PBIO/BIOL2240L: The Dynamic Genome
Freshman College Summer Experience
Summer 2009

Table of Contents

Syllabus
Grading Policy

Syllabus Summer 2009

HHMI

Syllabus   Grading   Data

	Lecture Topic	Lab Work	Due dates	Downloads
Mon, July 6	• Assessment • Course Overview • Mechanics	Pipettors		Start Survey Handout
Tues, July 7	Info transfer	DNA Extraction		Exp. 1
Weds, July 8	Intro PCR	PCR, pour gel		
Thurs, July 9	• Review • Analyze results	Gel of PCR		
Fri, July 10	• Quiz • Discussion Making of Fittest			

Mon, July 13	mPing background	Observe Plants		mPing Handout Slides
Tues, July 14	Review	Extract DNA		
Weds, July 15	PubMed	PCR		PubMed Handout
Thurs, July 16	blantn Analyze results	Gel of PCR		
Fri, July 17	Quiz/Mid-Term Projects			
Mon, July 20	Maize Browser Dr. Schuster			
Tues, July 21	Primer Design Finish Annotation			Browser and Primer Design Handout
Weds, July 22	Greenhouse Tour Sequencing Facility Tour			
Thurs				

Thurs, July 23		PCR, pour gel		
Fri, July 24	Quiz	Gels, more PCR		
Mon, July 27		Lab		
Tues, July 28		Lab		Homework 4
Weds, July 29		Lab		
Thurs, July 30	Design Poster Course Review	Lab		

Fri, July 31	Final
	
	Presentation
		• Surveys:
		• Post-Survey
		• CURE
		  ○ user: cureresp
		  ○ password: blue2white
		• FCE
		  ○ 4 FCE evaluations		

BIO/PBIO 2240L The Dynamic Genome
Summer 2009
Dr Susan Wessler and Dr Jim Burnette
Eleanor Kuntz, TA
Course website: http://www.dynamicgenome.org/classes/fcse/
User: dynamicgenome
Password: tesjump

	Dr. Susan Wessler	Dr. Jim Burnette	Eleanor Kuntz
Office	Plant Sciences 4510	Plant Sciences 1506	Plant Sciences 4505
Phone	706-542-1870	706-542-4581	706-542-1857
Hours	By appointment	By appointment	By appointment
E-mail	sue@plantbio.uga.edu	jburnette@plantbio.uga.edu	ekuntz@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 pulled 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	Friday, July 30, 2009
Final Presentation	Friday, July 31, 2009

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

Quizzes: 45 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	20%
	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.

What is the Genome?

Chromatin

1400 nm

Metaphase chromosomes

DNA double helix

2 nm

eight
histone proteins

Nucleosome

30 nm

Chromatin

300 nm

700 nm

Sugar–phosphate backbone

Base pair

Base + Ribose or deoxyribose = Nucleoside + P Phosphate = P Nucleotide

Pyrimidines

Cytosine (C)   Thymine (T)   Uracil (U)

Purines

Adenine (A)   Guanine (G)
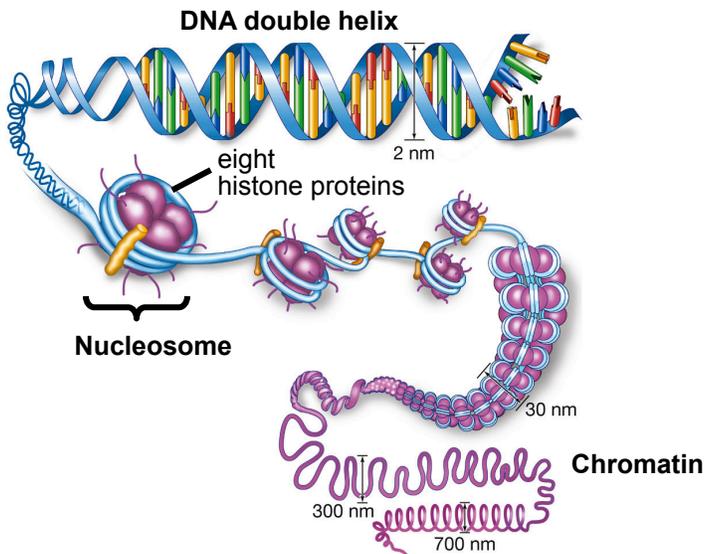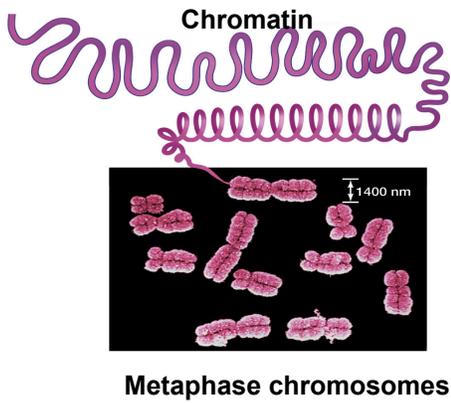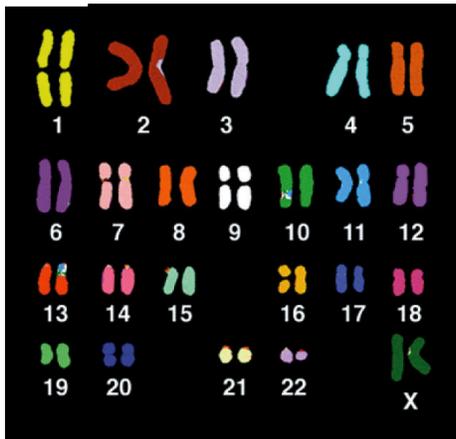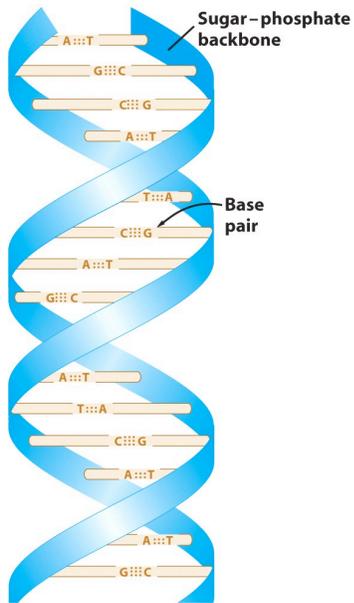
LIFE 8e, I
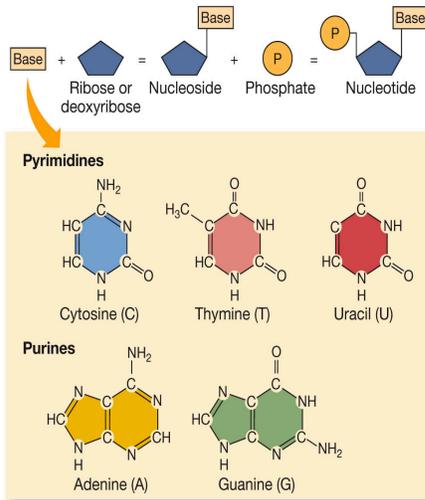
A

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.

C

Adenine   Thymine   Guanine   Cytosine

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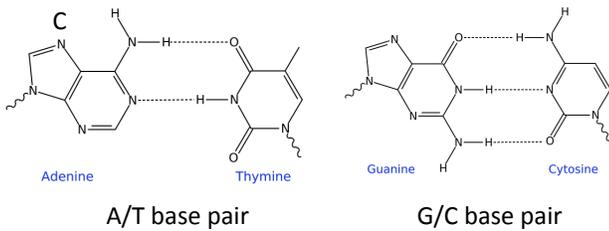
The two strands are held together by hydrogen bonding between the A/T base pairs and the G/C base pairs.

How is DNA and RNA similar? Different?

DNA

Pyrimidine base
Deoxyribose
Purine base

5′ end

3′ end

Phosphate

5′ end

Hydrogen bond

RNA

3′ end

Phosphate

Ribose

Phosphodiester linkage

5′ end
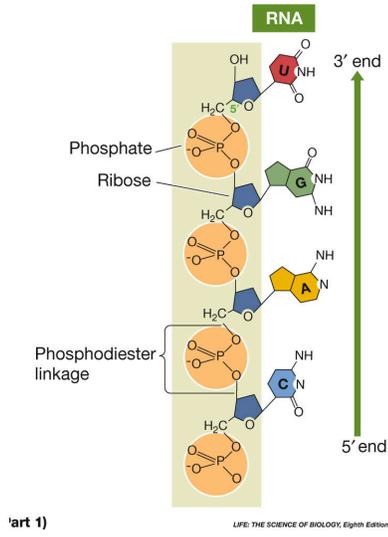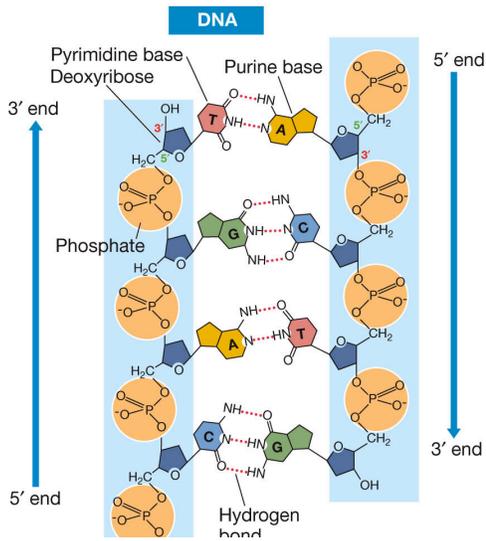
(Part 1)

LIFE: THE SCIENCE OF BIOLOGY, Eighth Edition © 2007 Sinauer Associates, Inc. and W. H. Freeman & Co.
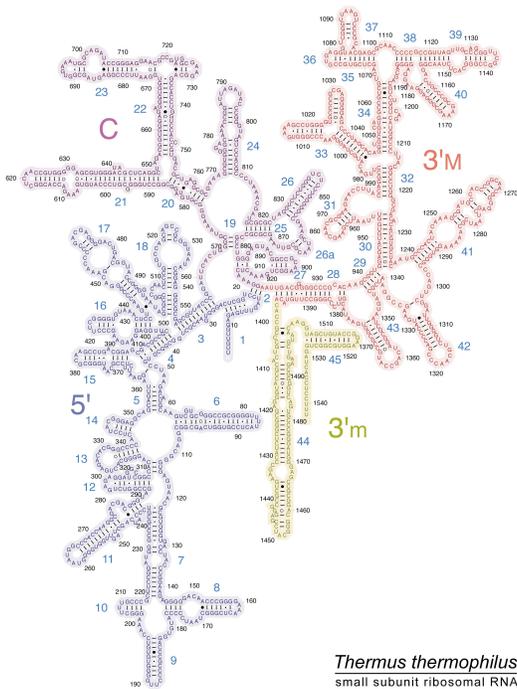
Thermus thermophilus
small subunit ribosomal RNA

aa₁   aa₂   aa₃

Amino end

Carboxyl end

H — N — C — C — N — C — C — N — C — C — OH + 2(H₂O)

aa₁   aa₂   aa₃

Peptide bond   Peptide bond

Secondary structure

Hydrogen bonds between amino acids at different locations in polypeptide chain

α helix

Pleated sheet
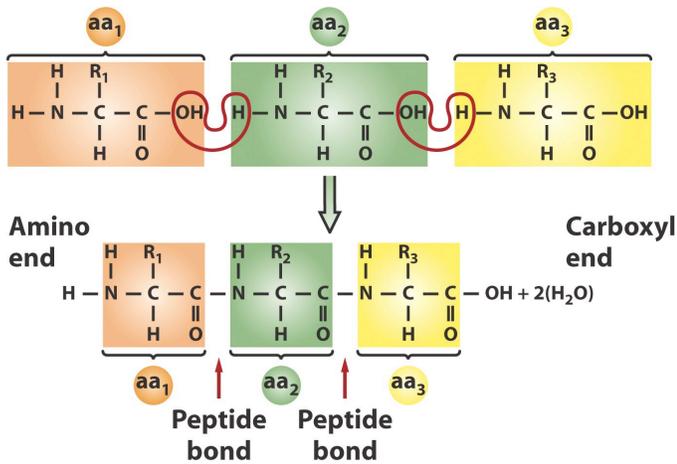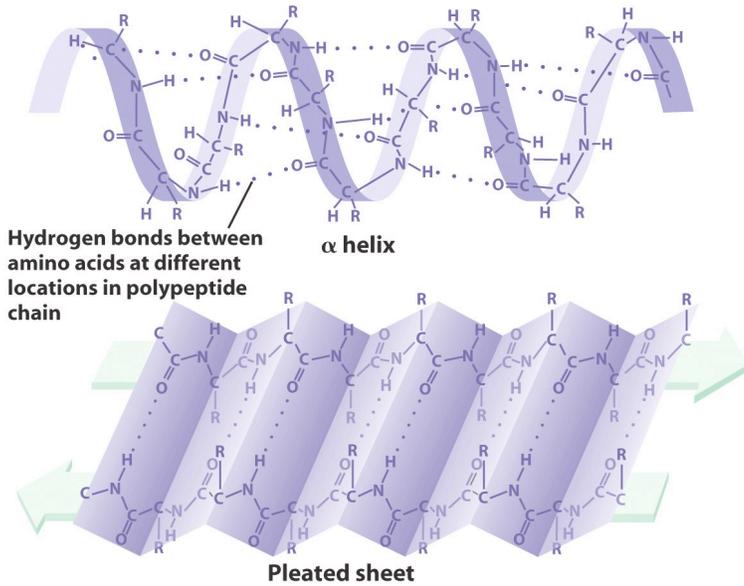
Tertiary structure

Heme

β polypeptide

Quaternary structure

β   β

Heme group

α   α

Replication   DNA   Transcription   RNA   Translation   Protein

Reverse
Transcription

DNA
replication

Mitosis

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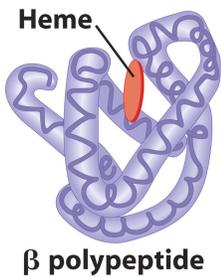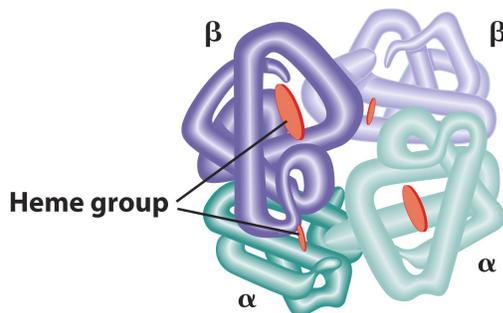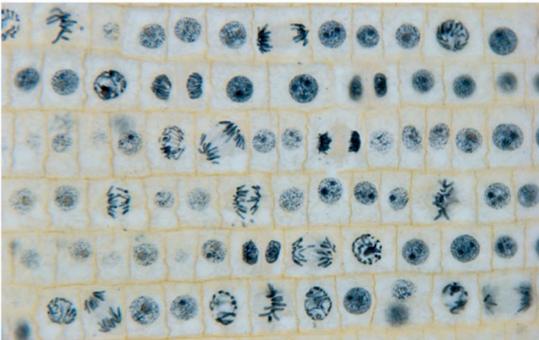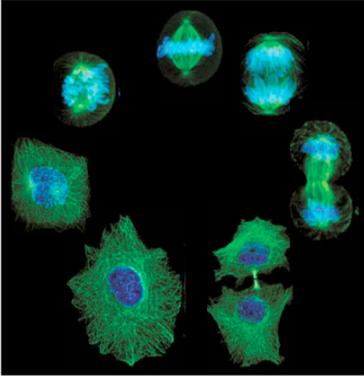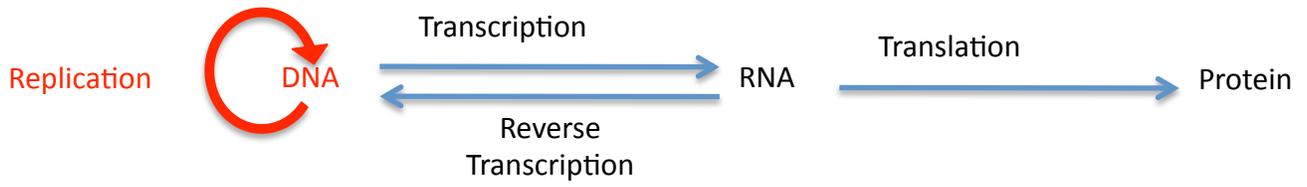
Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

Replication

Transcription

DNA ⇌ RNA

Reverse Transcription

Translation → Protein

6

The two strands of the parental double helix unwind, and each specifies a new daughter strand by base-pairing rules

Old

New

Topoisomerase

Replication fork movement

Helicase

Next Okazaki fragment will start here.

RNA primer

Primase

RNA primer

Okazaki fragment

Single-strand binding proteins

Clamp

DNA polymerase I

Leading strand

DNA polymerase III dimer

Lagging strand

Ligase

DNA template strand

5′   3′

DNA template strand

5′   3′

PP$_i$

T   5′

T   5′

Replication   Transcription   RNA   Translation   Protein
                        DNA
                   Reverse
                   Transcription

Gene 1   Gene 2   Gene 3

3' ⌇⌇⌇ 5'   3' ⌇⌇⌇ 5'

5'   3'

3'   5'

5' ⌇⌇⌇ 3'

Template strand

Unwinds   Rewinds   Template strand
                                                        of gene 2
                                        RNA
                        3'   5'   G T A G C C C A T   A
3'   U   C A U C G G G U A   A G
5'   U A C G G A U G C A   A U C
DNA   A T G C C T A C G T   3'
                                                5'
RNA   Template strand   RNA   Nontemplate strand   RNA
polymerase   of gene 1   polymerase   of gene 2   polymerase

        Gene 1   Gene 2

Coding strand 5' — CTGCCATTGTCAGACATGTATACCCCGTACGTCTTCCCGAGCGAAAACGATCTGCGCTGC — 3'   ⎫ DNA
Template strand 3' — GACGGTAACAGTCTGTACATATGGGGCATGCAGAAGGGCTCGCTTTTGCTAGACGCGACG — 5'   ⎭

5' — CUGCCAUUGUCAGACAUGUAUACCCCGUACGUCUUCCCGAGCGAAAACGAUCUGCGCUGC — 3' mRNA

Replication   DNA   Transcription   RNA   Translation   Protein

Reverse
Transcription

EUKARYOTE

Nucleus

3'   Primary
transcript

5'

Processing

CAP   mRNA

Transport

Computer model

Polypeptide

50S

E **P** **A**

5'   **30S**

mRNA   **3'**

Replication   Transcription   DNA   RNA   Translation   Protein

Reverse
Transcription

3'
OH

Amino acid
attachment site

5' p

mG

UH₂

DHU loop

UH₂

m₂G

UH₂

Anticodon loop

3'

Anticodon loop

5' Codon for alanine 3' mRNA

Codon	`AUU	GCU	CAG	CUU	GAC

Nonoverlapping   AUU   GCU   CAG   . . .
code   aa₁   aa₂   aa₃

Second letter

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

Third letter

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

Putting it all together….

Replication — DNA — Transcription → RNA — Translation → Protein

Reverse Transcription

Nucleus
Gene

DNA

Transcription

Transcription completed

Pre-mRNA

Processing

mRNA

Inside of cell

Translation

Polypeptide — Ribosome

What are the components of a gene?

Promoter

Exon

Intron

Promoter   Transcription unit

5′
3′

Start point   DNA

RNA polymerase

❶ Initiation

5′
3′

Unwound   RNA   Template strand
DNA   transcript   of DNA

❷ Elongation

Rewound
DNA

5′
3′

5′

RNA
transcript

❸ Termination

5′
3′

5′   Completed RNA transcript   3′

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

Genomic DNA

Transcription

pre-mRNA

Splicing

Spliced pre-mRNA

Cap and poly-A addition

7-methyl-G-Cap AAAAAAAA mRNA

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

Details of pre-mRNA and mRNA
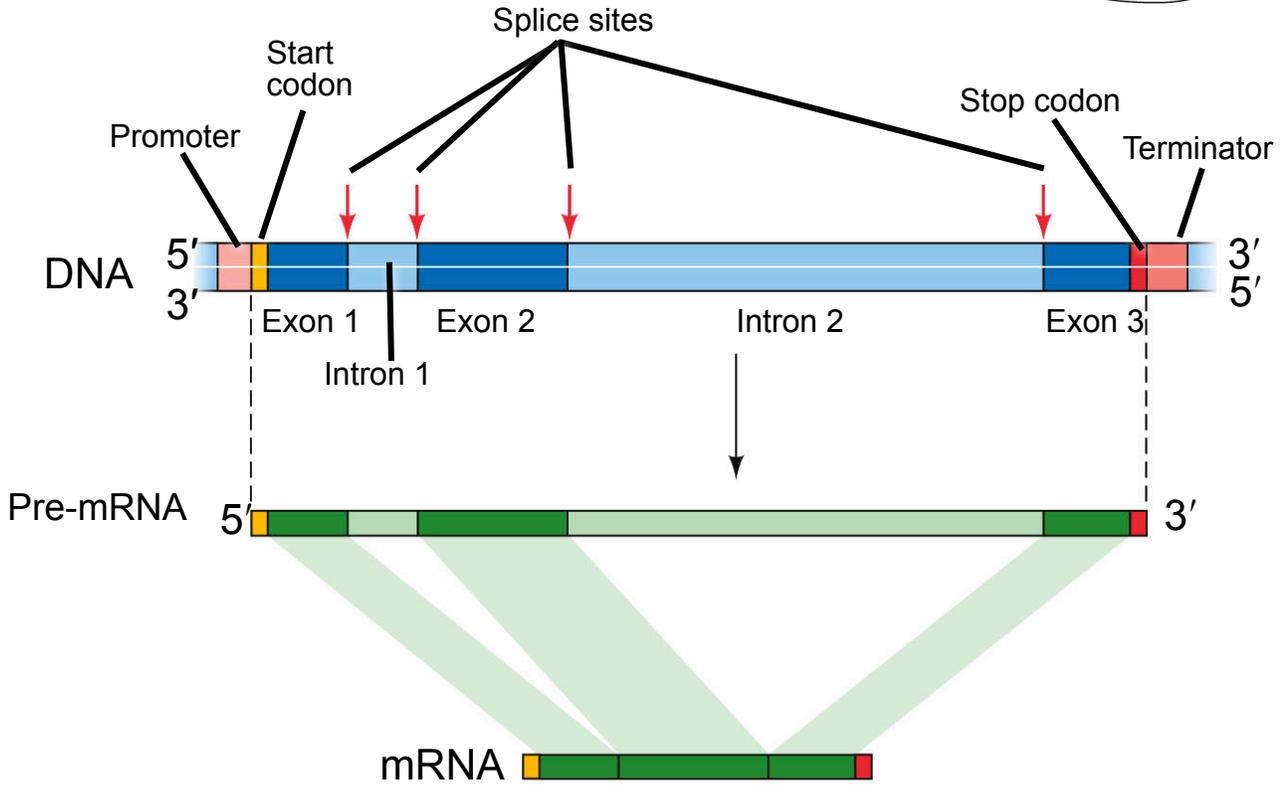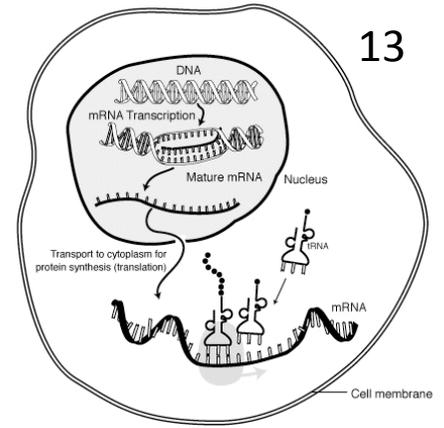
Experiment I: Analysis of the Actin gene of maize.

You will use the actin gene to investigate the differences between the DNA gene sequence and mRNA sequence. The actin gene encodes the actin protein and is found in all eukaryotes. Actin polymerizes to forms long strands that can contract. Actin has a large role in cell division and cell movement. Because the protein of actin is so important the location of 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.

You will analyze both the actin gene from DNA and the mRNA product of the actin gene. By comparing the results of the PCR and the actual nucleotide sequence you will be able to determine the location of the introns in the actin gene.

There are four steps in the experiment to analyze the DNA: DNA and RNA extraction, polymerase chain reaction (PCR) amplification of the actin gene, gel analysis, sequencing.

You will be using PCR amplification in most of your experiments. For this reason, it is summarized below.

Figure 20-14
Introduction to Genetic Analysis, Ninth Edition
© 2008 W. H. Freeman and Company
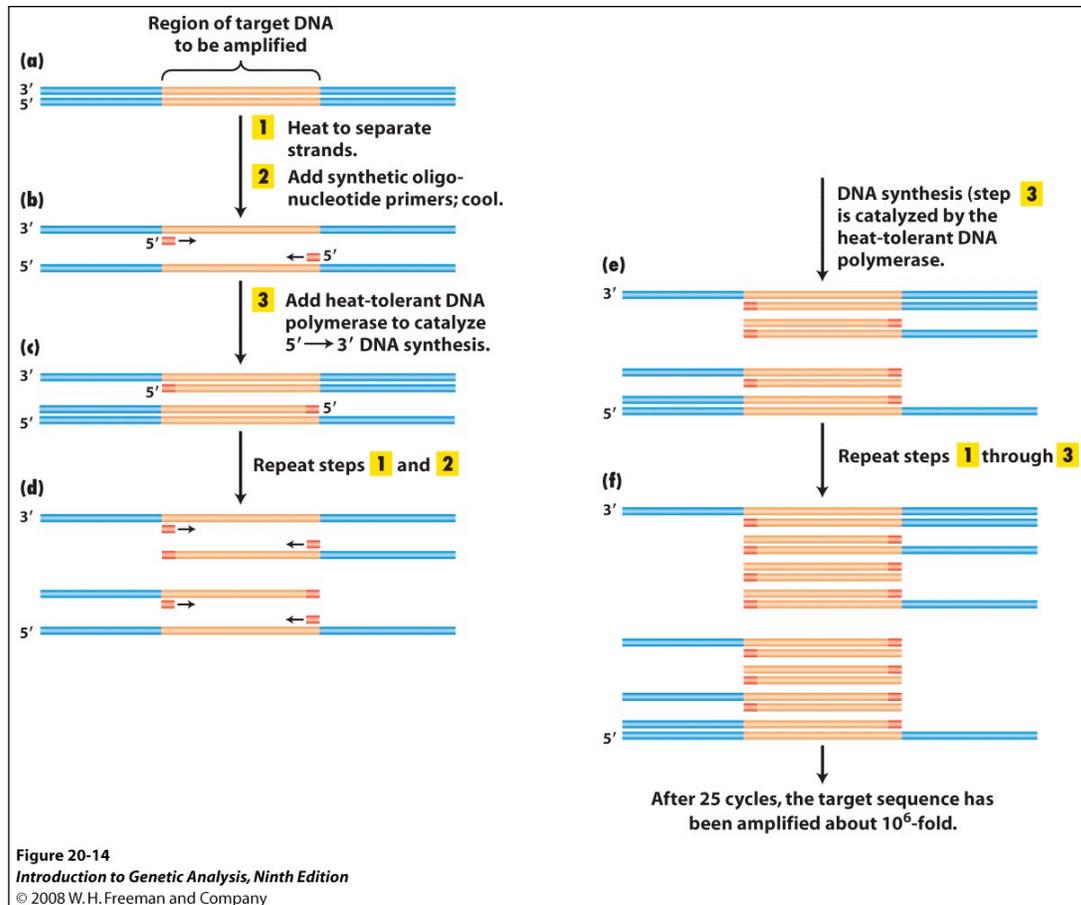
Figure 5: Details of PCR. See below.

<u>Polymerase Chain Reaction.</u> Known familiarly as 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 copied.

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

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

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 or dNTPs) that are in the test tube.

D,E and F - The procedure is repeated beginning with denaturation then cooling, annealing, extension etc.

Oligonucleotide primer. A primer is a short nucleic acid strand or a related molecule 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. (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

mRNA analysis by PCR
The polymerases used for PCR require a DNA template. That means we cannot use mRNA directly in a PCR reaction. We must first isolate RNA from a tissue and then reverse transcribe the RNA into DNA using the enzyme reverse transcriptase (RT). You will learn about the natural function of

reverse transcriptase later in the course. 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. This first strand of DNA is referred to as cDNA.

Due to time constraints you will not be able to isolate RNA and make cDNA. You will amplify the actin gene using PCR and primers for the actin gene from genomic DNA you extract and from cDNA made by the instructors. You will analyze the sizes of the PCR products from genomic DNA and cDNA.

You will then analyze the sequence of the to PCR products to determine the exon/intron boundaries of the first intron of the actin gene. This is how the gene structure of any gene can be determined experimentally. Later in the course you will learn about computer programs that predict gene structure.

Exercises:
1. Does mRNA contain the exons or introns from a gene?

2. Will the PCR band from the cDNA be larger or smaller than the band from genomic DNA?

Step I: Tuesday, July 7, 2009 DNA Extraction

You will extract DNA from two lines of maize: B73 and the popcorn Palomero. Be sure to store the DNA in the freezer because you will use it this week and in the last two weeks of the semester.

Damon Lisch's All Natural (no organics) Genomic Miniprep

Materials list:
Extraction Buffer
10% SDS
5M KOAC
100% Isopropanol
70% Ethanol
Ice Bucket with ice
liquid nitrogen
65°C water bath
sterile 1.5 ml tubes (2 for each prep)

1) Label 3 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 120 μl of 10% SDS. Mix by inverting.

5) Place at 65°C for 10 minutes. Repeat steps 1-5 for the second sample.

6) Add 300 μl 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 μl 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 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.

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

10) 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.

11) Resuspend the DNA in 50 μl water. Let sit at RT for about 5 minutes, then mix by pipetting.

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 1.0 µl of a stock solution of 100 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 15 µl of DNA from step 11 into each tube.

8) Add 3 µl of 6x loading dye to the tube. Tap the tube gently to mix.

9) Load all 18µl on your gel. Keep track of which sample went in which lane.

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

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

12) Photograph the gel.

Step II: Wednesday, July 8,2009 Amplify genomic DNA using PCR.

Today you will amplify the DNA using 1 pair of primers for the actin gene.

Each person will do PCR on the B73 and Palomero genomic DNA isolated on Tuesday and on cDNA provided by the instructor for a total of four DNA samples to analyze. We also need an additional negative control that will be water in place of DNA, giving you five reactions in total. You need to make a master mix of six reactions to make sure you have enough for the five tubes.

10 fold dilution of DNA with water
1) Label a tube with DNA name and '1/10 dil'.

2) Add 90µl of water to the tube.

3) Add 10 μl of DNA to the tube.

4) Repeat for each DNA sample.

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 '6X (stands for 6-fold).' Keep this on ice.

	1X (μl)	6X
2x Q Master Mix	12.5	75.0
H₂O	5.5	33.0
Forward Primer	1.0	6.0
Reverse Primer	1.0	6.0
DNA	5.0	------
Total	25.0	120.0 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. The mix is green because it contains a DNA loading dye.

4) Put 20.0 μl of master mix in 5 of the PCR tubes.

5) Add 5.0 μl of correct **diluted** genomic 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 340 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 22. You will need one gel per group.

Sequences of the Primers used:

```
>Actin F1
gtgacaatggcactggaatg
>Actin R1
ggagaactggcatcacacc
```

Sizes expected:

    Genomic DNA: 371 bp
    cDNA: 289 bp

Step III. Thursday, July 9, 2009. Gel analysis.

1. Load 15 μl of each reaction to a gel lane. Remember to add a lane with 7 μl of DNA ladder (molecular weight markers). Run the gel at 130 V for ~30 minutes.

O'GeneRuler 1 KB Plus DNA Ladder (Fermentas)

bp	ng/0.5µg	%
20000	20.0	4.0
10000	20.0	4.0
7000	20.0	4.0
5000	75.0	15.0
4000	20.0	4.0
3000	20.0	4.0
2000	20.0	4.0
1500	80.0	16.0
1000	25.0	5.0
700	25.0	5.0
500	75.0	15.0
400	25.0	5.0
300	25.0	5.0
200	25.0	5.0
75	25.0	5.0
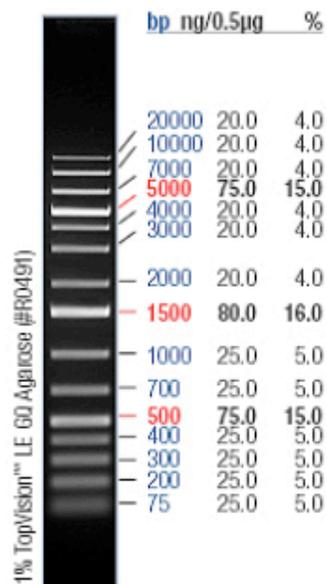
0.5µg/lane, 8cm length gel,
1XTAE, 7V/cm, 45min

Range
15 fragments (in bp): 20000, 10000, 7000, 5000, 4000, 3000, 2000, 1500, 1000, 700, 500, 400, 300, 200, 75.

Step IV. Thursday, July 9, 2009. Sequence Analysis.

In order to complete the analysis of the actin gene, we need to analyze the sequence of the PCR products. To do this you would cut out the PCR bands from the gel and purify the DNA. This would be sent to a company called Genewiz that would sequence the DNA for us and e-mail the results. Due to time constraints, you will be provided the DNA sequence for analysis.

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 use a computer program called Blast2sequences. This is a program that is part of a free suite of related programs and is available at the National Institutes of Health (NIH) website. Later in the course you will learn more about the NHI Blast website.

```
>Actin Genomic DNA PCR Product
ctgttacagaatagttgagaatggctgacgaggatatccagcctatcgtatgtgacaatggcact
ggaatggtcaaggttgttatctcgttcagaagtcttttttcaacaaagcaactctactcct
gtgcctaattgttgctcaactcctcaatatttacaggccggtttcgctggtgatgatgcg
ccaagagctgtcttccccagcattgtgggaagaccacgccacaccggtgtcatggtcggc
atgggccaaaaggatgcctacgtaggtgatgaggctcaggccaagagaggcatcctgaca
ctgaagtacccgattgagcatggcattgtcaacaactgggatgacatggagaactggcat
cacacc

>Actin cDNA PCR Product
ctgttacagaatagttgagaatggctgacgaggatatccagcctatcgtatgtgacaatggcact
ggaatggtcaagggccggtttcgctggtgatgatgcgccaagagctgtcttccccagcat
tgtgggaagaccacgccacaccggtgtcatggtcggcatgggccaaaaggatgcctacgt
aggtgatgaggctcaggccaagagaggcatcctgacactgaagtacccgattgagcatgg
cattgtcaacaactgggatgacatggagaactggcatcacacct
```

1. Open Firefox and go to the Blast Website and click on the 'nucleotide blast' link.
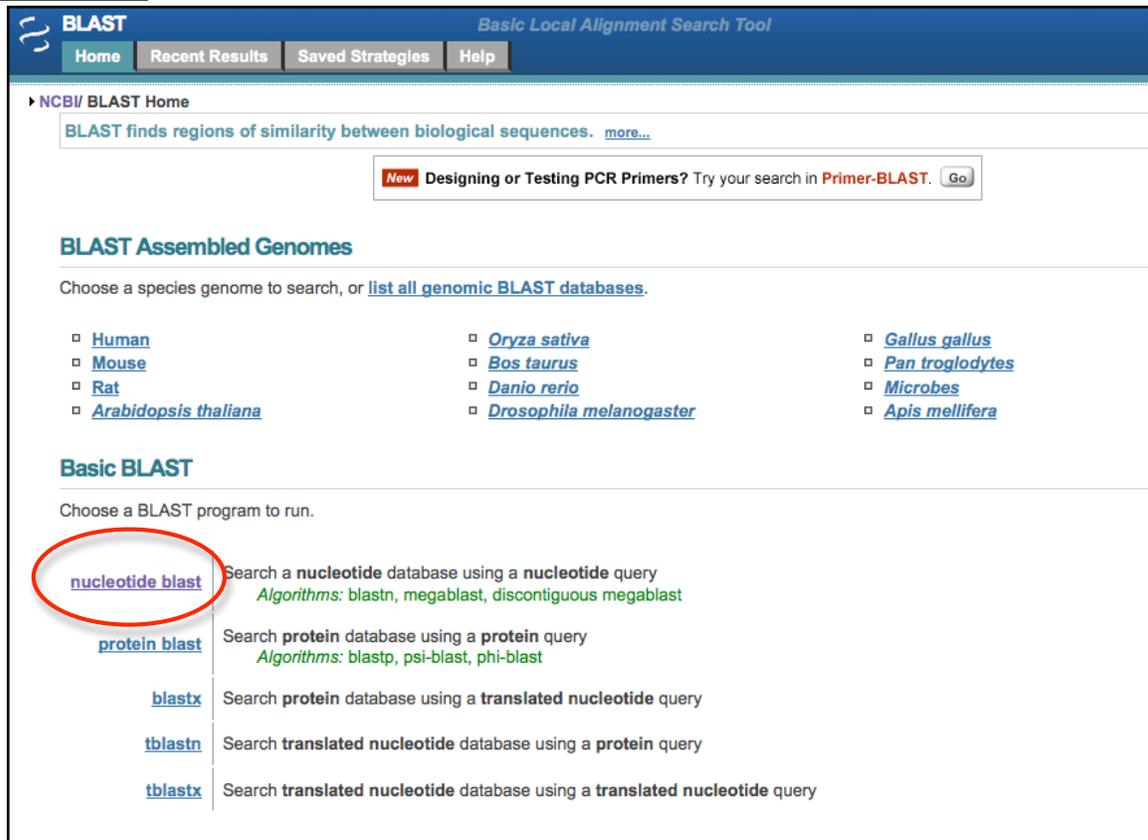
2. Check the 'Align two or more sequences' checkbox.
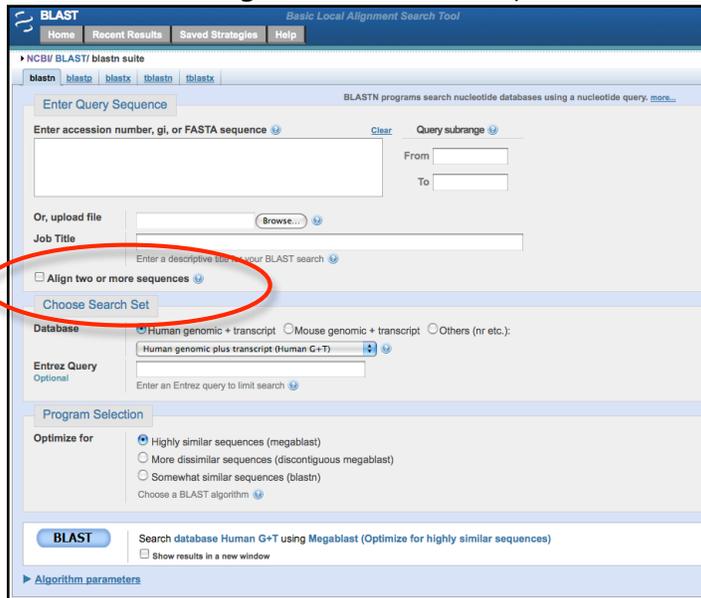
3. A) Enter the 'Actin Genomic DNA sequence' in the Query (top) textbox and B) the Subject (bottom) 'Actin cDNA sequence' in the bottom text box. C) Click 'Blast.'

A.

B.

C.
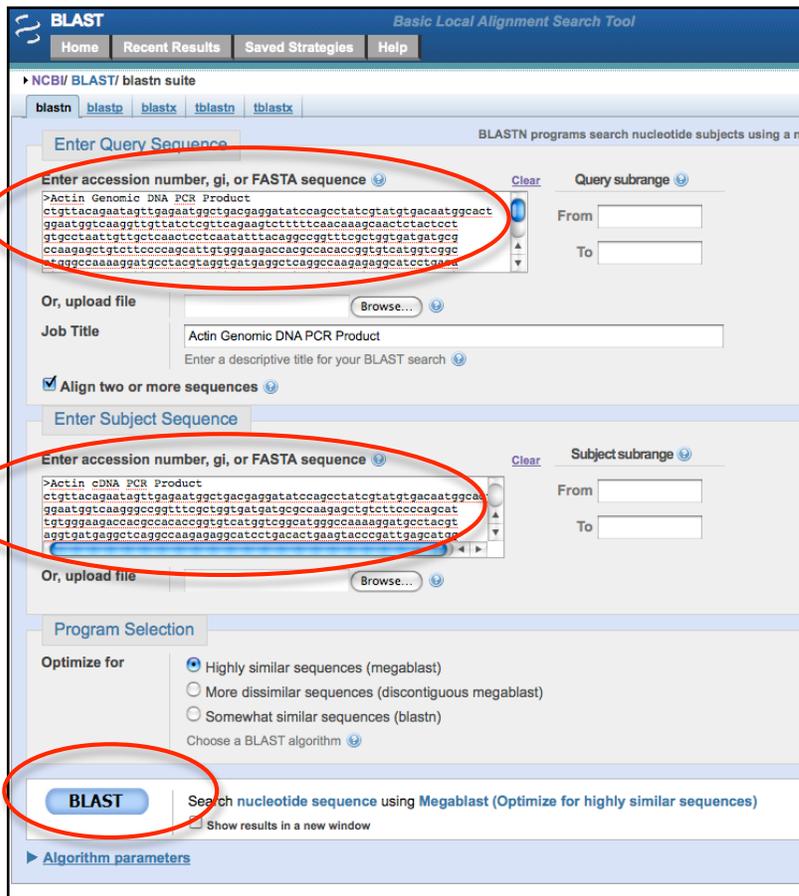
4. The results are presented in this page.

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?
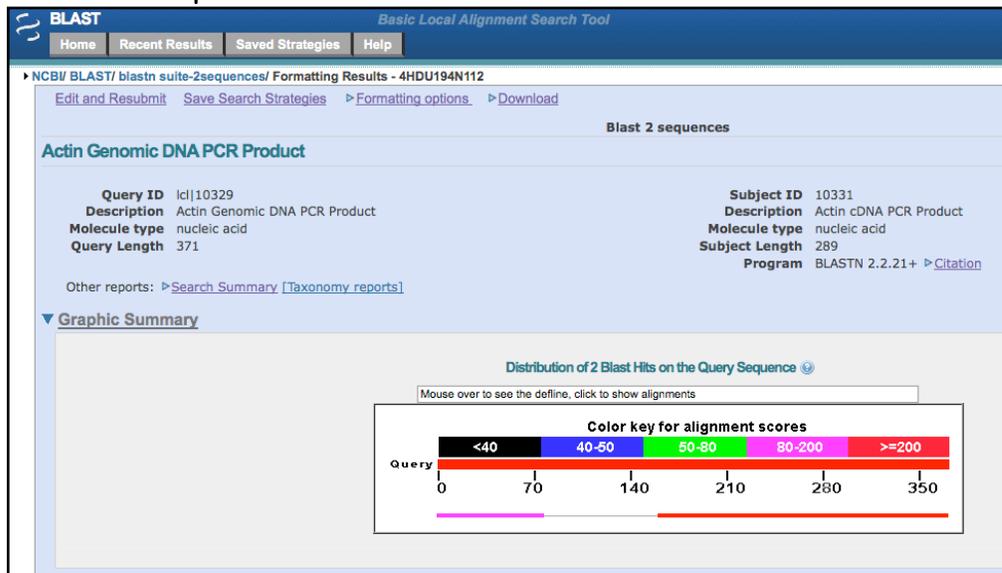
The bottom half of the results page shows the two hits nucleotide-by-nucleotide. When bases at the same position are the same a vertical line is placed between them. The alignments are ordered by score from highest to lowest. Notice that the second alignment starts with 1 in both query (genomic DNA) 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?

Homework 1 Due Friday, July 10 at 9:15 A.M.

Below is sequence of two PCR products from the actin gene using a different set of primers on genomic DNA and cDNA. The PCR products were sequenced and the sequence is shown below.

>Actin genomic PCR product 2
```
agatcaccctgtgctgctgaccgaggcccctctcaaccccaaggccaacagagagaaat
gacgcagattatgtttgaaaccttcgaatgcccagcaatgtatgttgctatcgaggctgt
tctttcgttgtacgctagtgggcgaacaactggtaagatggctgtcgcctcagatatata
tagtgatatgcactacaaagatcatacatacatacggcaatcctgacacggtctctttgt
caaatatctctgtgtgcaggtattgtgatggattcaggtgatggtgtgagccacacggtt
ccaatttatgaaggatacacacttcctcatgctattctccgtttggatcttgcggggcgt
gacctcaccgaccacctaatgaagatcctcacagagagagggtactccctcactacgagc
gctgagcgggagattgtcagggacataaaggagaagcttgcctacgttgcccttgattat
```

```
gaacaggagctggagactgccaagagcagctcctctgtcgagaagagctacgagatgcct
gatggtcaggtcatcaccattgggtcagaaaggttcaggtgccccgaggtgttgttccaa
ccatcccttgttggcatggaatcgcctagcgtccatgaggccacgtacaactccatcatg
aagtgtgatgttgatatcaggaaggatttgtatggtaacgttgtcctcagtgggggtttt
accatgtttcctgggattgccgatcgtatgagcaaggagatcacgtccctagttcctagc
agcatgaaggttaaagtagttgcgccacctagaaggaagtacagtgtctggatcggtggc
tctatttttggcttctctcagcacttttcagcaggtcttctctttcttactttattcttc
ataagcagtaggcttttatcattcaactcctcttcatcgagaacctacagcagccttgac
catgttgggtgtgtgtgcttttttttgtagttaaagtgcgaaataacccattatataccaa
ctcaatttggttaaactggcttgtgatgcactctttttttacctaaaaacgcaattggaat
acaaatatattcttgcttcagtcagtcactccagttttctagtatcaacttctctcaggg
cccatttggaacgaaggaaatatgaaggattttttgaaggattttcgaaggattcatttcc
tacatagaagttttctatgtgaccctttggaacacaggatttaattctatcaatttcttt
gaaatttctatggattgctttccttcatacaagttttggaggaaaataacgagctcatct
cttggaaaatttcctatgactcttatctctctattcaaattcatgtatagcacccaaagg
ctaggtgagaaattttgctacgttctcaatttctgtaggattgcaattcataggcaatta
tagtcatgtgtttttctctattcctgtgtttcgtattcttctgttcaaaaggggggctagt
aggcactgcacagtgctccacctccgcgtgtgttaacatacgtgttaatcattacagaca
tagacccaaacccgatagtaacatcttttttgaggcaaatgaaaagagcccaaggcctttt
tcttcgacgatagatataagcgggatgtggaaatggaattactagtgtgatagcaccgaa
tattttcttgcattgccattacttgcactcattattttaatttgcatttctttgacaat
gggcaactgtgttggcagatgtggatctccaagggagagtatgacgaaactggtccgggc
attgtccacatgaagtgcttctaa
```

>Actin cDNA PCR product 2

```
agatcaccctgtgctgctgaccgaggcccctctcaaccccaaggccaacagagagaaaat
gacgcagattatgtttgaaaccttcgaatgcccagcaatgtatgttgctatcgaggctgt
tctttcgttgtacgtattgtgatggattcaggtgatggtgtgagccacacggttccaatt
tatgaaggatacacacttcctcatgctattctccgtttggatcttgcggggcgtgacctc
accgaccacctaatgaagatcctcacagagagagggtactccctcactacgagcgctgag
cgggagattgtcagggacataaaggagaagcttgcctacgttgcccttgattatgaacag
gagctggagactgccaagagcagctcctctgtcgagaagagctacgagatgcctgatggt
caggtcatcaccattgggtcagaaaggttcaggtgccccgaggtgttgttccaaccatcc
cttgttggcatggaatcgcctagcgtccatgaggccacgtacaactccatcatgaagtgt
gatgttgatatcaggaaggatttgtatggtaacgttgtcctcagtgggggttttaccatg
tttcctgggattgccgatcgtatgagcaaggagatcacgtccctagttcctagcagcatg
aaggttaaagtagttgcgccacctagaaggaagtacagtgtctggatcggtggctctatt
ttggcttctctcagcacttttcagcaggacaatgggcaactgtgttggcagatgtggatc
tccaagggagagtatgacgaaactggtccgggcattgtccacatgaagtgcttcta
```

Using the same sequence analysis, determine:
1) How many introns are there?
2) What are the start and stop position of the intron(s).
3) Draw a diagram of the genomic and cDNA products showing the exon/intron structure of the gene before and after splicing.
4) How much smaller will the cDNA PCR product be? Why?

Answer these questions.
4) Why is it important to determine the exon/intron structure of a gene?
5) What are the components of a PCR?
6) What are the temperature steps of a PCR cycle?

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 <u>*Dissociation*</u> -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 <u>*Activator*</u>). 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 descendents 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**.
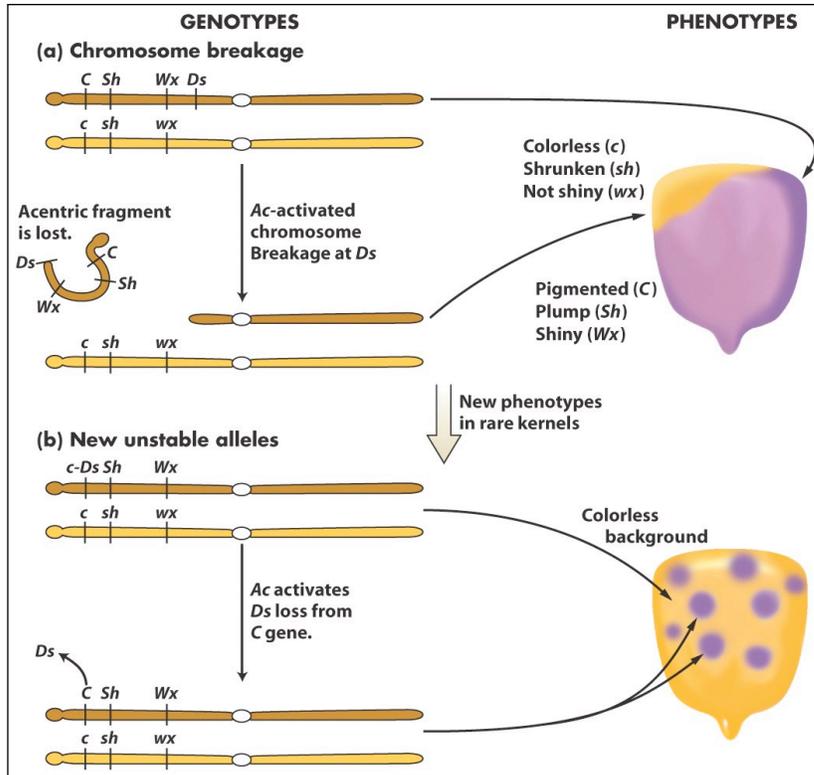
Figure 1. New phenotypes in corn are produced through the movement of the *Ds* transposable element on chromosome 9. (a) A chromosome fragment is lost through breakage at the *Ds* locus. Recessive alleles on the homologous chromosome are expressed, producing the colorless sector in the kernel. (b) Insertion 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:

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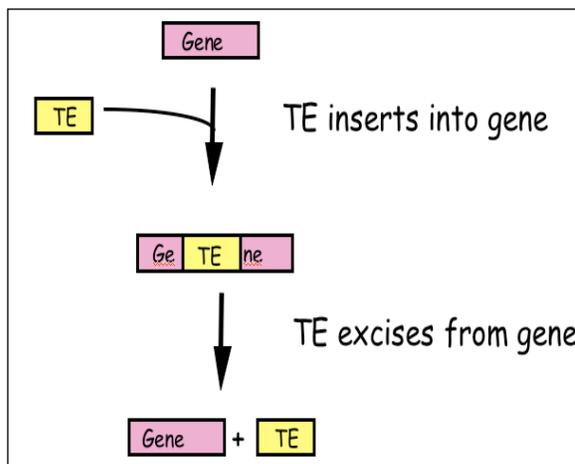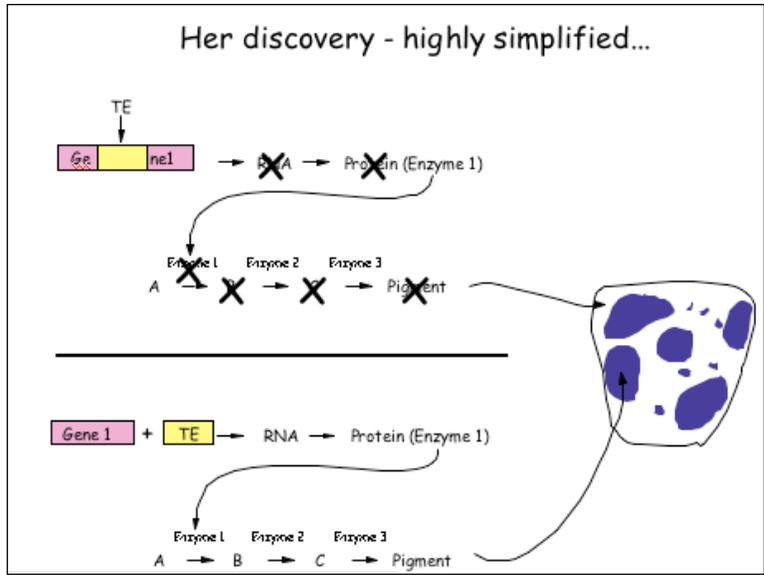
Figure 2: McClintock hypothesized that TEs were a source of "reversible" mutation. Their ability to transpose allowed them to excise from mutant genes leading to phenotypic reversion.

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 <u>autonomous element</u> while Ds was called a <u>non-autonomous element</u>.
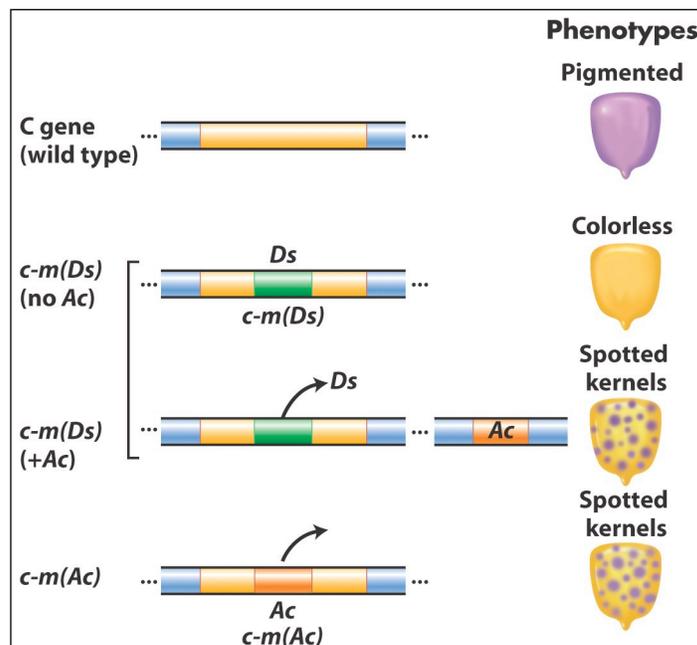
Figure 3 Summary of the main effects of transposable elements in corn. *Ac* and *Ds* are used as examples, acting on the *C* gene controlling pigment. In maize (but not many other organisms), normal alleles are capitalized and mutant alleles are written in lower case. In addition, McClintock designated 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.
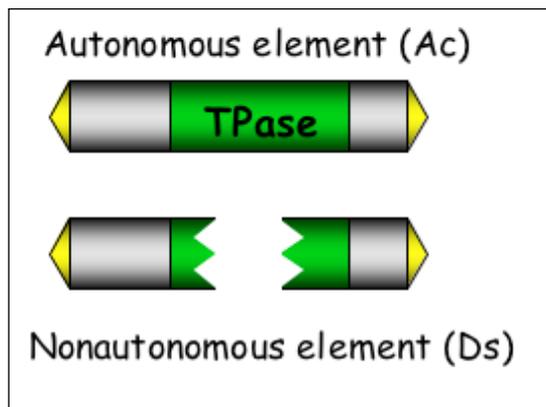
Figure 4: Molecular structure of Ac and Ds.
Ac: Tpase is the gene encoding the transposase enzyme, which is necessary for movement of both Ac and Ds.
Ds: Ds requires Ac for movement because it is a defective version of Ac where the Tpase gene has been deleted.
Yellow arrows at the ends are the terminal inverted repeats - this site 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.
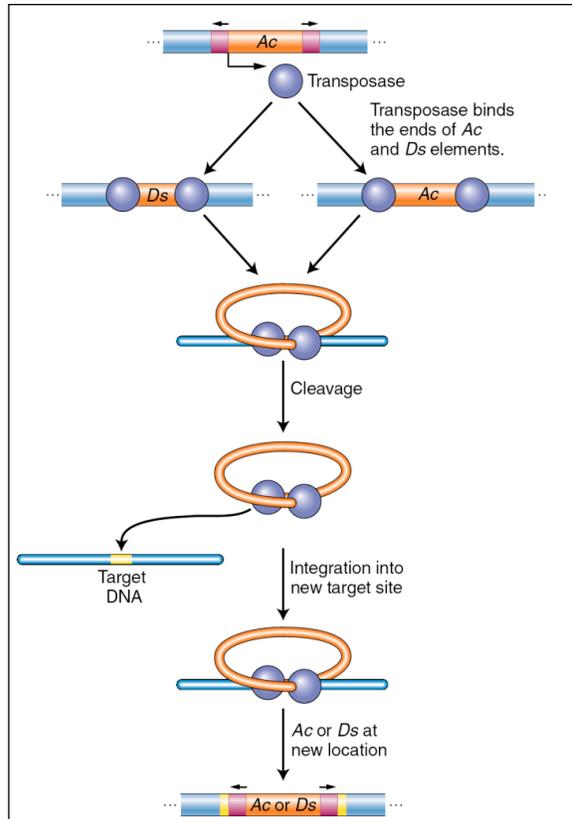
Figure 5 Ac transposase catalyzes excision and insertion (also called integration).
The maize *Ac* element encodes a transposase that binds its own ends or those of a *Ds* element, excising the element, cleaving the target site, and allowing the element to insert elsewhere in the genome.

Like many other proteins, the transposase protein can multi-task. First, it is a <u>DNA binding protein</u> 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 <u>enzyme</u>. 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 <u>target site duplication</u> (TSD) that is created during insertion of virtually all TEs. How it is generated is shown below in Fig 6.
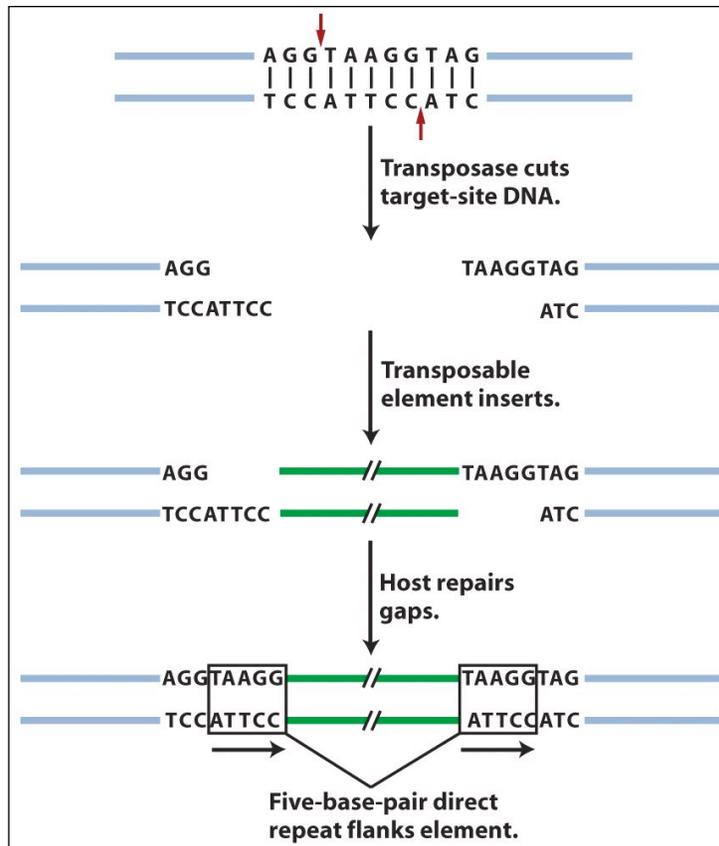
Figure 6: An inserted element is flanked by a short repeat. A short sequence of DNA is duplicated at the transposon insertion site. The recipient DNA is cleaved at staggered sites (a 5-bp staggered cut is shown), leading to the production of two copies of the five-base-pair sequence flanking the inserted 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*
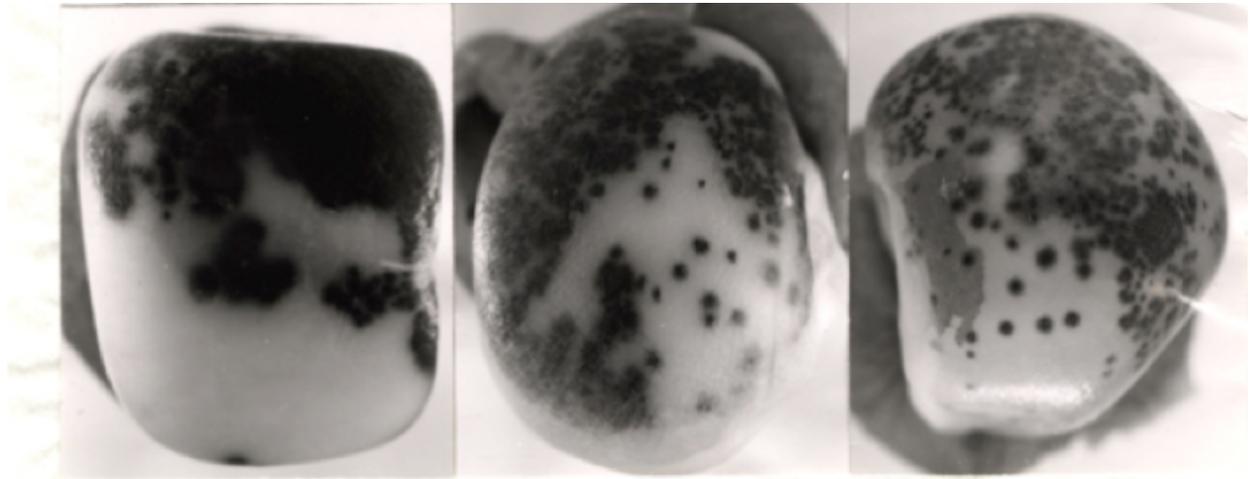
The Wessler lab needed to create an experimental system that mimics the one used by McClintock with TEs inserted into pigment genes and expressed in the kernel. What was needed was a visual assay to test for movement of a rice TE in Arabidopsis.

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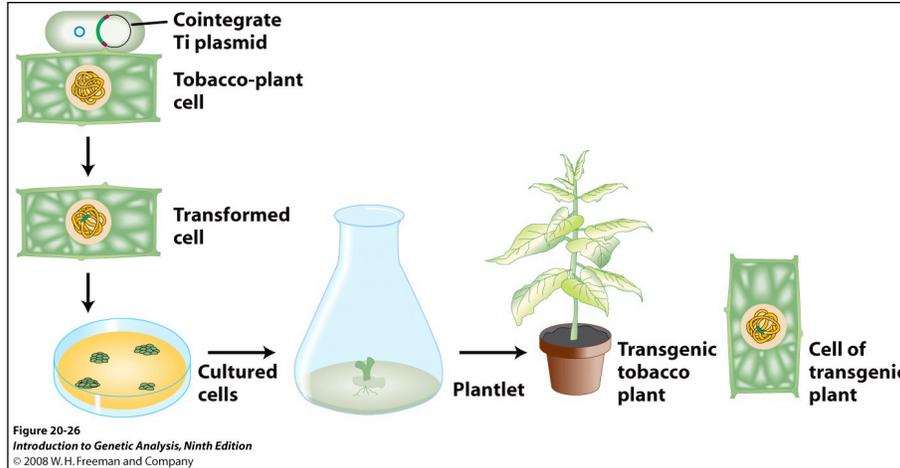
Figure 7: Schematic of how Agrobacterium has been exploited to deliver foreign DNA into plant chromosomes.

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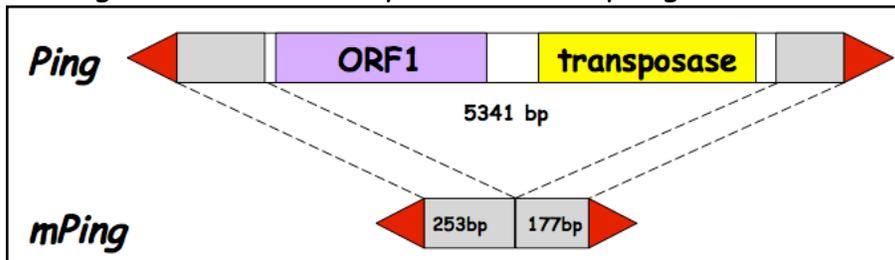
Figure 7: Ping encodes two genes: the transpoase (TPase) and ORF1 (open reading frame) (function unknown at this time). The red arrows are the terminal inverted repeats (14bp). mPing is 253bp + 177bp long.

Design of the experiment and controls

Now up until this point Ping and mPing were considered active TE <u>candidates,</u> - 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.

How your GFP reporter was constructed

How your rice Ping genes were constructed
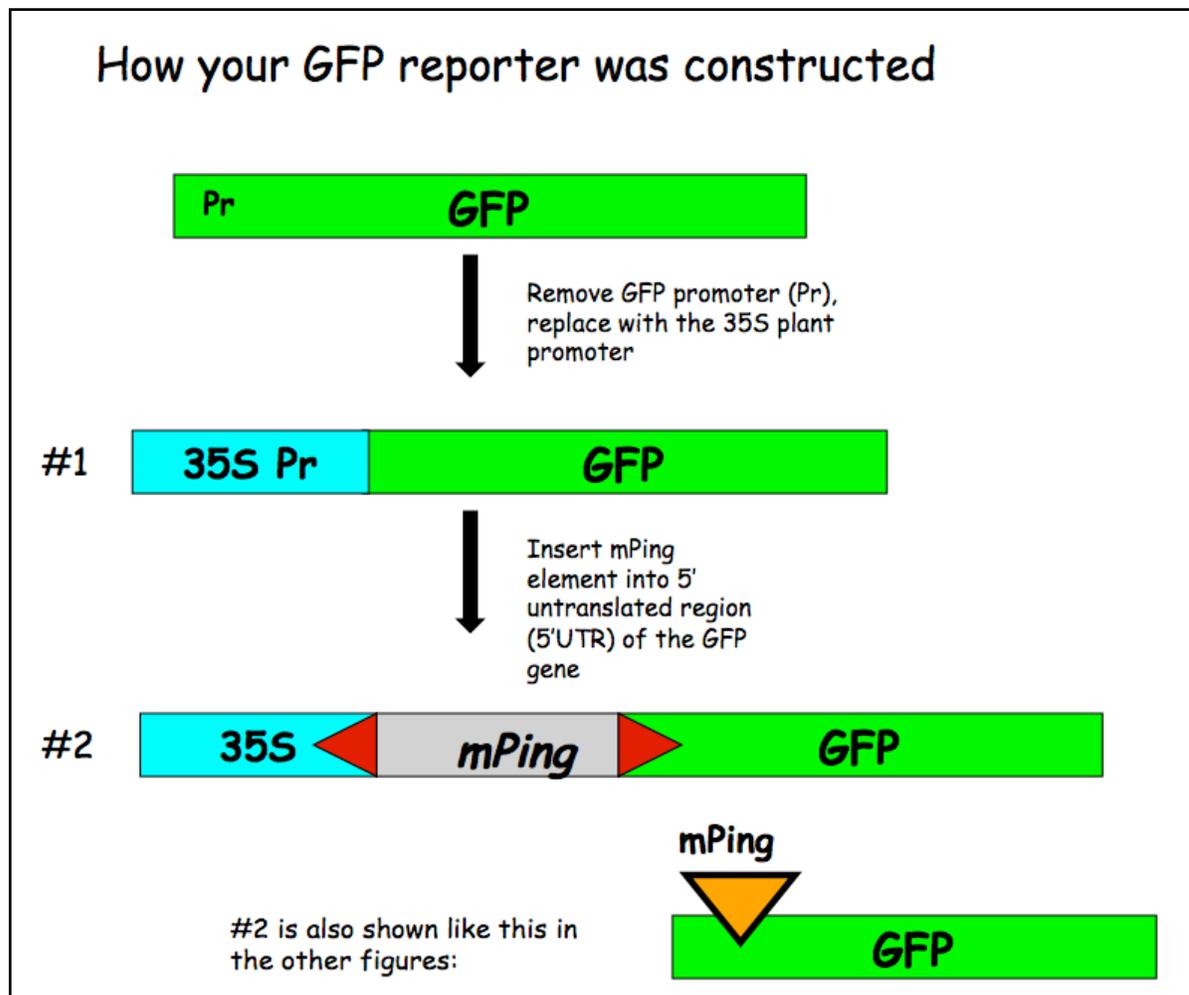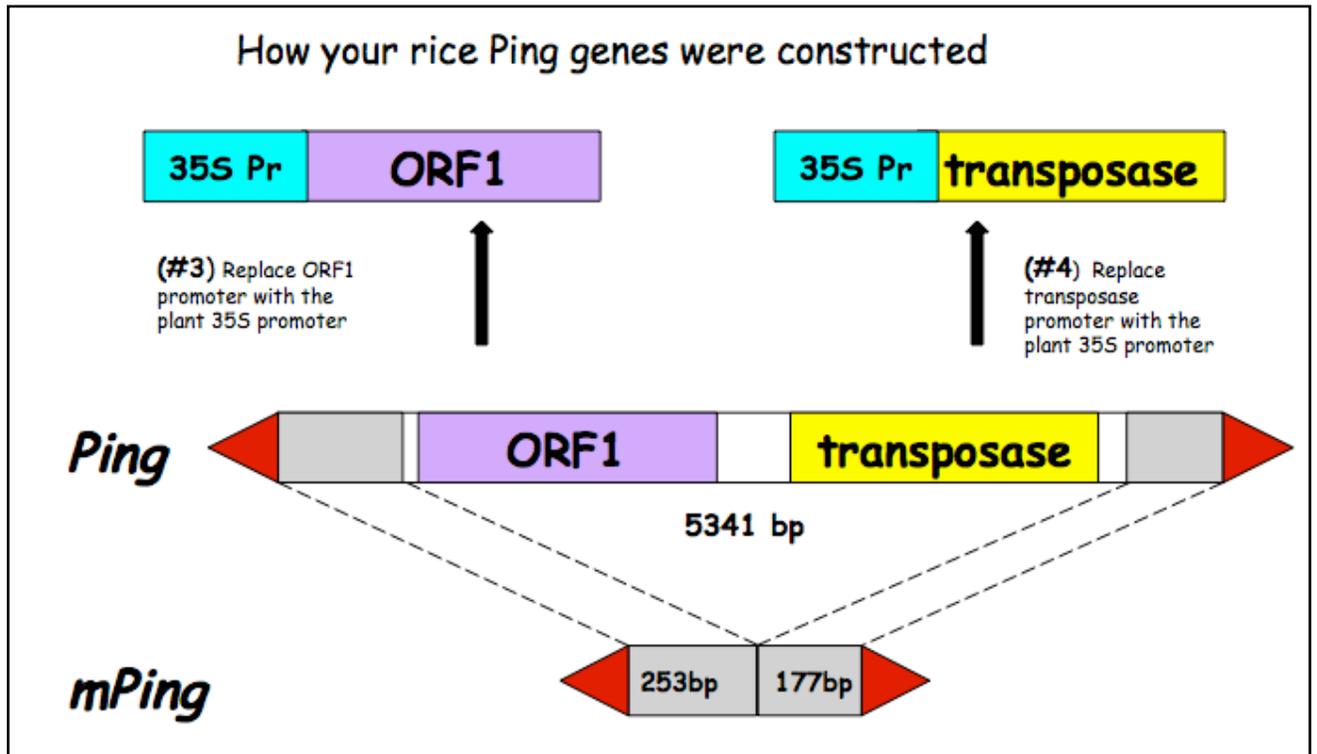
In experiment #1 we are testing whether the Ping encoded protein(s) can catalyze the transposition of mPing. So, there are actually three questions we will be attempting to answer:

--Can ORF1 protein by itself excise (move) mPing?

--Can Tpase protein by itself move mPing?

--Can both proteins work together to move mPing?

To address these questions you will analyze mPing excision in transgenic Arabidopsis plants containing <u>one, two, or three</u> of the following T-DNA constructs:
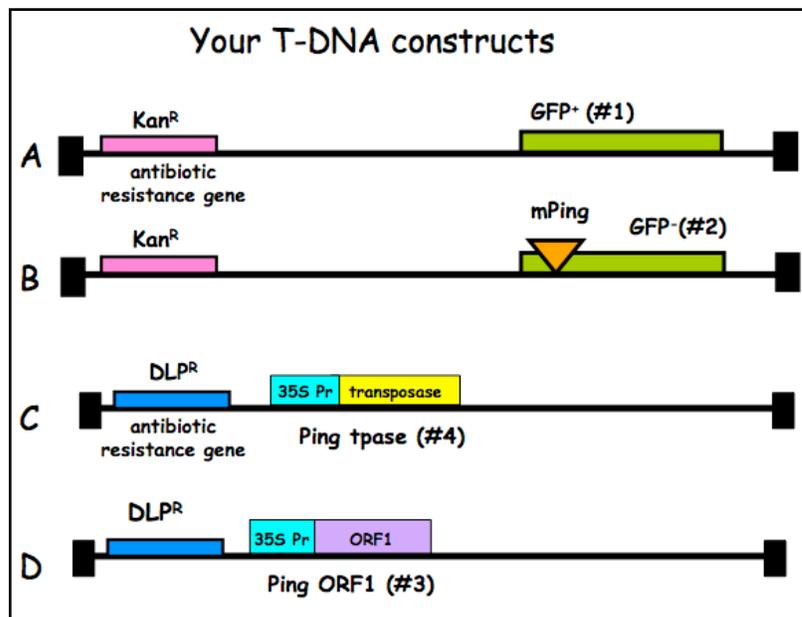
Figure 10: The transgenic Arabidopsis plants used in this experiment contain one or two of the 4 T-DNA insertions in their genome.

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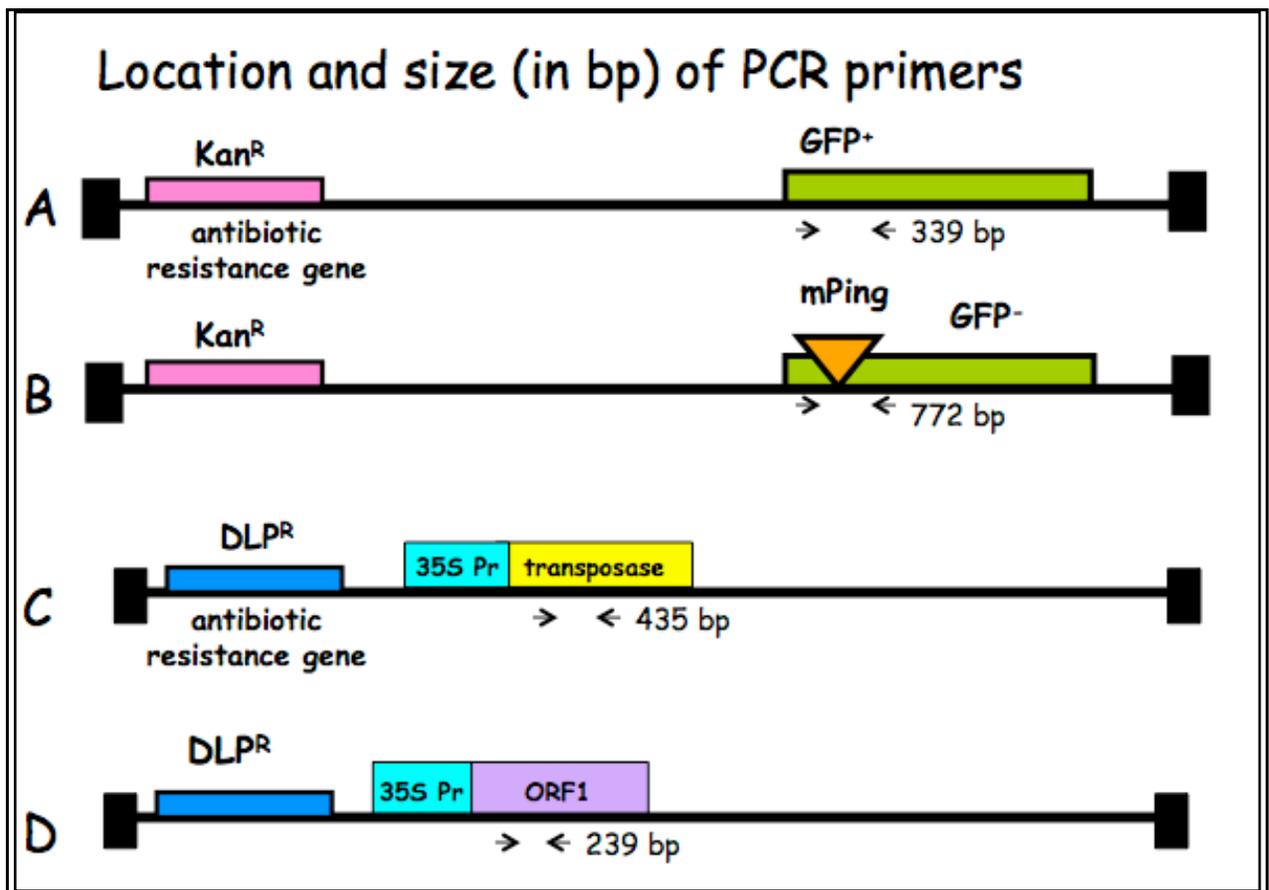
Experiment 2 Examination of mPing excision from *A. thaliana* leaf DNA

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. Monday, July 13, 2009. Examine leaves under microscope.

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

III. Tuesday, July 14, 2009. 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 μl 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 μl 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 μl 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) <u>Add 600 μl of isopropanol. Mix the contents thoroughly by inverting.</u>

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) <u>Spin for 5 minutes at top speed. Pipette off supernatant.</u>

10) <u>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 around 30 minutes with the caps open.</u>

11) <u>Resuspend the DNA in 50 μl water or TE</u>. 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) <u>Label new 1.5 ml tubes.</u>

13) <u>Put 15 μl of DNA from step 11 into each tube.</u>

14) <u>Add 3 μl of 6x loading dye to the tube. Tap the tube gently to mix.</u>

15) <u>Load all 18μl on your gel. Keep track of which sample went in which lane.</u>

16) <u>Load 7 μl of DNA Ladder in one empty well.</u>

17) <u>Run the gel at 130 Volts for ~20 minutes.</u>

18) <u>Photograph the gel.</u>

IV. Wednesday, July 15, 2009. 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 Tpase. The primers for ORF1 and Tpase 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+Tpase 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+TPase.

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 20µl into each well of the gel. Remember to add a lane with 7 µl 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. Thursday, July 16, 2009. Gel analysis.
We will analyze the results of your PCR reactions and relate them to your observed leaf phenotypes.

Introduction to the NCBI website: PubMed and Blast.

Biological sequence data and journal articles are collected, indexed, and made available by the <u>National Center for Biotechnology Information (NCBI)</u>. NCBI is a unit of the National Library of Medicine (NLM) at the National Institutes of Health (NIH). Because it is a part of the NIH, the collections of sequence data and journal articles are available free to anyone at <u>http://www.ncbi.nlm.nih.gov/</u>. This is what the NCBI home page (currently) looks like....
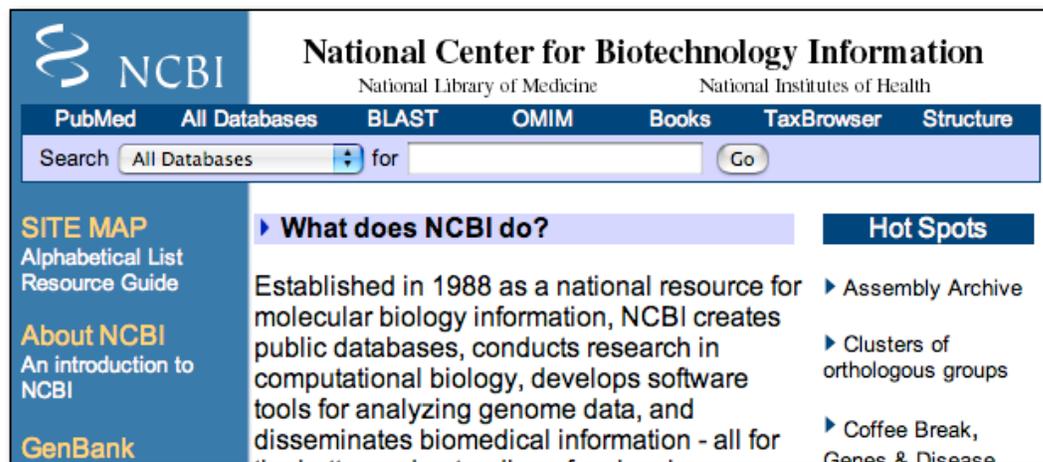
NCBI also provides tools for searching and downloading the databases it maintains through the web portal NCBI Entrez. While the search tool for the literature database is PubMed, sequence is searched though a number of tools collectively called <u>Blast</u>. PubMed indexes thousands of biological journals going back as far as 1950. It also contains thousands of full-length articles in PDF format available for free download in a collection called PubMed Central. Blast searches on sequence databases that are often referred to as GenBank. There are three public repositories for sequence data: NCBI, DDBJ (Japan), and EMBL (Europe), and all share data on a nightly basis. Although the file formats and search tools may differ between the three repositories, they are essentially redundant.

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

Let's begin our tour by visiting the PubMed site and then move on to the BLAST site where we will be spending a great deal of our time today and in the rest of the course.

I. **PubMed**: Literature searches about a biological problem are very easy. You can use PubMed (and Blast) from any computer and with any Internet connection; you are not required to be on the University network. To access many full-length articles and download PDF copies you will need to be using UGA's network.

Steps for a PubMed search:

1. Open NCBI in a web browser by going to the NCBI home page and click on PubMed in the bar. To get to the PubMed home page click on PubMed which is part of the main menu at the top.
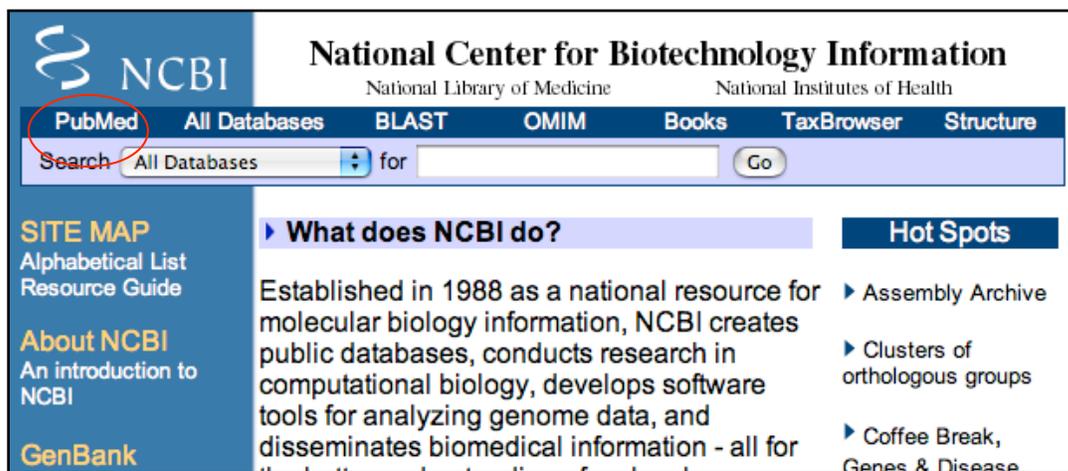
This is the PubMed homepage…
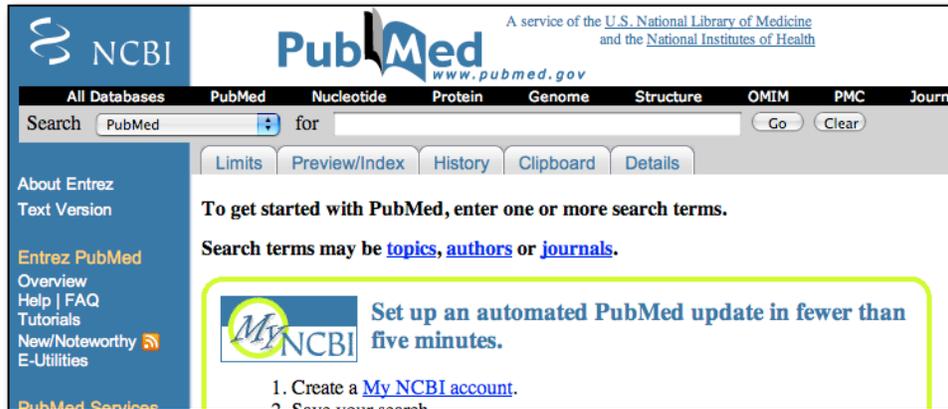
2. Think about the topic you want to search. You can use keywords, author last names, journal titles, publication year, or institution.

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

3. Before we discuss the results, click on the 'Details' tab. This will show you the details of the search that was performed by PubMed's search engine.

While you thought you were just searching "transposable element" PubMed was actually using these search terms (with added comments):

("dna transposable elements"[TIAB] NOT Medline[SB])   •Search titles and abstracts, not the medline subset

OR "dna transposable elements"[MeSH Terms]   •Seach Medline Subject Headings

OR transposable element[Text Word]   •Search all text

As you can see the simple query 'transposable element' is expanded into a more structured query by PubMed. MeSH is a controlled vocabulary for indexing PubMed. Curators at NCBI and journal editors assign these keywords based on suggestions by authors. Because of this query expansion it is a good idea to check the 'Details' tab whenever a search gives no results or unintended results.

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

5. Click on the underlined authors (in blue) and detailed information about the article is provided including the abstract. The abstract provides a detailed summary of the paper. On the right are two icons. Clicking on either of those will take you to a download page for the full article. The Related Links section is also a useful area to help refine searches and will be discussed in class.

II. Introduction to **Blast**:

You will use Blast a lot this semester in different forms. 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.

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:

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. We will focus on nucleotide blast this semester.

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.' <u>Cut and paste the following sequence in the Query text window (Enter accession number....)</u>:

```
>mPing
ggccagtcac aatgggggtt tcactggtgt gtcatgcaca tttaataggg gtaagactga
ataaaaaatg attatttgca tgaaatgggg atgagagaga aggaaagagt ttcatcctgg
tgaaactcgt cagcgtcgtt tccaagtcct cggtaacaga gtgaaacccc cgttgaggcc
gattcgtttc attcaccgga tctcttgcgt ccgcctccgc cgtgcgacct ccgcattctc
ccgcgccgcg ccggattttg ggtacaaatg atcccagcaa cttgtatcaa ttaaatgctt
tgcttagtct tggaaacgtc aaagtgaaac ccctccactg tggggattgt ttcataaag
atttcatttg agagaagatg gtataatatt ttgggtagcc gtgcaatgac actagccatt
gtgactggcc
```

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

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

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

5. When your search is complete a results page will be presented. We
    will discuss this page in detail in class.
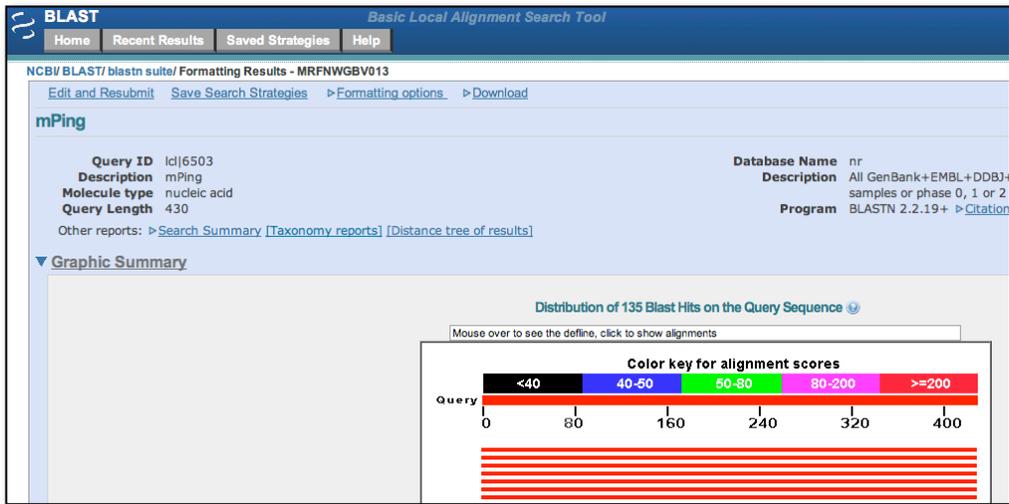
6. Details of the Alignment (to be discussed in class)
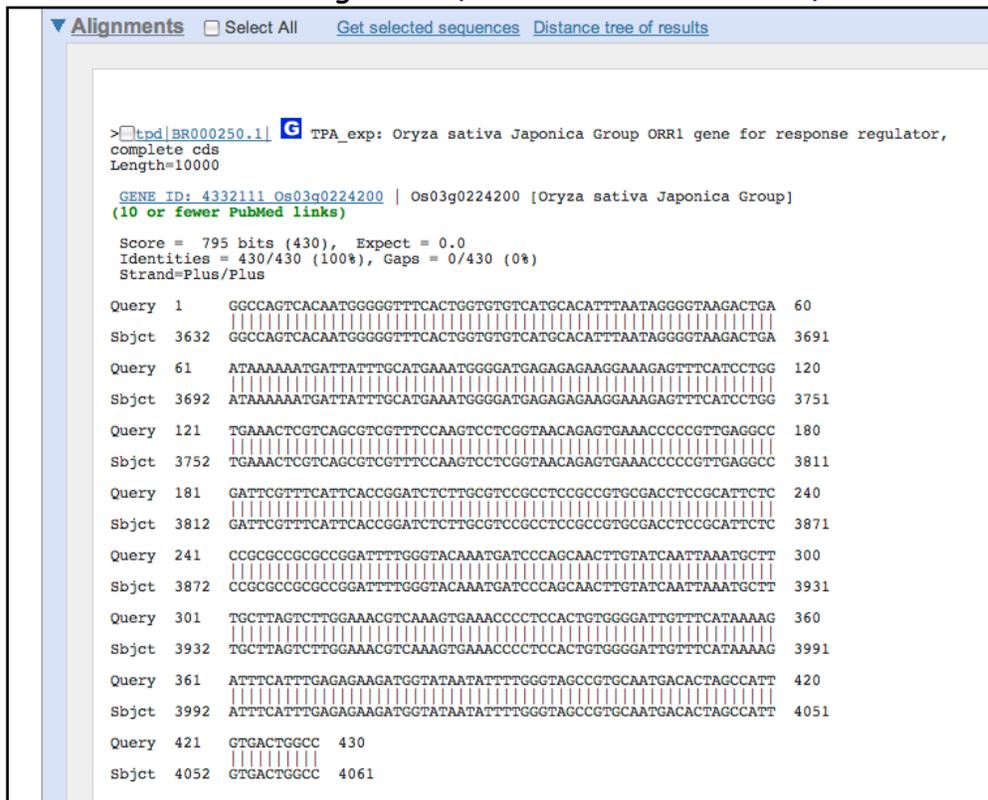
A short discussion on how Blast works.

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.

Maize Genome Browser
The maize genome is very large and has many features: genes, simple repeats, transposable elements, ESTs, 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.

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 cover the features that are relevant to your project.
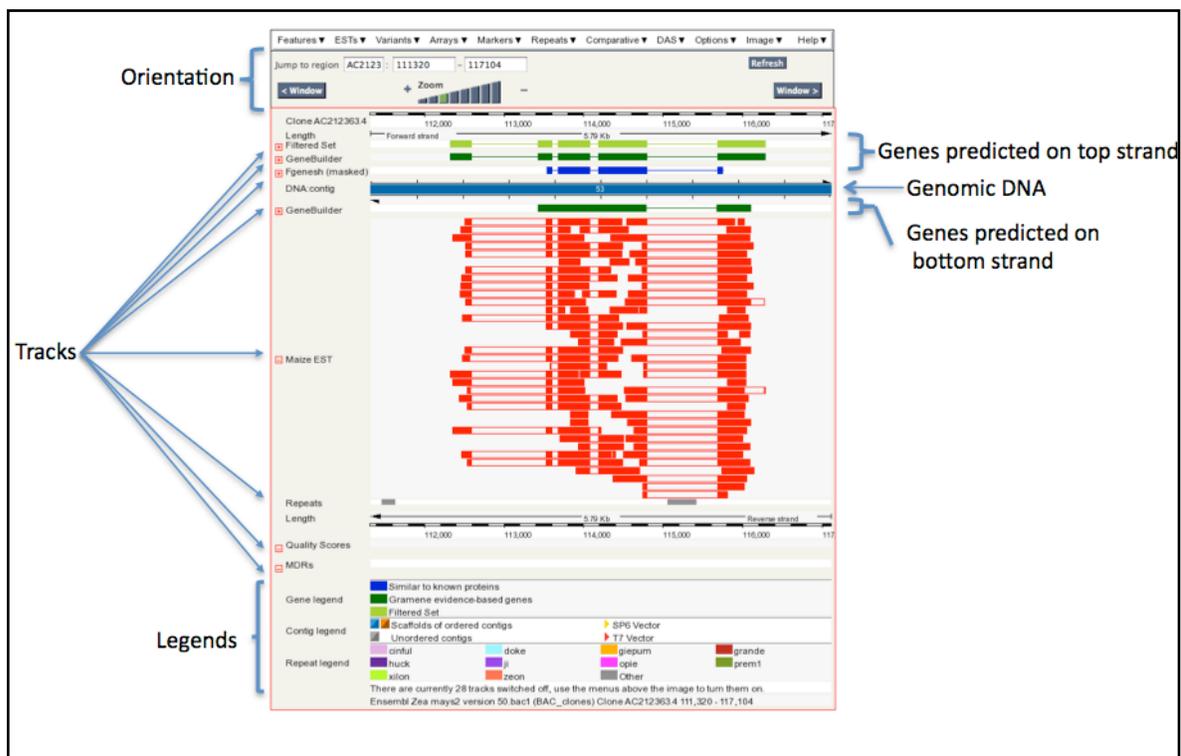
The genomic DNA represented by the blue rectangle towards the top. Note that this rectangle can be many colors. Above and below this rectangle is a series of features organized into "Tracks." For example immediately above and below the genomic DNA representation is the Gene Model Tracks. These "genes" are predictions by computer programs. Since they are predictions we refer to these genes as gene models. Gene models above the track are on the Top strand of the DNA. That means that the start of the gene is to the left of the model. Gene models below the genomic DNA rectangle are on the

bottom strand. This means that the start of the gene is to the right of the model.

An interesting track in this region of the genome is called "Maize EST." EST stands for Expressed Sequence Tag. Basically an EST is a type of cDNA. If there is EST evidence for a gene model then we have some confidence that the gene is expressed and is real (not the figment of some computer program's predictor). Because actin is widely expressed in the maize plant there are a lot of EST "hits."

Another interesting Track for this class is the "Repeats." Remember that TEs 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. It is a MITE similar to the ones you will be studying.

Annotating a TE using the Maize Genome Browser
There are two steps to annotating a TE using the genome browser.
     1. Use Blast to find genomic location.
         a. The alignment of the blast hit will give you the genomic coordinates.
         b. The Contig viewer will show you the features in the location of the TE you are working with.
     2. Use Blast2Sequences to find the ends of the TE. This will give you the terminal invereted repeats of the TE.

We will use a Blast search as our entry point into the maize genome.

Click on Blast located in the Left-hand bar. 1. Copy-and-paste the Pack-MULE sequence in the Query window and select the 2. 'Sequenced Clones' for the database.

```
>ZmarMITE.64
TACTCCCTCCGTACTCAGAATTAAGGTCGTTTAGGACATCATTTGTGATACCAAGGAGTG
ATTAATTGCATGGTAATATCCTGTGTATGCCCCTATTAAATAAGGTATCGGTCCATTTTG
AACATTATTATGTTCCTGTCCGCCTGGTTGTTTGCTCGCTGCTGGTAGCTCAGGTCACAT
GTTCGATCCCCCACAGAACCATTTTTTGGGGTTGTGCGCACTTGATTTTTTCCGGTTGCG
TTTTCGCATGCGGCAAGCAAAGTGAGCAACAGCAAGCACCAAAAGCTGGCGGCGTGTTTT
TCTTGCATGGGAGCTCTCGACGCAATAAAGTCCGTATGGGCAAAACAGTCCGGAATACGA
CTCCACAACATAGAACGCCTCCAAATACGAAAACCAGCTATACAGCCAGGACGTCCTTAT
TTCTGAGTACGGAGGGAGTA
```

Notes on the Databases:

     a. Filtered Gene Set: A set of 40,000 genes predicted by gene prediction software but with Pseudogenes removed. This set is useful if you are looking to see if your TE is inserted into a gene or if you are looking for the parental gene for a captured gene segment.

     b. Sequenced Clones: This is the genome sequence.

     c. Sequence Clones (RepeatMasked): This is the genome sequence with the repeats like TEs masked or hidden. Not very useful for TE research!

     d. Working Gene Set: Another set of maize genes but before filtering.

The Blast results are shown graphically in two windows.
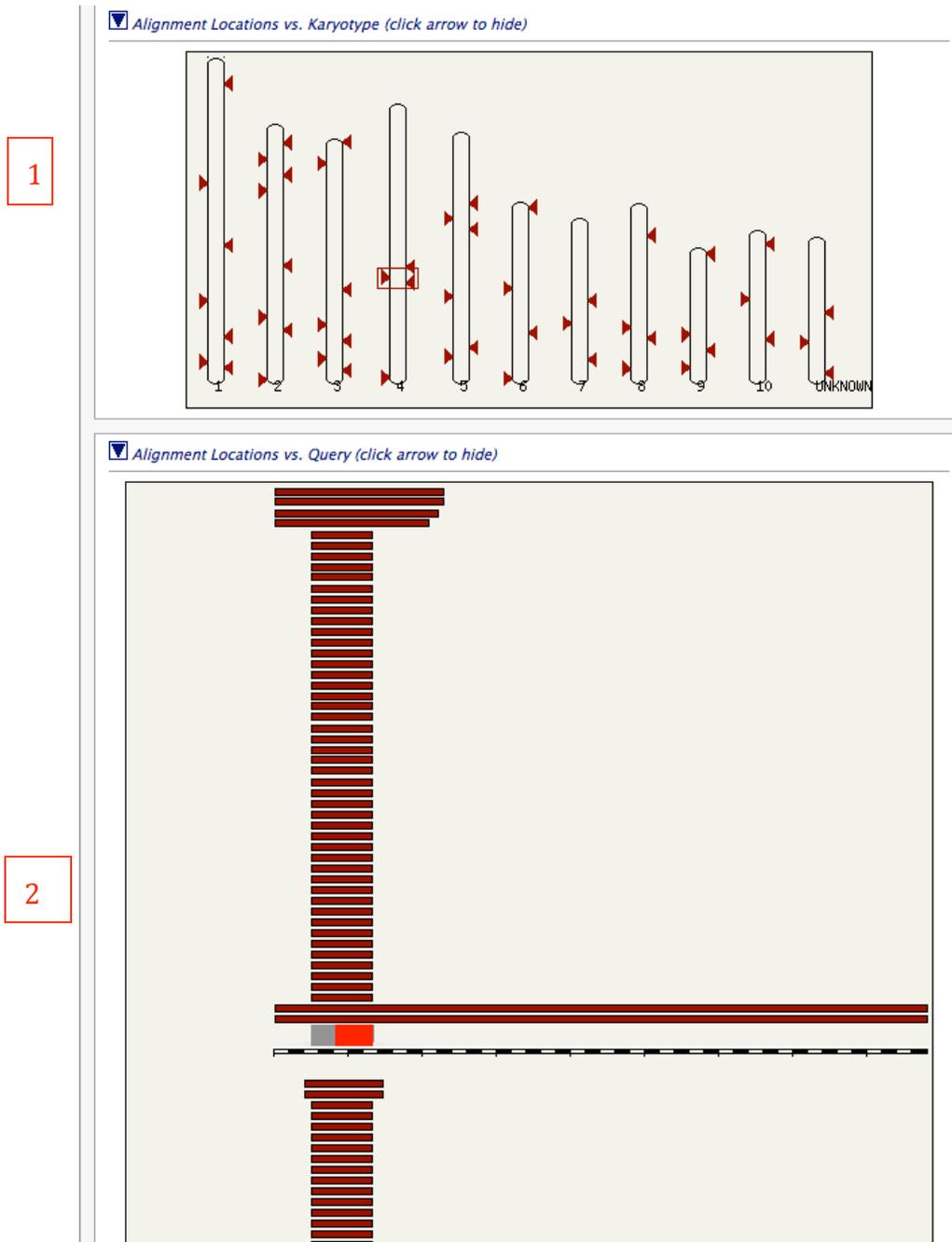
1. Blast hits by chromosome.

2. Blast hits based on the query sequence. The query is shown in the center as a black/white bar. The Blast hits (subjects) are shown on the top and bottom of the query bar indicating the strand of the hit. The hits closest to the center bar have the lower (better) e-values. In this case we are looking for hits to the full element. There is only one just above the black/white bar.

Click on that hit and a contextual menu opens. You will see that the accession the hit is on is AC195796.3 Click on Alignment and new tab will open with the alignment.
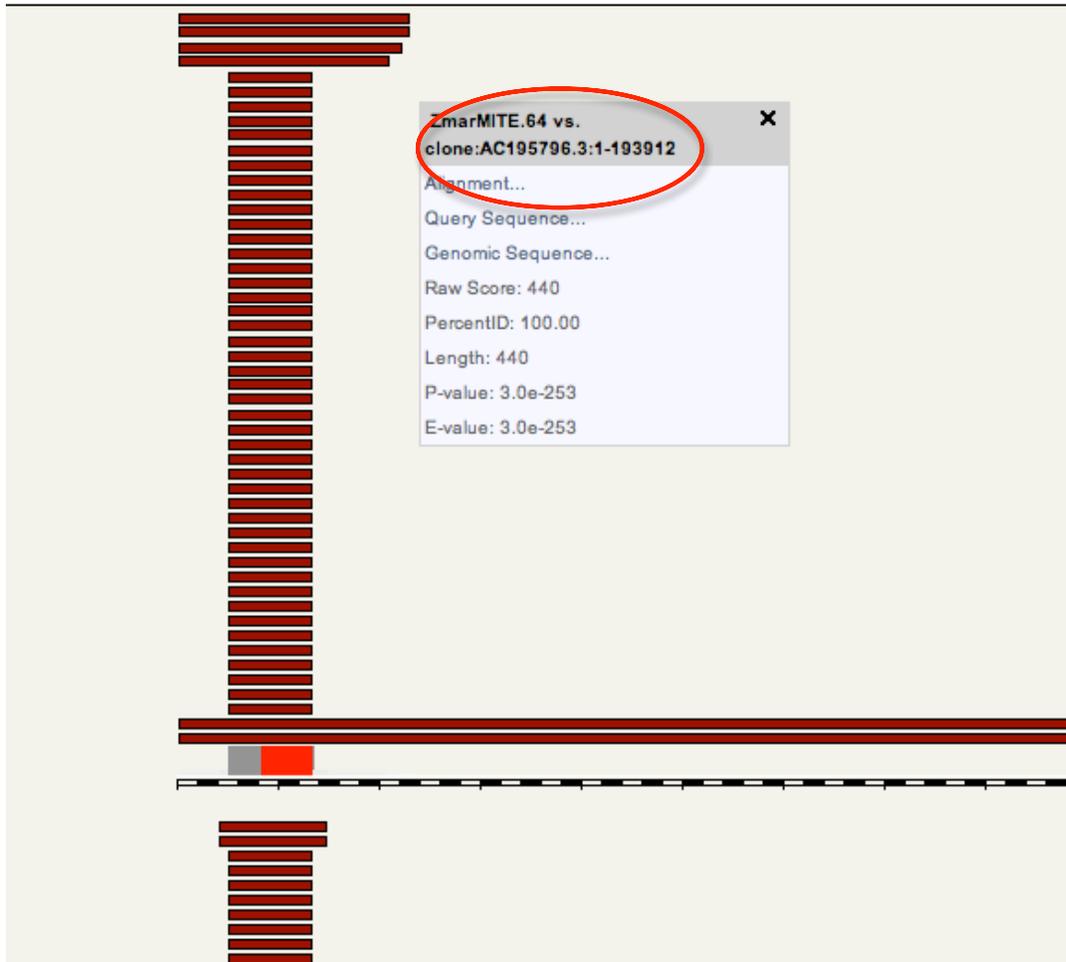
To get to the contig view for that hit we need to first add accession numbers to the Alignment Summary. To do this,

      1 select Name in the Subject window and

      2. click Refresh.

Your screen should look like this:

      3. Click the [C] in the first line. This will open a new tab with the Contig Viewer.
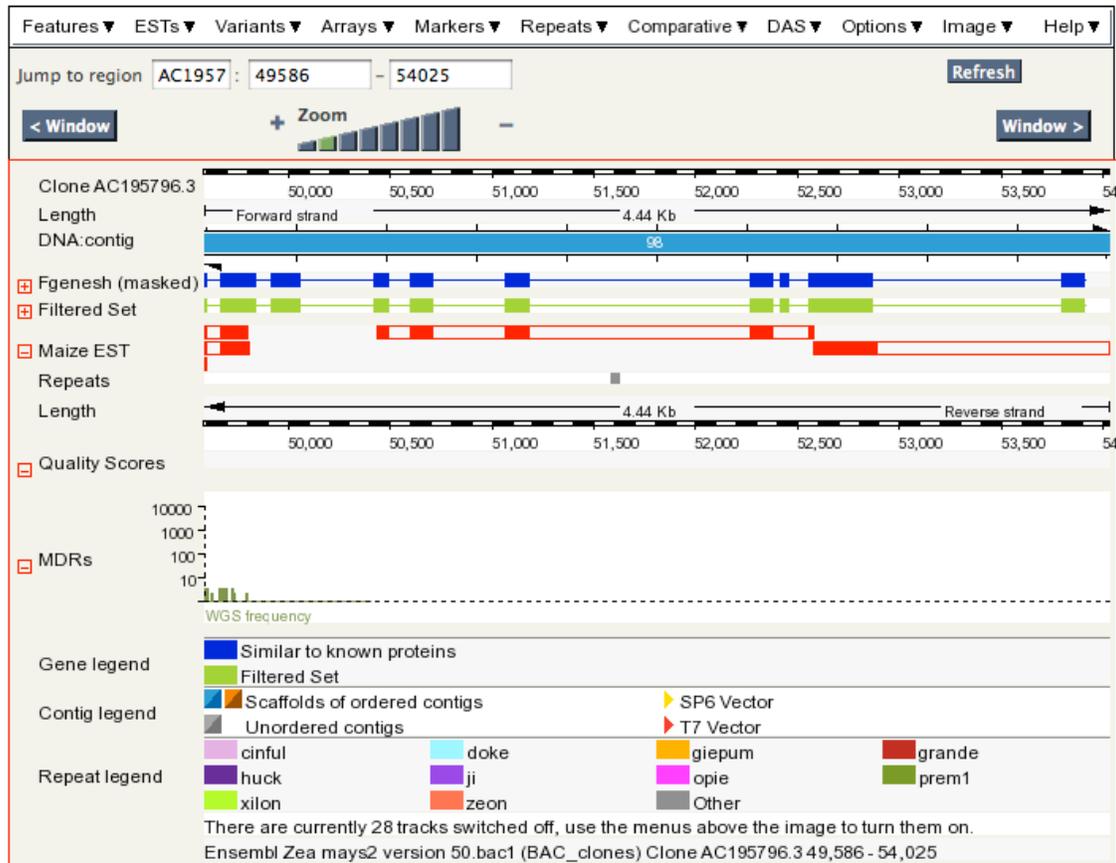
So in this case, the Blast hit does not appear as a Track. Sometimes it will, sometimes it won't…the joys of JavaScript.

We want to annotate the location of this element. First take a screen shot of the Contig view by pressing "Apple + Shift + 4." Drag the cross-hair cursor over the region you want to take a picture of. To annotate the element we need:

    1. Coordinates of the element on the Contig.
    2. The coordinates of the ends of the element.

1. The coordinates of the element location can be found by looking at the alignment between the element (called the Query) and the contig (called the Subject). This element starts at 51,586 and ends at 52,025 bp.

```
Query location       : ZmarMITE.64                        1 to       440 (+)
Database location    : clone:AC195796.3:1-193912      51586 to     52025 (+)
Genomic location     : 4                           206488734 to 206489173 (+)


Alignment score      : 440
E-value              : 3.0e-253
Alignment length     : 440
Percentage identity: 100.00

Query:       1 TACTCCCTCCGTACTCAGAATTAAGGTCGTTTAGGACATCATTTGTGATACCAAGGAGTG 60
               || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 51586 TACTCCCTCCGTACTCAGAATTAAGGTCGTTTAGGACATCATTTGTGATACCAAGGAGTG 51645


Query:      61 ATTAATTGCATGGTAATATCCTGTGTATGCCCCTATTAAATAAGGTATCGGTCCATTTTG 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 51646 ATTAATTGCATGGTAATATCCTGTGTATGCCCCTATTAAATAAGGTATCGGTCCATTTTG 51705


Query:     121 AACATTATTATGTTCCTGTCCGCCTGGTTGTTTGCTCGCTGCTGGTAGCTCAGGTCACAT 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 51706 AACATTATTATGTTCCTGTCCGCCTGGTTGTTTGCTCGCTGCTGGTAGCTCAGGTCACAT 51765


Query:     181 GTTCGATCCCCCACAGAACCATTTTTTGGGGTTGTGCGCACTTGATTTTTTCCGGTTGCG 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 51766 GTTCGATCCCCCACAGAACCATTTTTTGGGGTTGTGCGCACTTGATTTTTTCCGGTTGCG 51825


Query:     241 TTTTCGCATGCGGCAAGCAAAGTGAGCAACAGCAAGCACCAAAAGCTGGCGGCGTGTTTT 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 51826 TTTTCGCATGCGGCAAGCAAAGTGAGCAACAGCAAGCACCAAAAGCTGGCGGCGTGTTTT 51885


Query:     301 TCTTGCATGGGAGCTCTCGACGCAATAAAGTCCGTATGGGCAAAACAGTCCGGAATACGA 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 51886 TCTTGCATGGGAGCTCTCGACGCAATAAAGTCCGTATGGGCAAAACAGTCCGGAATACGA 51945


Query:     361 CTCCACAACATAGAACGCCTCCAAATACGAAAACCAGCTATACAGCCAGGACGTCCTTAT 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 51946 CTCCACAACATAGAACGCCTCCAAATACGAAAACCAGCTATACAGCCAGGACGTCCTTAT 52005


Query:     421 TTCTGAGTACGGAGGGAGTA 440
               ||||||||||||||| ||||||
Sbjct: 52006 TTCTGAGTACGGAGGGAGTA 52025
```

Now we can draw the element onto the Contig.
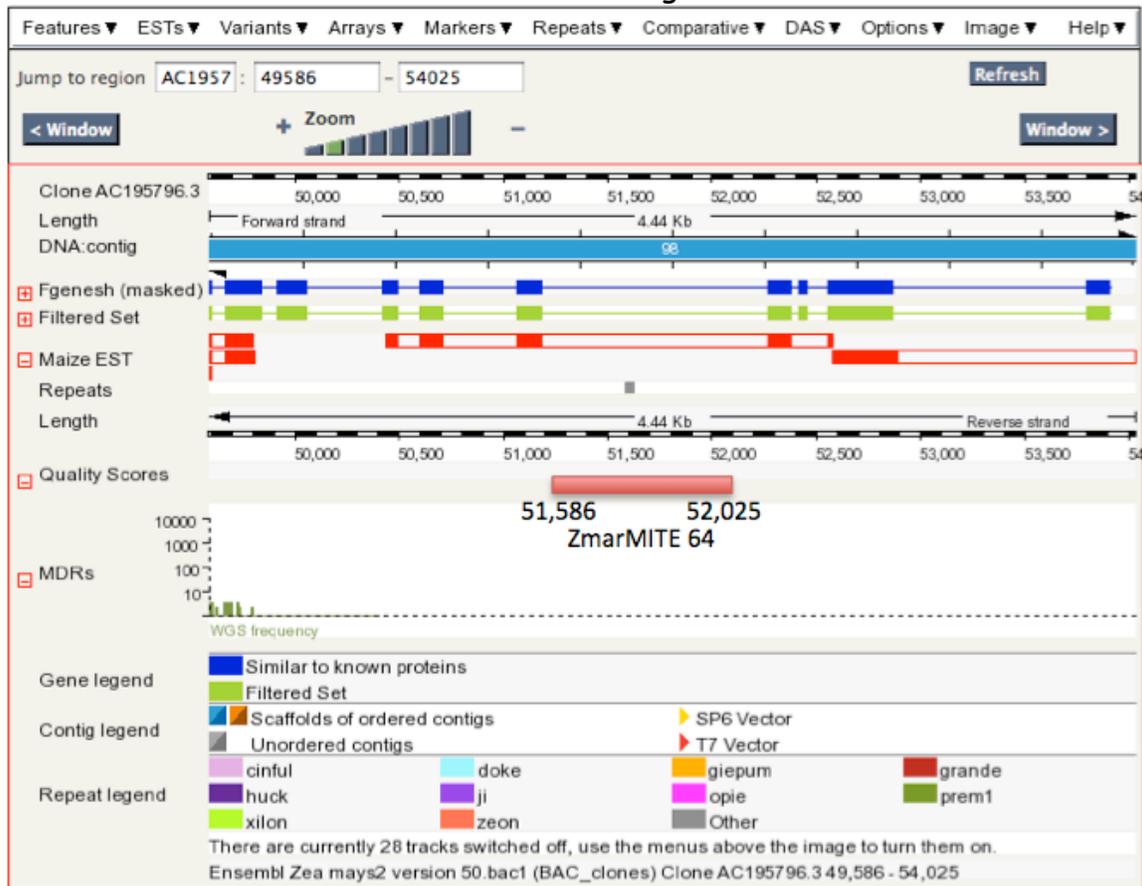
Find the ends of the element.

1. In the "Alignment Summary" click on the [G] link for the genomic region.

2. Change the limits to 1000. This will give you the TE sequence with 1000 nucleotides added to the beginning and the end. Click "Update."

3. Copy-and-paste the sequence into a Word document. Save it.

4. Draw a sketch of your element and label the new coordinates. The TE sequence begins at 1001. The TE ends at 1001 plus the length of the element.
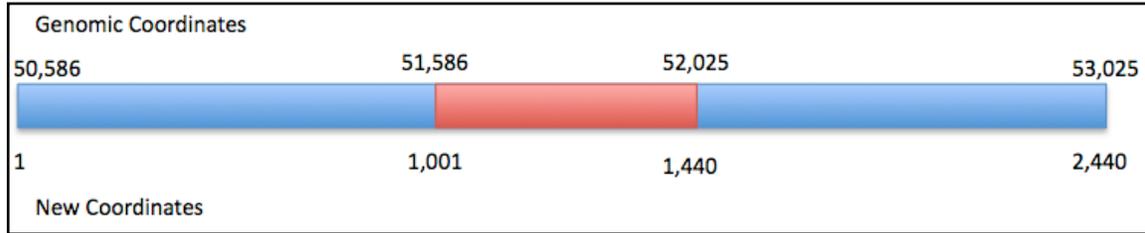
To find ends of the TE sequence, you compare the sequence to itself. We can use blast2seq to do this quickly.

5. Open a nucleotide blast and select the "Align 2 or more sequences." Copy and paste the TE sequence with the 1000 nucleotides on each side into BOTH windows.
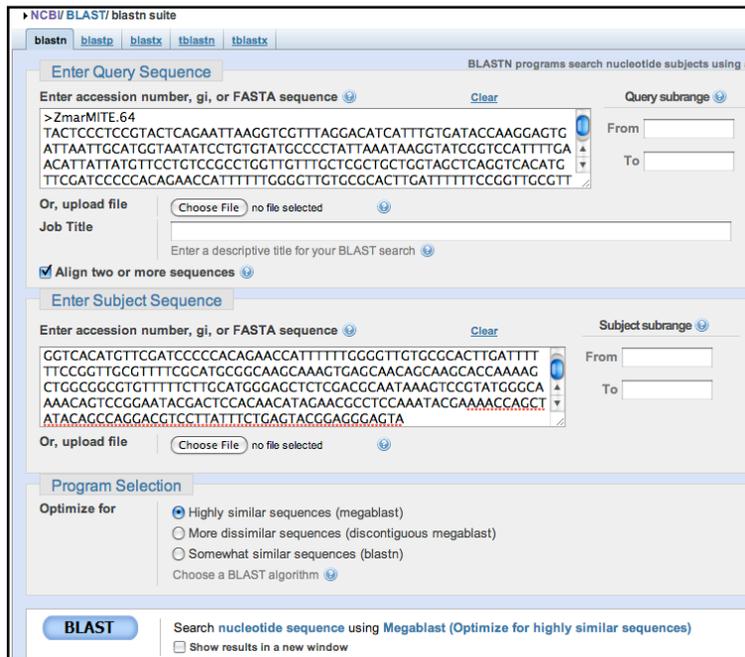
6. In the algorithm parameters window:
    a. Click the Arrow to open the window.
    b. Change "Word size" to 16.

a.
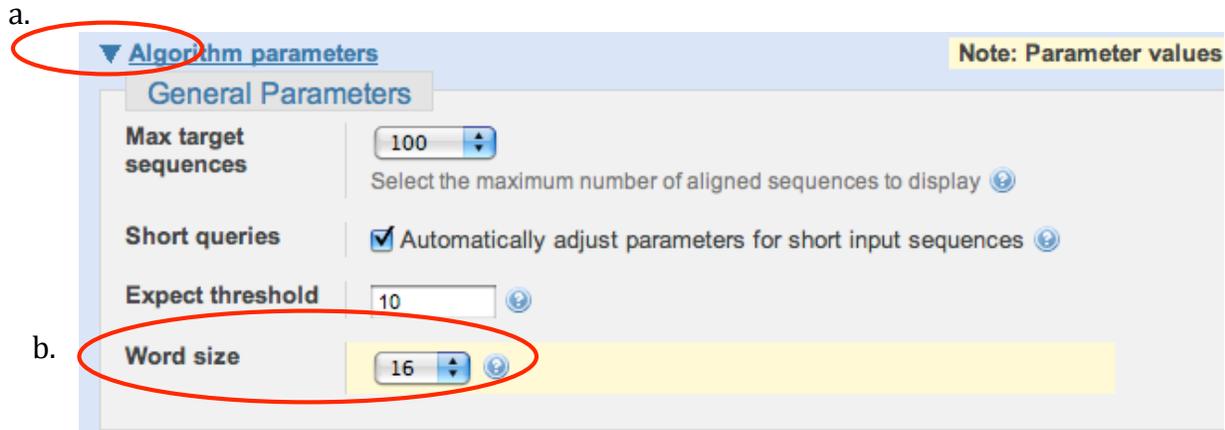
7. Click Blast.

8. In the results, click on the Dot Matrix View. You will get a graph like this:
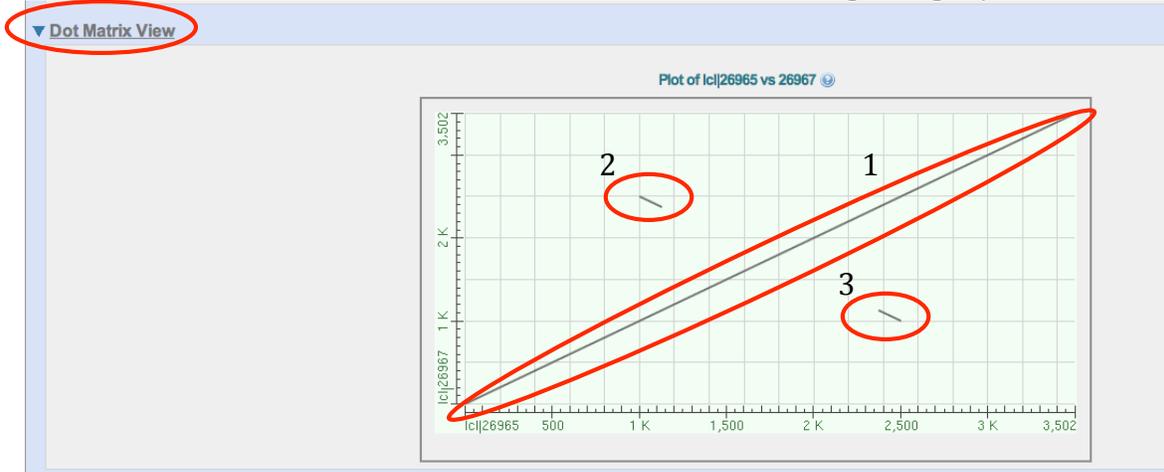
The DNA sequence is laid out on the axis of the graph. Each position in the sequence is compared to every position in the opposite sequence. If there is the same nucleotide then a dot is placed on the graph. Stretches of similarity between the two sequences show up as diagonals.

There are three diagonals on the graph.
    1. Becusae you compared the sequence against itself, it matches itself perfectly. This long diagonal shows this. It has a positive slope because the DNA is an exact copy.

2 and 3. These diagonals are matches between the end of the element. These diagonals have a negative slope because the ends of the element are the reverse complement of each other. These are the TIRs. If you notice diagonal 2 starts at position 1000.

Scroll down to the alignments. The first alignment is the exact match represented by diagonal 1. You can skip over this alignment. The alignments for the two TIR matches are below.

```
Score = 49.1 bits (26),  Expect = 1e-08
Identities = 32/35 (91%), Gaps = 0/35 (0%)
Strand=Plus/Minus

Query  1411  ACGTCCTTATTTCTGAGTACGGAGGGAGTACAGAA  1445
             ||| ||||| ||||||||||||||||||||| ||
Sbjct  1030  ACGACCTTAATTCTGAGTACGGAGGGAGTACTGAA  996


Score = 49.1 bits (26),  Expect = 1e-08
Identities = 32/35 (91%), Gaps = 0/35 (0%)
Strand=Plus/Minus

Query  996   TTCAGTACTCCCTCCGTACTCAGAATTAAGGTCGT  1030
             ||| ||||||||||||||||||||||| |||| ||
Sbjct  1445  TTCTGTACTCCCTCCGTACTCAGAAATAAGGACGT  1411
```

The limits of both matches are the same. So the left TIR goes from 996 to 1030 and the right TIR goes from 1411 to 1445. The TIRS are 35 bp long. You can include these regions on your diagram of the element.

Genomic Coordinates			
50,586	51,586	52,025	53,025
1	1,001	1,440	2,440
New Coordinates	960—1,030	1,411—1,445	

Experimental Design

Your goal in this project is to determine whether the MITEs you characterize in the B73 genome are also present in the same place in the Palomero genome. To do this you will design a PCR experiment shown in the figure below.
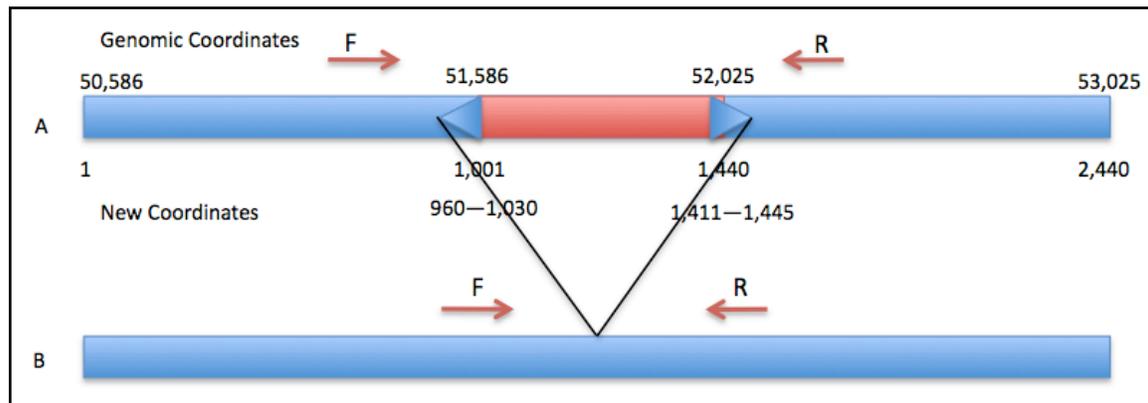
You will need to design a primer pair that flank the MITE. The PCR result using B73 DNA will be band that is a large size. If the MITE is present in Palomero the band will be the same size. What size will the band be if the MITE is absent? PCR primer design is easy when you use the website Primer3Plus.

Primer Design using the web program Primer 3 Plus

1. Open the Primer3Plus website: http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi.
    1. Give the search a name in "Sequence Id."
    2. Copy the genomic sequence into the large textbox.
    3. In "Targets" enter 1000, 439. This means that the PCR product must include the nucleotides 1000-1439.

    Since nucleotide 1001 is the start of the TE and the TE is 439 bp long, these primers will flank the whole TE. You will need to change the "439" to be the length of the TE you are studying.

2. Click on the "General Settings" Tab.

3. <u>In the field for "Product Size Ranges" delete the 150-250, 100-300, 301-400, 401-500, and 501-600 ranges.</u> The smallest product we want is around 600 bp. Leave the rest of the settings alone.

4. <u>Click on the "Advanced Settings" Tab. Set the GC Clamp to 1</u>. This will require that the 3' nucleotide be a G or C.

5. <u>Click the green "Pick Primers" button.</u>

6. Results come back quickly. The 'best' primer pair is listed at the top. Other good pairs are listed further down the page. The first pair is good and would be the ones ordered. The "Left" or "Forward" primer is highlighted in purple and the "Right" or "Reverse" primer is highlighted in yellow.

Save this page. On a Mac you can create a PDF by selecting File->Print. In the Print dialog box select save as PDF. Select a location save the file.

Typically primers are designed in pairs. A good pair of primers will meet these criteria:

> 1. Each primer should be 18-25 base pairs long.
> 2. The 3' nucleotide of the primer should be a G or C. This is called the G/C clamp.
> 3. The G/C content should be around 50%.
> 4. The melting temperature (T_m) of each primer should be close to 60°C. The difference between the Tms of a primer pair should 5°C or less.
> 5. Neither primer should form a hairpin.
> 6. Primers should not be self-complementary.
> 7. Partners should not complement.
> 8. All primer sequences are reported in 5' to 3' direction.
> 9. The product size should be between 200 and 1000 nt for Taq DNA Polymerase.

Primer 3 rarely fails to find good pairs. However, if Primer3 cannot find primers you can use the statistics table on the results page to modify and repeat the search. The statistics tell you how many primers were considered and why they were rejected. This information can be used to modify the parameters of Primer 3.

Statistics:	
Left Primer:	considered 13422, GC content failed 567, GC clamp failed 1853, low tm 4667, high tm 3141, high end compl 15, long poly-x seq 69,high 3' stability 222, ok 2888
Right Primer:	considered 13442, GC content failed 567, GC clamp failed 1855, low tm 4549, high tm 3300, high end compl 3, long poly-x seq 74,high 3' stability 224,high template mispriming score 1, ok 2869
Primer Pair:	considered 1492, unacceptable product size 1475, ok 17

7. To order the primers, go to the course Data webpage. There you will notice a short form to fill out. You must enter all fields of the form for each primer.

Primer Order Form

Enter each primer separately. All fields

* Required

Student Name *

Primer Name *
Do not use spaces. End primer with '_F

Primer Sequence *

Comments *
Short description of purpose.

Target *
Gene or TE primer targets

(Submit)

Powered by Google Docs

Terms of Service - Additional Terms