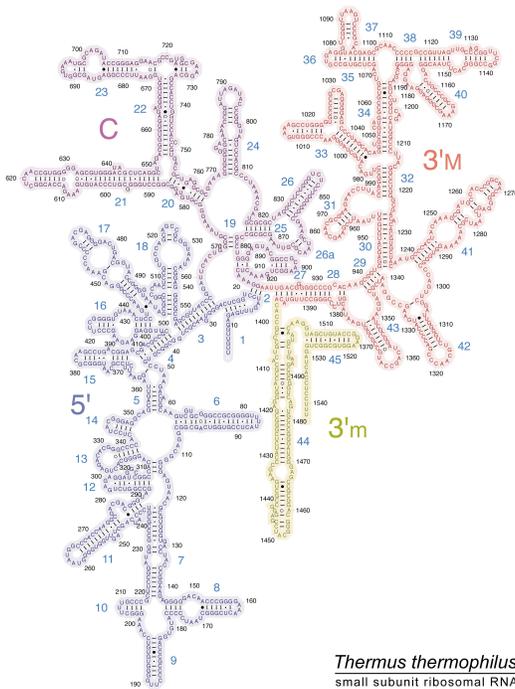
The two strands are held together by hydrogen bonding between the A/T base pairs and the G/C base pairs.

How is DNA and RNA similar? Different?

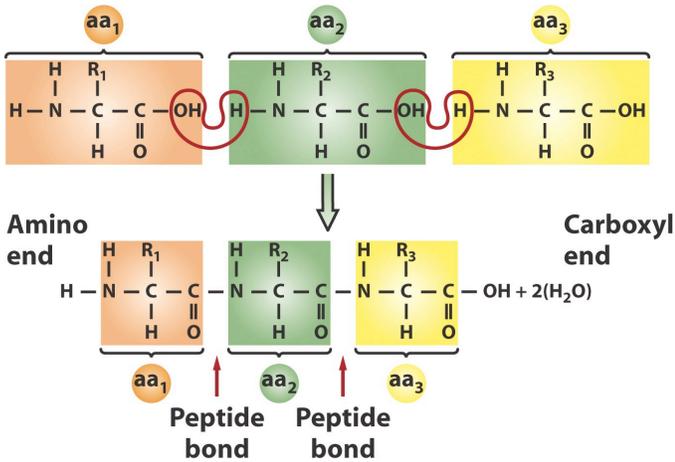


part 1)

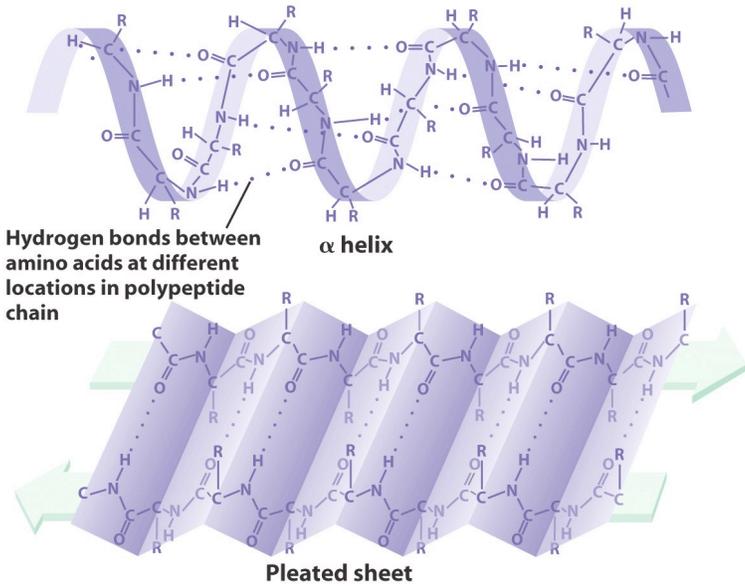
LIFE: THE SCIENCE OF BIOLOGY, Eighth Edition © 2007 Sinauer Associates, Inc. and W. H. Freeman & Co.



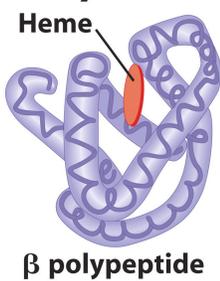
How does the amino acid sequence determine protein structure? 4



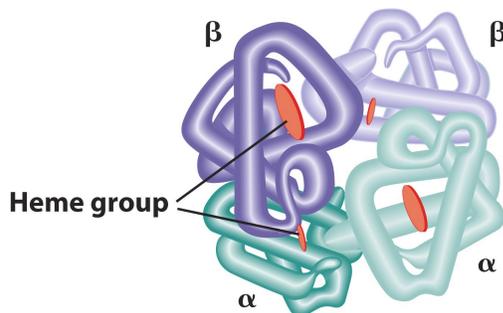
Secondary structure



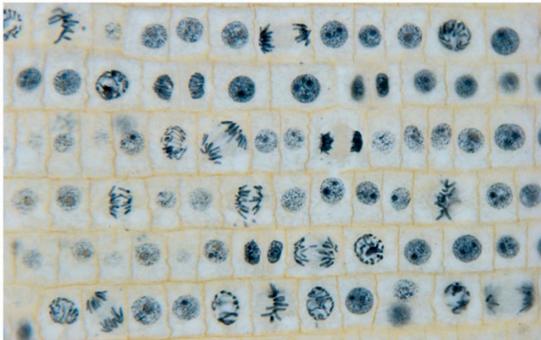
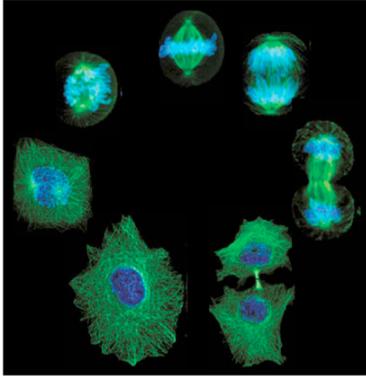
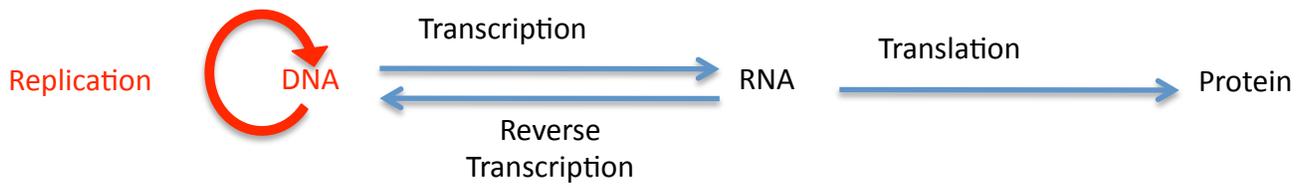
Tertiary structure



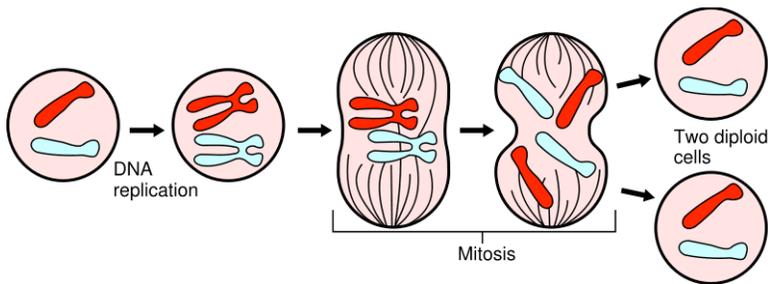
Quaternary structure

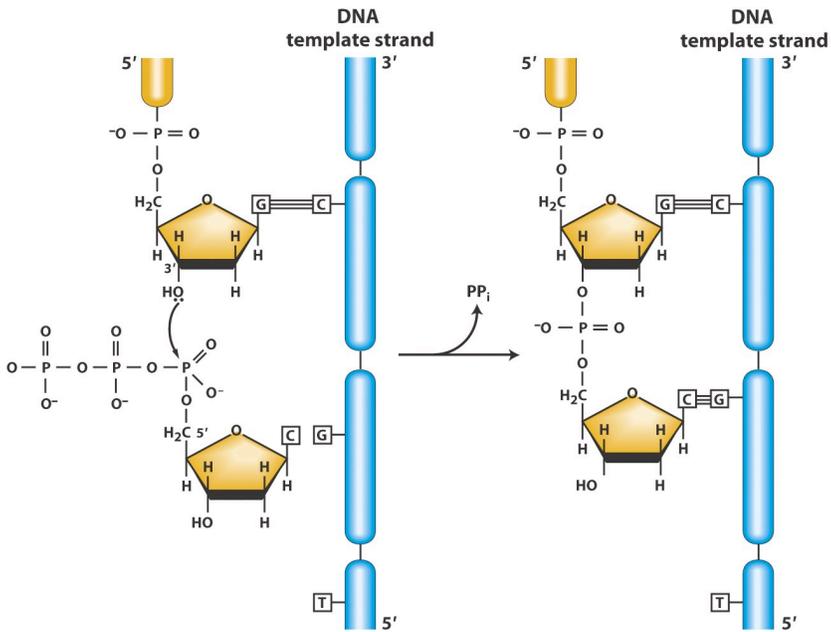
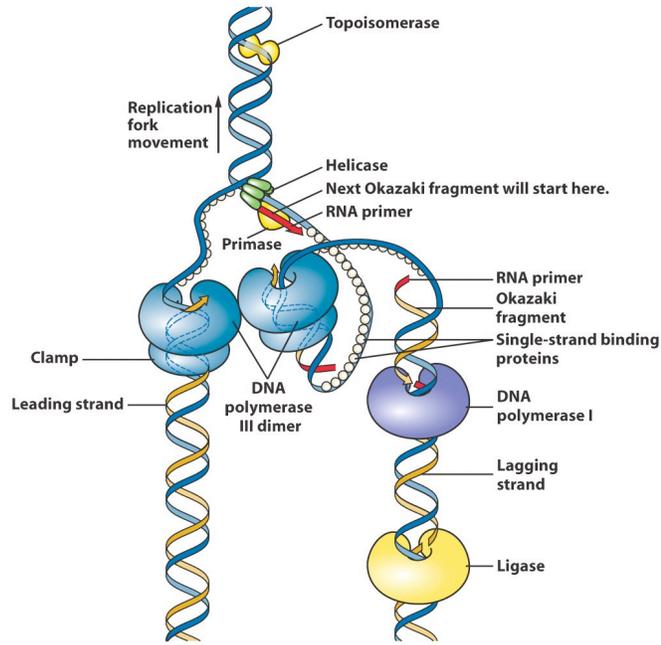
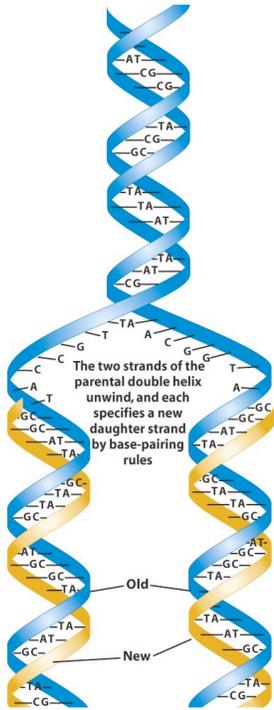
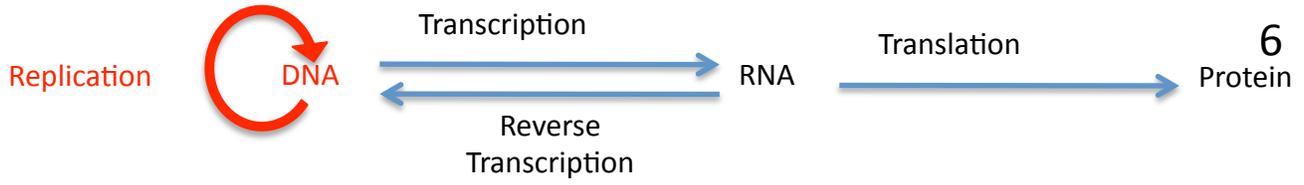


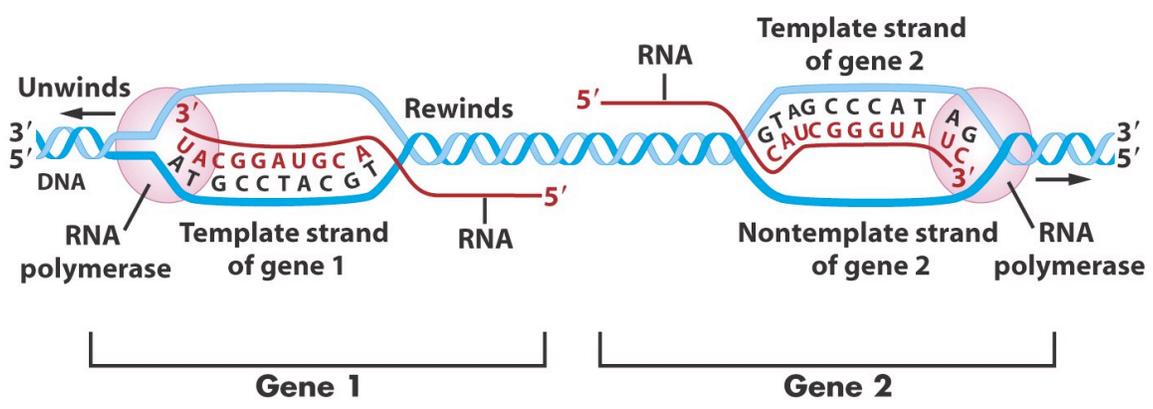
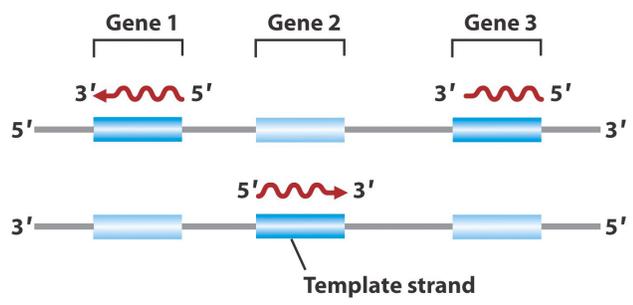
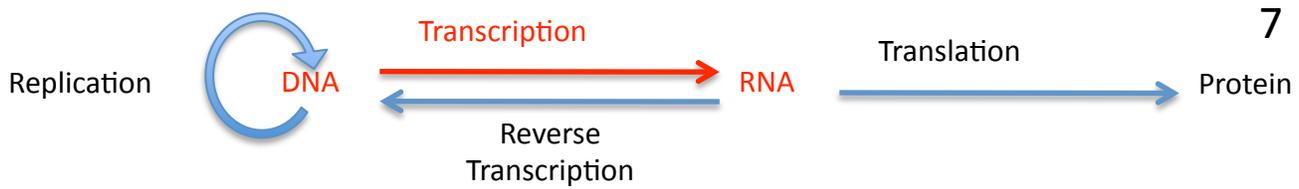
How is information transferred in cells?



Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.



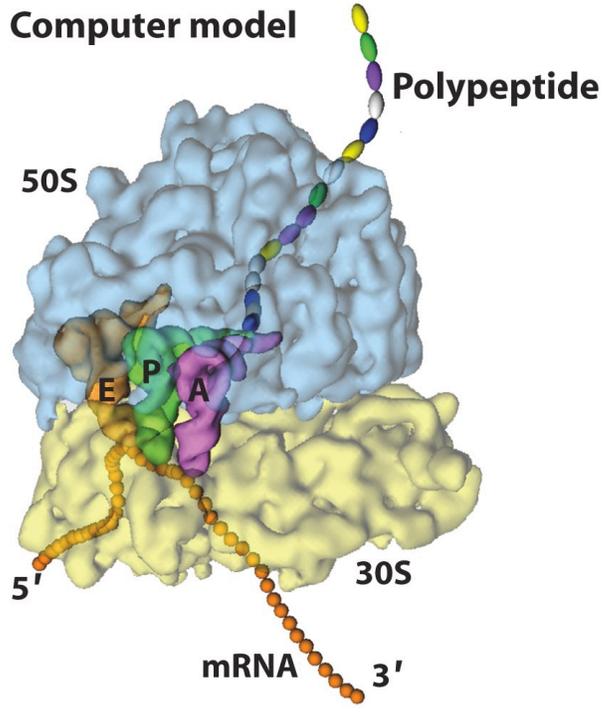
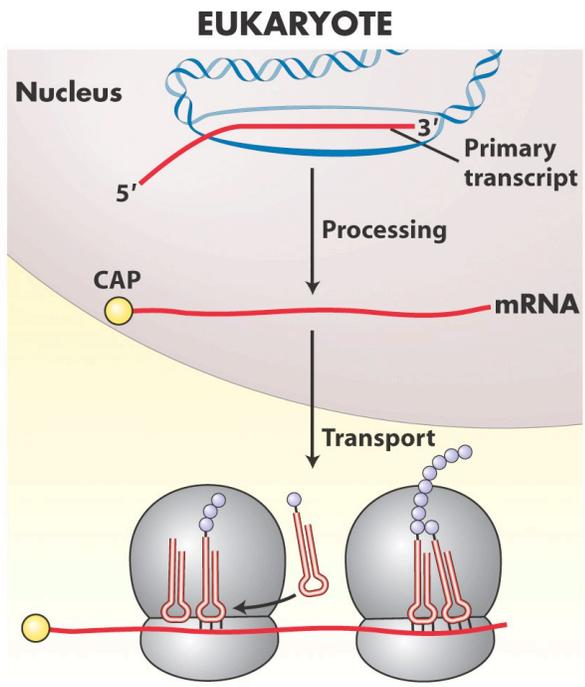
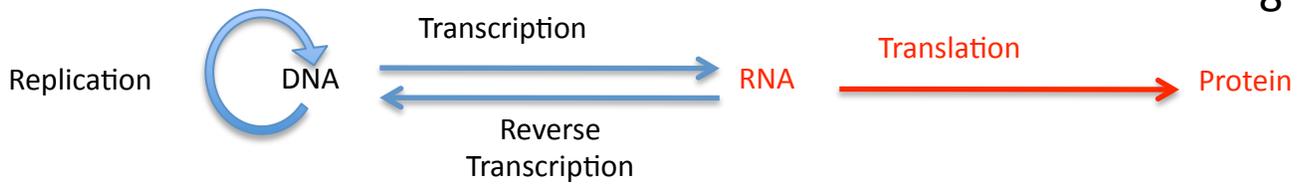


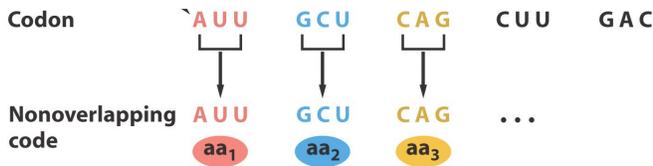
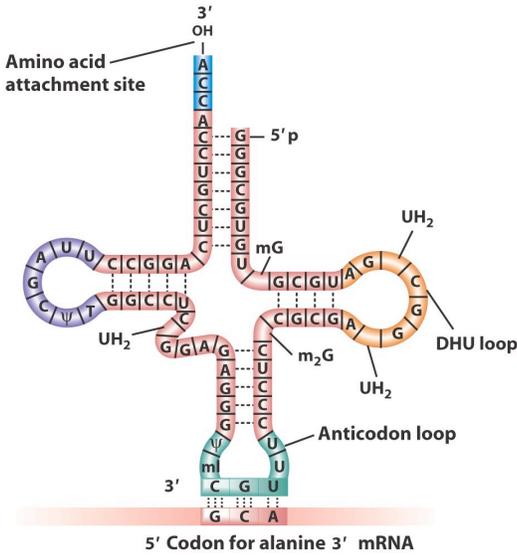
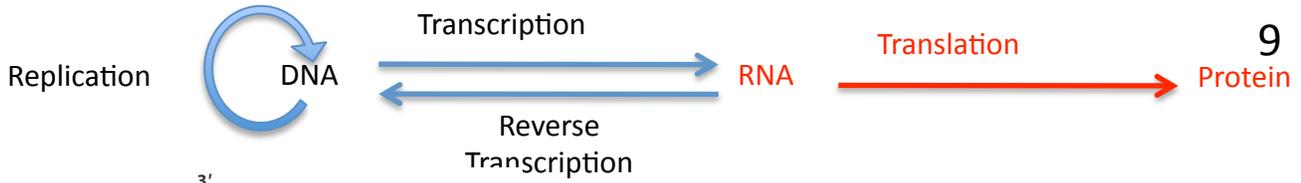


Coding strand 5' — CTGCCATTGTCAGACATGTATACCCCGTACGCTTCCCGAGCGAAAACGATCTGCGCTGC — 3' } DNA

Template strand 3' — GACGGTAAACAGTCTGTACATATGGGGCATGCAGAAGGGCTCGCTTTTGCTAGACGCGACG — 5' }

5' — CUGCCAUUGUCAGACAUGUAUACCCCGUACGUCUUCCCGAGCGAAAACGAUCUGCGCUGC — 3' mRNA

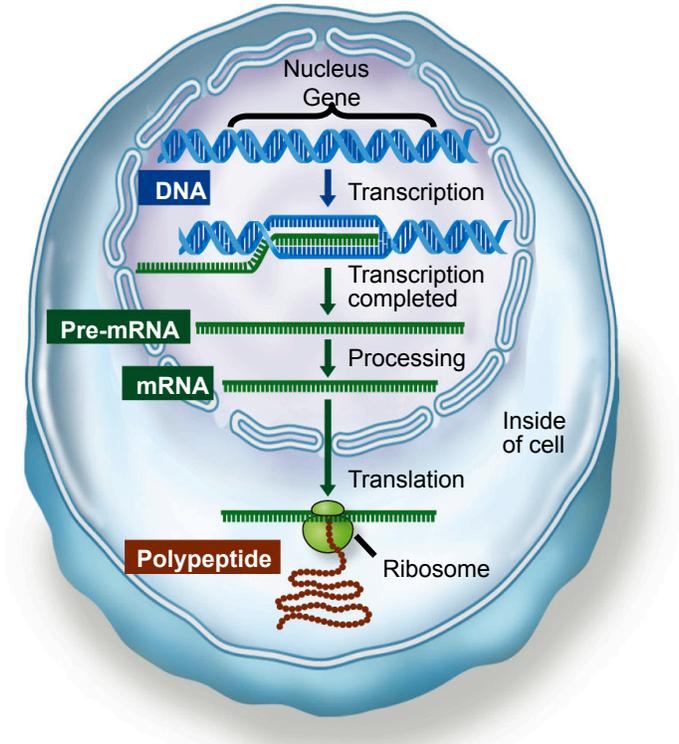
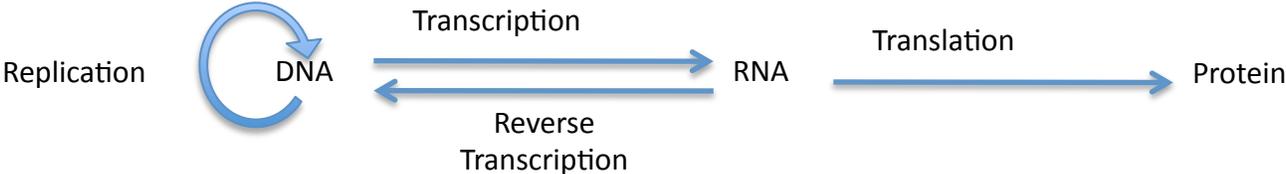




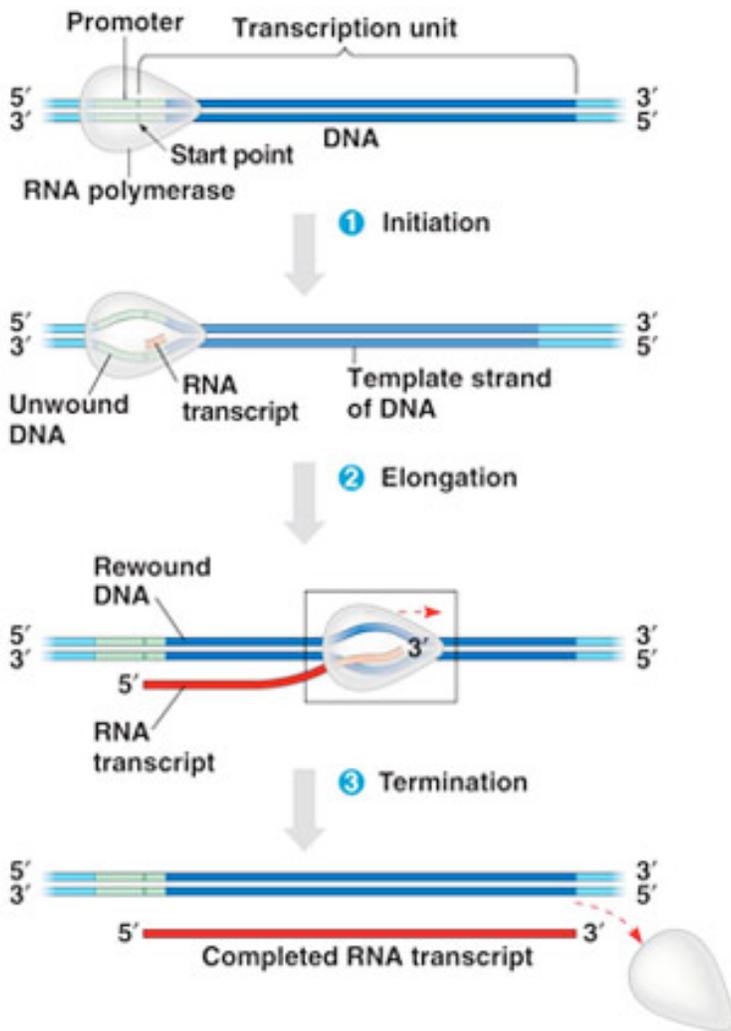
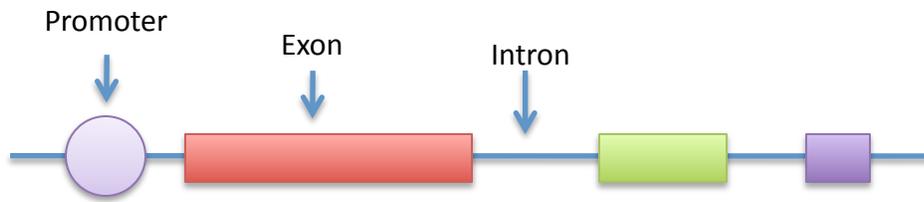
		Second letter				
		U	C	A	G	
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } Ser UCC } UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U C A G
	C	CUU } Leu CUC } CUA } CUG }	CCU } Pro CCC } CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } Arg CGC } CGA } CGG }	U C A G
	A	AUU } Ile AUC } AUA } AUG Met	ACU } Thr ACC } ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
	G	GUU } Val GUC } GUA } GUG }	GCU } Ala GCC } GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } Gly GGC } GGA } GGG }	U C A G

Translate this mRNA sequence
AUG GAA CUA GUA AUC UCU AUU UCG GAU GAG GCG GAU UGA

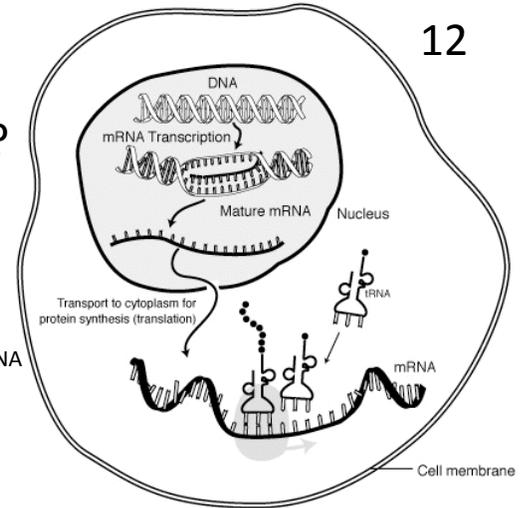
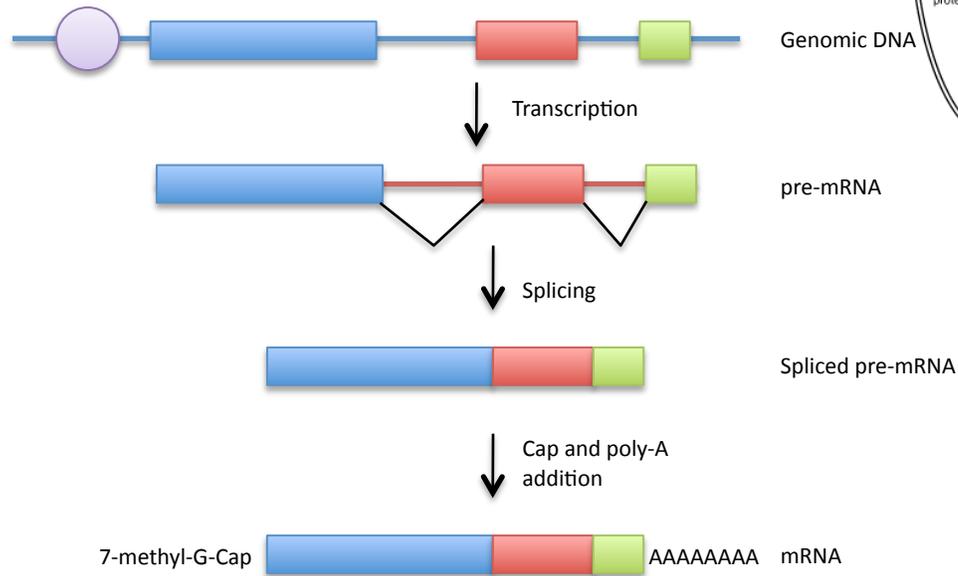
Putting it all together....



What are the components of a gene?

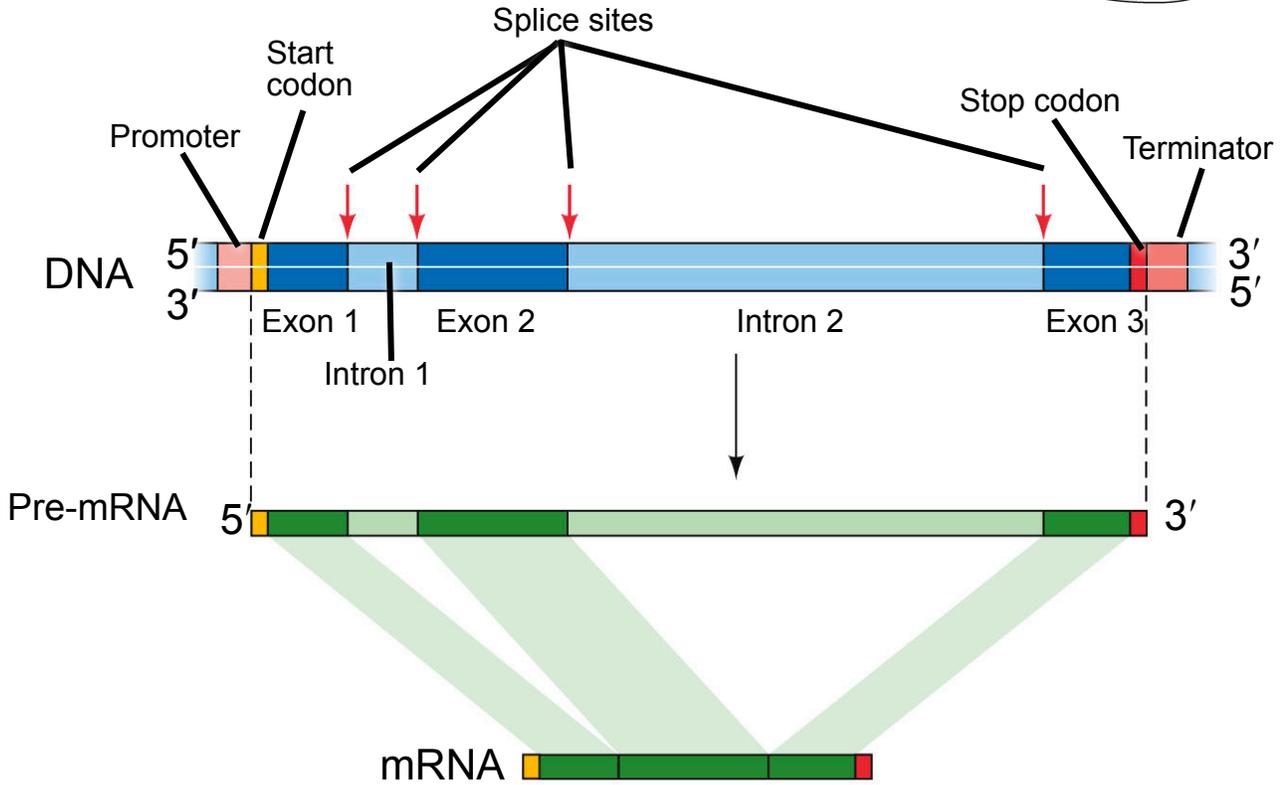
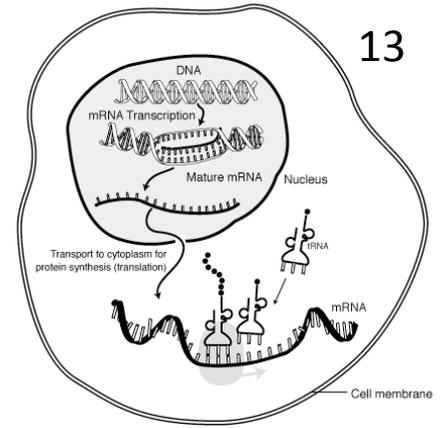


What are the components of a mRNA?
How is mRNA processed from pre-mRNA?



What are the components of a mRNA?
How is mRNA processed from pre-mRNA?

Details of pre-mRNA and mRNA



Experiment I: Analysis of the Actin gene of maize.

You will use the actin gene to investigate the differences between the DNA gene sequence and mRNA sequence. This exercise will also demonstrate how gene structure (exon and introns) is determined experimentally. The process is one step in genome annotation or giving meaning to the billions of nucleotides that make up a genome.

The actin gene encodes the actin protein and is found in all eukaryotes. Actin polymerizes to forms long strands that can contract and plays a large role in cell division. Because the role of actin is so important the chromosomal location of the gene is known in many organisms. This makes it a great gene to use for learning about gene structure and as a control for many experiments. While planning an experiment it is important to research what is already know about the subject. To do this you need to do literature search. Fortunately for a biologist this does not mean going to the library! We simply go online to the National Center for Biotechnology Information (NCBI) that is a unit of the National Library of Medicine (NLM) which is an institute in the National Institutes of Health (NIH) all funded by the US government.

Using the NCBI Website

The NCBI website is a portal for a lot of information including a literature reference (PubMed), a repository for free journal articles (PubMed Central), repository for DNA (GenBank) and protein sequences, and well as databases for the general public (DailyMed and MedlinePlus). The site also provides tools for accessing the information in many ways. Throughout Experiment 1 you will learn parts of the NCBI website.

To do a literature search

1. Go to www.ncbi.nlm.nih.gov to open the NCBI website. The page (currently) looks like this:

The screenshot shows the NCBI homepage with the following elements:

- Resources (Left Sidebar):** A vertical list of categories including NCBI Home, All Resources (A-Z), Literature, DNA & RNA, Proteins, Sequence Analysis, Genes & Expression, Genomes, Maps & Markers, Domains & Structures, Genetics & Medicine, Taxonomy, Data & Software, Training & Tutorials, Homology, Small Molecules, and Variation.
- Welcome to NCBI (Center):** A central banner with the text: "The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information." Below it, a "Genome" section states: "1000 prokaryotic genomes are now completed and available in the Genome database." A "How To..." section lists tasks like "Obtain the full text of an article" and "Retrieve all sequences for an organism or taxon".
- Popular Resources (Right):** A list of links including PubMed (circled in red), PubMed Central, Bookshelf, BLAST, Gene, Nucleotide, Protein, GEO, Conserved Domains, Structure, and PubChem.
- NCBI News (Right):** A section for news updates, including "November and October News" and "NCBI News - September 2009".

2. Click on the PubMed link (circled in red above). The PubMed homepage (currently) looks like this:

The screenshot shows the PubMed homepage with the following elements:

- Search Bar:** Located at the top, containing the text "actin" (circled in red). It includes a "Search" button and a "Clear" button.
- Welcome to PubMed (Banner):** A central banner with the text: "PubMed comprises more than 19 million citations for biomedical articles from MEDLINE and life science journals. Citations may include links to full-text articles from PubMed Central or publisher web sites."
- Using PubMed (Left Column):** A list of links including PubMed Quick Start, New and Noteworthy, PubMed Tutorials, Full Text Articles, and PubMed FAQs.
- PubMed Tools (Middle Column):** A list of links including Single Citation Matcher, Batch Citation Matcher, Clinical Queries, and Topic-Specific Queries.
- More Resources (Right Column):** A list of links including MeSH Database, Journals Database, Clinical Trials, E-Utilities, and LinkOut.

3. Type actin in the search box and click Search. Parts of this page will be explained in class.

The screenshot shows the PubMed website interface. At the top, there is a navigation bar with 'NCBI Resources' and 'How To' dropdown menus. The main header features the 'PubMed.gov' logo and the text 'U.S. National Library of Medicine National Institutes of Health'. A search bar is located in the top right, with 'PubMed' in the dropdown menu and 'actin' entered in the input field. To the right of the search bar are links for 'RSS', 'Save search', 'Advanced search', and 'Help'. Below the search bar, there are 'Search' and 'Clear' buttons. Underneath the search bar, the 'Display Settings' are shown as 'Summary, 20 per page, Sorted by Recently Added'. The main content area displays 'Results: 1 to 20 of 70512'. Three search results are visible, each with a checkbox, a title link, authors, journal information, and PMID. The first result is 'Isoform-Specific Contributions of alpha-Actinin to Glioma Cell Mechanobiology' by Sen S, Dong M, Kumar S. The second is 'PPARgamma agonists inhibit TGF-beta-PKA signaling in glomerulosclerosis' by Zou R, Xu G, Liu XC, Han M, Jiang JJ, Huang Q, He Y, Yao Y. The third is 'Cell-penetrating peptides with intracellular actin remodeling activity in malignant fibroblasts' by Delaroché D, Cantrelle FX, Subra F, Van Heijenoort C, Guittet E, Jiao CY, Blanchoin L, Chassaing G, Lavielle S, Auclair C, Sagan S.

As you can see there are over 70,000 articles that mention actin. Scroll down the page until you see a box on the right side called Search Details. There you will see the actual search used by PubMed: "actins"[MeSH Terms] OR "actins"[All Fields] OR "actin"[All Fields]. If a search does not produce what you expect it is a good idea to look at the Search Details to see what PubMed actually used.

4. We need to narrow the search so let's try "actin and maize." Now there are less than 200 articles.

The screenshot shows the PubMed website interface with the search term 'actin and maize'. The search bar now contains 'actin and maize'. The 'Display Settings' remain 'Summary, 20 per page, Sorted by Recently Added'. The main content area displays 'Results: 1 to 20 of 159'. One search result is visible, with a checkbox, a title link, authors, journal information, and PMID. The result is 'Live imaging of rapid chromosome movements in meiotic prophase I in maize' by Sheehan MJ, Pawlowski WP.

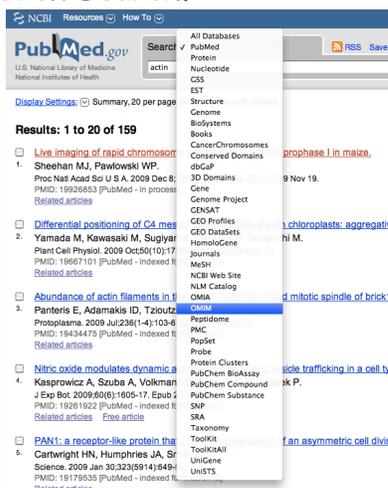
5. Click on the link of the first article. On this page you will see the abstract of the article as well as related articles. This is often the most useful way to find articles that you are really interested in.

The screenshot shows the PubMed website interface. At the top, there is a search bar with 'PubMed' entered and a 'Search' button. Below the search bar, the article title 'Live imaging of rapid chromosome movements in meiotic prophase I in maize.' is displayed, along with the authors 'Sheehan MJ, Pawlowski WP.' and their affiliation 'Department of Plant Breeding and Genetics, Cornell University, Ithaca, NY 14853.' The abstract text follows, describing the study of chromosome dynamics in maize. On the right side, there is a 'Full Text Article at www.pnas.org' link circled in red, and a 'Related articles' section with several links to other articles. At the bottom, there is a 'Recent activity' section with 'Turn Off' and 'Clear' buttons.

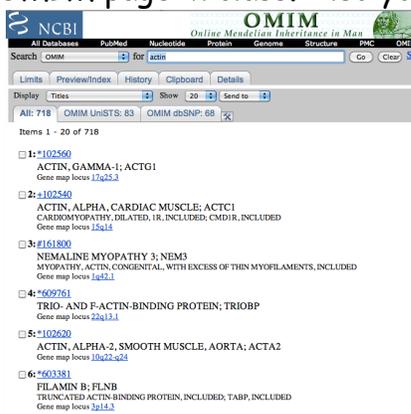
6. On the upper right hand side you will see an icon for the journal's webpage where you can download the article. While PubMed (and everything on the NCBI website) is freely available everywhere you may have to be on UGA's campus network to access the actual article. If the article is also in PubMed Central then you can download the full article from any Internet connection. A PubMed Central icon will appear next to the journal's icon if the article is in the NCBI database.

Since actin is found in all eukaryotes you might be interested to find out whether actin is associated with any diseases of animals or humans. NCBI has two hand curated databases that contain such information. To search for human diseases you use OMIM or Online Medelian Inheritance in Man and for animals you use OMIA.

7. Type actin into the search field and select OMIM from the pop up menu. Click Search.



The results of the OMIM search will look similar to this. We will discuss the OMIM page in class. Also you can do a similar search with OMIA.



As you can easily see, NCBI and PubMed make it possible to do all the background research you need for a biomedical and most plant topics without leaving the comfort of your own bed. Imagine how useful this would be writing a paper for class or researching a disease for medical school! A word of caution about NCBI: You may have noticed that the look and feel of the NCBI website changes depending on the database you are using. The home page and PubMed recently got a face lift. The other sections of NCBI will in time get the same treatment. So while the screen shots in the these pages may change, the basics of finding information on NCBI will remain the same.

Back to the Experiment...

You will analyze the sequence of the maize actin genomic region and the actin mRNA sequence. What differences do you expect to find? The sequence of mRNA cannot be determined directly because there are no convenient techniques to determine RNA sequence. Instead, DNA is synthesized from RNA templates. This is done by using an enzyme called reverse transcriptase (RT for short) that catalyzes the synthesis of DNA from RNA (called reverse transcription, can you guess why?). DNA synthesized in this way is called complementary DNA or cDNA. cDNA synthesis is described in detail below. To obtain the sequence of a molecule of DNA you must first create sufficient quantities of just the region you are interested in. To do this you use a technique called the Polymerase Chain Reaction (PCR). Once you have enough DNA for sequencing it will be shipped off to a company called Genewiz that will sequence the DNA. You can then compare the sequences of the genomic DNA and the cDNA to determine the gene structure and compare that to the structure predicted in the fully sequenced maize genome.

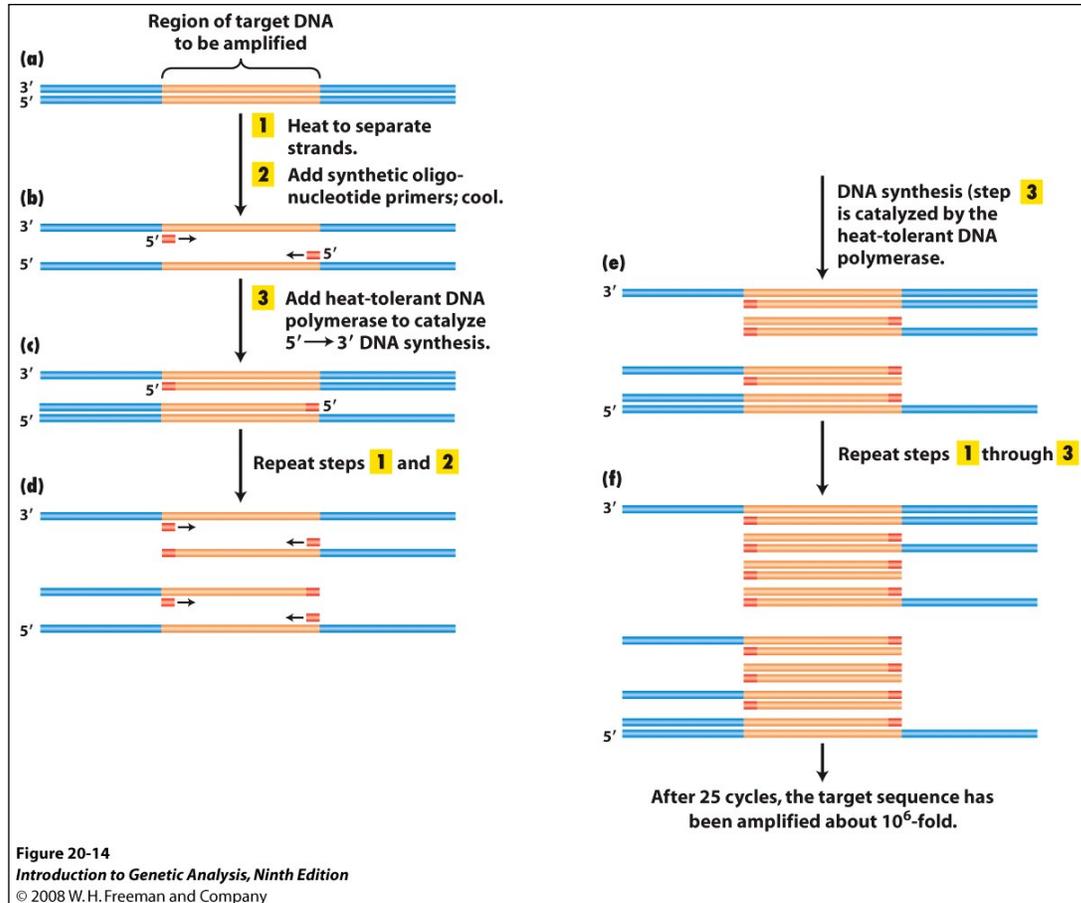


Figure 1:
Details of PCR.
See below.

Polymerase Chain Reaction. Better known by its initials - PCR: a technique enabling multiple copies to be made of sections of DNA molecules. It allows isolation and amplification of such sections from large heterogeneous mixtures of DNA such as whole chromosomes and has many diagnostic applications, for example in detecting genetic mutations and viral infections. The technique has revolutionized many areas of molecular biology—and won a Nobel Prize for Kary Mullis.

The reaction starts with a double-stranded DNA fragment. A part of it is to be amplified (see Figure 1).

Denaturation: A to B. The two DNA strands are separated (denatured) by heating to 95°Celsius (C).

Annealing: B. After cooling, short oligonucleotide primers (see below) that are complementary to the ends of the region to be amplified anneal with each strand.

Extension: C. When the temperature is raised to 72° C the DNA polymerase (the heat-stable *Taq* polymerase) begins to catalyze DNA synthesis from the ends of the primer using the denatured DNA as template (the extension

phase) and the nucleotide triphosphates (A, G, C, and T -collectively called dNTPs for deoxyNTPs) that are in the test tube.

D, E and F - The procedure is repeated (for many cycles) beginning with denaturation then annealing, extension etc.

Oligonucleotide primer. A primer is a short nucleic acid strand that serves as a starting point for DNA replication. A primer is required because most DNA polymerases (enzymes that catalyze the replication of DNA), cannot copy one strand into another from scratch, but can only add to an existing strand of nucleotides. (Recall from your lecture courses that in most natural DNA replication, the ultimate primer for DNA synthesis is a short strand of RNA. This RNA is produced by primase, and is later removed and replaced with DNA by a DNA polymerase.) The primers used for PCR are usually short, chemically synthesized DNA molecules with a length of about 20-30 nucleotides.

Denaturation: separation of the two DNA strands of a double helix by heating them to a very high temperature. This breaks the hydrogen bonds holding the double helix together.

Annealing: when DNA or RNA strands pair by hydrogen bonds to complementary strands, forming a double-stranded molecule. The term is also used to describe the reformation (renaturation) of complementary strands that were separated by heat.

Extension: enzymatically extending the primer sequence—copying DNA.

Watch this animation to help you understand how PCR works:

<http://www.dnalc.org/ddnalc/resources/pcr.html>

The PCR products are analyzed by gel electrophoresis. Watch this animation to learn how this technique works.

<http://www.dnalc.org/ddnalc/resources/electrophoresis.html>

cDNA Synthesis

The polymerases used for PCR require a DNA template. That means we cannot use mRNA directly in a PCR reaction. We must extract RNA from a tissue and reverse transcribe the RNA into DNA using the enzyme reverse transcriptase (RT). RT creates the DNA complement of every mRNA strand in the reaction. This DNA strand is referred to as cDNA. cDNA can then be used in a PCR sample as the template for Taq polymerase.

cDNA is made from mRNA using RT, dNTPs, and a primer that binds to the poly-A tail of the mRNA. The DNA primer is a short DNA strand that is fifteen T bases in a row and is called oligo-dT. The oligo-dT primer binds to the poly-A tail and the RT extends the oligo-dT primer to create a strand of DNA that is complementary to the mRNA strand.

Due to time constraints you will not isolate RNA and make cDNA for this first experiment. Instead you will perform two different PCRs - (i) you isolate maize genomic DNA in class and, using specific primers, amplify the actin gene and (ii) you will amplify the actin cDNA from cDNA made by the instructors. You will then analyze the sizes of the PCR products from (i) and (ii). Finally, you will analyze the sequence of the PCR products to determine the exon/intron boundaries of the first intron of the actin gene. This will provide you with an example of how the gene structure of any gene can and often is determined.

Step I: DNA Extraction

You will extract DNA from two maize inbred lines. Be sure to store the DNA in the freezer because you will use it through out the semester.

This protocol should be written up in your lab notebook. You will use your lab notebook in lab, not the printed course book.

Damon Lisch's All Natural Genomic Miniprep

Materials list:

Extraction Buffer

RNase A (500 μ g/ml)

10% SDS

5M KOAC

100% Isopropanol

70% Ethanol

Ice Bucket with ice

liquid nitrogen

37 $^{\circ}$ C water bath

65 $^{\circ}$ C water bath

sterile 1.5 ml tubes (2 for each prep)

- 1) Label 2 tubes for each plant with plant name and your initials.
- 2) Harvest a piece of maize leaf about the length of your hand. Rip it into pieces small enough to fit in the mortar. Ask for liquid nitrogen to be put in the mortar. Grind vigorously with the pestle.
- 3) Add 1 ml of Extraction Buffer, and grind some more in the buffer. Pour the slurry into the appropriately labeled tube.
- 4) Add 0.5 μ l RNase A to the tube. Use only the pipette labeled for RNase A use. Incubate for 20 minutes at 65 $^{\circ}$ C. Invert to mix. RNase A is an enzyme that degrades RNA strands into single bases. Repeat steps 1-4 for the next sample.
- 5) Add 120 μ l of 10% SDS. Mix by inverting.

6) Incubate at 65°C for 10 minutes.

7) Add 300 µl 5M KOAc. Mix well by inverting several times (important!), then incubate on ice for 10 minutes.

8) Spin for 5 minutes at top speed in microfuge. Squirt 700 µl of the supernatant through miracloth into the second tube. (make small funnel, place tip directly onto the miracloth at the tip of the funnel and squirt through - do not allow the whole funnel to get soaked).

9) Add 600 µl of isopropanol. Mix the contents thoroughly by inverting.

DNA precipitate may or may not be visible at this point; don't worry if you don't see much. However, a really good prep (excellent grinding of tissue) should result in visible DNA at this stage.

10) Spin for 5 minutes at top speed. Pipette off supernatant.

11) Add 500 µl 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.

12) Resuspend the DNA in 50 µl water.

Store your DNA samples in a box with your name the freezer (-20°C).

Visualize genomic DNA on a 1.5% agarose gel:

A 1.5% agarose gel contains 1.5 grams in 100 ml of gel buffer TAE (1.5/100 x 100% = 1.5%).

1. Weigh out 1.5g agarose and add to a 250 ml flask.

2. Add 100 ml 1X TAE buffer (available in a big jug) to the flask with agarose. (TAE = 40mM Tris acetate, 1mM EDTA pH 8.4)

3. Heat contents in the microwave until boiling (2-3 min). Be very careful, as superheated liquids can boil over and burn you.

4. Swirl to make sure that the agarose is completely melted.

5. Add 7.0 μ l of a stock solution of 10 mg/ml ethidium bromide (EtBr). (This binds to the DNA allowing it to be visualized under UV light. Do not let this stuff touch your skin.)

6. Swirl again to mix and pour into a gel-casting stand with a comb (this will be demonstrated in the lab.) The gel should cool and solidify within 10-15 minutes at which time it is ready to place the gel in the electrophoresis apparatus and add enough TAE buffer to completely immerse the gel.

After the gel solidifies

7) Put 10 μ l of DNA into a tube.

8) Add 2 μ l of 6x loading dye (blue dye) to the tube. Tap the tube gently with your finger to mix.

9) Load all 12 μ l on your gel. Keep track of which sample went in which lane.

10) Load 7 μ l of DNA Ladder in one empty well.

11) Run the gel at 130 Volts for 30 minutes.

12) Photograph the gel.

Determine the DNA concentration

DNA concentration is determined by measuring the amount of UV light (260 nm) absorbed by the DNA. The absorbance is converted into concentration by using the relationship that an absorbance reading of 1 corresponds to a concentration of 50 μ g/ml of pure DNA. So a solution of DNA that has an absorbance of $A_{260} = 0.5$ has a concentration of 25 μ g/ml ($50 \times 0.5 = 25$). What is the concentration in μ g/ μ l and ng/ μ l?

Other molecules also absorb at 260 nm and inflate the DNA concentration value. You can also use the Nanodrop spectrometer to determine the purity of the DNA solution by taking readings at 280 (protein) 230 (EDTA, carbohydrates, and phenol) and 320 (this is visible light and measures particulates). Purity is determined by taking the ratio of the A260 value to the A280 or A320. A260/280 ratio for pure DNA is 1.8. Your values should be between 1.7 and 2. The A260/A320 ratio for pure DNA is between 2.0 and 2.2.

The Nanodrop spectrophotometer will scan the DNA sample from 200 nm to 400 nm and report values for A230, 260, 280, 320, the concentration, and ratios. You must record these numbers in your notebook. You will be shown how to use the Nanodrop in class.

Dilute the DNA sample

You need to add between 50 and 100 ng of DNA into each PCR. You will most likely need to dilute your DNA sample. A good concentration to use is 20 ng/ μ l. Calculating dilutions is an important skill that you need to know for lab. You should remember from chemistry that a dilution is made using the formula $c_i v_i = c_f v_f$. In biology lab we re-arrange the equation like this:

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

So if you have a DNA concentration of 500 ng/ μ l and you want to make 100 μ l of solution that is 20 ng/ μ l you set up the formula this way

$$\frac{20 \text{ ng}/\mu\text{l}}{500 \text{ ng}/\mu\text{l}} \times 100 \mu\text{l} = 4 \mu\text{l}$$

In a labeled tube you would place 96 μ l of sterile water and 4 μ l of the concentrated DNA solution.

How many microliters of the diluted DNA do you need for 50 ng?

Step II: Amplify DNA using PCR.

Each person will perform PCR using as template the B73 genomic DNA and one inbred line you isolated in class and the cDNA provided by the instructor. We also need an additional negative control that will be water in place of DNA, giving you five reactions. Instead of mixing the PCR reagents separately for each reaction we first make a cocktail of all reagents in common (Taq, dNTPs, Primers, and water). To ensure we have enough cocktail for five reactions we add one extra.

Set up PCR

1) Label a 1.5 ml tube. This is for making the PCR mix.

2) Label a strip of 0.2 ml PCR tubes. The instructors will show you where to label the tubes. The label may be rubbed off in the machine if you put it in the wrong place.

3) Mix the following in your tube using the volumes in the column labeled x4. Keep this on ice. You will need to calculate the number of μl of DNA to add and then adjust the volume of water.

	x1 (μl)	X6
2x Master Mix	25.0	
H ₂ O		
Forward Primer	1.0	
Reverse Primer	1.0	
DNA 100 ng		-----
Total	50.0	

The 2x Master Mix is supplied by a company (NEB) and contains Taq enzyme, buffer, and deoxynucleotide triphosphates (dNTPs) in a 2x concentration. This means that it must be diluted by half for the working concentration (1x). This tube should be kept on ice to protect the enzyme from degradation.

4) Put 45.0 μl of master mix in 5 of the PCR tubes.

5) Add the DNA to the PCR tubes. Add 5 μl of sterile water in tube 5.

6) Seal the tubes tightly with a strip of caps. Keep PCR tubes on ice until everyone is finished.

After everyone is done, your samples will be placed in a thermocycler or 'PCR machine' and cycled with the following conditions:

1 cycle for:	initial denaturation	94°C	3 min
40 cycles for:	denaturation	94°C	30 sec
	annealing	50°C	30 sec
	extension	72°C	1 min

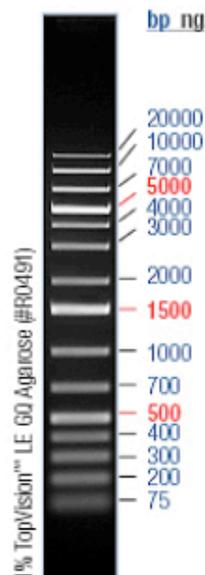
[Note: "40 cycles" means all steps— denaturation, annealing, and extension—are repeated 40 times before going on to the next step]

1 cycle for:	final extension:	72°C	10 minutes
--------------	------------------	------	------------

7. After you finished setting up the PCR, you should pour a 1.5% agarose gel. See page 20-21. You will pour one gel per group.

Step III. Gel analysis.

1. Pipette 20 μ l of each PCR reaction into a new tube. Add 4 μ l loading dye. Load 20 μ l into a gel lane. Remember to add a lane with 7 μ l of DNA ladder. Run the gel at 130 V for 30 minutes. The sizes of the DNA bands in the ladder are shown in the figure below.



Step IV. Prepare samples for sequencing

If there is only one band in a lane, the PCR sample can be used for sequencing after an enzymatic cleanup. Two enzymes are used: Exonuclease I that degrades the primers to single nucleotides and shrimp alkaline phosphatase that removes the phosphate group from unincorporated dNTPs. The mix is called ExoSAP-IT.

Keep the ExoSAP-IT on ice.

Instructions will be given in class.

How your DNA samples will be sequenced

DNA sequencing is the process of determining the nucleotide order of a given DNA fragment. Most DNA sequencing is currently being performed using the chain termination method developed by Frederick Sanger. [Sanger is particularly notable as the only person to win two Nobel prizes in chemistry - his second in 1980 for developing this DNA sequencing method and his first in 1958 for determining the first amino acid sequence of a protein (insulin)]. His technique involves the synthesis of copies of your input DNA by the enzyme DNA polymerase. However, one difference between this reaction and PCR, for example, is the use of modified nucleotide substrates (in addition to the normal nucleotides), which cause synthesis to stop whenever they are incorporated. Hence the name: "chain termination".

Chain terminator sequencing (Sanger sequencing)

Your samples were sent to Genewiz along with information about the sequencing primer to be used (recall that DNA polymerase needs a primer to start DNA synthesis of a template strand). The reaction contains your DNA sample, the sequencing primer, DNA polymerase and a mixture of the 4 deoxynucleotides that are "spiked" with a small amount of a chain terminating nucleotide (also called dideoxy nucleotides, see below).

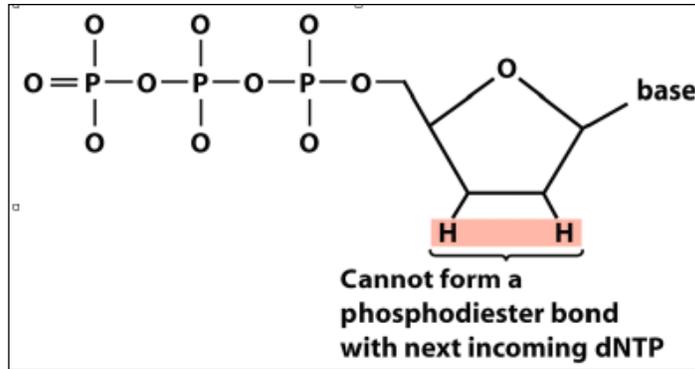


Figure 2. A chain-terminating nucleotide triphosphate (called a di-deoxynucleotide or ddNTP). Because it has a "H" instead of a "OH" at the 3' position, it is not a substrate for the addition of another NTP and DNA synthesis terminates.

Limited incorporation of the chain terminating nucleotide by the DNA polymerase results in a series of related DNA fragments that are terminated only at positions where that particular nucleotide is used.

ATGGGATAGCTAATTGTTTACCGCCGGAGCCA	3'	Template DNA clone
CGGCCTCGGT	5'	Primer for synthesis
←		
*ATGGCGGCCTCGGT	5'	Dideoxy fragment 1
*AATGGCGGCCTCGGT	5'	Dideoxy fragment 2
*AAATGGCGGCCTCGGT	5'	Dideoxy fragment 3
*ACAAATGGCGGCCTCGGT	5'	Dideoxy fragment 4
*AACAAATGGCGGCCTCGGT	5'	Dideoxy fragment 5
*ATTAACAAATGGCGGCCTCGGT	5'	Dideoxy fragment 6
*ATCGATTAAACAAATGGCGGCCTCGGT	5'	Dideoxy fragment 7
*ACCCTATCGATTAAACAAATGGCGGCCTCGGT	5'	Dideoxy fragment 8

The fragments are then size-separated by electrophoresis in a slab polyacrylamide gel, or more commonly now, in a narrow glass tube (capillary) filled with a viscous polymer.

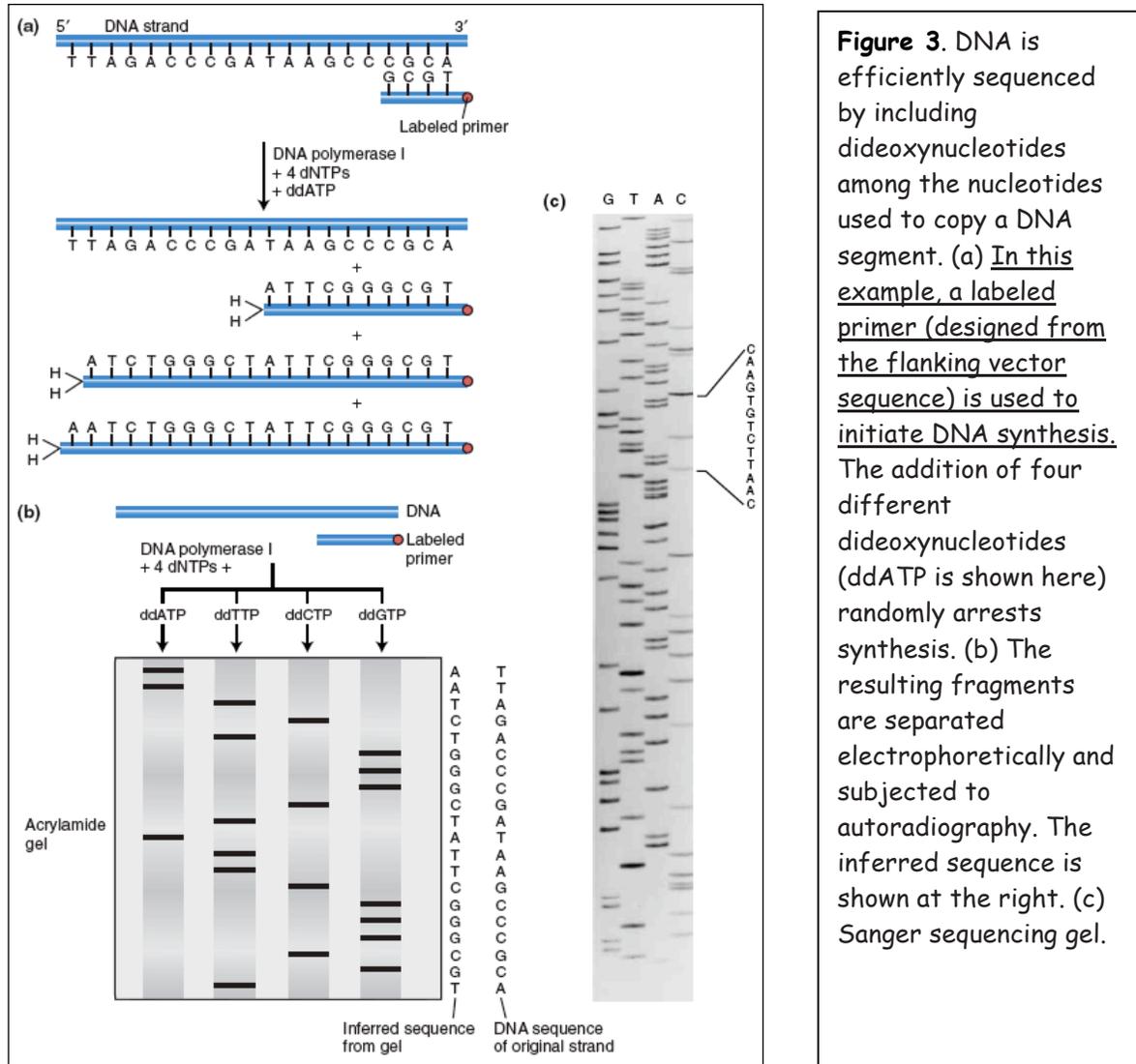


Figure 3. DNA is efficiently sequenced by including dideoxynucleotides among the nucleotides used to copy a DNA segment. (a) In this example, a labeled primer (designed from the flanking vector sequence) is used to initiate DNA synthesis. The addition of four different dideoxynucleotides (ddATP is shown here) randomly arrests synthesis. (b) The resulting fragments are separated electrophoretically and subjected to autoradiography. The inferred sequence is shown at the right. (c) Sanger sequencing gel.

Modifying DNA sequencing to automation: dye terminator sequencing (this is how your DNA samples will be sequenced)

An alternative to the labeling of the primer is to label the dideoxy nucleotides instead, commonly called 'dye terminator sequencing'. The major advantage of this approach is the complete sequencing set can be performed in a single reaction, rather than the four needed with the labeled-primer approach. This is accomplished by labeling each of the dideoxynucleotide chain-terminators with a separate fluorescent dye, which fluoresces at a different wavelength.

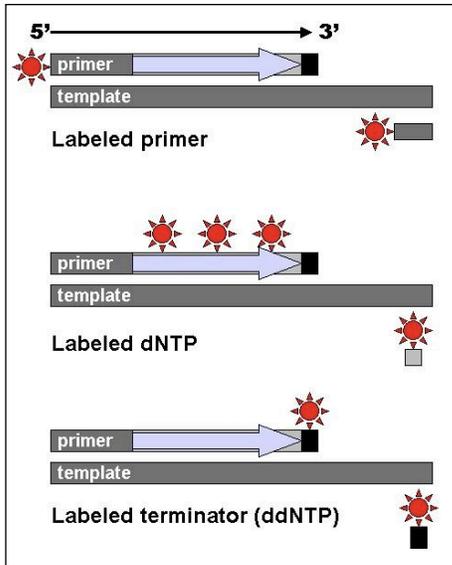


Figure 4: DNA fragments can be labeled by using a radioactive or fluorescent tag on the primer (1), in the new DNA strand with a labeled dNTP, or with a labeled ddNTP.

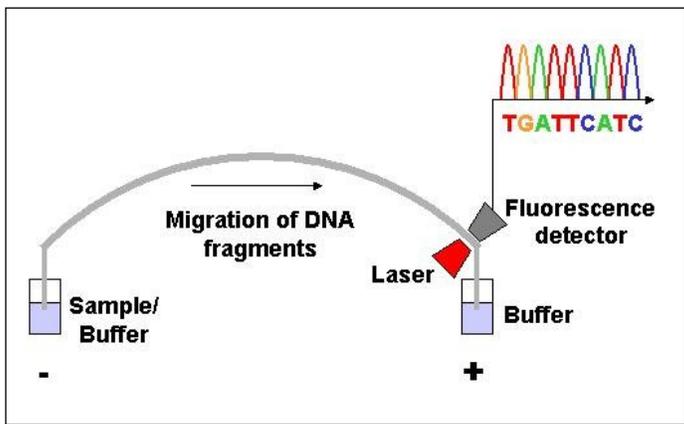


Figure 5. Modern automated DNA sequencing instruments (DNA sequencers) can sequence up to 384 fluorescently labelled samples in a single batch (run) and perform as many as 24 runs a day. However, automated DNA sequencers carry out only DNA size separation by capillary electrophoresis, detection and recording of dye fluorescence, and data output as fluorescent peak trace chromatograms.

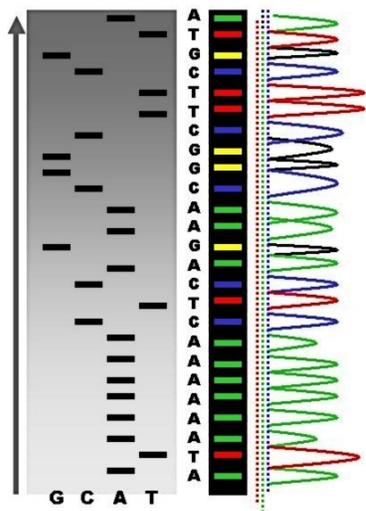


Figure 6: Sequence ladder by radioactive sequencing compared to fluorescent peaks

This method is now used for the vast majority of sequencing reactions, as it is both simpler and cheaper. The major reason for this is that the primers do not have to be separately labeled (which can be a significant expense for a single-use custom primer), although this is less of a concern with frequently used 'universal' primers.

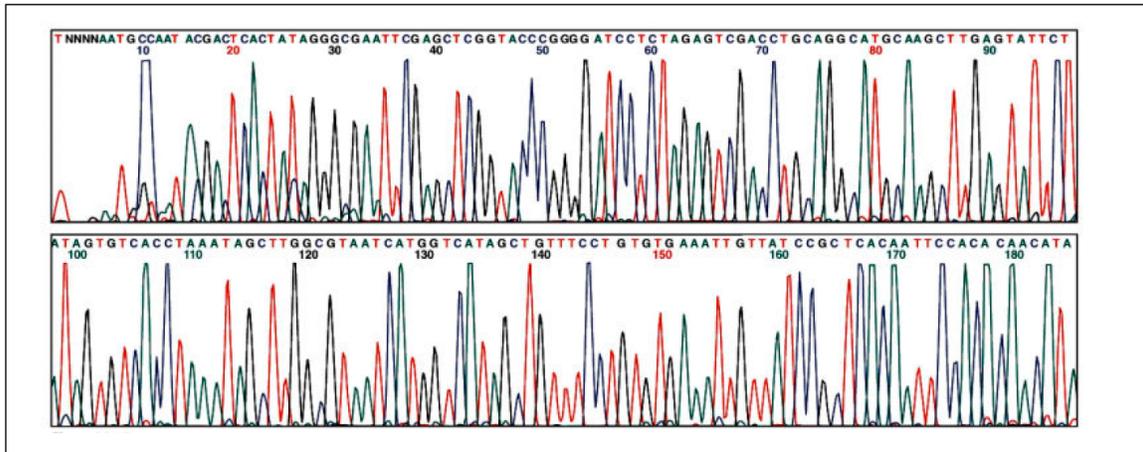


Figure 7. An example of a chromatogram file of a Sanger sequencing read. The four bases are detected using different fluorescent labels. These are detected and represented as 'peaks' of different colors, which can then be interpreted to determine the base sequence, shown at the top.

Step V. Sequence Analysis.

Now that you have DNA sequence from the PCR bands you need to analyze that sequence. We will do the following analyses:

- I. Verify the sequences are from the actin gene - Page 40.
Bioinformatics technique: blastn
- II. Compare the two sequences to each other - Page 44.
Bioinformatics technique: blast2sequences
- III. Compare the sequences to the maize genome - Page 49.
Bioinformatics technique: Genome Browsers
- IV. Compare the class generated sequences to each other - Page 59.
Bioinformatics technique: Multiple Sequence Alignment
- V. Find related sequences in the maize genome and other organisms - Page 72.
Bioinformatics technique: protein blast, tblastn, TARGeT

Before we analyze the sequence we need to obtain the sequence and also we need to create a way to document the annotation you are doing.

Obtain sequences from Genewiz Website.

Before you can analyze sequence you must obtain the sequence from the Genewiz website and check the quality.

1. Open the Genewiz website and login in:

<https://clims3.genewiz.com/default.aspx>

user: jburnette@plantbio.uga.edu

password: 1503A

2. Click the tracking number provided in class.

GENEWIZ DNA Sequencing Service - Online System

Welcome Jim Burnette

Place an Order

- Create Sequencing Order
- Create Primer Order
- Create Plasmid DNA Preparation Order

Pending Orders

No orders are found on this system.

Step 1: Please remember to print out your order receipt that has the barcode
Tracking number (optional)

Step 2: Send your samples along with the order receipt to GENEWIZ

Support

- File-Based Excel Order Form (Excel, .xls)
- File-Based Excel Order Form (Excel, .xlsx)
- Custom Excel Order Form (Excel, .xls)
- Sequence Annotation Guidelines
- GENEWIZ Universal Primers

Order Date	Tracking Number	Description	Count
09-09-2009	10-30827306		135
08-18-2009	10-10242355		15
07-14-2009	10-28981258		3
07-09-2009	10-28853308		83

3. Find your samples on the spreadsheet. I will explain the naming in class.

4. Find a sample where the QS (for quality score) and CRL (contiguous read length) are both in black. These numbers help you assess the quality of the sequence. We will discuss this in class.

5. Click on "View" in the Trace File column.

GENEWIZ DNA Sequencing Service - Online System

Welcome Jim Burnette

Order Results*

QS and CRL Description

Tracking Number: 10-30827306

Samples in the Order

Total number of free repeats: Free re

Number	Tube Label	QS	CRL	Failure Cause	Sample Name	SEQ File	Trace File	Seq File	Trace File	Download Seq Files	Download Trace Files	Repeat
1	JB01	25	553	Early Termination	JAN-A-M13R	View	View	Download	Download	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Free Repeat: Half price repeat with alternative protocol
2	JB02	37	601	Early Termination	JAN-B-M13R	View	View	Download	Download	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Free Repeat: Half price repeat with alternative protocol
3	JB03	39	693	Early Termination	JAN-C-M13R	View	View	Download	Download	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Free Repeat: Half price repeat with alternative protocol
4	JB04	43	949	Early Termination	JAN-D-M13R	View	View	Download	Download	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Free Repeat: Half price repeat with alternative protocol
5	JB05	36	544	Early Termination	JAN-E-M13R	View	View	Download	Download	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Free Repeat: Half price repeat with alternative protocol
6	JB06	35	582	Early Termination	JAN-F-M13R	View	View	Download	Download	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Free Repeat: Half price repeat with alternative protocol
7	JB07	21	77	Early Termination	JAN-G-M13R	View	View	Download	Download	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Free Repeat: Half price repeat with alternative protocol
8	JB08	45	962	Early Termination	JAN-H-M13R	View	View	Download	Download	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Free Repeat: Half price repeat with alternative protocol
9	JB09	0	0	No Priming	JAN-I-M13R	View	View	Download	Download	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Free Repeat: Half price repeat with alternative protocol

6. A new window will pop up. On the top will be the trace file and on the bottom will be the sequence. We will discuss how the trace file is used to assess the quality of the sequence.

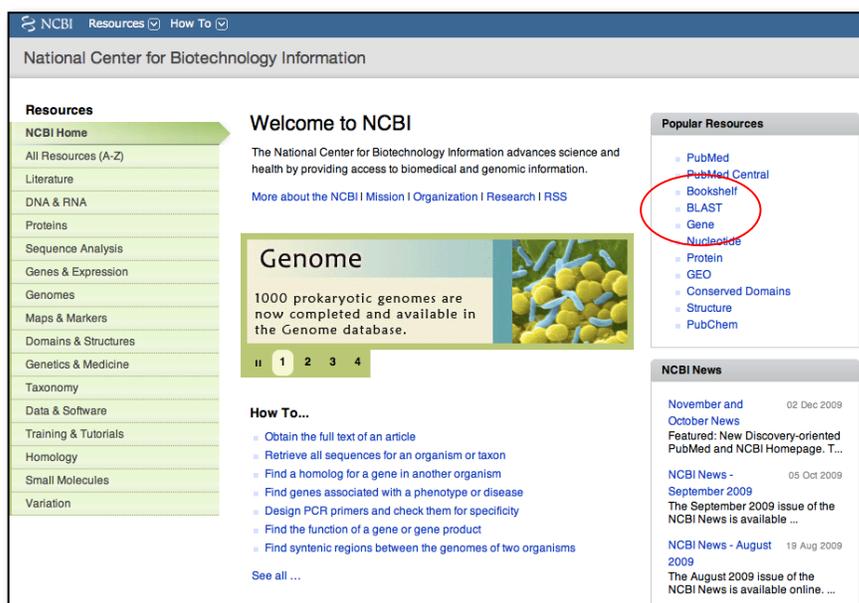
5. DNA and Protein sequence should be in FASTA format with the first line starting with ">" and containing the name of the image. The remaining lines are the sequence in fixed length. To clean up sequence use this web tool: target.iplantcollaborative.org/fasta_formatter.html. This tool is also useful if you need a subsequence from a longer sequence.

I. Verify the sequences are from the actin gene

PCR is a very useful technique, but sometimes it generates artifacts, that is random sequences unrelated to the sequence of interest. We first want to make sure that the sequence from the PCR bands is what we want to study. There are several ways to do this, but we will use Blast from NCBI to analyze the sequence.

Introduction to Blast:

You will use Blast a lot this semester. It is the major biological sequence search tool for DNA, RNA, and protein databases. Whole genomes can be searched using Blast. Access Blast by clicking on the Blast link on the NCBI home page (<http://www.ncbi.nlm.nih.gov/>).



The screenshot shows the NCBI homepage with a navigation menu on the left and several content sections. The 'Popular Resources' section on the right contains a list of links: PubMed, PubMed Central, Bookshelf, BLAST, Gene, Nucleotide, Protein, GEO, Conserved Domains, Structure, and PubChem. The 'BLAST' link is circled in red. Below this, the 'NCBI News' section lists recent news items with dates.

The Blast link will take you to the Blast page and to the Basic Blast Menu which will also be used frequently in this course:

BLAST Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

NCBI/BLAST Home

BLAST finds regions of similarity between biological sequences. [more...](#)

[Learn more](#) about how to use the new BLAST design

BLAST Assembled Genomes

Choose a species genome to search, or [list all genomic BLAST databases](#).

- [Human](#)
- [Mouse](#)
- [Rat](#)
- [Arabidopsis thaliana](#)
- [Oryza sativa](#)
- [Bos taurus](#)
- [Danio rerio](#)
- [Drosophila melanogaster](#)
- [Gall](#)
- [Pan](#)
- [Micr](#)
- [Apis](#)

Basic BLAST

Choose a BLAST program to run.

- [nucleotide blast](#) Search a **nucleotide** database using a **nucleotide** query
Algorithms: blastn, megablast, discontinuous megablast
- [protein blast](#) Search **protein** database using a **protein** query
Algorithms: blastp, psi-blast, phi-blast
- [blastx](#) Search **protein** database using a **translated nucleotide** query
- [tblastn](#) Search **translated nucleotide** database using a **protein** query
- [tblastx](#) Search **translated nucleotide** database using a **translated nucleotide** query

There are six different versions of BLAST because you can use a nucleotide sequence or protein sequence to query nucleotide or protein databases. The different versions are summarized in the screenshot above. Today we will give nucleotide blast a test drive. We will discuss protein blast and tblastn later when we need to use them.

- A. **Nucleotide Blast:** This is the most straightforward type of search. You begin with a nucleotide sequence you want to know more about (the query) and "blast" it against a nucleotide database (the subject). You can learn a lot about your query sequence with a blast including:
- a. Are there publications that already report information about this sequence (have you been "scooped")?
 - b. Where is the sequence located in the genome (more on location in class)?
 - c. Is the sequence found in genomes of closely related organisms?
 - d. Does it code for an RNA and/or a protein? If so is anything known about its function?

1. Select 'nucleotide blast.' Cut and paste the following sequence in the Query text window (Enter accession number....). For the in class example we will use the actin genomic DNA sequence. You can use this sequence or your sequence.

>Actin Genomic Sequence

```
GTGACAATGGCACTGGAATGGTCAAGGTTGTTATCTCGTTCAGAAGTCTTTTTCAACAAA
GCAACTCTACTCCTGTGCCTAATTGTTGCTCAACTCCTCAATATTTACAGGCCGGTTTCG
CTGGTGATGATGCGCCAAGAGCTGTCTTCCCCAGCATTGTGGGAAGACCACGCCACACCG
GTGTCATGGTCGGCATGGGCCAAAAGGATGCCTACGTAGGTGATGAGGCTCAGGCCAAGA
GAGGCATCCTGACACTGAAGTACCCGATTGAGCATGGCATTGTCAACAACCTGGGATGACA
TGGAGAACTGGCATCACAC
```

The screenshot shows the NCBI BLAST Basic Local Alignment Search Tool interface. The 'Enter Query Sequence' field is populated with the actin genomic DNA sequence. The 'Job Title' field is set to 'Actin Genomic Sequence'. The 'Align two or more sequences' checkbox is unchecked.

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.

The screenshot shows the 'Choose Search Set' section of the BLAST interface. The 'Database' section has three radio buttons: 'Human genomic + transcript', 'Mouse genomic + transcript', and 'Others (nr etc.)'. The 'Others (nr etc.)' radio button is selected and circled in red. Below the radio buttons, a dropdown menu is open, showing 'Nucleotide collection (nr/nt)' selected. The 'Organism' and 'Entrez Query' sections are optional and currently empty.

3. The next section gives you three options for a nucleotide blast. Choose megablast (default).

Program Selection

Optimize for

Highly similar sequences (megablast)
 More dissimilar sequences (discontiguous megablast)
 Somewhat similar sequences (blastn)
 Choose a BLAST algorithm

4. Select the "Blast" button. What you see below is called the queue page:

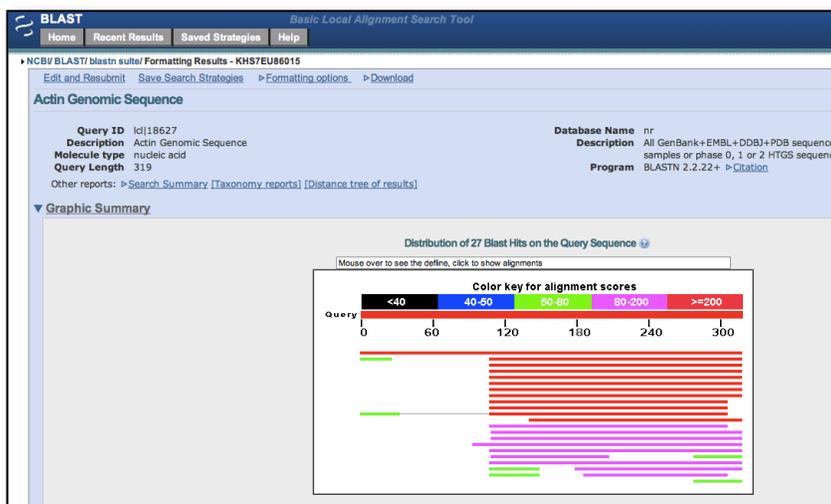
▸ NCB/BLAST/blastn/ Formatting Results - S9WZE65Y013 [\[Formatting options\]](#)

Job Title: Icd|21740 (430 letters)

Request ID	S9WZE65Y013
Status	Searching
Submitted at	Wed Jan 9 11:18:54 2008
Current time	Wed Jan 9 11:18:56 2008
Time since submission	00:00:01

This page will be automatically updated in 10 seconds

5. When your search is complete a results page will be presented. We will discuss this page in detail in class.



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

The screenshot shows a BLAST alignment interface. At the top, there are tabs for 'Alignments', 'Select All', 'Get selected sequences', and 'Distance tree of results NEW'. Below the tabs, the query is identified as 'gb|J01238.1|MZEACT1G Maize actin 1 gene (MAc1), complete cds' with a length of 2424. The alignment statistics are: Score = 590 bits (319), Expect = 2e-165, Identities = 319/319 (100%), Gaps = 0/319 (0%), and Strand=Plus/Plus. The alignment itself is shown in a table format with columns for Query, Subject (Sbjct), and positions. The sequences are identical, with vertical bars indicating the alignment. The alignment starts at position 1 for the query and 252 for the subject, and ends at 319 for the query and 570 for the subject.

```

>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      GTGACAATGGCACTGGAATGGTCAAGGTTGTTATCTCGTTCAGAAGTCTTTTTCACAAA 60
           |||
Sbjct 252     GTGACAATGGCACTGGAATGGTCAAGGTTGTTATCTCGTTCAGAAGTCTTTTTCACAAA 311

Query 61     GCAACTCTACTCCTGTGCCTAATGTTGCTCAACTCCTCAATATTACAGGCCGGTTTCG 120
           |||
Sbjct 312     GCAACTCTACTCCTGTGCCTAATGTTGCTCAACTCCTCAATATTACAGGCCGGTTTCG 371

Query 121    CTGGTGATGATGCGCCAAGAGCTGTCTTCCCCAGCATTGTGGGAAGACCACGCCACCCG 180
           |||
Sbjct 372     CTGGTGATGATGCGCCAAGAGCTGTCTTCCCCAGCATTGTGGGAAGACCACGCCACCCG 431

Query 181    GTGTCATGGTCGGCATGGGCCAAAAGGATGCCTACGTAGGTGATGAGGCTCAGGCCAAGA 240
           |||
Sbjct 432     GTGTCATGGTCGGCATGGGCCAAAAGGATGCCTACGTAGGTGATGAGGCTCAGGCCAAGA 491

Query 241    GAGGCATCCTGACTGAAGTACCCGATTGAGCATGGCATTGTCAACAACCTGGGATGACA 300
           |||
Sbjct 492     GAGGCATCCTGACTGAAGTACCCGATTGAGCATGGCATTGTCAACAACCTGGGATGACA 551

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

A short discussion on how Blast works.

Blast takes the query sequence and divides it into "words" based on the word size parameter (the default is usually "fine"). For a megablast query the default (and minimum) is a size of 28. The algorithm then takes these "words" and runs them against a hash database where the large database is cut into 28 bp words. When an exact match occurs, the program attempts to extend the alignment in each direction on the full sequence. If the alignment extends then a score is calculated and as long as the score remains above a threshold the alignment continues. If a mismatch occurs the score decreases, but as long as the score remains above threshold the mismatch is allowed. Word size can be changed. Long word sizes increase stringency.

The threshold is determined by the Expect value in the "Algorithm Parameters" tab on the Blast page. The default Expect value is 10. This means that you expect to find 10 matches to your query in randomly generated sequence. Blast uses this value, the size of the query sequence, and the size of the database (called the search space) to calculate a threshold on 10 random matches and then reports only hits that score

better than the random model. Lowering the Expect value increases the stringency of the search.

While extending the alignment Blast may encounter a series of mismatched nucleotides. Blast will try to skip over the mismatch region (called opening a gap) to see if the alignment begins again. If the alignment begins again, Blast will continue. If the alignment does not begin again, the alignment process stops and Blast reports the hit. Opening a gap is penalized heavily. Extending a gap is also penalized. The process of opening gaps is necessary to allow for small insertion mutations that occur fairly frequently in a genome.

II. Compare the cDNA to the genomic DNA sequences.

We need to line up the two sequences to determine where they are similar and where they differ. We could do this by hand for short PCR sequences, but it would be very time consuming. Computer programs are very good at this type of analysis and are extremely fast. We will modified version of blastn called Blast2Sequences.

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

The screenshot shows the NCBI BLAST website interface. At the top, there are navigation tabs: Home, Recent Results, Saved Strategies, and Help. Below the navigation, there is a section for 'BLAST Assembled Genomes' with a list of species: Human, Mouse, Rat, Arabidopsis thaliana, Oryza sativa, Bos taurus, Danio rerio, Drosophila melanogaster, Gallus gallus, Pan troglodytes, Microbes, and Apis mellifera. The 'Basic BLAST' section is visible, and the 'nucleotide blast' link is circled in red. The description for 'nucleotide blast' is: 'Search a nucleotide database using a nucleotide query. Algorithms: blastn, megablast, discontinuous megablast'.

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

The screenshot shows the NCBI BLAST search form. The 'Align two or more sequences' checkbox is circled in red. The form includes fields for 'Enter accession number, gi, or FASTA sequence', 'Query subrange' (From and To), 'Or, upload file', 'Job Title', 'Choose Search Set' (Database: Human genomic + transcript, Mouse genomic + transcript, Others (nr etc.); Human genomic plus transcript (Human G+T)), 'Entrez Query', 'Program Selection' (Optimize for: Highly similar sequences (megablast), More dissimilar sequences (discontinuous megablast), Somewhat similar sequences (blastn)), and a 'BLAST' button. The 'Align two or more sequences' checkbox is currently unchecked.

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.

BLAST Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

NCBI/ BLAST/ blastn suite

blastn blasto blastx tblastn tblastx

BLASTN programs search nucleotide subjects using a nucleotide query sequence.

Enter Query Sequence

Enter accession number, gi, or FASTA sequence Clear Query subrange

>Actin Genomic DNA PCR Product
 ctgttacagaatggtggaatggtgacgaggatataccagcctatcgatgtgacaatggcaact
 ggaatgccaagggttctgctcagagctcttttccaaagaaactatctctct
 gtgctaatgttctcacaactcctcaatatttacagccggtttcctggtgatgacg
 ccaagagctcttcccaagcattgtgggaagaccacgccaaccggttcattggtggc
 atggccaaaaggatgctcactagctaggctgaggtcagcccaagagagcactcctgaca

Or, upload file Browse

Job Title
 Enter a descriptive title for your BLAST search

Align two or more sequences

Enter Subject Sequence

Enter accession number, gi, or FASTA sequence Clear Subject subrange

>Actin cDNA PCR Product
 ctgttacagaatggtggaatggtgacgaggatataccagcctatcgatgtgacaatggcaact
 ggaatgccaagggttctgctcagagctcttttccaaagaaactatctctct
 ttgtggaagaccacgccaaccggttcattgctgcatggccaagagatgctcactg
 aggtgatgaggtcaggccaagagagcactcctgacactgaagtaccgattgacatgg

Or, upload file Browse

Program Selection

Optimize for

Highly similar sequences (megablast)

More dissimilar sequences (discontiguous megablast)

Somewhat similar sequences (blastn)

Choose a BLAST algorithm

Search nucleotide sequence using Megablast (Optimize for highly similar sequences)

Show results in a new window

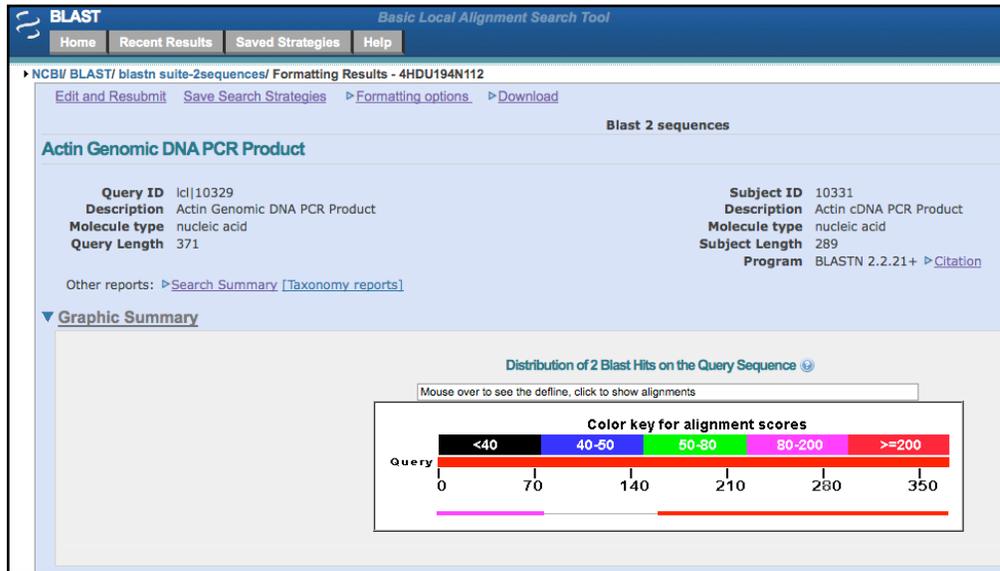
Algorithm parameters

4. The results.

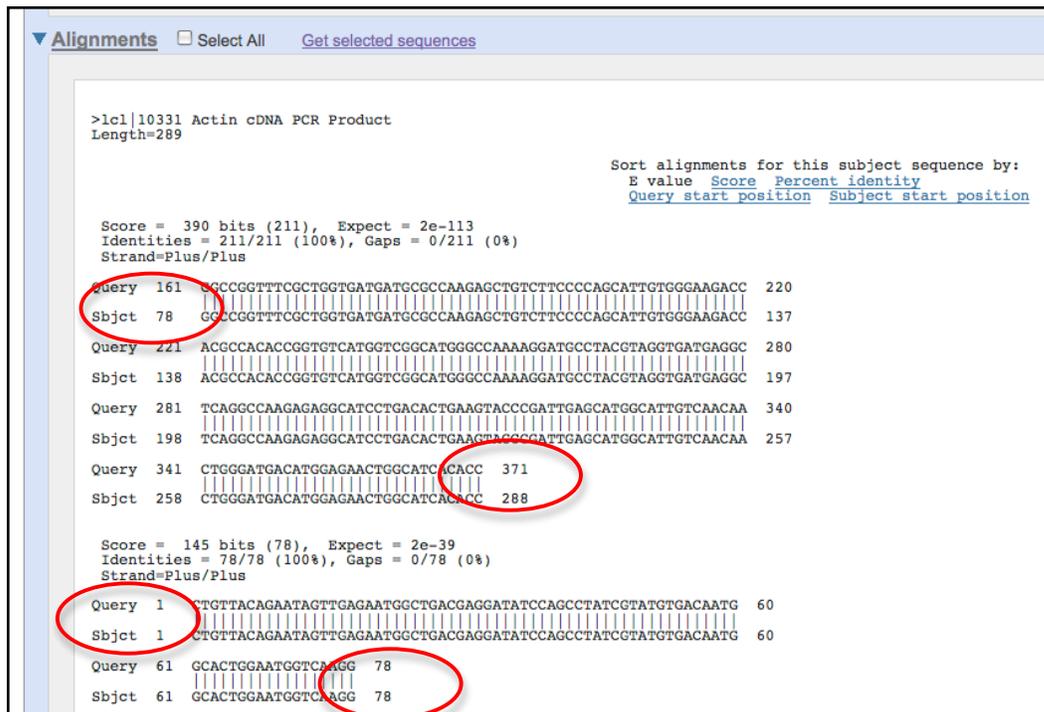
In the top half of the page the results are presented diagrammatically. The query is shown as a red, thick rectangle. Any similarity between the query and the subject is shown as thin rectangles below the query. The color of the rectangle indicates the hit score. The higher the score the better the hit.

The Query is the Genomic DNA sequence and the Subject is the cDNA sequence.

In this case there are two hits between the query and the subject. What do these hits represent?



The bottom half of the results page shows the two hits nucleotide-by-nucleotide called alignments. When bases at the same position are the same a vertical line is placed between them. The 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.



So the cDNA sequence matches the genomic DNA from nucleotides 1 to 78 and then from nucleotides 161 to 371.

1. What do these two alignments represent?
2. What is the sequence from 79 to 160 in the genomic DNA? Why isn't this sequence present in the cDNA?

III. Finding the Actin gene in the Maize Genome

Now that you have verified that the sequences you have are from the actin gene you need to compare the sequence to the published maize genome sequence. To do this we will use blast and the maize browser.

The B73 strain of maize was sequenced to high quality. This is now considered the reference genome for all maize genomic work. The genomes of two other maize strains Mo17 and Palomero (a popcorn) have also been sequenced but not to the level of quality of B73. You will start exploring a genome by finding the location of actin in the maize genome.

Genome sequencing is an involved process. Read pages 453-468 of Chapter 13 from *Introduction to Genetic Analysis*. This can be downloaded from the course web page.

“BLASTing” the Maize Genome

While the blast programs were written by scientists at NCBI, the programs are freely available and are used by many DNA sequence repositories. For the maize browser we use blast to find the location of our sequences on the maize chromosomes.

1. Open the maize browser (www.maizesequence.org). The home page contains information about the B73 genomic sequence. You should read this page and the information pages to learn more about the sequence and the tools that are available on the website. For now, click on BLAST in the upper right hand corner.

The screenshot shows the MaizeSequence.org website interface. At the top, there is a navigation bar with links for 'Home', 'BLAST', 'Login', 'Register', 'FAQ', 'Archive', 'Docs', and 'Help'. The 'BLAST' link is circled in red. Below the navigation bar, there is a search form titled 'Search MaizeSequence' with a search box and a 'Go' button. The search box contains the text 'e.g. gene bt1 or 1.8k-18k or Ramosa 1'. Below the search form, there is a 'Welcome to MaizeSequence.org' message and a 'Publication of the Maize Genome Project' section. The publication section includes the date 'November 19, 2009' and a paragraph of text: 'Today Science published the maize genome paper, titled, The B73 Maize Genome: Complexity, Diversity, and Dynamics. Alongside the main paper, 9 companion papers were published in *Science* and 10 companion papers were published in *PLoS Genetics*. The papers represent the culmination of the four-year project and the hard work of countless individuals. Congratulations to all who contributed. [USDA press release]'. Below the text is a circular genome map showing the chromosomes of the B73 maize strain.

2. Paste your genomic DNA sequence in the textbox. This is your query. We are using the programs with all defaults so click "Run."

New Setup Configure Results Display

Enter the Query Sequence

Either Paste sequences (max 10 sequences) in FASTA or plain text:

```
>Actin genomic DNA
gtgacaatg gcactggaat ggtcaagggt gttatctcgt tcagaagtct
ttttcaacaa agcaactcta ctctgtgcc taattgtgc tcaactctc aatattaca
ggccgatttc gctggtgatg atgcgccaag agctgtcttc cccagcattg tgggaagacc
acgccacacc ggtgtcatgg tcggcatggg ccaaaaggat gcctacgtag gtgatgagcc
```

Or Upload a file containing one or more FASTA sequences
Choose File no file selected

Or Enter a sequence ID or accession (EMBL, UniProt, RefSeq)
Retrieve

Or Enter an existing ticket ID:
Retrieve

dna queries
 peptide queries

Select the databases to search against

dna database Genomic sequence (masked)
 peptide database All Peptides

Select the Search Tool

BLASTN
BLASTP
BLASTX
TBLASTN

Configure Run >>

Search sensitivity:
Optimise search parameters to find the following alignments
Near-exact matches

3. The blast results are show three ways:

A) By karyotype (chromosome)

Displaying *Actin* sequence alignments vs *Zea_mays* LATESTGP_MASKED database

Showing top 100 alignments of 82, sorted by Raw Score

Alignment Locations vs. Karyotype (click arrow to hide)

SUMMARY

Setup

- Genomic sequence (masked)
- BLASTN
- Low sensitivity

Configure

- E: 10
- B: 100
- filter: dust
- RepeatMasker
- W: 15
- M: 1
- N: -3
- Q: 3
- R: 3

Results

Display

Not yet initialised

B) Alignment by query (similar to what you see on the NCBI results page)

Alignment Locations vs. Query (click arrow to hide)

HSPs

coverage

HSPs

50 100 150 200 250 300

C. Alignment summary.

Alignment Summary (click arrow to hide)

Select rows to include in table, and type of sort (Use the 'ctrl' key to select multiples)

Refresh

Query	Subject	Chromosome	Clone	Contig	Stats	Sort By
off Name	_off_ Name	_off_ Name	_off_ Name	_off_ Name	_off_ Score	<Contig
Start	Start	Start	Start	Start	E-val	>Contig
End	End	End	End	End	P-val	<Score
Links	Query	Chromosome	Stats			
	Start End Ori	Name Start End Ori	Score E-val %ID Length			
[A] [S] [G] [C]	1 319 -	Chr:8 99200176 99200498 +	303 4.4e-177 98.45 323			
[A] [S] [G] [C]	107 307 +	Chr:5 160814310 160814510 +	105 6.5e-55 88.06 201			
[A] [S] [G] [C]	107 307 +	Chr:8 165324191 165324391 +	97 1.3e-52 87.06 201			
[A] [S] [G] [C]	107 307 +	Chr:7 32861638 32861838 +	93 2.7e-44 86.57 201			
[A] [S] [G] [C]	108 307 -	Chr:2 137288505 137288704 +	88 2.2e-33 86.00 200			
[A] [S] [G] [C]	107 307 -	Chr:3 174154535 174154735 +	77 1.9e-36 84.58 201			

The Alignment Summary is the most useful of these windows. Usually the summary will look similar to this, but from time to time you may need to use the menus to turn on various columns of information.

For this query we want to look at the one that has the highest %ID or the lowest E-val. As you can see the first row of the results is 98.45% and is located on Chromosome 8. The percent match of your sequence may be different but it should still be on chromosome 8. The 4 letters on the left are links to useful information:

[A] - This link will take you to the alignment. Click on this link to see why the sequence is only a 98.45% match and not 100%.

[S] - This is the sequence of the query.

[G] - This link will take you to the genomic sequence. This is where you can download the B73 sequence that matches your sequence.

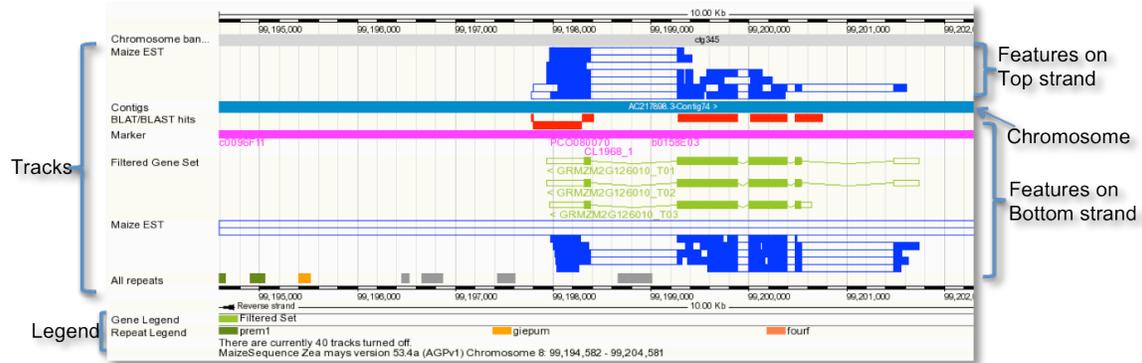
[C] - This link takes you to the Contig viewer that is explained in detail below. Click on the [C] now.

Maize Genome Browser

The maize genome is very large and has many features: genes, simple repeats (CACACACA or TTATTATTATTA, etc.), transposable elements, ESTs (similar to cDNA), and many other things. Genome browsers were developed to visualize all of the features of a region of the genome in a single window. Many of the features can be clicked on for coordinates and detailed information.

We will be using the 4a.53 version of the B73 sequence. The browser is a bit clunky at times so just ask questions if the browser on your screen does not match the browser on paper.

The figure below shows the region of the maize genome that contains the actin gene. This is called the "Contig View." There is a lot of information on this figure. We will discuss most of it in class.



The chromosome is represented by the long light blue rectangle and it is labeled 'contig.' Note that this rectangle may be colored in different regions of a chromosome. Above and below this rectangle is a series of features organized into "tracks." Features are also called annotations. Features above the contig rectangle are on the top or plus DNA strand and features below the contig are on the bottom or minus strand. There is a predicted gene here on the bottom strand. It is represented by the gene rectangles. The filled in green rectangles are predicted to be coding while the open green rectangles are predicted to be non-coding. This "gene" is a prediction by several computer programs. Since they are predictions we refer to these genes as gene models. Since this gene model is on the bottom strand the first exon is to the right.

In red is the Blast/Blat track. This is where the results of your blast will show up. So in this case the cDNA sequence of actin matches very closely to the exons predicted by the computer program.

More evidence that this gene model is real is found in the Maize EST track. EST stands for expressed sequence tag. ESTs are cDNAs that are sequenced *en masse* to provide evidence that a gene is expressed. ESTs may come from the whole plant or specific tissues. Because actin is widely expressed in the maize plant there are a lot of ESTs found for actin.

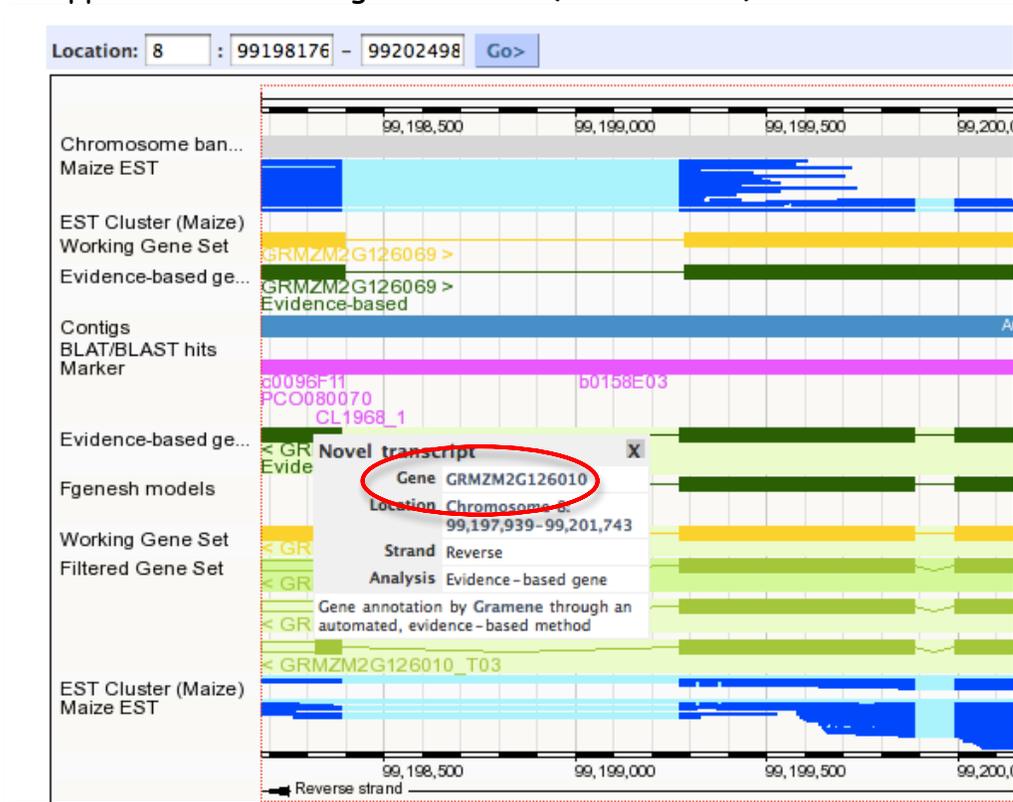
Another interesting Track for this class is the "All Repeats." Transposable elements (which we will study next) are clumped into the general category of DNA Repeats. Click on the little gray rectangle that is in the fourth exon of the Actin gene and you will find that it is a DNA TE called Castaway. The repeat track is always at the bottom of the contig viewer and strandedness does not apply.

Instructions for controlling visible tracks are found at the end of this chapter pages 52-53.

Drilling down for more information

Each track in the contig view has supporting information that you can obtain. In this example you will find the genomic coordinates for the exons of the actin gene.

1. Click on one of the gene models in the Contig Viewer. A contextual menu will appear. Click on the gene number (GRMZM2G...).



2. Some details about the gene model will appear in a new tab. To get exon information click on one of the Transcript IDs.

maizesequence

Home > Location: 8:99,198,176-99,202,498 Gene: GRMZM2G126010

Gene: GRMZM2G126010 Gene: GRMZM2G126010

Gene summary
Splice variants (3)
Supporting evidence
Sequence
External references (0)
Regulation
Comparative Genomics
Genomic alignments (0)
Gene Tree
Gene Tree (text)
Gene Tree (alignment)
Orthologues (5)
Paralogues (22)
Protein families (0)
Genetic Variation
Variation Table
Variation Image
External Data
ID History
Gene history

protein_coding
Location Chromosome 8: 99,197,939-99,201,743 reverse strand.
Transcripts There are 3 transcripts in this gene: [hide transcripts](#)

Name	Transcript ID	Protein ID	Description
Novel	GRMZM2G126010_T01	GRMZM2G126010_P01	protein_coding
Novel	GRMZM2G126010_T02	GRMZM2G126010_P02	protein_coding
Novel	GRMZM2G126010_T03	GRMZM2G126010_P03	protein_coding

Transcript and Gene level displays
In Ensembl a gene is made up of one or more transcripts. We provide displays at two levels:
▶ Transcript views which provide information specific to an individual transcript such as the cDNA and CDS sequences and protein domain annotation
▶ Gene views which provide displays for data associated at the gene level such as orthologues and paralogues, regulatory regions and splice variants
This view is a gene level view. To access the transcript level displays select a Transcript ID in the table above and then navigate to the information you want page. To return to viewing gene level information click on the Gene tab in the menu bar at the top of the page.

Gene summary [help](#)

Gene type protein coding
Prediction Method Gene annotation by [Gramene](#) through an automated, evidence-based method

Transcripts

Configuring the display

3. A new tab will open. On the left is a menu list. Click on the General Identifiers under "External Information" heading. Here you will see EST and other evidence that suggests that this is the maize Actin 1 gene.

Home > Location: 8:99,198,176-99,202,498 Gene: GRMZM2G126010 Trans: GRMZM2G126010_T01

Transcript-based displays
Transcript summary
Supporting evidence (37)
Sequence
Exons (5)
cDNA
Protein
External references
General identifiers (4)
Cloning probes (2)
Gene ontology (2)
Genetic Variation
Population comparison
Comparison image
Protein Information
Protein summary
Domains & features (13)
Variations (1)
External Data
ID History
Transcript history
Protein history

protein_coding
Location Chromosome 8: 99,197,939-99,201,743 reverse strand.
Gene This transcript is a product of gene GRMZM2G126010 - There are 3 transcripts in this gene: [hide transcripts](#)

Name	Transcript ID	Protein ID	Description
Novel	GRMZM2G126010_T01	GRMZM2G126010_P01	protein_coding
Novel	GRMZM2G126010_T02	GRMZM2G126010_P02	protein_coding
Novel	GRMZM2G126010_T03	GRMZM2G126010_P03	protein_coding

Transcript and Gene level displays
In Ensembl a gene is made up of one or more transcripts. Views in Ensembl are separated into Gene based views and Transcript based views associated. This view is a transcript level view. To flip between the two sets of views you can click on the Gene and Transcript tabs in the menu bar

Transcript summary [help](#)

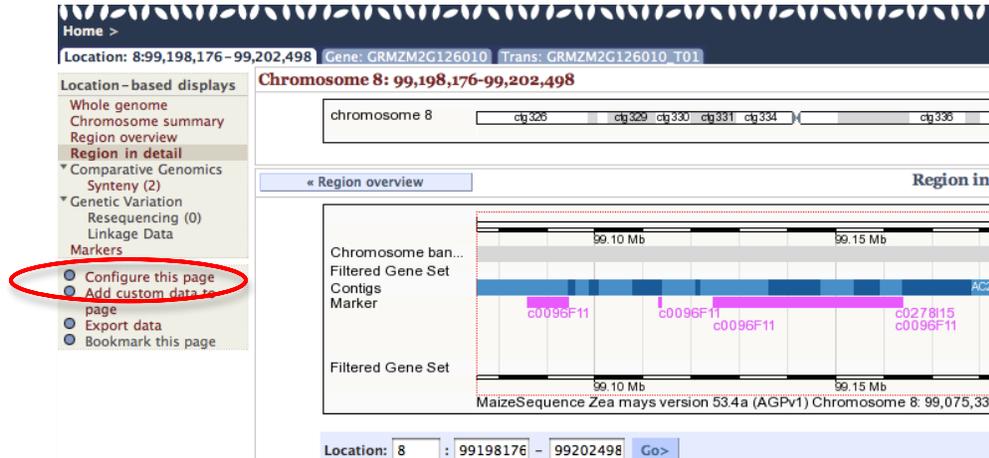
Reverse strand

Statistics Exons: 5 Transcript length: 1,784 bps Translation length: 376 residues
Type Putative protein coding
Prediction Method Unknown

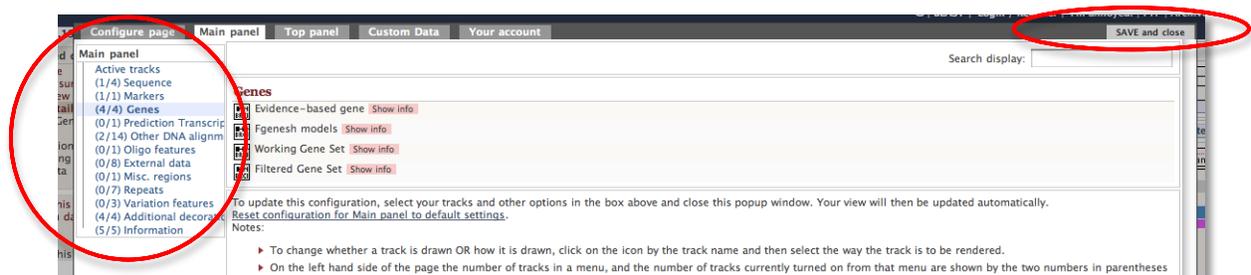
NSF • USDA • DOE • WU • AGI • ISU • CSHL • Gramene
MaizeSequence Release 4a.53 - October 2009

Appendix: Adding and removing tracks from the Contig Browser.

Currently there are over 30 tracks of information available for viewing in the browser. To select which tracks are visible use the "Configure this Page" link in the left hand menu bar.



An overlay window will appear and you can select the categories of tracks in the left hand menu bar. Most categories will have several sub-categories. Select what you want on or off and then click the Save and Close button in the upper right of the window. The Contig Browser will reload with the new selections. You may want to register for an account as this may remember your settings. This is not a guaranteed behavior though. Register before you come to class. Write your user name and password on this page.



Background

The Discovery of Transposable Elements in Maize by Barbara McClintock



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." You know what it looks like—those ears with richly colored kernels that we associate with Thanksgiving and that we 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 [~ 2500 Mb (million bases), about the same size as the human genome] 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**.

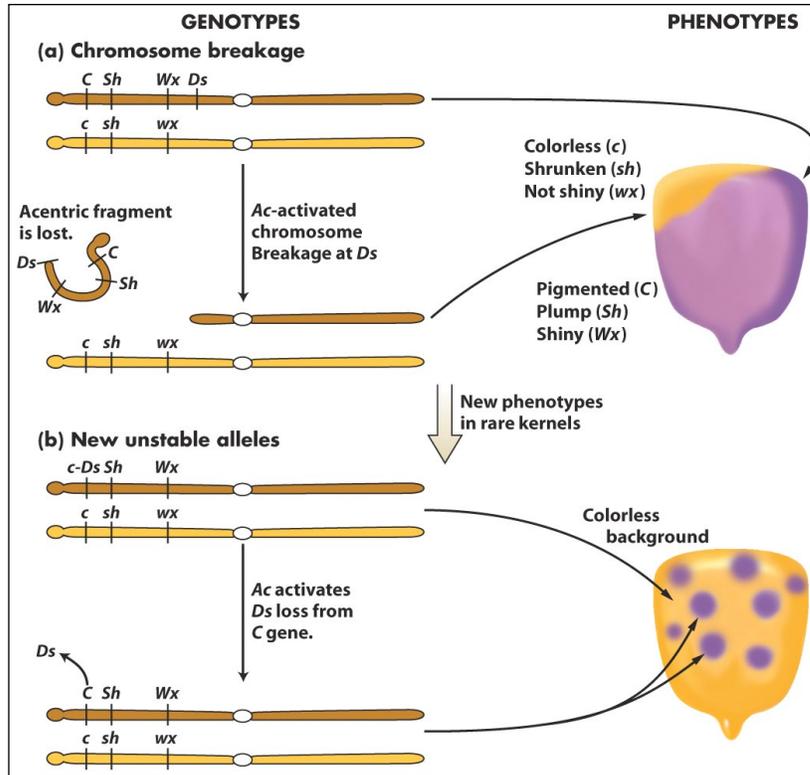
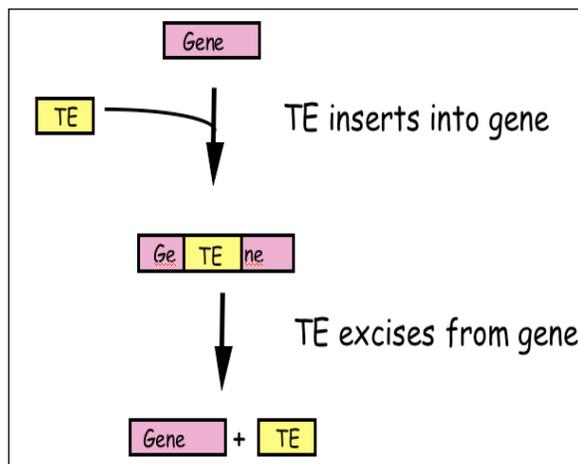


Figure 1. New phenotypes in corn are produced through the movement of the *Ds* transposable element on chromosome 9. (a) A chromosome fragment is lost through breakage at the *Ds* locus. Recessive alleles on the homologous chromosome are expressed, producing the colorless sector in the kernel. (b) Insertion 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 are probably more familiar with like base pair changes and 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:*



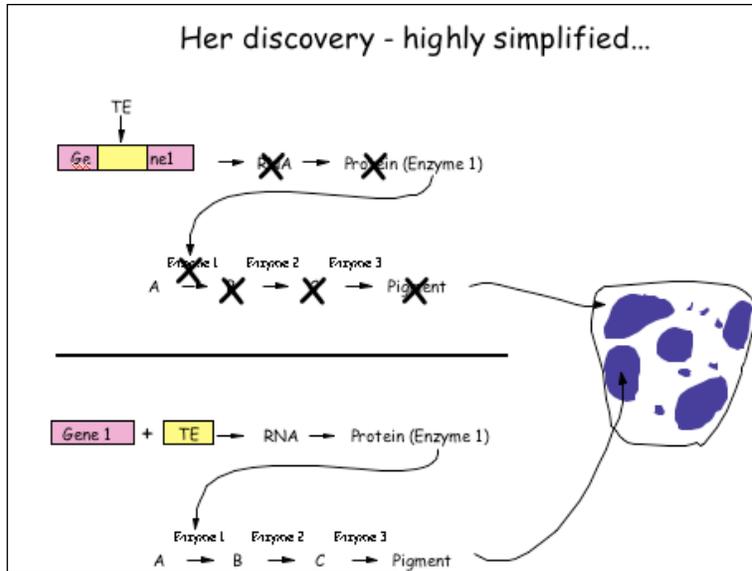


Figure 2: McClintock hypothesized that TEs were a source of "reversible" mutation. Their ability to transpose allowed them to excise from mutant genes leading to phenotypic reversion.

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

As you have seen 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.

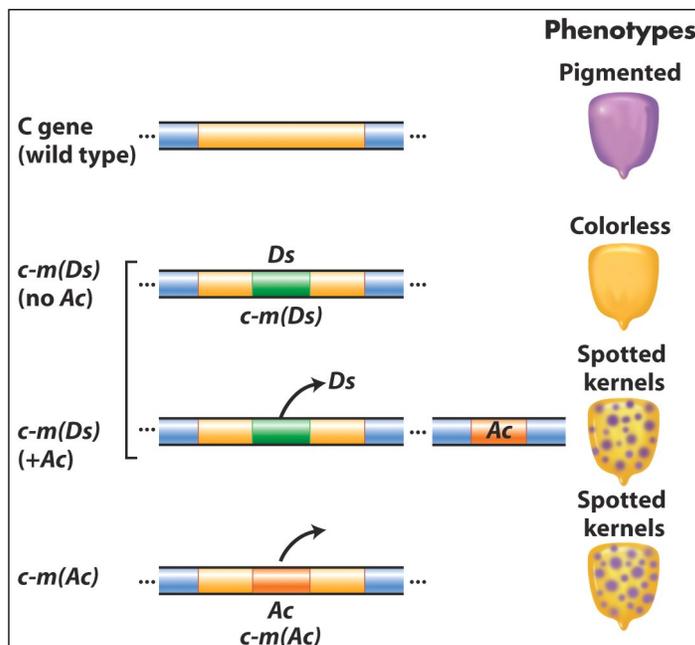


Figure 3 Summary of the main effects of transposable elements in corn. *Ac* and *Ds* are used as examples, acting on the *C* gene controlling pigment. In maize (but not many other organisms), normal alleles are capitalized and mutant alleles are written in lower case. In addition, McClintock designated 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.

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. They say that a picture is worth a thousand words. So... here is a simplified figure showing what Ac and Ds look like at the DNA level.

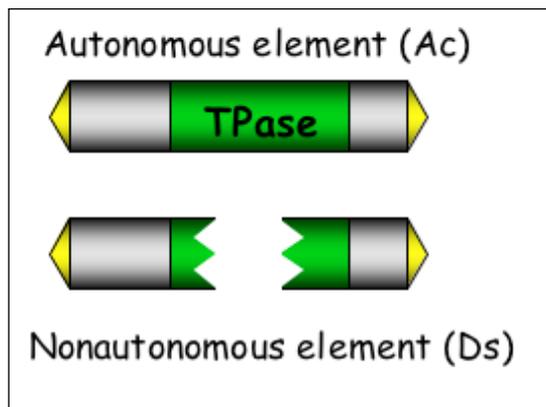


Figure 4: Molecular structure of Ac and Ds.

Ac: T_pase 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 T_pase gene has been deleted.

Yellow arrows at the ends are the terminal inverted repeats - this site is where transposase binds and cuts the element out of the surrounding genomic

Ac contains a single gene - that encodes the transposase protein. Figure 5 shows how this protein catalyzes the movement of Ac and Ds.

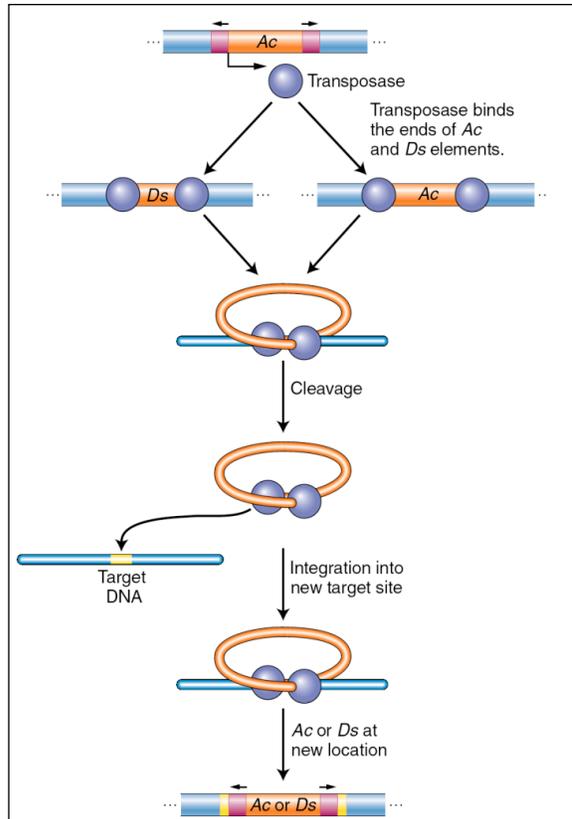


Figure 5 Ac transposase catalyzes excision and insertion (also called integration). The maize *Ac* element encodes a transposase that binds its own ends or those of a *Ds* element, excising the element, cleaving the target site, and allowing the element to insert elsewhere in the genome.

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 Fig 6.

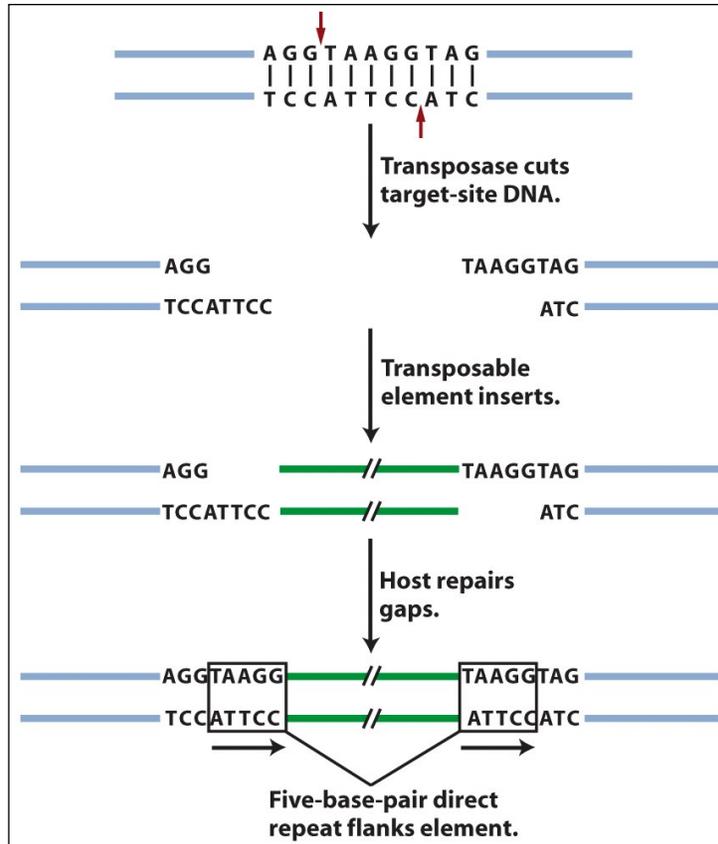


Figure 6: An inserted element is flanked by a short repeat. A short sequence of DNA is duplicated at the transposon insertion site. The recipient DNA is cleaved at staggered sites (a 5-bp staggered cut is shown), leading to the production of two copies of the five-base-pair sequence flanking the inserted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

What transposable elements look like to the bioinformaticist

As you know, the 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, drosophila, worm, dog, mouse, rat, chimp; plants - including *Arabidopsis thaliana*, rice, cottonwood (a tree), and maize]. 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.

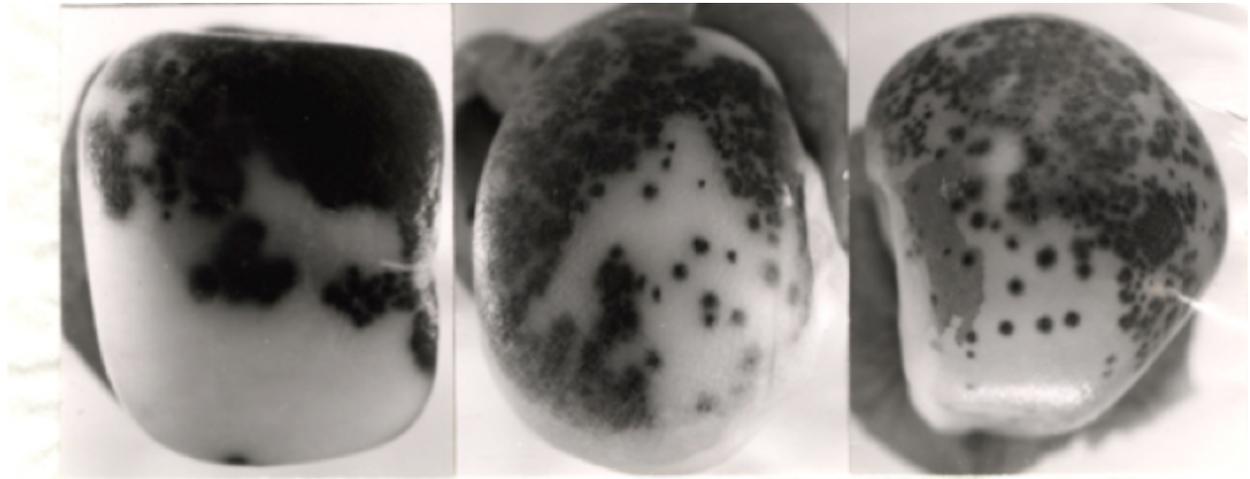
In the Introduction to Bioinformatics Lesson, you will learn about how computers are used to annotate (give meaning to) genomic DNA and how to identify TEs. For many scientists, an experiment ends with the identification of TEs in genomic sequence. They will write up their results and submit them to a journal for publication. However, for other scientists

(like those in the Wessler lab), this is only the beginning. Identifying a TE in genomic DNA does not mean that this TE is still active (capable of moving around). In order to see if a TE sequence is active we have to do an experiment - which is precisely what you will be doing this week. However, you need to be introduced to a few things before we let you in the lab (!).

INTRODUCTION TO THE COMPONENTS OF THE EXPERIMENT:

Overview: In this experiment you will be using the same materials used by the Wessler lab to determine whether a TE sequence that was identified by computational analysis from the rice genome can function (that is, can it move from one place in the genome to another). This experiment will not be done in rice but in the model plant, the humble mustard weed, *Arabidopsis thaliana*.

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



The Wessler lab needed to create an experimental system that mimics the one used by McClintock with TEs inserted into pigment genes and expressed in the kernel. What was needed was a visual assay to test for movement of a rice TE 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 semester, called Green Fluorescent Protein or GFP.

GFP comes from the jellyfish *Aequorea victoria* and fluoresces green when exposed to certain wavelengths of 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.

In our experiments, the GFP reporter gene will substitute for the maize pigment gene. A rice TE (which is described below) has been engineered into the GFP gene so that it cannot produce fluorescent protein. If the TE excises from the GFP gene it will be able function again.

This is all done in *Arabidopsis* - so let's begin with some background on the botanical "lab rat".

Arabidopsis thaliana



In your previous classes you have probably discussed model organisms and their desirable features. 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*. The human genome is almost 20X bigger than this at ~2500 Mb).

***Agrobacterium tumefaciens*: introducing foreign DNA into plants (how scientists construct transgenic 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. 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), which contains a fragment of DNA (called T-DNA). Proteins encoded by the Ti-plasmid facilitate the transfer of the bacterial T-DNA into plant cells and ultimately, insertion of the T-DNA 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 (plant) DNA.

As shown in Figure 7, 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.

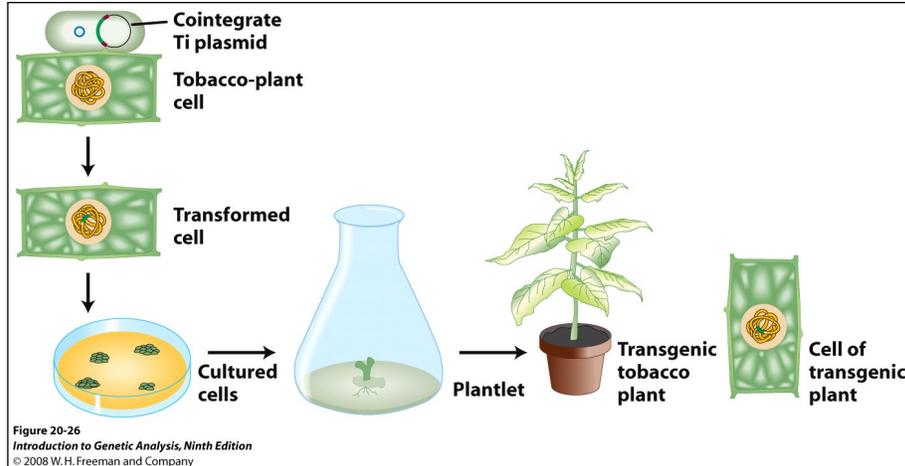


Figure 7: Schematic of how *Agrobacterium* has been exploited to deliver foreign DNA into plant chromosomes.

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.

Rice and the discovery of TE sequences in its genome

Rice (*Oryza sativa*) has the smallest genome of all cereal grasses at 450 million base pairs (Mb). By contrast, the maize (and human) 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. 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. Here is a figure showing the element that you will be analyzing....

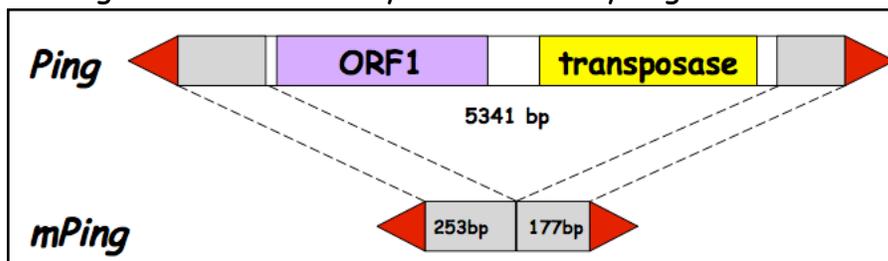
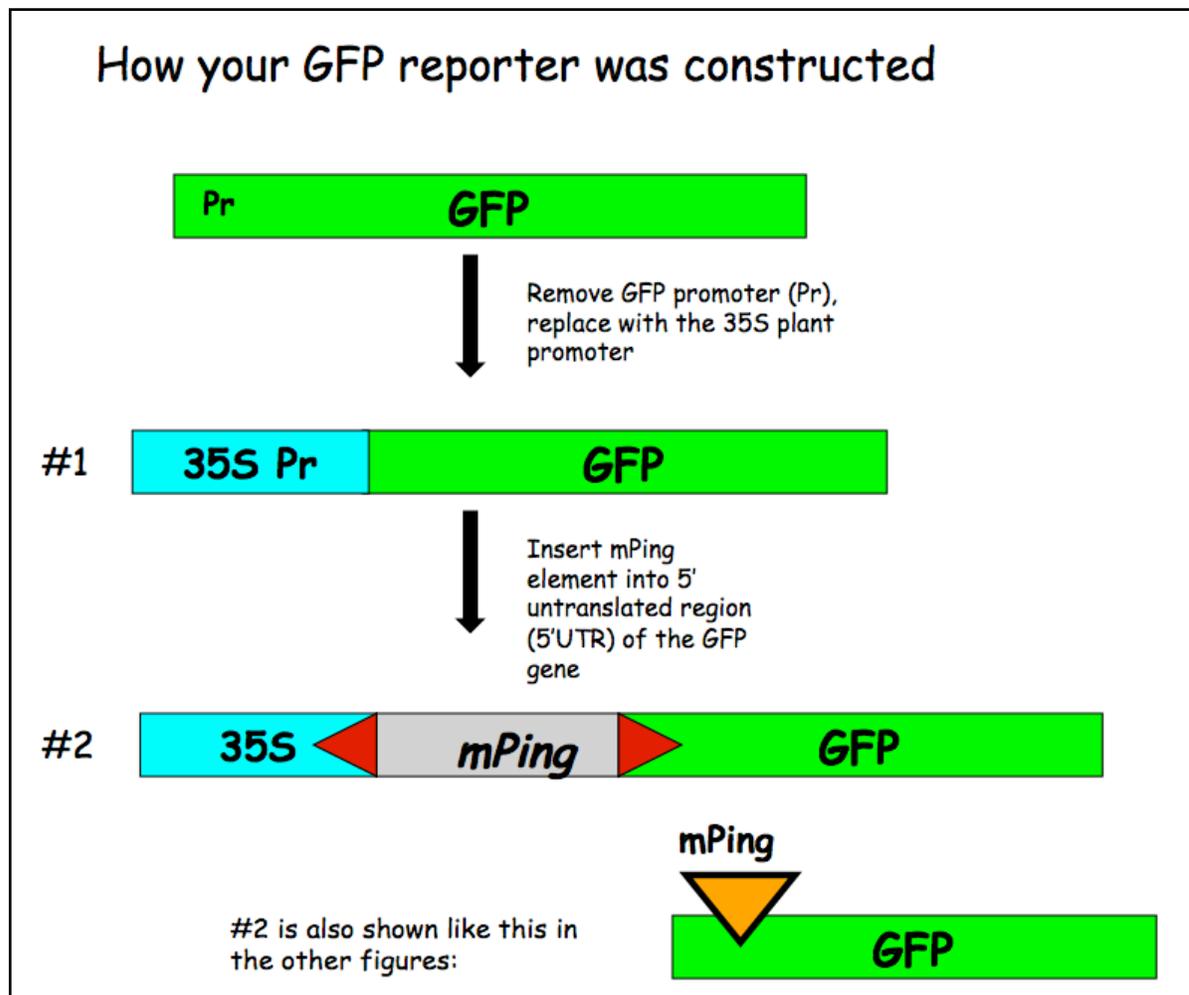
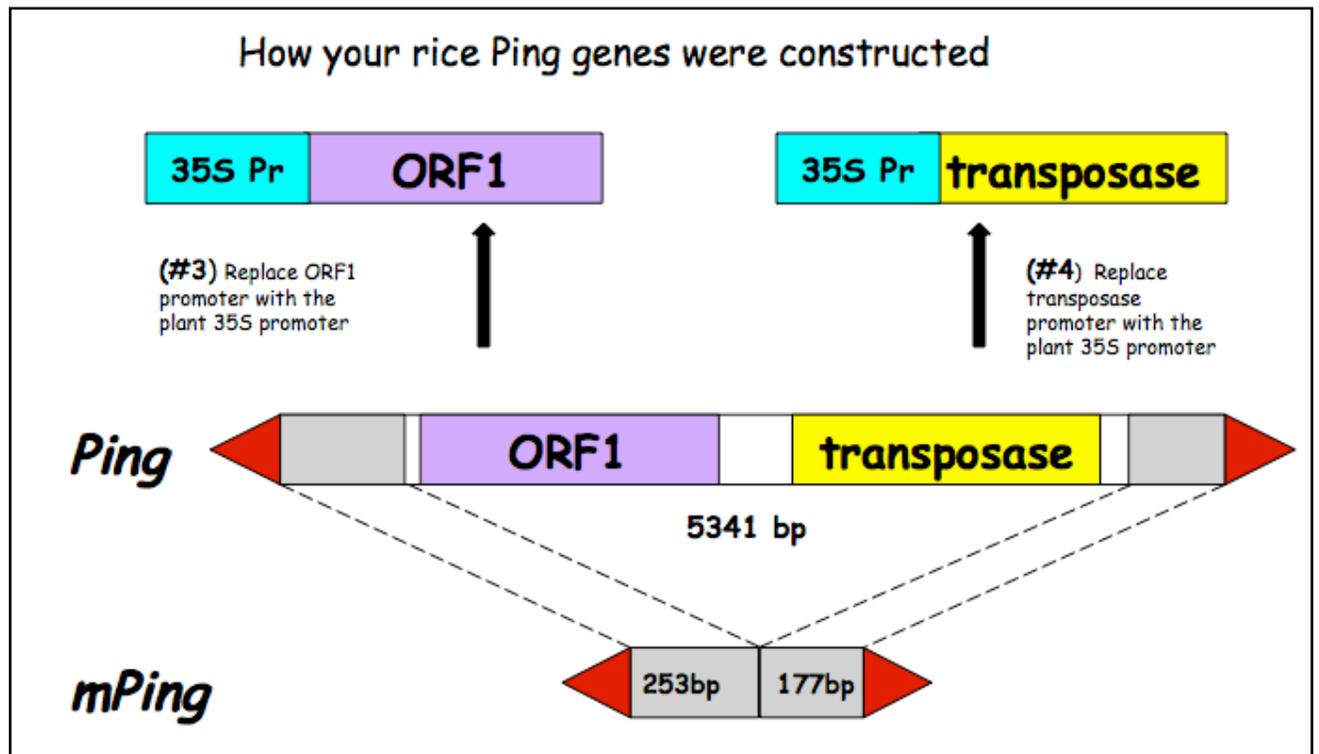


Figure 7: Ping encodes two genes: the transposase (TPase) and ORF1 (open reading frame) (function unknown at this time). The red arrows are the terminal inverted repeats (14bp). mPing is 253bp + 177bp long.

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 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.





In experiment #1 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 T_{pase} 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 one, two, or three of the following T-DNA constructs:

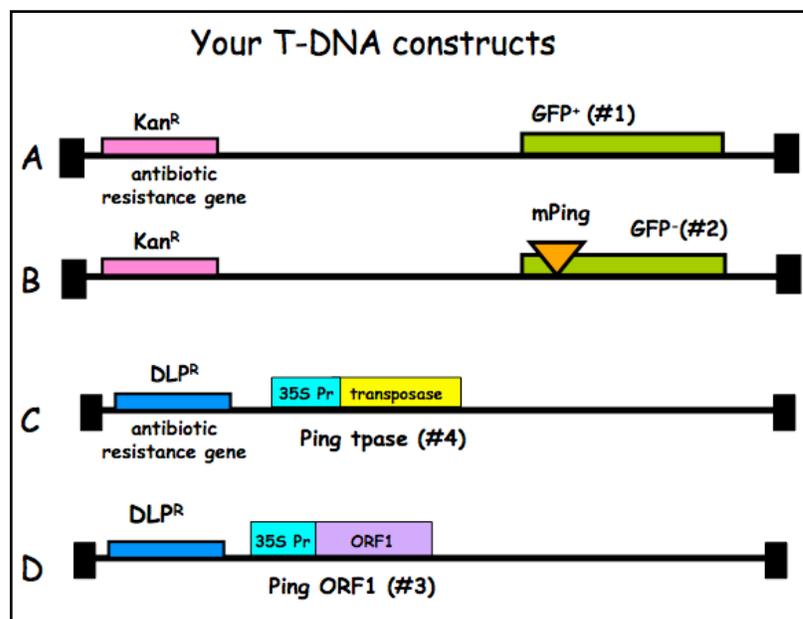


Figure 10:

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.

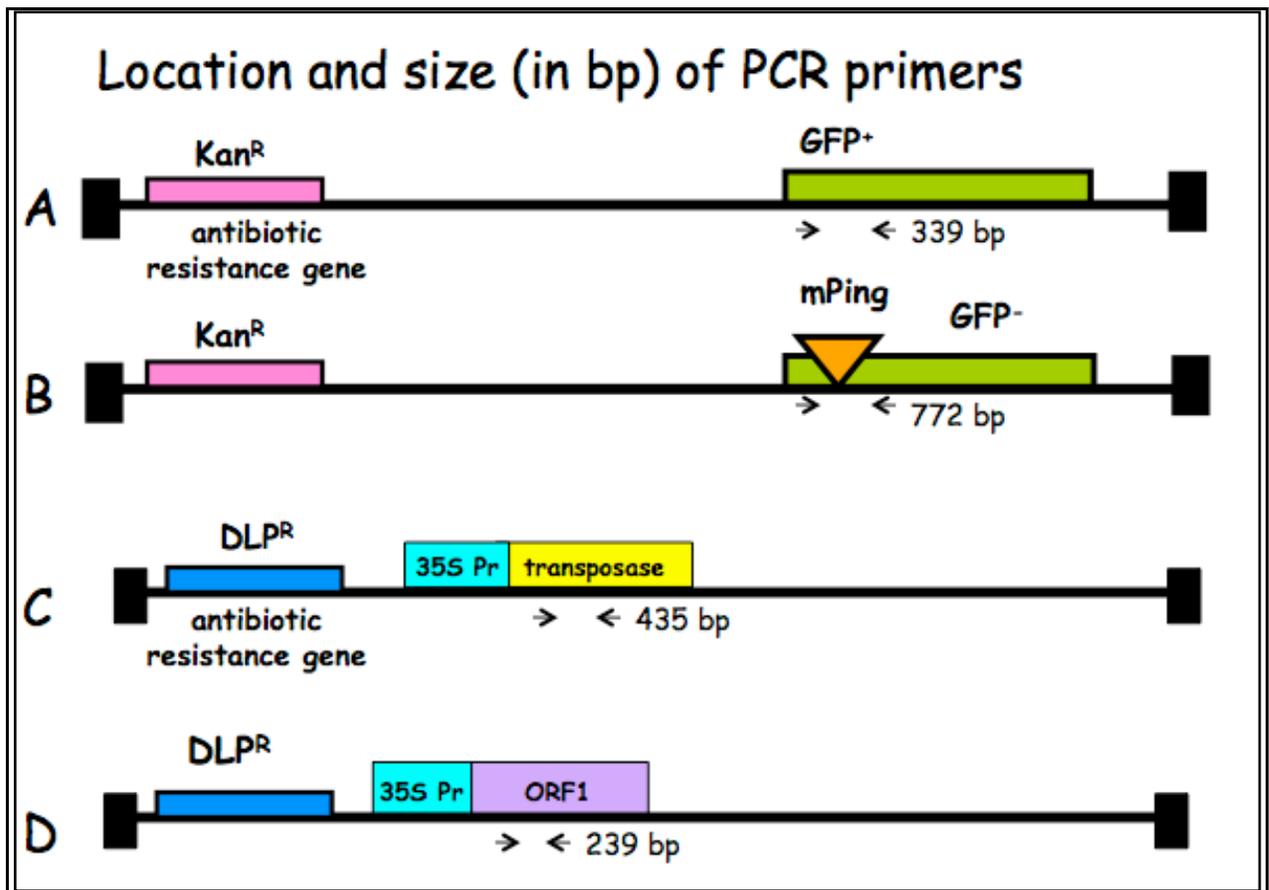
(E) Not shown - NO T-DNA at all. This is the wild type control.

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 Experiment 1

The regions to be amplified by PCR in Expt 1 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.



Experiment 2 Examination of mPing excision from *A. thaliana* leaf DNA

Overview: *In this experiment you will test the hypothesis that the Ping element of rice produces one or two proteins (transposase plus ORF1 protein) that can catalyze the excision the mPing element in the model plant Arabidopsis thaliana.*

You will work in groups of two. Each group will analyze a full set of plants.

I. Plating of Arabidopsis seeds.

One week before actually doing PCR with leaf DNA, the instructors will start growing the plants we will use in this experiment by plating Arabidopsis seeds on petri dishes containing MS salt media. Plate means more or less the same thing as plant, except in a petri dish. The plates were put into a growth chamber where they germinated for ~5 days. The reason that you will not be doing this part of the experiment is because it is very easy for a novice to contaminate the plates with bacteria and/or fungus.

II. Examine leaves under microscope.

You will view and photograph the seedling under a microscope using visible and UV light.

III. Extract genomic DNA from seedlings.

Nucleic acids are extracted from the seedlings using a simple protocol. After you extract genomic DNA you will visualize it on an agarose gel.

Damon Lisch's All Natural (no organics) Genomic Miniprep
Modified for Arabidopsis seedlings

Materials list:

Extraction Buffer

10% SDS

5M KOAC

100% Isopropanol

70% Ethanol

Ice Bucket with ice

liquid nitrogen

65°C heating block

sterile 1.5 ml tubes (2 for each prep)

sterile sticks for grinding

- 1) Label 2 tubes for each plant. Set one set of tubes aside.
 - 2) Harvest 5-6 seedlings. Grind tissue to a fine powder in a 1.5 ml tube dipped in liquid nitrogen.
Put a little liquid nitrogen in a mortar and dip the end of the tube in it. Grind the frozen tissue with a sterile stick.
 - 3) Add 1 ml of Extraction Buffer, and grind some more in the buffer.
 - 4) Add 120 μ l of 10% SDS. Mix by inverting.
- Prepare all samples to this step. Keep them on ice until all are ready for step 6.*
- 5) Put at 65°C for 20 minutes.
 - 6) Add 300 μ l 5M KOAc. Mix well by inverting several times (important!), then place on ice 10 minutes.
 - 7) Spin for 5 minutes at top speed in microfuge. Squirt about 700 μ l of the supernatant through miracloth into second tube. (make small funnel, place tip directly onto the miracloth at the tip of the funnel and squirt through - do not allow the whole funnel to get soaked).

8) Add 600 μ l of isopropanol. Mix the contents thoroughly by inverting.

DNA precipitate may or may not be visible at this point; don't worry if you don't see much. However, a really good prep (excellent grinding of tissue) should result in visible DNA at this stage.

9) Spin for 5 minutes at top speed. Pipette off supernatant.

10) Add 500 μ l 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 around 30 minutes with the caps open.

11) Resuspend the DNA in 50 μ l water or 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.

Visualize genomic DNA on an agarose gel:

12) Label new 1.5 ml tubes.

13) Put 15 μ l of DNA from step 11 into each tube.

14) Add 3 μ l of 6x loading dye to the tube. Tap the tube gently to mix.

15) Load all 18 μ l on your gel. Keep track of which sample went in which lane.

16) Load 7 μ l of DNA Ladder in one empty well.

17) Run the gel at 130 Volts for ~20 minutes.

18) Photograph the gel.

IV. Amplify genomic DNA using PCR.

Today you will amplify the DNA using 3 pairs of primers: one pair for GFP, one pair for ORF1, and one pair for T_pase. The primers for ORF1 and T_pase will be mixed and used in a single PCR reaction. This is called duplex PCR.

For each group, one person should set up the GFP PCR reaction and the other should set up the ORF1+T_pase reaction. Each person will have six DNA samples to analyze. We also need an additional negative control that will be water in place of DNA, giving you seven reactions in total. You need to make a master mix of eight reactions to make sure you have enough for the seven tubes.

1) Label a 1.5 ml tube with the primer set you are using: either GFP or ORF1+T_pase.

2) Label a strip of 0.2 ml PCR tubes. The instructors will show you where to label the tubes. The label may be rubbed off in the machine if you put it in the wrong place.

3) Mix the following in your tube using the volumes in the column labeled '8X (stands for 8-fold).' Keep this on ice.

	1X (μl)	8X
2x Master Mix	12.5	100.0
H ₂ O	5.5	44.0
Forward Primer	1.0	8.0
Reverse Primer	1.0	8.0
DNA	5.0	-----
Total	25.0	160.0 μl

The 2x Master Mix is supplied by a company (NEB) and contains Taq enzyme, buffer, and deoxynucleotide triphosphates (dNTPs) in a 2x concentration. This tube should be kept on ice to protect the enzyme from degradation.

4) Put 20.0 μl of master mix in 7 of the PCR tubes.

5) Add 5.0 μl of correct genomic DNA to the PCR tubes. Put 5 μl of sterile water in tube 7.

6) Seal the tubes tightly with a strip of caps. Keep PCR tubes on ice until everyone is finished.

After everyone is done, your samples will be placed in a thermocycler or 'PCR machine' and 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
[Note: "30 cycles" means all steps— denaturation, annealing, and extension—are repeated 30 times before going on to the next step]			
1 cycle for:	final extension:	72°C	10 minutes

7. After you finished setting up the PCR, you should pour a 1.5% agarose gel. Use two combs per gel. You will need one gel per group.

8. When the PCR finishes, load 20 μ l into each well of the gel. Remember to add a lane with 7 μ l of DNA ladder (molecular weight markers). Run the gel at 130 V for ~30 minutes.

Sequences of the Primers used:

GFP Primers (772 and/or 339 bp amplimers)

GFP-R 5'-AGA CGT TCC CAA CCA CGT CTT CAA AGC-3'

S35-F 5'-CCT CTC CAC TGA CAG AAA ATT TGT GC-3'

ORF 1 Primers (239 bp amplimer)

PING-ORF1-FOR 5'- CAC TGG TCA AGG TTG AAG TCA GCG ATC TCT G -3'

PING-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'

V. Gel analysis.

We will analyze the results of your PCR reactions and relate them to your observed leaf phenotypes.

Investigating polymorphism in intron sequence among maize strains

Introns can accumulate DNA sequence polymorphisms at a higher rate than exons. By now you should be able to explain why this is true. During your project you will focus on an intron in a maize gene, amplify the sequence using PCR, and sequence the PCR products from several strains. You will then learn how to analyze the data to determine whether there are any polymorphisms in the intron you are studying.

Earlier you learned how to use GBrowse to find a gene in the maize genome and we will return to GBrowse soon. In this project we will use the DNA Subway an integrated genome annotation and exploration platform. You will recognize GBrowse as one of the tools in the DNA Subway. The DNA Subway will step you through the process of annotating DNA sequence and help you find the information you need to complete your project.

Here are the steps you will use:

- I. Create a project. Research gene.
- II. Genome Annotation: Repeat Mask
- III. Genome Annotation: Gene Prediction
- IV. Genome Annotation: Experimental support for exons
- V. Genome Annotation: Blast to find gene of interest
- VI. Genome Annotation: Extract gene sequence
- VII. Design PCR Primers
- VIII. Amplify intron and analyze DNA Sequence

For illustration purposes these instructions will continue to use the maize Actin 1 gene.

```
>Intron4 Actin1
GTCTTCTCTTTTCTTACTTTTATTCTTTCATAAGCAGTAGGCTTTTATCATTCAACTCCTCT
TCATGAGAACCTACAGCAGCCTTGACCATGTTGGGTGTGTGTGCTTTTTTTGTAGTTAAA
GTGGAATAACCCATTATATACCACTCAATTTAGGTTAAACTGGCTTGTGATGCACTCT
TTTTTACCTAAAAACGCAATTGGAATACAAATATATTCTTGTCTTAGTCAGTTACTCCAG
TTTTCTAGTATCAACTTCTCTTAGGGCCATTTGGAACGAAGGAAATATGAAGGATTTTT
GAAGGATTCATTCCCTACATAGAAGTTTCCTATGTGACCCTTTGGAACACAGGATTTAAT
CCTATCAATTCCTTTGAAATTTCTATGGATTGCTTTCCTCCATACAAGTTTTGGAGGAAA
ATAACAAGAGCTCCAATCTCTTGGAAAATTTCCCTATGACTCTTATCTCTCTATTCAAATT
CATGTATAGCACCCAAAGGCTAGGTGAGAAAATTTGCTACGTTCTCAATTTCTGTAGGAT
TGCAATTCATAGGCAATTATAGTCATGTGTTTTCTCTATTCCCTGTGTTTTTCATATTCTTC
TGTTCTAAAGGGGGCTTAGTAGGCACTGCACATGTTGCTCCATCCTCCGCGTTGTGTT
TAATCATTACGTTGTTAATCATTACAGACATAGATTTAAATTTGATAGTAACATCTTTT
TGAGGAAATGAAAAGATTTTAAAGGCTTTTTCTTCACGATAGATATAAGCGGGATGTGGA
AATGGAATTACTAGTGTGATAGCACCGAATATTTCTTGCATTGCCATTACTTGCACCTC
ATTATTTTAATTTCAATTTCTTTGACAATGGGCAACTGTGTTGGCAG
```

Preliminary Set-up: Create DNA Subway Account

The instructors have pre-screened introns for you to study. The introns have the following characteristics:

- a. Relatively short sequence. This is a necessary because we do not want the PCR products to be too long. If they are too long, it makes determining the cycling conditions

of the PCR very difficult. Ideally we want a PCR product to be at least 300 bp and no longer than 1500 bp (for using the standard Taq polymerase).

b. The intron is in an interesting gene. Many genes are simply predicted by computer programs and have no known biological function. While these would be interesting to study to determine the function, that could take years! So the instructors have tried to limit the genes to ones with know function.

c. The gene has some experimental evidence to support the exon and intron predictions.

You also need to create a Google Document to record all of your work. Remember the bioinformatics portion of the project is just as important as the bench work. You need keep a notebook of what you do so someone else can reproduce your work. You should include the steps you use and make screen shots to include in the document. On an Apple computer you can get a screen shot by pressing "Shift + Apple (Command)+4" at the same time. Drag the cross-hair cursor over the area you want to capture. An image will be placed on the Desktop that you can upload to your Google Doc notebook.

1. Open <http://dnasubway.iplantcollaborative.org/>:

FAST TRACK TO GENE ANNOTATION AND GENOME ANALYSIS

Optimized FOR PLANTS animals coming soon

ID:
 Password:
 Log In Enter As Guest
 Forgot Password? Register

D N A
SUBWAY

Annotate a Genomic Sequence
 Find Repeats
 Predict Genes
 Search Databases
 Build Models
 Prospect Genomes Using TARGET
 Search Genomes
 Alignment & Tree Viewer
 Browsers & Transfer

This site ties together key bioinformatics tools and databases used to annotate genes and analyze genome data. Roll over any of the "stations" on the subway map to find out more about the analysis steps. Start a project by selecting one of the "subway lines" (red, yellow). Register and login to be able to save or share your results.

Tour Manual
 All content is © 2009 by the iPlant Collaborative
 About Credits Resources Feedback

2. Click on the Register link and fill out the form. In a few minutes a email will be sent to you. Follow the instructions in the e-mail to complete registration. **While you wait, watch the Tour of the DNA Subway.** Write your username and password here.

Username: _____

Password: _____

3. Open the DNA Subway and look at the map. There are several important things for you to know:

FAST TRACK TO GENE ANNOTATION AND GENOME ANALYSIS

Jim Burnette

Log Out

Projects

Optimized FOR PLANTS animals coming soon

D N A

SUBWAY

Annotate a Genomic Sequence

Find Repeats

Predict Genes

Search Databases

Build Models

Prospect Genomes Using TARGeT

Search Genomes

Alignment & Tree Viewer

Browsers & Transfer

This site ties together key bioinformatics tools and databases used to annotate genes and analyze genome data. Roll over any of the "stations" on the subway map to find out more about the analysis steps. Start a project by selecting one of the "subway lines" (red, yellow). Register and login to be able to save or share your results.

Tour Manual

All content is © 2009 by the iPlant Collaborative

About Credits Resources Feedback

- A. Once you have projects you can click here to see a list to access and manage your projects.
- B. The Annotate a Genomic Sequence "Red Line" is the line you will be "riding" for this project.
- C. The Prospect Genomes using TARGeT or "Yellow Line" will be saved for a different project.
- D. Download the Manual and refer to it for further instruction.

Save it to your computer. Using bioinformatics tools you already know, find out what gene the intron is in and research the gene function.

Step I. Create a project on the Red Line in the DNA Subway.

A. Click on the red square to start riding the Red Line.

FAST TRACK TO GENE ANNOTATION AND GENOME ANALYSIS

Jim Burnette

Log Out

Projects

D N A

SUBWAY

Optimized FOR PLANTS animals coming soon

Annotate a Genomic Sequence

Find Repeats

Predict Genes

Search Databases

Build Models

Alignment & Tree Viewer

Browsers & Transfer

Prospect Genomes Using TARGeT

Annotate a Genomic Sequence

Input a DNA sequence of up to 100,000 nucleotides to identify the genes and transposons in it, and annotate (associate) the sequence with this information.

This site ties together key bioinformatics tools and databases used to annotate genes and analyze genome data. Roll over any of the "stations" on the subway map to find out more about the analysis steps. Start a project by selecting one of the "subway lines" (red, yellow). Register and login to be able to save or share your results.

B. Fill out the Project form:

FAST TRACK TO GENE ANNOTATION AND GENOME ANALYSIS

Jim Burnette

Log Out

Map

My Projects

Public Projects

Annotate a Genomic Sequence

Prospect Genomes Using TARGeT

Find Repeats

Predict Genes

Search Databases

Build Models

Browsers & Transfer

D N A

SUBWAY

Select Sequence Source *

Upload a sequence file in FASTA format (max 150kb):

Choose File actin region

Enter a sequence in FASTA format (max 150kb):

Select a sample sequence:

Arabidopsis thaliana (mouse-ear cress) Chr1, 100.00 kb

Arabidopsis thaliana (mouse-ear cress) Chr1, 14.00 kb

Arabidopsis thaliana (mouse-ear cress) Chr1, 12.00 kb

Arabidopsis thaliana (mouse-ear cress) Chr1, 50.00 kb

Arabidopsis thaliana (mouse-ear cress) Chr1, 2.93 kb

Arabidopsis thaliana (mouse-ear cress) Chr2, 100.00 kb

Name Your Project *

Project title:

Intron Polymorphism

Organism *

Scientific name (genus species):

Zea mays

Common name:

corn

Class:

Monocotyledons

Other

Dicotyledons

I don't know

Description

Total characters (max.140): 51

Annotate a gene and determine intron polymorphism.

Continue

* Required information

Tour Manual

All content is © 2009 by the iPlant Collaborative

About Credits Resources Feedback

A. Click the "Choose File" button and use the file box to select your genomic sequence file.

B. Enter a project description.

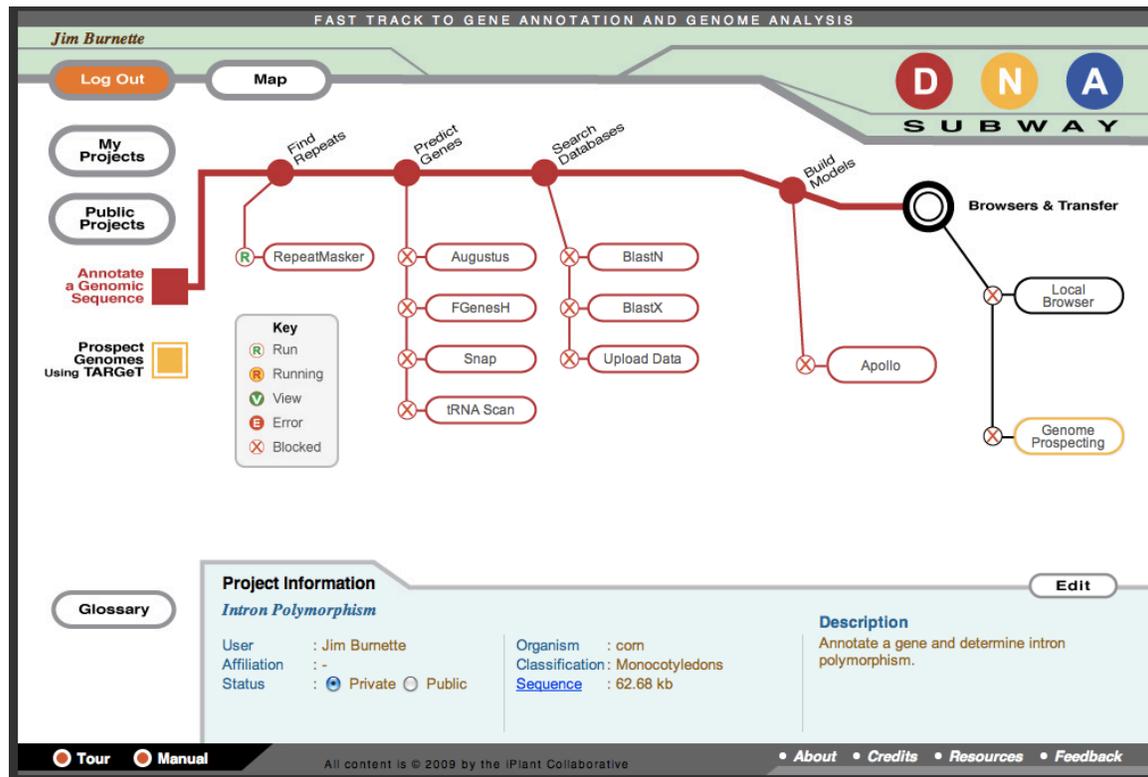
C. Enter a meaningful Project title.

D. Enter *Zea mays* for the scientific name and "corn" for the common name. (Maize is not allowed.)

E. Choose Monocotyledons. This is the large group of plants that maize belongs to.

F. Click "Continue" to create this project.

Step II. Genome annotation: Repeat Masking



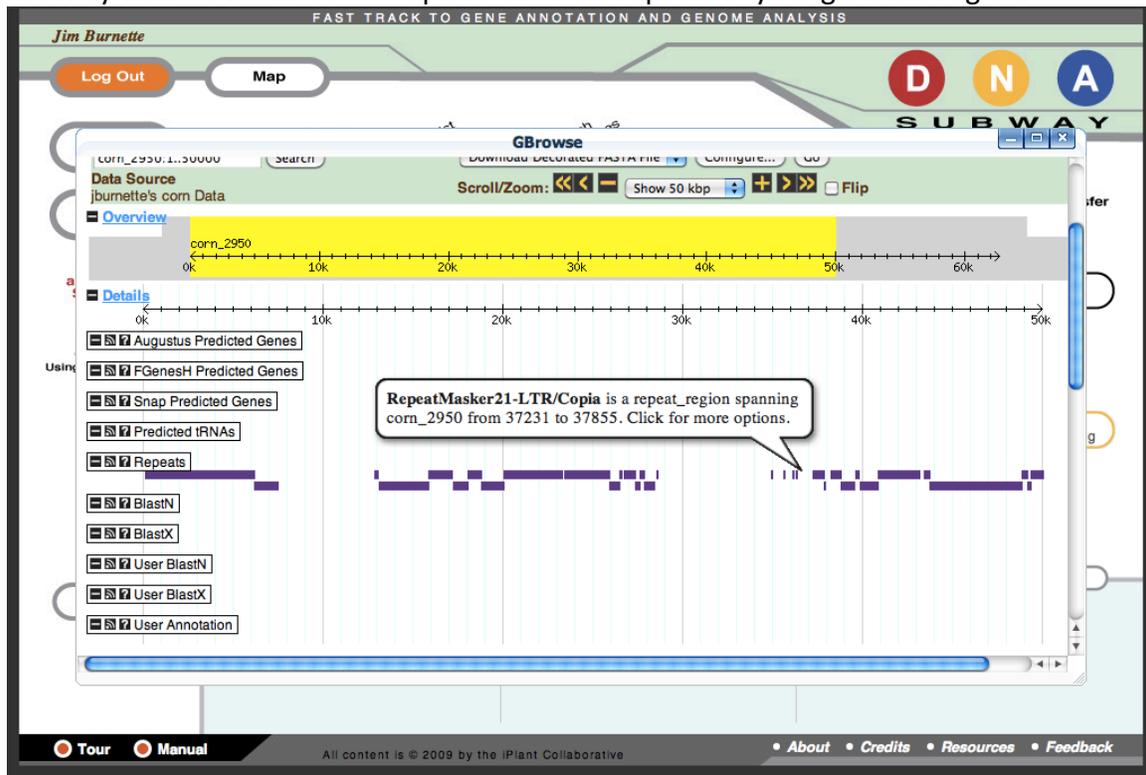
This is the main annotation page. You will see that there are stops (circles) and sidelines along the main Red Line. One circle has a green “R” for Run, while all the others have red “X’s.” which means blocked. You have to start by clicking the green “R” in the RepeatMasker station.

Masking Repeats

While the repeat masking program runs let me explain what it is doing. The genome is only 2% genes. The majority of the rest is poorly understood repeated sequences. Some of these sequences are repeats such as “TA” or “CGA” repeated hundreds of times. All di-nucleotide and tri-nucleotide repeats are presents in the genome. Other repeats are transposable elements that you will learn about. These repeated sequences bog down other programs that we need to use to annotate DNA sequence. The RepeatMasker program has a database of experimentally identified repeats. If genomic sequence matches a repeat in the database the sequence is “masked” by replacing it with “N’s.” All of the other programs know to skip over this sequence, and that speeds them up.

When RepeatMasker finishes the station will show the “View” symbol. Click on the RepeatMasker button and you can see the results in a spreadsheet. Very exciting stuff here. Don’t you wish you could see this graphically in a genome browser? Guess what, you can by clicking the Local Browser station at the right. You will get a new window

with GBrowse and the repeats loaded. Mouse over the repeats to see what they are. You may want to zoom out to explore all of the repeats in your genomic region.



The majority of the Actin 1 genome neighborhood is repeats! Most are Copia and Gypsy sequences which are transposable elements. If there most of the sequence is repeats, where do you think the genes will be?

Close the GBrowse window using the 'X' button on the upper right.

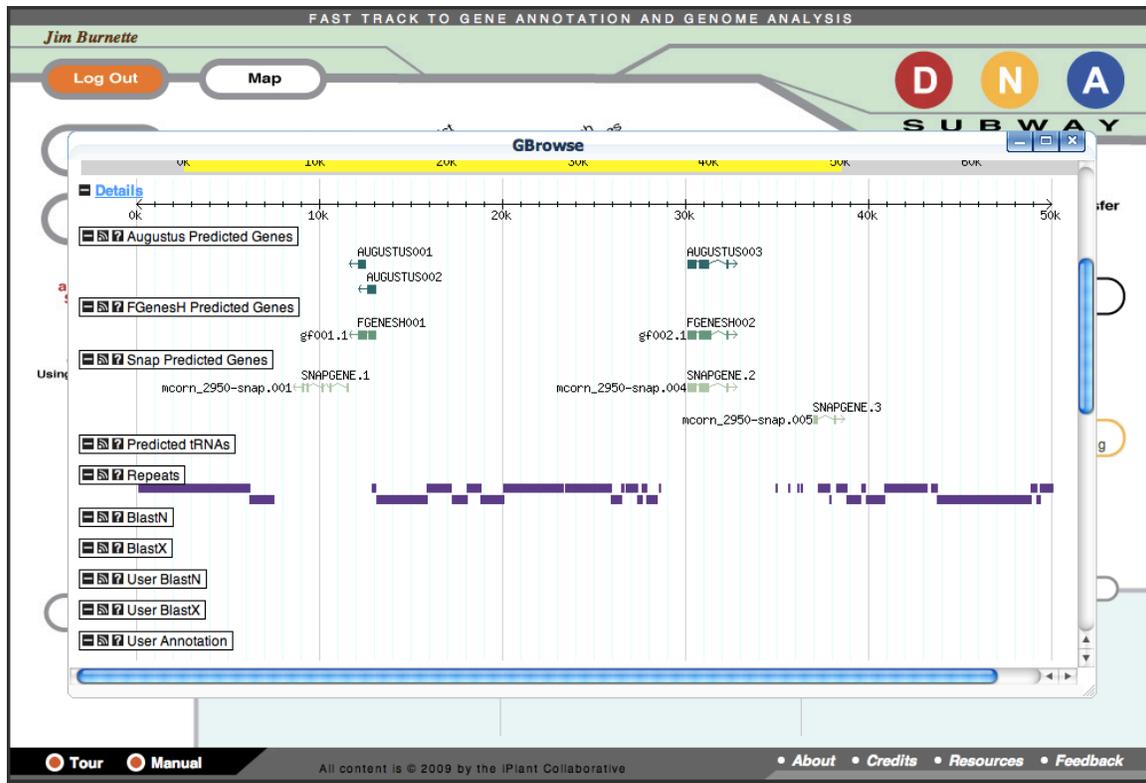
Step III. Genome Annotation: Gene prediction.

Back on the Subway Map click all of the gene prediction stops: Augustus, FGenesH, Snap, tRNA Scan.

These programs read in DNA sequence and try to predict genes. tRNA Scan is specifically looking for tRNA genes while the other three are looking for protein encoding genes. Predicting protein encoding genes is a very difficult task. Partly this is due to the fact that scientists do not fully understand the significant sequences that characterize genes. Partly this is due to the rule proteins use to bind to promoters, splicing signals, transcription.

Watch and study the Dolan Gene Annotation Animation
<http://www.dynamicgene.org/index.html>.

After the programs finish running and display the “View” station symbol, click on the Local Browser station to see the results. You can also see the actual results by clicking on the station for each gene prediction program.

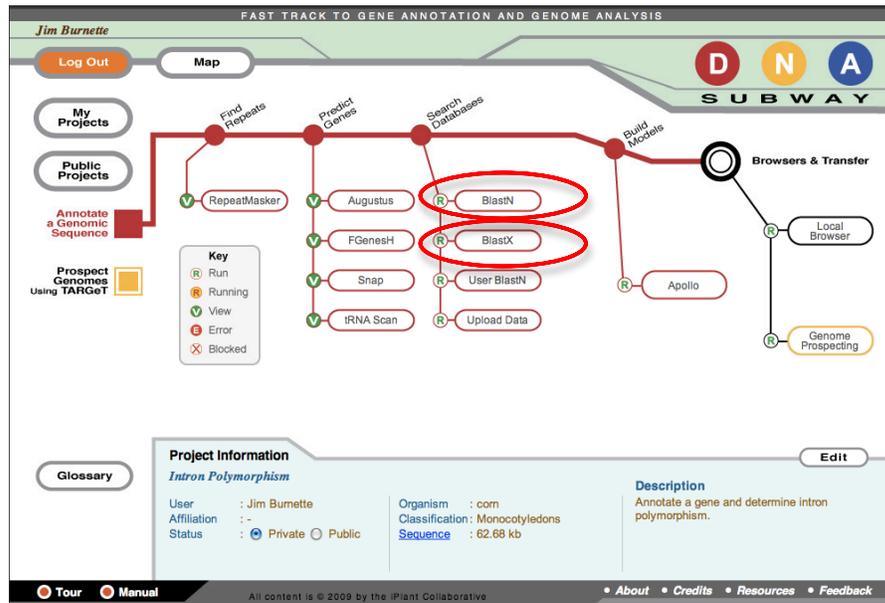


The gene predictions or models are shown in shades of green. You can see that Augustus and FgenesH predictions are similar while Snap predictions are very different. One reason for this is that Augustus and FgenesH predictions start at a start codon and stop at a stop codon. Snap attempts to predict untranslated regions of genes. Which gene is Actin 1 or the gene you are interested in? You can find out using the intron sequence you were provided.

Step IV. Genome Annotation: Experimental support for exons

The genes that were predicted in Step V have no experimental support. They are best guesses made by computer program. We need to provide experiment evidence to support the exons. One way to do this is to include experimental evidence from NCBI. The DNA Subway has two databases: UniGene and UniProt. These databases are compilations of well-characterized experimentally derived gene and protein sequences.

Click the BlastN and BlastX stations on the Search Databases line. BlastN will blast your genomic regions DNA against the UniGene database and BlastX will blast against UniProt.

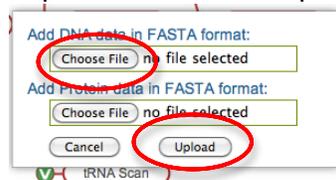


Open the Local Browser and view the results. You will see there is plenty of experimental evidence for Actin 1. You can also see that the experimental evidence shows there is evidence for exons not predicted by Augustus, FgenesH, or Snap.

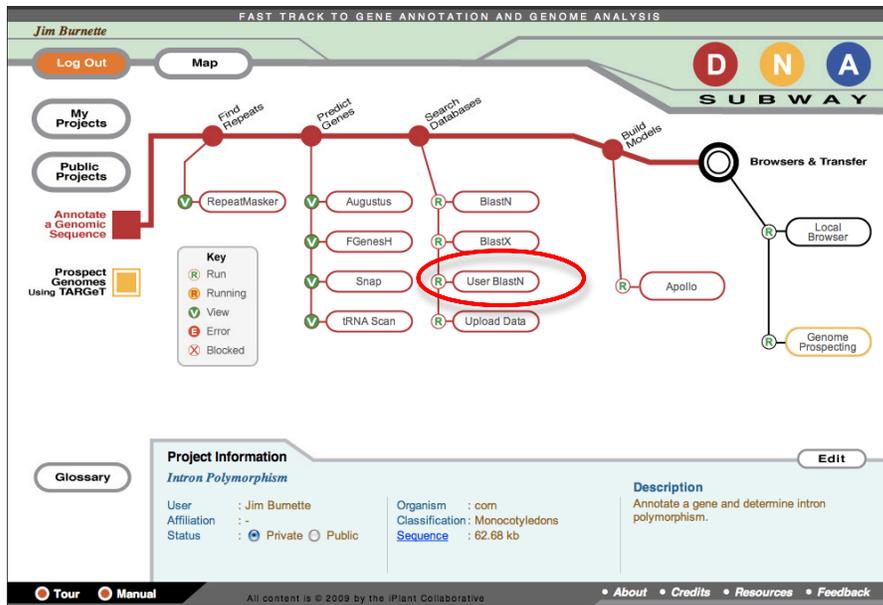
Step V. Find the gene of interest.

To find the gene your intron belongs to you will BLAST the gene region you just annotated. Although the form looks different the BLAST is the same program that you used on NCBI.

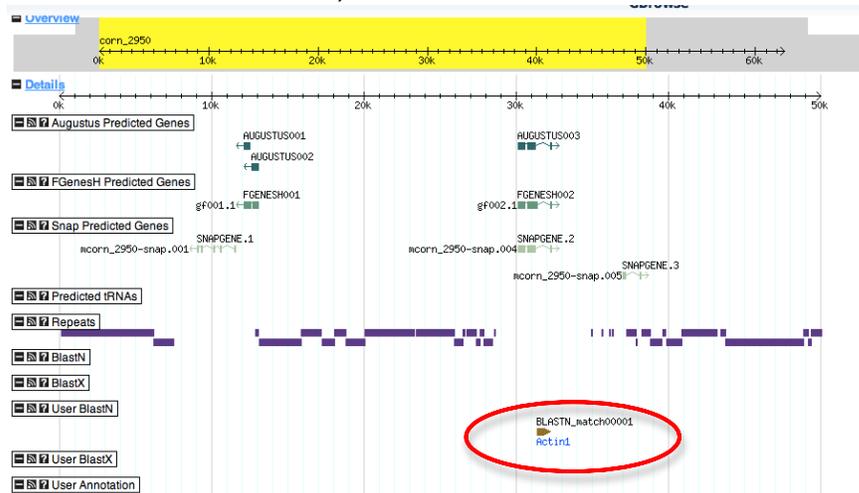
A. On the Search Databases sideline click the “Upload Data” station. Click on the “Choose File” button and the File Chooser select the file that contains the intron sequence. Then click Upload button.



B. A new station will appear on the Search Databases sideline called “User BlastN.” The intron sequence is now a Query sequence and you can now do a nucleotide BLAST against the genome. Click the User BlastN button to run the BLAST.

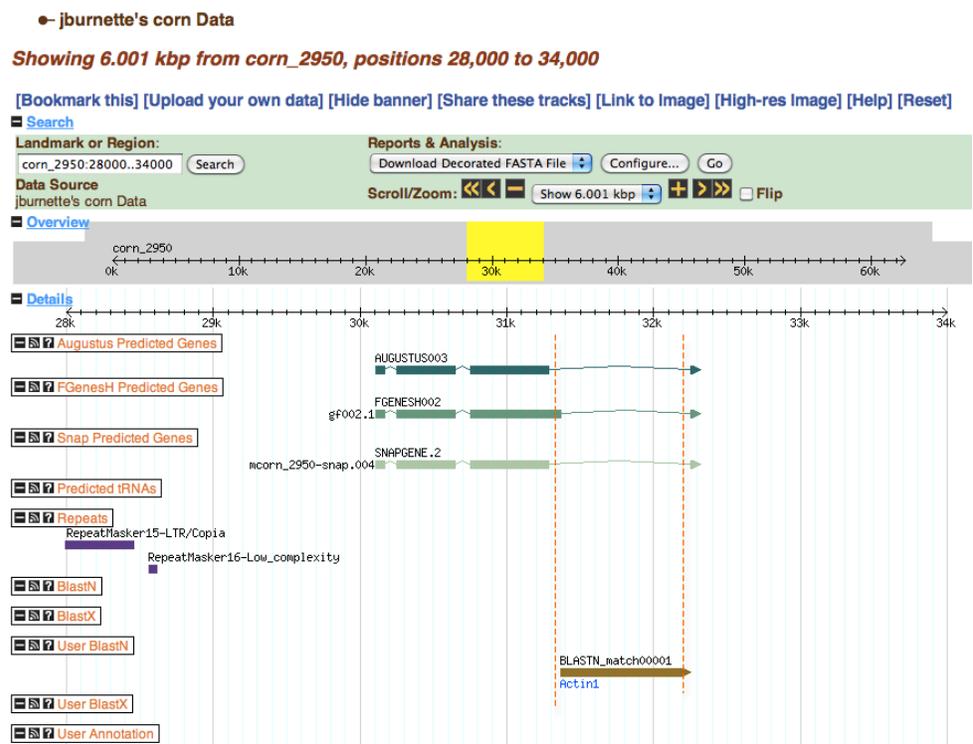
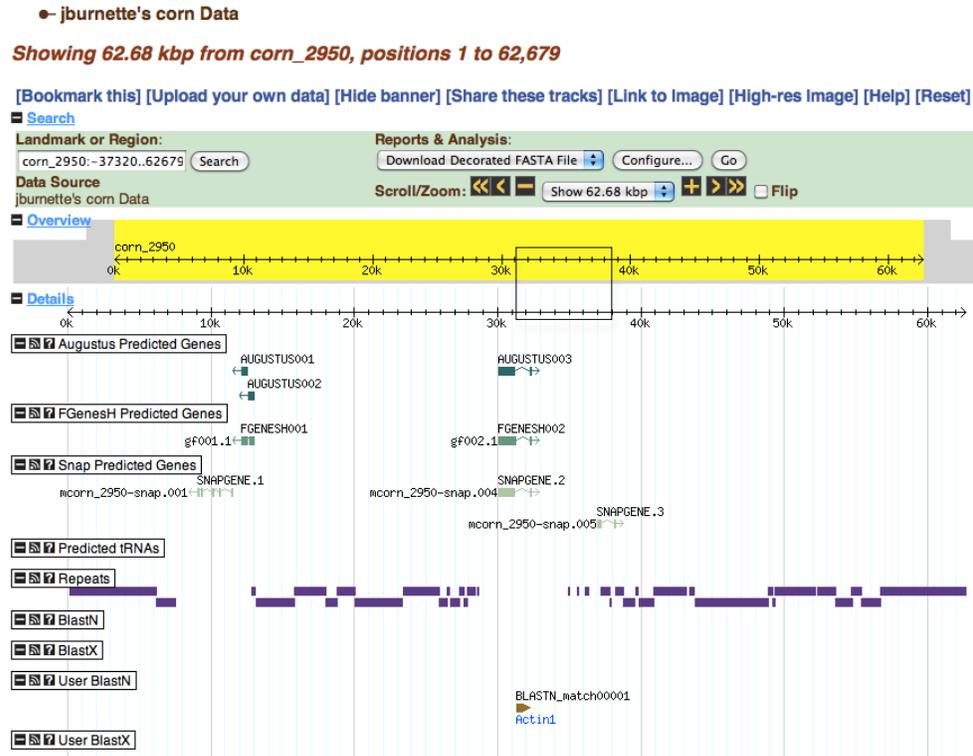


C. When the Blast finishes, view the result in the Local Browser.



So it looks like AUGUSTUS003, FGENESH002, and SNAPGENE.2 are Actin 1. But which intron are we studying? We need to zoom in on the gene region.

D. To zoom in Click and Drag the mouse in the Chromosome Track under Overview. See Black box. When you release the mouse GBrowse will refresh and zoom in on the region you selected. This refresh may take a minute so be patient.



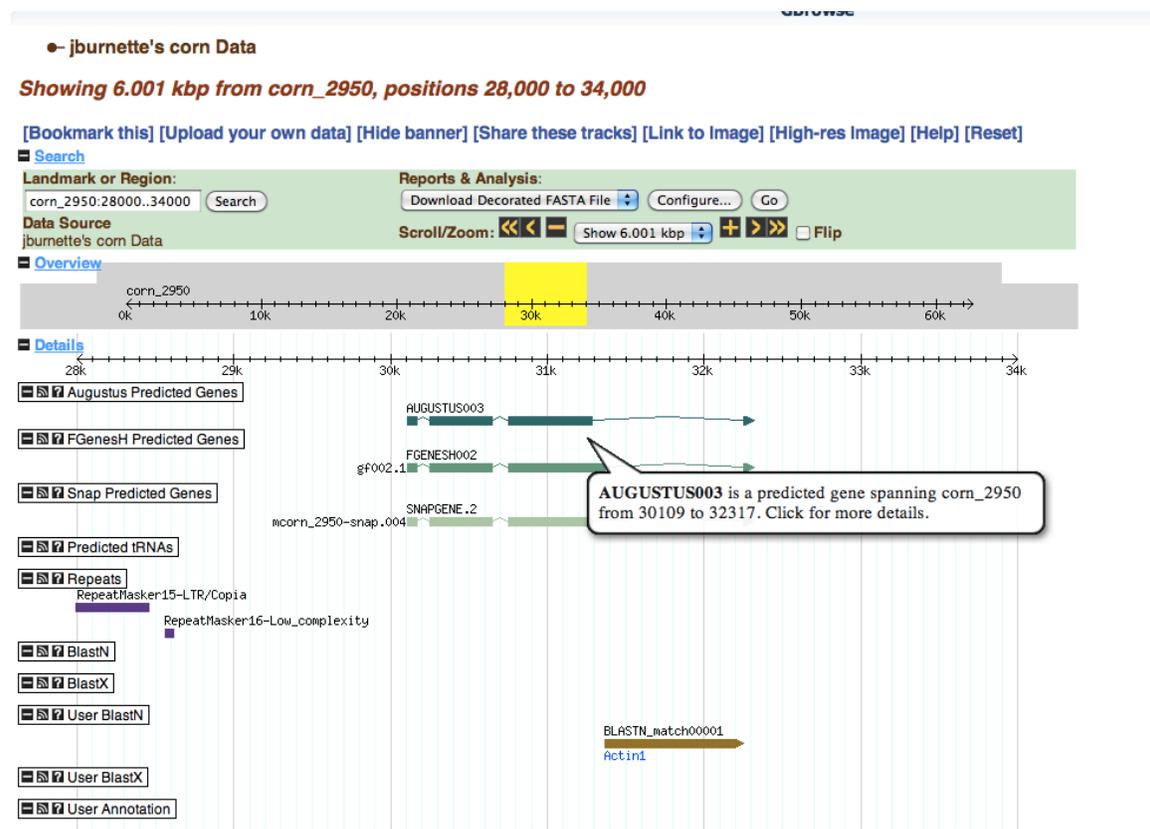
In the zoomed in view you can see that we are working with Intron 4 of Actin 1 as demonstrated by the dashed red lines.

Step VI. Obtain the DNA sequence of the gene.

We have come a long way. We started with some raw, unannotated sequence. We masked the “uninteresting” stuff. (It’s in quotes because it is going to get very interesting very soon.) We ran computer programs that predict genes. We skipped over the step called “Build Models” using Apollo. If you look closely in the zoomed in view you can see that the exon 3 predicted by FgenesH is slightly longer than the other two predictions as represented by the longer box. In another module you may learn how to include some experimental evidence and use Apollo to build what you think is the best gene model. This is the process of genome finishing and can only be done by humans.

Now we are ready to go back to the DNA sequence of Actin 1 and design our PCR experiment to explore intron polymorphism. We want to design PCR primers to the exons that flank the intron as shown in the next figure.

A. In the GBrowse window mouse over a gene prediction for Actin 1. You will see a bubble pop up.



B. Click the Bubble and then click “Show details.” The next window will take a moment to show up.

GBrowse

← jburnette's corn Data

Showing 6.001 kbp from corn_2950, positions 28,000 to 34,000

[Bookmark this] [Upload your own data] [Hide banner] [Share these tracks] [Link to Image] [High-res Image] [Help] [Reset]

Search

Landmark or Region: corn_2950:28000..34000 Search

Reports & Analysis: Download Decorated FASTA File Configure... Go

Data Source: jburnette's corn Data

Scroll/Zoom: <<< << < Show 6.001 kbp > >> >>> Flip

Overview

Details

Augustus Predicted Genes

FBESH002

SNRPGENE.2

RepeatMasker15-LTR/Copia

RepeatMasker16-Low_complexity

BlastN

BlastX

User BlastN

User BlastX

User Annotation

BLASTN_match00001

Act1nl

C. In the new window you will see much information about the gene prediction. Scroll down to towards the end of the report and you will see the sequence in purple and pink. The purple is the exon sequence and the pink is the intron sequence. Copy and paste the sequence of the exons you need in your bioinformatics notes (in the case of Actin 1 we need exons 3 and 4). Also copy the sequence including the intron into your notes. Be sure to label each sequence using a '>' comment line. Change the font to "Courier New" and the font size to 8pt.

```

1 2 3 4 5 6 7
>exon3
GTATTGT GATGGATTCC GGTGATGGT TGAGCCACAC GGTCCAATC TATGAAGGAT ACACACTTC TCATGCTATT
CTCCGTTTGG ACCTTGGCGG GCGTGACCTC ACCGACCACC TAATGAAGAT CCTCAGAG AGAGGGTACT CCCTCACTAC
GAGCCCTGAG CGGGAGATTG TCAGGGACAT AAAGGAGAAG CTTGCCTACG TTGCCCTTGA TTATGAACAG GAGCTGGAGA
TTGCCAGGAG CAGCTCCTCT GTCGAGAAGA GCTACGAGAT GCCTGATGGI CAGGTCATCA CCATTGGGTC AGAAAGTTF
AGGTGCCCGC AGGTGCTGTT CCAACCAATC CTTGTGGCA TGGAAATGCC TGGCGTCCAT GAGGCCACGT ACAACTCCAT
TATGAAGTGT GATGTTGATA TCAGGAAGGA TTTGTATGTT AACGTGTGCC TCAGTGGGGG ATCTACCATG TTCCCTGGGA
TTGCCGATCG TATGAGCAAG GAGATCACGT CCCTGGCTCC TAGCAGCATG AAG
>intron3-4
GTAAAG TGATTGCGCC ACCTGAAAGG
AAGTACAGTG TCTGGATCGG TGGCTCTATT TTGGCTCTCC TCAGCACITF TCAGCAGGTC TTCTCTTCT TACTTTATTC
TTTCATAAGC AGTAGGCTTT TATCATTCAA CTCCCTCTCA TGAGAACCTA CAGCAGCCTT GACCATGTTG GGTGTGTGTG
TTTTTTTTGT AGTTAAAGTG GAAATAACCC ATTATATACC AACTCAATTT AGGTTAAACT GGCTTGGAT GCACCTTTTT
TTACCTAAAA ACGCAATTGG AATACAATA TATTCTTGCT TTAGTCAGTT ACTCCAGTTF TCTAGTATCA ACTCTCTTA
GGGCCCATTT GGAACGAAGG AAATATGAAG GATTTTTGAA GGATTCATTC CCTACATAGA AGTTTCCTAT GTGACCCITF
GGAAACAGG ATTTAATCCT ATCAATTCCT TTGAAATTC TATGATTCG TTTCCCTCCAT ACAAGTTTTG GAGGAAATA
ACAAGAGCTC CAATCTCTTG GAAATTTCC TATGACTCTT ATCTCTCTAT TCAAATTCAT GTATAGCACC CAAAGGCTAG
GTGAGAAATF TTGCTACGTT CTCATTTCT GTAGGATTCG AATTCATAGG CAATTATAG CATGTGTTT CTCATTCCT
GTGTTTTTAT ATTCTCTGTF TCTAAAGGGG GGCTTAGTAG GCACCTGCACA TGTGCTCCA TCCTCCGCGT TGTGTTTAA
TCATTACGTT GTTAATCAT TACAGACATA GATTTAAATF TGATAGTAA ACCTTTTTGA GGAATGAAA AGATTTTAA
GCCTTTTCT TCACGATAGA TATAAGCGGG ATGTGGAAT GGAATTAATA GTGTGATAG ACCGAATATT TTTCTGCA
TGCCATTA CTGACTCATT ATTTAATTT CATTTCITG ACAATGGCA ACTGTGTTGG CAG
>exon3
ATGTGGA TCTCCAAGG
AGAGTATGAC GAAACTGGTC CGGGCATTGT CCACATGAAG TGCTTCTAA

```

(Note: In the DNA Subway, to return to the GBrowse window use the "Back" button of the web browser window.)

Step VI. Design Primers

You will need to amplify across the intron you are studying. To do this you need to pick primers in the exons that flank the intron. For Actin 1 we need to amplify the third intron so we will try to pick primers in Exons 3 and 4 as shown in the figure below.



You should have the gene sequence in your notebook. You need to calculate the length of the intron and the exons.

A. Open the TARGeT Class Portal: http://target.iplantcollaborative.org/class_index.php in a new browser tab. Click on the “Format Sequence and Count Nucleotides” link.

TARGeT: Tree Analysis of Related Genes and Transposons

Yujun Han, James Burnette, and Susan Wessler
Department of Plant Biology, University of Georgia, Athens, Georgia.

[Nucleic Acids Research Article](#)



Choose your TARGeT start page:

- [TARGeT using blastn with DNA query](#)
- [TARGeT using tblastn with Protein query](#)

Tools for Class

[Format Sequence and Count Nucleotides](#)
Formats sequences to FASTA format and number of nucleotides. Can also reverse and/or complement a sequence.

[Multiple Sequence Alignment](#)
Use MUSCLE to align 2 or more sequences.

[Primer3Plus](#)
Software for designing PCR primers.

Tools based on TARGeT

[Identify conserved regions of a multiple sequence alignment.](#)
This tool can help you to choose the sequences from the conserved region for a TARGeT search or other uses.

[Check whether a sequence is a transposable element](#)
This program can use the ends (100 bps for each end) of the query sequences to search a database. Homologs together with flanking sequences will be retrieved and aligned. You need to look at the multiple alignment file to determine whether it a TE or not. For example, there should be clear boundaries and the ends of a TE, and some unique TE structures such as terminal inverted repeat (TIR) and target site duplication (TSD) structures may also be found.

[Extract a DNA sequence \(with flanking sequences\) based on its position.](#)
Based on the position information that you input, this tool will give you the sequence with flanking regions from the genome database. This information is also available from the PHI step of TARGeT

Number of searches by 469 visitors.

Search	Number	Total
blastn	1.4x10 ⁵	3.5x10 ⁷ nt
tblastn	1.2x10 ⁵	4.7x10 ⁶ aa

Contact [TARGeT Help](#) with questions and comments.

Hosting and sysadmin provided by the iPlant Collaborative.



iPlant Collaborative
Empowering A New Plant Biology

Visitor Locations



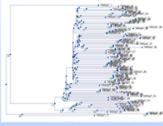
ClustrMaps™ Click to see

D. Go to the TARGeT Class Portal and click on the Primer3Plus link.

TARGeT: Tree Analysis of Related Genes and Transposons

Yujun Han, James Burnette, and Susan Wessler
Department of Plant Biology, University of Georgia, Athens, Georgia.

[Nucleic Acids Research Article](#)



Choose your TARGeT start page:

- [TARGeT using blastn with DNA query](#)
- [TARGeT using tblastn with Protein query](#)

Tools for Class

[Format Sequence and Count Nucleotides](#)
Returns sequence in FASTA format and number of nucleotides. Can also reverse and/or complement a sequence.

[Multiple Sequence Alignment](#)
Use MUSCLE to align 2 or more sequences.

[Primer3Plus](#)
Software for designing PCR primers.

Tools based on TARGeT

[Identify conserved regions of a multiple sequence alignment.](#)
This tool can help you to choose the sequences from the conserved region for a TARGeT search or other uses.

[Check whether a sequence is a transposable element](#)
This program can use the ends (100 bps for each end) of the query sequences to search a database. Homologs together with flanking sequences will be retrieved and aligned. You need to look at the multiple alignment file to determine whether it is a TE or not. For example, there should be clear boundaries and the ends of a TE, and some unique TE structures such as terminal inverted repeat (TIR) and target site duplication (TSD) structures may also be found.

[Extract a DNA sequence \(with flanking sequences\) based on its position.](#)
Based on the position information that you input, this tool will give you the sequence with flanking regions from the genome database. This information is also available from the PHI step of TARGeT

Number of searches by 469 visitors.

Search	Number	Total
blastn	1.4x10 ⁵	3.5x10 ⁷ nt
tblastn	1.2x10 ⁵	4.7x10 ⁶ aa

Contact [TARGeT Help](#) with questions and comments.
Hosting and sysadmin provided by the iPlant Collaborative.

E. The Primer3Plus page looks like this.

Primer3Plus

pick primers from a DNA sequence

[Primer3Manager](#) [Help](#)
[About](#) [Source Code](#)

Task: Detection *Select primer pairs to detect the given template sequence. Optionally targets and included/excluded regions can be specified.* Pick Primers Reset Form

Main
 General Settings
 Advanced Settings
 Internal Oligo
 Penalty Weights
 Sequence Quality

Sequence Id:

Paste source sequence below Or upload sequence file: no file selected

GTATTGTGATGGATCCGGTGTGTTGAGCCACAGGTTCCAAATCTATGAAGGATACACACTTCTCATGCTATTCTC
 CGTTTGGATCTTCGGGGCTGACCTCAGCAGCACCTAATGAGATCTCAGAGAGAGGGTACTCCCTGACTACGAG
 CGCTGAGGGGAGATTTGACGGACATAAAGGAGAGCTTGCCTACGTGCCCCGATGATGAACAGGAGCTGGAGACTG
 CCAGGAGAGCTCTCTCTCGAGAGAGCTACGAGATCCCTGATGCTCAGGTCATCACCATTGGGTGAGAAAGGTTTAG
 TGCCCGAGGTGCTTCCAAACATCCCTTGTGGCATGGAATCGCCTGGCGTCCATGAGGCCACGTACAACTCCATCAT
 GAGTGTGATTTGATACAGGAAGATTGTATGGTAACCTTGTCTCAGTGGGGATCTACCATGTTCCCTGGGATTG
 CGATCTGTGAGGAGGAGATCACTGCTGGCTCTCAGGAGCATGAGTAAAGTGGTTCGCCACCTGAAAGGAG
 TACAGTCTGATGGTGGCTCTATTGCTCTCTCAGCACTTTCAGCAGGCTCTCTCTTCTTACTTATTGCTTT
 CATAAGCAGTAGGCTTTATCATTCACCTCTCTCATGAGAACCTACAGCAGCTTGACCATGTTGGGTGTGTGCTT
 TTTTGTAGTTAAGTGGAAATACCCATATATACCACCTCAATTAGGTTAAACTGGCTTGTGATGCACTCTTTTTTA
 CCTAAACCCCATTTGGAAATACAAATATCTCTCTTGTAGTGGTTCAGTTCTTCTGATATCAACTCTCTTGGG
 CCCATTTGGAAGGAGAAATGAGGATTTTGAAGGATTCATCCCTACATAGAAGTTCCCTATGTGACCCCTTGGGA

Mark selected region:

Excluded Regions: < >

Targets: []

Included Region: { }

Pick left primer or use left primer below.

Pick hybridization probe (internal oligo) or use oligo below.

Pick right primer or use right primer below (5'→3' on opposite strand).

a. Copy and paste the sequence for the exons and intron from the GBrowse window into the large textbox. Copying from Google Docs to the web form can create problems so use GBrowse. Enter a short Sequence Id in the small textbox.

b. Enter the Targets coordinates. You need to tell Primer3Plus the location of the intron so that the primers will flank the intron. The first number is the first nucleotide of the intron, 531 for the case of Actin 1. The next number is the length of the intron, which is 970 for intron 3 of Actin 1.

c. Click the “General Settings” tab.

Primer3Plus
pick primers from a DNA sequence

[Primer3Manager](#) [Help](#)
[About](#) [Source Code](#)

Task: Select primer pairs to detect the given template sequence. Optionally targets and ~~included~~excluded regions can be specified.

Main **General Settings** **e Advanced Settings** **Internal Oligo** **Penalty Weights** **Sequence Quality**

d **Product Size Ranges**

Primer Size Min: Opt: Max:

Primer Tm Min: Opt: Max: [Max Tm Difference:](#)

Primer GC% Min: Opt: Max: [Fix the](#) [prime end of the primer](#)

[Concentration of monovalent cations:](#) [Annealing Oligo Concentration:](#)

[Concentration of divalent cations:](#) [Concentration of dNTPs:](#)

[Mispriming/Repeat Library:](#)

Load and Save
[Please select special settings here:](#) (use Activate Settings button to load the selected settings)
To upload or save a settings file from your local computer, choose here:
 no file selected

d. Delete all of the numbers in the Product Size Ranges. Enter 500-2000 as the range of PCR product sizes. This means the smallest product can be 500 nt and the largest can be 2000.

e. Click on the “Advanced Settings” tab.

Primer3Plus
pick primers from a DNA sequence

Primer3Manager Help
About Source Code

Task: Detection Select primer pairs to detect the given template sequence. Optionally targets and included/excluded regions can be specified. g Pick Primers Reset Form

Main General Settings **Advanced Settings** Internal Oligo Penalty Weights Sequence Quality

Max Poly-X: 5 Table of thermodynamic parameters: Breslauer et al. 1986
 Max #N's: 0 Salt correction formula: Schmitt and Lifson 1965
 Number To Return: 5 **f** GC Clamp: 2
 Max Self Complementarity: 8.00 Max 3' Self Complementarity: 3.00
 Max 3' Stability: 9.0
 Max Repeat Mispriming: 12.00 Pair Max Repeat Mispriming: 24.00
 Max Template Mispriming: 12.00 Pair Max Template Mispriming: 24.00
 Left Primer Acronym: F Internal Oligo Acronym: IN
 Right Primer Acronym: R Primer Name Spacer: -

Product Tm Min: Opt: Max:

Use Product Size Input and ignore Product Size Range Warning: slow and expensive!
 Product Size Min: 100 Opt: 200 Max: 1000

Liberal Base Do not treat ambiguity codes in libraries as consensus Use Lowercase Masking

Sequencing
 Lead Bp: 50 Spacing Bp: 500
 Accuracy Bp: 20 Interval Bp: 250
 Pick Reverse Primers

f. Change the “GC Clamp from ‘0’ to ‘2.’ This tells Primer3Plus that the last two nucleotides at the 3’ end must be a ‘G’ or ‘C.’

g. Click the “Pick Primers” button. With any luck Primer3Plus will be successful at finding primers that meet our criteria.

Primer3Plus Results

< Back

Pair 1:

Left Primer 1: Actin1_exon3-4_F

Sequence: ATGAGGCCACGTACAACCTCC

Start: 376 Length: 20 bp Tm: 60.0 °C GC: 55.0 % ANY: 4.0 SELF: 0.0

Right Primer 1: Actin1_exon3-4_R

Sequence: GCACCTCATGTGGACAATGC

Start: 1560 Length: 20 bp Tm: 60.1 °C GC: 50.0 % ANY: 5.0 SELF: 2.0

Product Size: 1185 bp Pair Any: 4.0 Pair End: 2.0

Send to Primer3Manager Reset Form

1	GTATTGTGAT	GGATTCCGGT	GATGGTGTGA	GCCACACGGT	TCCAATCTAT
51	GAAGGATACA	CACCTCCCTCA	TGCTATTCTC	CGTTTGGATC	TGCGGGGCG
101	TGACCTCACC	GACCACCTAA	TGAAGATCCT	CACAGAGAGA	GGGTACTCCC
151	TCACTACGAG	CGCTGAGCGG	GAGATTGTCA	GGGACATAAA	GGAGAAGCTT
201	GCCTACGTTG	CCCTTGATTA	TGAACAGGAG	CTGGAGACTG	CCAGGAGCAG
251	CTCCTCTGTC	GAGAAGAGCT	ACGAGATGCC	TGATGGTCAG	GTCATACCCA
301	TTGGGTCAGA	AAGGTTTAGG	TGCCCCGAGG	TGCTGTTCCT	ACCATCCCTT
351	GTTGGCATGG	AATCGCCTGG	CGTCCATGAG	GCCACGTACA	ACTCCATCAT
401	GAAGTGTGAT	GTTGATATCA	GGAAGGATTT	GTATGGTAAC	GTTGTCCCTCA
451	GTGGGGGATC	TACCATGTTC	CCTGGGATTG	CCGATCGTAT	GAGCAAGGAG
501	ATCACGTCCC	TGGCTCCTAG	CAGCATGAA	GTTAAAGTGA	TTGCGCCACC
551	TGAAAGGAAG	TACAGTGTCT	GGATCGGTGG	CTCTATTTTG	GCTTCTCTCA
601	GCACTTTTCA	GCAGGTCTTC	TCTTCTTAC	TTTATTCTTT	CATAAGCAGT
651	AGGCTTTTAT	CATTCAACTC	CTCTTCATGA	GAACCTACAG	CAGCCTTGAC
701	CATGTTGGGT	GTGTGTGCTT	TTTTTGTAGT	TAAAGTGAA	ATAACCCATT
751	ATATACCAAC	TCAATTTAGG	TTAAACTGGC	TTGTGATGCA	CTCTTTTTTA
801	CCTAAAAACG	CAATTGGAAT	ACAAATATAT	TCTTGCTTTA	GTCAGTACT
851	CCAGTTTCT	AGTATCAACT	TCTCTTAGGG	CCCATTGGA	ACGAAGGAAA
901	TATGAAGGAT	TTTTGAAGGA	TTTATTCCCT	ACATAGAAGT	TTCCATATGTG
951	ACCCTTTGGA	ACACAGGATT	TAATCCTATC	AATTCCTTTG	AAATTTCTAT
1001	GGATTGCTTT	CCTCCATACA	AGTTTGGAG	GAAAATAACA	AGAGCTCCAA
1051	TCTCTTGGA	AATTCCTAT	GACTCTTATC	TCTCTATTCA	AATTCATGTA
1101	TAGCACCCAA	AGGCTAGGTG	AGAAATTTTG	CTACGTTCTC	AATTTCTGTA
1151	GGATTGCAAT	TCATAGGCAA	TTATAGTCAT	GTGTTTCTC	TATTCCTGTG
1201	TTTTCATATT	CTTCTGTTCT	AAAGGGGGGC	TTAGTAGGCA	CTGCACATGT
1251	TGCTCCATCC	TCCGCGTTGT	TGTTAATCA	TTACGTTGTT	TAATCATTAC
1301	AGACATAGAT	TTAAATTTGA	TAGTAACATC	TTTTTGAGGA	AATGAAAAGA
1351	TTTTAAGGCC	TTTTTCTTCA	CGATAGATAT	AAGCGGGATG	TGGAAATGGA
1401	ATTACTAGTG	TGATAGCACC	GAATATTTTT	CTTGCATTGC	CATTACTTGC
1451	ACTCATTATT	TTAATTTTCA	TTCTTTGACA	ATGGGCAACT	GTGTTGGCAG
1501	ATGTGGATCT	CCAAGGGAGA	GTATGACGAA	ACTGGTCCGG	GCATTGTCCA
1551	CATGAAGTGC	TTCTAA			

If Primer3Plus is successful at finding primer sequences based on the criteria you specified in addition to dozens of thermodynamic modeling parameters you will see a page similar to the one above. At the top is the primer pair that Primer3Plus considers the best. The location on the sequence is also shown. The Left primer (also called the Forward primer) sequence is the same as the sequence highlighted in purple. The Right Primer (also called the Reverse primer) is the reverse complement of the sequence highlighted in yellow. The sequence highlighted in green is the region that we skipped over and represents intron 3. Why do you think the reverse primer is reverse

complemented? You will order this primer pair using instructions provided by your instructor.

You should copy and paste the sequences of the primer pair in your Google Doc and also note all of the parameters you changed.

Step VII. Design the PCR Experiment.

When you receive your primers you will design your PCR experiment. Things to think about when designing your experiment:

1. What positive controls do you need to include?
2. What negative controls do you need to include?
3. How many inbred lines should you include?
4. How many reaction is will you make the cocktail for?