

PBIO/BIOL3240L: The Dynamic Genome

Fall 2009



Table of Contents

Page numbers continue from Summer 2009 Course Book

Syllabus

Grading Policy

Chapter 1: Experiment 1: Alu		83-86
A. Background	83	
B. Protocol	83-86	
Chapter 2: Reading guide for scientific papers		87-90
Chapter 3: B2 MITE Mini-project		91-130
A. Background	91	
B. Protocols	92-101	
C. DNA Sequencing and analysis	102-119	
D. Introduction to Phylogenetics	120-125	
E. Using TARGeT to infer trees	126-130	
Chapter 4: Project: Annotating intron MITEs		130-137

# Syllabus Fall 2009



Syllabus

Grading

Data

	Lecture	Lab	Download	
Tues., Aug. 18	<ul style="list-style-type: none"> <li>• Course Mchanics</li> <li>• Discuss summer</li> <li>• Discuss RNA elements/Alu</li> </ul>	<ul style="list-style-type: none"> <li>• Extract Cheek cell DNA</li> </ul>		
Thurs., Aug 20	<ul style="list-style-type: none"> <li>• Discuss Retro elements</li> </ul>	<ul style="list-style-type: none"> <li>• PCR for Alus</li> <li>• Pour Gel</li> </ul>		
Tues., Aug. 25	<ul style="list-style-type: none"> <li>• Review Retros</li> <li>• Alus in Genome</li> </ul>	<ul style="list-style-type: none"> <li>• Run Gel</li> </ul>	<ul style="list-style-type: none"> <li>• Batzer and Deininger (2002) Alu Repeats and Human Diversity.Nature Reviews Genetics</li> <li>• Cordaux et al (2007) In search of</li> </ul>	

			polymorphic Alu insertions with restricted geographic distributions. Genomics.	
<b>Thurs., Aug 27</b>	Discuss Alu results, wrap up project  Paper discussion	<ul style="list-style-type: none"> <li>• Clean PCR</li> <li>• Digest DNA</li> <li>• Run gel</li> </ul>		
<b>Tues., Sept. 1</b>	<ul style="list-style-type: none"> <li>• Discuss Alu results</li> <li>• Paper discussion</li> <li>• B2 project</li> </ul>	PCR with B2 primers		
<b>Thurs., Sept. 3</b>		<ul style="list-style-type: none"> <li>• Run Gels</li> <li>• Excise bands</li> <li>• Clean</li> <li>• Clone</li> <li>• Transform</li> </ul>		
<b>Tues., Sept. 8</b>	Discussion	<ul style="list-style-type: none"> <li>• Minipreps</li> <li>• Restriction Digest</li> <li>• Prepare sequencing rxns</li> </ul>		

Thurs., Sept. 10	<ul style="list-style-type: none"> <li>Analyze Sequence</li> <li>Discuss DNA sequencing</li> </ul>	<ul style="list-style-type: none"> <li>Repeat Restriction Digest</li> </ul>		B2 Handout
Tues., Sept. 15	Analyze sequence	run gel		Sequencing Handout  Sequencing Talk
Thurs., Sept. 17	Analyze EcoRI digests by graphing  Analyze sequence	"plant seeds"		MITE review  Handout
Tues., Sept. 22	Analyze sequence  Begin MITE annotation			
Thurs., Sept. 24	Finish MITE annotation	Extract DNA		
Tues., Sept. 29	Primer Design	Gel Genomic DNA		
Thurs., Oct 1	Primer Design, Order Primers	<ul style="list-style-type: none"> <li>Spec DNA</li> <li>Actin PCR</li> </ul>		

<b>Tues., Oct 6</b>	Finish annotation	gel PCR with new Primers		
<b>Thurs., Oct 8</b>	Mid-Term	gel of PCR		
<b>Tues., Oct 13</b>	<ul style="list-style-type: none"> <li>• Discuss results</li> <li>• Transposon display</li> <li>• B2 writing assignment</li> </ul>	<ul style="list-style-type: none"> <li>• PCR repeats</li> <li>• Collect DNAs</li> <li>• Run Panels</li> </ul>		
<b>Thurs., Oct 15</b>	Discuss Nature 2003 paper	<ul style="list-style-type: none"> <li>• Analyze PCR results</li> <li>• Excise Bands for sequencing</li> </ul>		
<b>Tues., Oct 20</b>	B2 writing assignment	<ul style="list-style-type: none"> <li>• Clean bands</li> <li>• More PCR</li> <li>• Germinate Seeds for RNA extraction</li> </ul>	Monday B2 writing assignment draft due by 9AM.	
<b>Thurs., Oct 22</b>	Plan RNA experiment		Friday: B2 writing assignment due by 5 PM.	

<b>Tues., Oct 27</b>	Retreat Abstract			
<b>Thurs., Oct 29</b>				
<b>Tues., Nov. 3</b>				
<b>Thurs., Nov. 5</b>				
<b>Tues., Nov. 10</b>				
<b>Thurs., Nov. 12</b>				
<b>Tues., Nov. 17</b>				
<b>Thurs., Nov. 19</b> <b>Fri., Nov. 20</b>	Plant Center Retreat			
<b>Tues. Nov 23 &amp; Thurs. Nov 26</b>	Thanksgiving Break			

<b>Tues., Dec 1</b>	Epigenetics: TE silencing			
<b>Thurs., Dec 3</b>	Epigenetics: TE silencing			
	Final			

### **Protocols for Detecting Alu polymorphism in human DNA**

In this lab you will determine the genotype for three loci in your own DNA: 225, PV92, and Yg. One locus has three Alu alleles, another locus has one allele, and one insertion locus is fixed for the population (meaning all members of the species has this one allele). You will need to analyze your data and the data from the group to determine the variation at each locus.

Re-read pages 98-102 of *Making of the Fittest* by Thursday class time.

#### **Material List**

##### **Buccal (Cheek Cell) DNA Extraction**

Heating block/water bath at 95°C

1 swab

2 1.5 ml tubes

sterile water

50 mM sodium hydroxide (NaOH)

1 M Tris, pH 8.0

#### **Extract Genomic DNA.**

1. Rinse mouth with water.
2. Label a 1.5 ml tube with your initials.
3. Put 500  $\mu$ l of sterile water into the tube.
4. Swab each cheek 20 times.
5. Put swab in tube. Twirl the swab vigorously. Remove swab and discard in beaker of bleach. Snap the tube shut.
6. Flick tube with finger. The solution should be cloudy.
7. Centrifuge at maximum speed for 2 min. You will see a pellet of cells and the liquid 'supernatant' on top. The pellet contains the cells you want.
8. Pipette off the liquid.



9. Add 200  $\mu$ l of 50mM NaOH to the pellet. (Careful, this is a strong base.)  
Flick to resuspend the cells.

10. Heat tube for 5 min. at 95°C. The cells will lyse during the incubation.

11. Add 20  $\mu$ l of 1 M Tris, pH 8.0. (Tris is a buffer and will neutralize the NaOH.)

**Make a 1:10 dilution of the DNA. (Dilution is critical to success of the PCR.)**

12. Label a second tube with your initials and '1/10'.

13. Pipette 90  $\mu$ l of water into the tube.

14. Pipette 10  $\mu$ l from your DNA sample into the new tube.

15. Close the tube and flick gently with your finger to mix. Keep on ice.

**Set up the PCR.**

In your notebook design the PCR mix for each Alu. You will do a 25 $\mu$ l reaction. Note: you cannot use the green Master Mix for this experiment!

Reagent	$\mu$ l
2X MM	12.5
Water	8.0
Your DNA	2.5
Primer F	1.0
Primer R	1.0

**Pour a Gel.** You will need a 1.5% gel.

## QIAquick PCR Purification Protocol

1. Add 5 volumes of Buffer PB to 1 volume of PCR. Mix by inverting.

If you have a 50  $\mu$ l PCR you will add 250  $\mu$ l of PB.

2. Pipette all liquid to a lilac colored spin column.

3. Centrifuge column at top speed for 60 secs.

During the spin the PCR products will bind to the column. The primers and primer dimers do not bind.

4. Discard the flow through. Pipette 750  $\mu$ l of PE to the column. Spin again.

5. Discard the flow through. Spin again.

6. Put column in a 1.5 ml tube. Add 30  $\mu$ l of EB to the column. Wait 1 minute.

7. Spin in centrifuge at top speed for 1 min. Throw away the column.

The PCR products are in the 1.5 ml tube. Keep this.

**Ban I digestion:**

To the 30  $\mu$ l recovered from the clean up add:

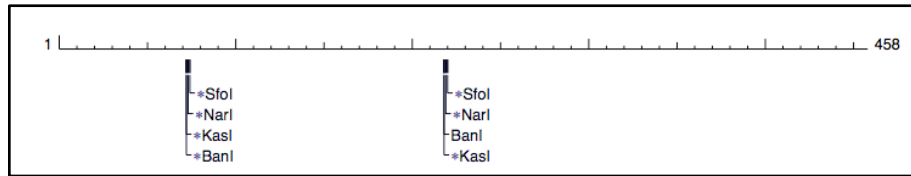
**Keep the enzyme on ice!**

10X Buffer 4	3.5 $\mu$ l
100X BSA	0.5 $\mu$ l
BanI	1.0 $\mu$ l

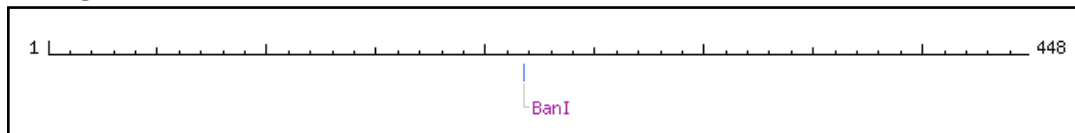
The total will be 35  $\mu$ l. Incubate the reaction at 37°C for 1 hour.

Pour a 3% gel.

### Digestion Pattern for S1



### Digestion Pattern for S2



Sizes	S1	S2	L	No Alu
Restriction	239	229	----	----
Fragments	146	219	----	----
(bp)	73		----	----
Total Size	458	448	474	135
(bp)				

### Alignment

```

225F
S1/1-458 GAGTCCAGCCCATTTTAGCATGGGACATTGAGTATGTTTTCATAACTGTTATCAAGAAGT
S2/1-448 GAGTCCAGCCCATTTTAGCATGGGACATTGAGTATGTTTTCATAACTGTTATCAAGAAGT
L/1-474 GAGTCCAGCCCATTTTAGCATGGGACATTGAGTATGTTTTCATAACTGTTATCAAGAAGT
      V   73
S1/1-458 ATTTTTATGCCGGGGCGCGGTGGCTCACGCCTGTAATCCAGCACTTTGGGAGGCCGAGGC
S2/1-448 ATTTTTATGCCGGGGTGCAGTGGCTCACGCCTGTAATCCAGCACTTTGGGAGGCCGAGGC
L/1-474 ATTTTTATGCCGGGGTGCAGTGGCTCACGCCTGTAATCCAGCACTTTGGGAGGCCGAGGC

S1/1-458 GGGTGGATCATGAGGTCAGGAGATCGAGACCATCCTGGCTAACAAAGGTGAAACCCCATCT
S2/1-448 GGGTGGATCATGAGGTCAGGAGATCGAGACCATCCTGGCTAACAAAGGTGAAACCCCGTCT
L/1-474 GGGTGGATCATGAGGTCAGGAGATCGAGACCATCCTGGCTAACAAAGGTGAAACCCCGTCT
      224
S1/1-458 CTACTAAAAATACAAAAAATTAGCCGGGCGCGGTGGCGGGCGCTGTAGTCCCAGCTACT
S2/1-448 CTACTAAAAATACAAAAAATTAGCCGGGCGCGGTGGCGGGCGCTGTAGTCCCAGCTACT
L/1-474 CTACTAAAAATACAAAAAATTAGCCGGGCGCGGTGGCGGGCACCTGTAGTCCCAGCTACT

S1/1-458 TAGGAGGCTGAGGCGGGGAGAAGGGCGTGAACCCGGAAGCGGAGCTTGCAGTGAGCCGAG
S2/1-448 TAGGAGGCTGAGGCGGGGAGAAGGGCGTGAACCCGGAAGCGGAGCTTGCAGTGAGCCGAG
L/1-474 TAGGAGGCTGAGGCGGGGAGAAGGGCGTGAACCCGGAAGCGGAGCTTGCAGTGAGCCGAG
      V
S1/1-458 ATCGCGCCACTGCAGTCCGAGTCCGGCCTGGGCGACAGAGCGAGACTCCGTCTC-----
S2/1-448 ATCGCGCCACTGCAGTCCGAGTCCGGCCTGGGCGACAGAGCAAGACTCCGTCTC-----
L/1-474 ATCGCGCCACTGCAGTCCGAGTCCGGCCTGGGCGACAGAGCAAGACTCCGTCTCAAAAA

S1/1-458 -----AAAAAAAAAAAAAAAAAAAAAAAAAAGAAGTATTTTTATATGCTTCACACT
S2/1-448 -----AAAAAAAAAAAAAAAAAAAAAAAAAAGAAGTATTTTTATATGCTTCACACT
L/1-474 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAGTATTTTTATATGCTTCACACT

S1/1-458 AATTTAAAGATGATAAAGTTTGATATGAGAAATTAATGACATGTTTGTGCTGGG
S2/1-448 AATTTAAAGATGATAAAGTTTGATATGAGAAATTAATGACATGTTTGTGCTGGG
L/1-474 AATTTAAAGATGATAAAGTTTGATATGAGAAATTAATGACATGTTTGTGCTGGG
      225R
    
```

## Reading Guide:

R. Cordaux et al., (2007) In search of polymorphic *A/u* insertions with restricted geographic distributions. *Genomics*. 90(1): 154-158.

### Paper structure:

**Abstract:** The abstract of a scientific paper is a brief summary of the entire paper. It introduces the topic, states the research question being addressed, mentions the approach used to address that question and reports the results.

**Introduction:** The introduction provides the necessary background information for the reader to understand the paper. The introduction will state the hypothesis and explain how past knowledge was built upon to form the question being addressed as well as why this question is important.

**Results and Discussion:** Although the results and discussion section is merged in this paper it is very common for there to be separate sections, one for the results and one for the discussion. The results section objectively reports precisely what is shown in the data without any interpretations or speculation. The discussion section is for interpreting the data presented in the results and using that interpretation to answer your hypothesis.

**Materials and Methods:** The materials and methods clearly explains the experimental procedures used in the study. A well written materials and methods section should contain enough detail so that a peer in the same field of study could reproduce the experiments.

**Figures and Tables:** The purpose of the figures and tables is to present the data described in the paper. You will also notice legends that accompany the figures and tables. The legend provides a brief description of the figure including details like what a color or a triangle represents. Figure legends do not include interpretations of the data - you find that in the results and discussion sections. Whenever the authors are discussing a figure or table they will reference it by including "(Table X)" or "(Figure X)" at the end of the sentence.

**Questions to consider:**

- The authors of this paper mention that *Alu* elements that have been recently integrated are more useful for determining a person's geographic origin since they are more likely to be polymorphic among human populations. Why would recent integrations be more likely to be polymorphic than more ancient integrations?
- In the results and discussion the authors state that some of the *Alu* loci were amenable to PCR (meaning that the primers were able to successfully create a band from that specific loci) while other loci were not. Based on their experimental approach why do you think that some *Alu* loci were able to be amplified by PCR while others were not? Think about your experiment with your MITE and the Zea mays inbred lines and why the primers designed for B73 did not always work in Palomero.

**Glossary:**

**Conserved target site duplications:** A target site is the genetic locus (the location in the genome) where a transposable element such as *Alu* inserts. When transposable elements insert into the genome they will often create a copy of the target site DNA sequence so that two copies (one original and one new) of the target site DNA will flank both sides of the inserted transposable element. How and why transposable elements can create target site duplications will be covered in class. When an element first inserts, both copies of the target site duplication will be identical. Over time the originally identical target sites will independently accumulate mutations. The term conserved target site duplication refers to when the sequence of the target sites flanking a transposable element have very similar (conserved) sequences. The conserved nature of these sequences indicates a recent insertion of the transposable element.

**Frequency distribution:** A frequency distribution is a method for displaying data in a table or graph. Figure 2 in this paper is an example of a frequency distribution. For creating a frequency distribution graph the data is broken up into groups, for example if we had collected data on the average

number of hours people sleep we might break the data into groups such as 1-3 hours, 3-6 hours and 6-8 hours (these group names would be found on the x-axis of a frequency distribution graph). Then we look to see how many individuals or what percentage of individuals from our data fall into each group. In the frequency distribution graph the height of the bar corresponds to the number or percentage of individuals from the data that fall into that group.

**Linker-mediated PCR:** Linker-mediated PCR is a useful method for finding locations in the genome where specific elements, such as the *AluYa8* element, have inserted. The linker-mediated PCR technique amplifies the region of DNA flanking the element at the location it has inserted in the genome. Once we have amplified this flanking region we can sequence it and then use the sequence to perform a BLAST search to determine where in the genome it is located which will tell us where the *Alu* element is inserted. A more in depth explanation of this procedure can be found in the linker-mediated PCR explained portion of this handout.

**Orthologous:** To say that a DNA sequence is orthologous means that the sequence exists in different species and that it originated from the same ancestor sequence. For example suppose the evolutionary ancestor of primates had an *Alu* element insertion in the intron of the actin gene. Millions of years later once humans and chimpanzees evolved from that common ancestor the *Alu* element in humans and the *Alu* element in chimpanzees located in the intron of the actin gene are orthologous to each other.

**Restriction enzyme:** A restriction enzyme is an enzyme that cuts DNA at specific nucleotide sequences known as restriction sites. Each restriction enzyme has a specific DNA sequence that it cuts. By knowing the sequence that a restriction enzyme cuts, if we set up a restriction digest reaction (restriction digest refers to the process of cutting DNA with restriction enzymes) by a single restriction enzyme we will know the DNA sequence at the ends of the pieces of DNA since it must be the sequence targeted by the restriction enzyme.

**Resources for Scientific Reading and Writing:**

A detailed discussion of scientific writing and the format of a scientific paper; useful for writing better scientific papers and as a guide for better understanding scientific writing.

<http://abacus.bates.edu/~ganderso/biology/resources/writing/HTWtoc.htm>

↓

This website explains the organization and structure of scientific papers and provides some useful tips for understanding the content as well as critically analyzing the findings of the paper.

<http://www.biochem.arizona.edu/classes/bioc568/papers.htm>

## MITE Mini-project

This mini project is designed to introduce you to the diversity of a MITE sub-family. We will start with MITE B2 that was found in the *waxy* gene of maize. Primers were designed this included the TIR of B2 plus 11 bp of the sub-TIR sequence.

```
>Waxy B2 MITE
TCCCAGGCAGGGCTTGGTTCGGTTAGGGCTGGATTGAGGGGGATTGGAGTGGATTAAATCCCCTTCTATAC
AAATTAAATAGGAGGGGATTTAATCCCCTCCAATCCCTCTCAAACCCCTTCAAACCGAACAAAGCCCGCA
AGGTGCGCGC
```

The TIR is highlighted in yellow and the priming sites of the primers are underlined. Using these primers you should amplify any B2-like MITEs. Because the PCR band will be a mixture of many different sequences we cannot sequence the band directly. We must first create a "library" by cloning the bands into a bacterial plasmid. The plasmids will be transformed into *E. coli* where each *E. coli* cell will take up one plasmid. From that one cell a colony will grow with over 10 million cells in it. Each colony that grows on a plate will have a different PCR product inside the plasmid. *E. coli* is a living factory that will make lots of plasmid for us. We will isolate plasmid from the cells, check that the plasmid has the PCR product insert and then have the insert sequenced.

The mini-project will also provide you with some new molecular and bioinformatics tools for use in your projects such as cloning PCR products, sterile technique, plasmid purification, restriction digestion, DNA sequencing, and DNA sequence analysis. The outline of the project:

Tuesday, September 1 PCR with B2 primers  
 Thursday, September 3 Analyze PCR products on gel, excise and clone bands.  
**Friday, September 4 Pick colonies**  
 Tuesday, September 8 Mini-prep plasmid isolation, digest plasmid, send for DNA sequencing.  
 Thursday, September 10 Analyze Sequence

**Tuesday, September 1**

We will do PCR on B73, Palomero, and B79 DNA.



Reagent	$\mu\text{l}$
2X MM	12.5
Water	6.5
DNA	5.0
B2 F	0.5
B2 R	0.5

Pour a 1.5% gel before you leave today. Use the wide-tooth combs.

### Thursday, September 3

Add loading dye to each sample and load 25  $\mu\text{l}$  on the gel. Run at 130 V for 30 minutes.

We will analyze your gel and then excise (cut-out) the bands for cloning.

#### Excising bands from a gel.

1. Weigh one 1.5 ml tube per band you will excise. Write the weight on the tube and label the tube with the band name.
2. Cut out the band and place in the tube.
3. Weigh the tube with the band and calculate the weight of the gel piece.

#### Purification of the PCR band using the QIAQuick Spin Column.

4. Add 3 volumes of Buffer QG to 1 volume of gel (100mg ~ 100 $\mu\text{l}$ ). For example, add 300  $\mu\text{l}$  of Buffer QG to each 100 mg of gel.
5. Incubate at 50°C for 10 min in a water bath (or until the gel slice has completely dissolved). To help dissolve gel, mix by inverting the tube every 2-3 min during the incubation.
6. After the gel slice has dissolved completely, add 1 gel volume of isopropanol to the sample and mix. For example, if the agarose gel slice is 100 mg, add 100  $\mu\text{l}$  isopropanol. This step increases the yield of DNA fragments < 500 bp and > 4kb.

7. To bind DNA to the column material, apply the sample to the QIAquick column and then spin at 13,000 rpm for 1 minute. The DNA is now in a high salt/non-polar solution. Under these conditions the DNA sticks to silica (the stuff in the column). The maximum volume of the column reservoir is 800  $\mu$ l. For sample volumes of more than 800  $\mu$ l, simply load again.

8. Discard flow-through and place QIAquick column back in the same collection tube.

9. Add 0.5 ml of buffer QG to QIAquick column and centrifuge for 1 min. Discard the flow through.

10. To wash any impurities (EtBr and agarose) from the DNA, add 0.75 ml of Buffer PE to QIAquick column and spin column at 13,000 rpm for 1min.

11. Discard the flow through and centrifuge for another 1 min at 13,000 rpm.

IMPORTANT: This spin is necessary to remove residual ethanol (Buffer PE).

12. Place QIAquick column in a clean 1.5 ml microcentrifuge tube.

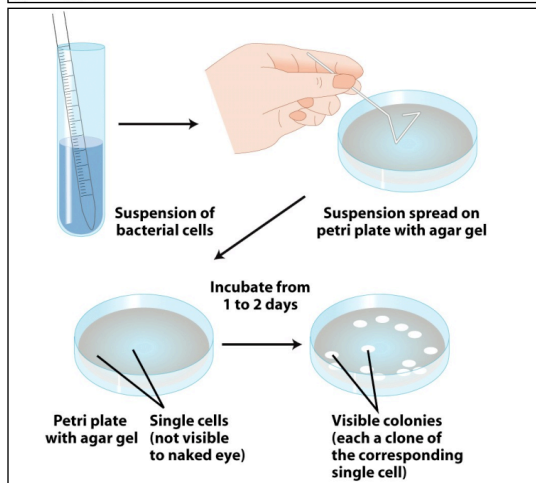
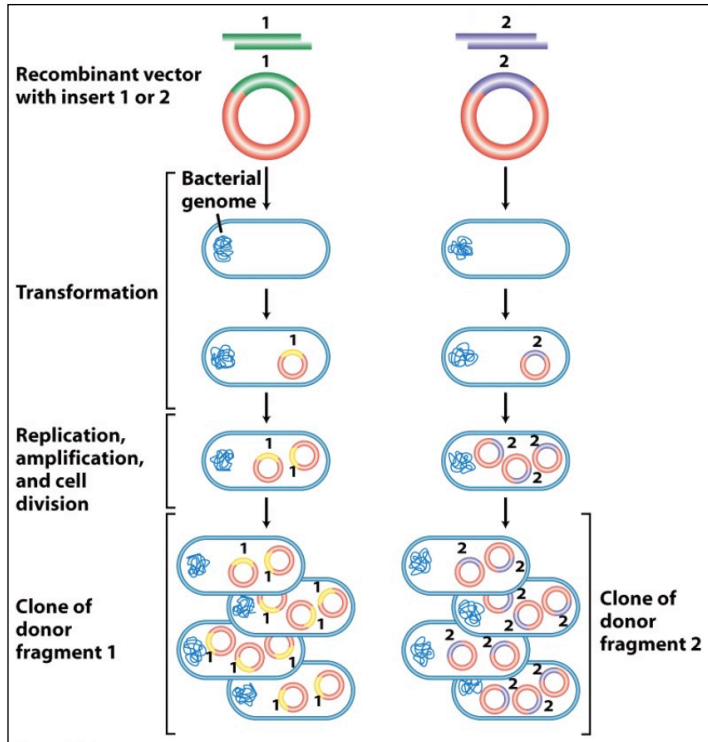
13. To elute DNA from the column, add 30ul water to the center of QIAquick membrane, leave column on bench for 1 min, and centrifuge the column for 1 min at 13,000 rpm.

IMPORTANT: Ensure that the elution buffer is dispensed directly onto the QIAquick membrane for complete elution of bound DNA.

### **Using *E.coli* to construct a "library" of DNA fragments**

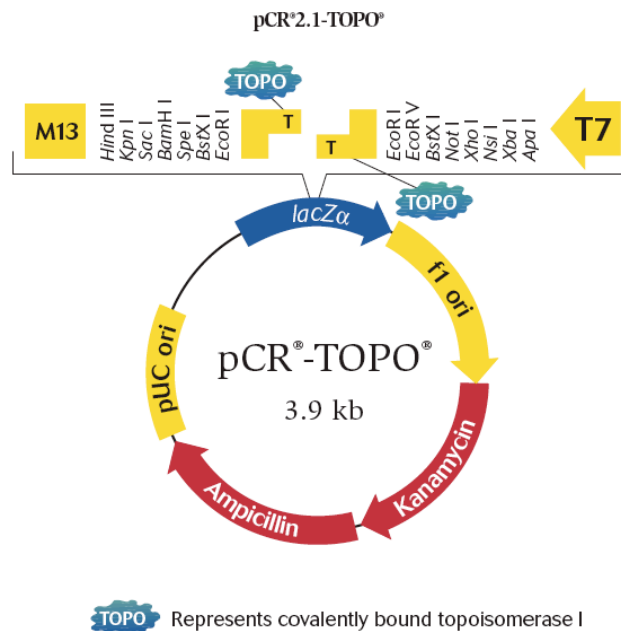
For this experiment, we will be using a patented system called TOPO Cloning from a company called Invitrogen. We will be ligating our PCR products into the TOPO plasmid (vector), transforming these plasmids into *E.coli*, picking *E.coli* colonies, isolating plasmid with our PCR inserts and then sending these to the sequencing facility. Here is some background to make all of this clearer.

To make a library, single DNA fragments are ligated into a plasmid vector and then transformed into competent *E.coli*. Individual cells with a single plasmid and insert grow into a single colony, which is grown up and used for plasmid DNA isolation.



The TOPO vector:

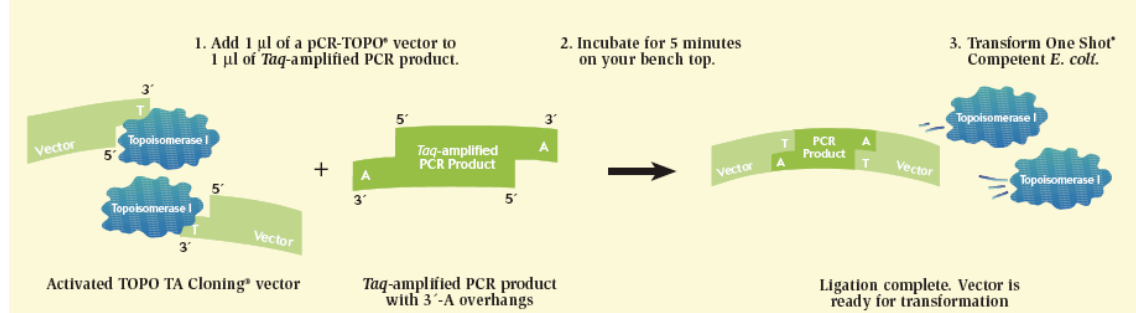
We will be using TOPO to clone the gel bands from the PCR of maize DNA then we will insert this purified DNA into the TOPO vector, transform *E. coli*, let the transformed bugs grow overnight, purify plasmid from bacterial colonies and send these purified plasmids to the sequencing facility. The idea behind TOPO cloning, according to the company's web site, is "to effectively clone DNA produced by a particular method (in your case, PCR) and to enable specific downstream studies (in your case, DNA sequencing)."



**Direct ligation with TA Cloning<sup>®</sup> Technology**

The TA Cloning<sup>®</sup> technology makes it possible to easily clone PCR products produced by *Taq* polymerase. *Taq* has a terminal transferase activity that adds a single 3'-A overhang to each end of the PCR product. TOPO TA Cloning<sup>®</sup> vectors contain 3'-T overhangs that enable the direct ligation of *Taq*-amplified PCR products (Figure 6)(2,3).

Figure 6 - How TOPO TA Cloning<sup>®</sup> works



### Materials

PCR2.1 TOPO Vector - Invitrogen

Top-10 chemically competent E. coli cells (this will be explained in class)

SOC medium (store at room temperature)

LB/Carb/S-Gal agar plates

Sterile glass beads

### Protocol

1. Place tube of Top-10 competent cells and PCR2.1 TOPO vector on ice to thaw.

2. Add the following to a 1.5 ml centrifuge tube. Pipet gently and do not mix vigorously.

4  $\mu$ l gel purified PCR product

1  $\mu$ l salt solution

1  $\mu$ l PCR2.1 TOPO Vector (add last)

3. Incubate for 10 min at room temperature (on your benchtop)

***From this point on you will be working with live E. coli bacteria. All contaminated tips, tubes, and plates must be disposed of properly (waste containers will be provided). Wash hands after handling.***

4. Transfer 2  $\mu$ l of the incubated mixture to the tube containing Top-10 competent cells (keep on ice). Pipette gently and **do not** mix vigorously.

5. Incubate the tube on ice for 20 min. (Only transformed *E.coli* cells can grow on LB selective plate. Furthermore, if S-gal is added, cells that have empty vectors will grow into black colonies and they can be easily discerned from cells with "loaded" vectors, whose colonies would be white.)

6. Incubate in a water bath for 30 sec at 42°C. (This is called the heat shock - it is when DNA is actually taken up into the bacteria from the surrounding liquid)

7. Immediately place cells on ice for 1 min.

8. Add 250  $\mu$ l SOC medium (keep sterile).
9. Incubate in a 37 °C shaker for 30-60 min.
10. Label the selective plates.
11. Pipet 100  $\mu$ l of bacterial solution onto one selective plate (work quickly to keep the plates closed as much as possible). Pour 3-5 sterile glass beads onto the plates, cover and shake horizontally to spread the liquid.
12. Dump the glass beads into the used beads container.
13. Incubate the plates **overnight** in an incubator at 37°C. Plates should be placed 'up-side-down' in the incubator. Plates will be placed at 4°C after 14-20 hours of incubation.

**Friday or Monday photograph gel and make "overnights."**

Pick bacterial colonies that have inserts (the white colonies) into test tubes containing liquid medium and grow them overnight as described below. Take a picture of each plate for your lab notebook.

Materials:

LB + Carb liquid growth medium (for growing bacteria)  
Sterile inoculating loops

Protocol:

1. Add 3 ml of liquid growth medium (LB/Carb) into sterile test tubes
2. Using a sterile loop touch a single white colony from the agar plate and twirl the loop (one per tube) into the test tube.
3. Incubate the test tubes in the air spinner overnight at 30°C.

**Tuesday, September 8: isolate plasmid from bacterial colonies, digest plasmid with EcoRI to check for inserts, send for sequencing.**

Materials:

QIAprep Miniprep kit  
Agarose gel (+EtBr), TAE buffer, 5X DNA loading buffer

Protocol:

1. Transfer 1.5 ml of your *E. coli* sample from the overnight culture to a labeled 1.5ml centrifuge tube (put tip in bacterial waste container after use).
2. Cap and centrifuge for 3 min at 8,000 rpm. Decant (dump) supernatant into the bacterial waste.
3. Add 250ul buffer P1 and vortex to re-suspend the pelleted bacterial cells. No cell clumps should be visible after re-suspension of the pellet.
4. Add 250ul buffer P2 and gently invert the tube 4-6 times to mix. Do not vortex. Invert until the solution turns a homogenous blue.
5. Add 350ul buffer N3 and invert the tube **immediately** but gently 4-6 times. Invert until the solution becomes cloudy and colorless.
6. Centrifuge for 10 min at 13,000 rpm. A compact white pellet will form.
7. Pipet ~ 800ul of the supernatant (not the white precipitate) and apply to a labeled QIAprep spin column.
8. Centrifuge for 30 sec. Discard the flow-through.
9. Add 0.75 ml PE buffer to the spin column and centrifuge for 1 minute. Discard the flow-through.
10. Centrifuge for an additional 1 min to remove residual buffer.
11. Transfer the QIAprep column to a clean labeled 1.5 ml centrifuge tube.

12. Add 50  $\mu$ l EB to the center of each QIAprep spin column, let stand for 1 min, and centrifuge for 1 min.

13. Discard the column (plasmid DNA will be in the liquid at the bottom of the tube).

14. Digest 5  $\mu$ l of the purified plasmid (the column flow-through) with EcoRI.

We will discuss how to set up the digest in class.

15. Run the digest on a 1.5% gel.

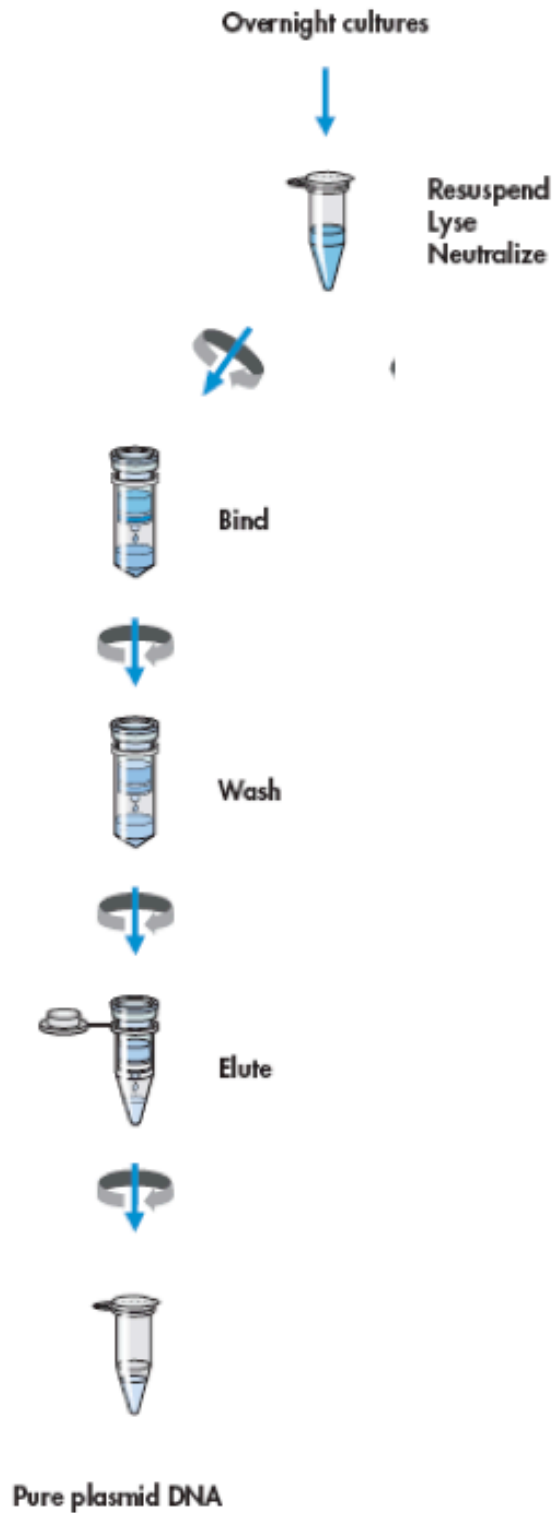
16. Pick Plasmids to send for sequencing. Measure the concentration of the plasmids with the Nanodrop. You need to put 600 ng of plasmid into a 0.2 ml tube (PCR tube). Make the total volume in each tube 8  $\mu$ l.

Calculate the amount of plasmid needed for 600 ng if the concentration is 200 ng/ $\mu$ l.

$$\frac{600 \text{ ng}}{200 \text{ ng}/\mu\text{l}} = 3 \mu\text{l in the tube. Add 5 } \mu\text{l water to make the volume 8 } \mu\text{l.}$$



# QIAprep Spin Procedure in microcentrifuges



**Preparation of cell lysates (this is a detailed summary of the “chemical logic” of the plasmid miniprep from the manufacturer)**

Bacteria are lysed under alkaline conditions. After harvesting and resuspension, the bacterial cells are lysed in NaOH/SDS (**Buffer P2**) in the presence of RNase A. SDS solubilizes the phospholipid and protein components of the cell membrane, leading to lysis and release of the cell contents while the alkaline conditions denature the chromosomal and plasmid DNAs, as well as proteins. The optimized lysis time allows maximum release of plasmid DNA without release of chromosomal DNA, while minimizing the exposure of the plasmid to denaturing conditions. Long exposure to alkaline conditions may cause the plasmid to become irreversibly denatured.

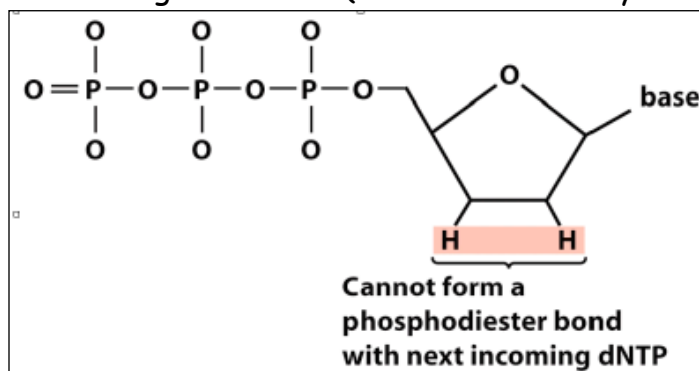
The lysate is neutralized and adjusted to high-salt binding conditions in one step by the addition of **Buffer N3**. The high salt concentration causes denatured proteins, chromosomal DNA, cellular debris, and SDS to precipitate, while the smaller plasmid DNA renatures correctly and stays in solution. It is important that the solution is thoroughly and gently mixed to ensure complete precipitation. *To prevent contamination of plasmid DNA with chromosomal DNA, vigorous stirring and vortexing must be avoided during lysis. Separation of plasmid from chromosomal DNA is based on coprecipitation of the cell wall-bound chromosomal DNA with insoluble complexes containing salt, detergent, and protein. Plasmid DNA remains in the clear supernatant.* Vigorous treatment during the lysis procedure will shear the bacterial chromosome, leaving free chromosomal DNA fragments in the supernatant. Since chromosomal fragments are chemically indistinguishable from plasmid DNA under the conditions used, the two species will not be separated on QIAprep membrane and will elute under the same low-salt conditions. *Mixing during the lysis procedure must therefore be carried out by slow, gentle inversion of the tube.*

## How your DNA samples were 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).

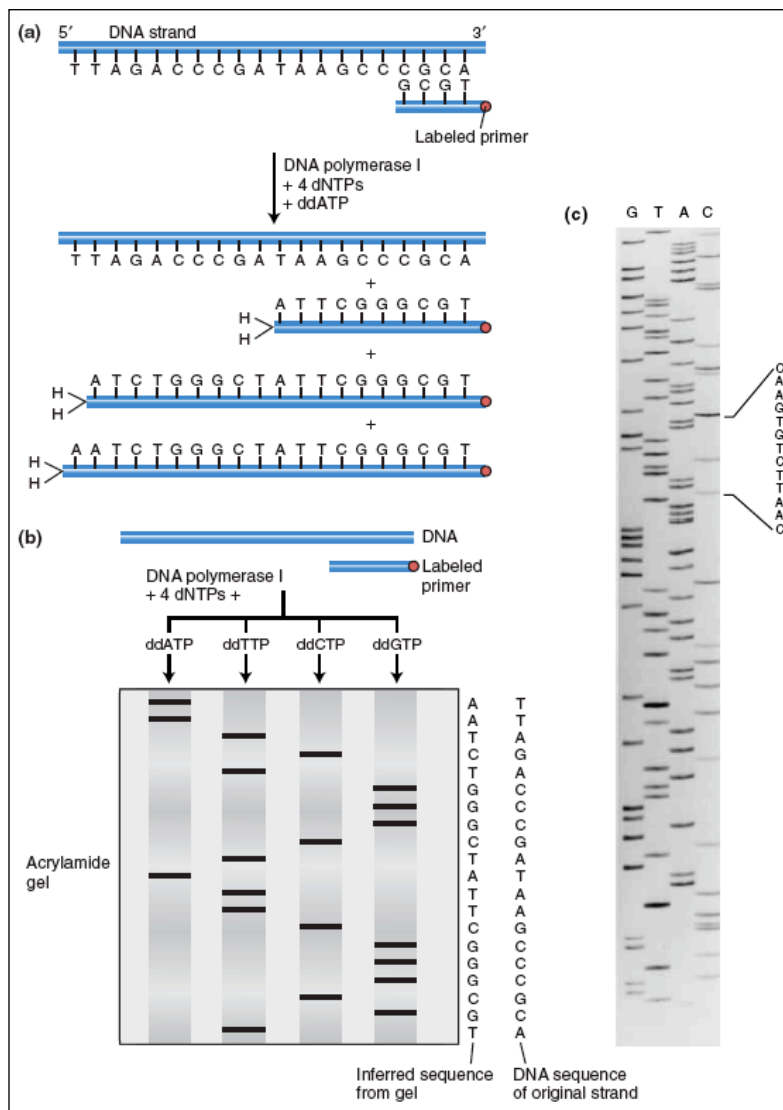


**Figure 1.** A chain-terminating nucleotide triphosphate (called a di-deoxynucleotide or ddNTP). Because it has a "H" instead of a "OH" at the 3' position, it is not a substrate for the addition of another NTP and DNA synthesis terminates.

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
← Direction of DNA synthesis		
*ATGGCGGCCCTCGGT	5'	Dideoxy fragment 1
*AATGGCGGCCCTCGGT	5'	Dideoxy fragment 2
*AAATGGCGGCCCTCGGT	5'	Dideoxy fragment 3
*ACAAATGGCGGCCCTCGGT	5'	Dideoxy fragment 4
*AACAAATGGCGGCCCTCGGT	5'	Dideoxy fragment 5
*ATTAACAAATGGCGGCCCTCGGT	5'	Dideoxy fragment 6
*ATCGATTAAACAAATGGCGGCCCTCGGT	5'	Dideoxy fragment 7
*ACCCTATCGATTAAACAAATGGCGGCCCTCGGT	5'	Dideoxy fragment 8

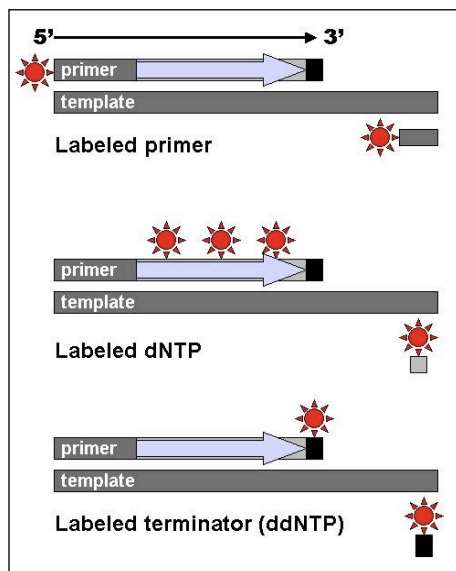
The fragments are then size-separated by electrophoresis in a slab polyacrylamide gel, or more commonly now, in a narrow glass tube (capillary) filled with a viscous polymer.



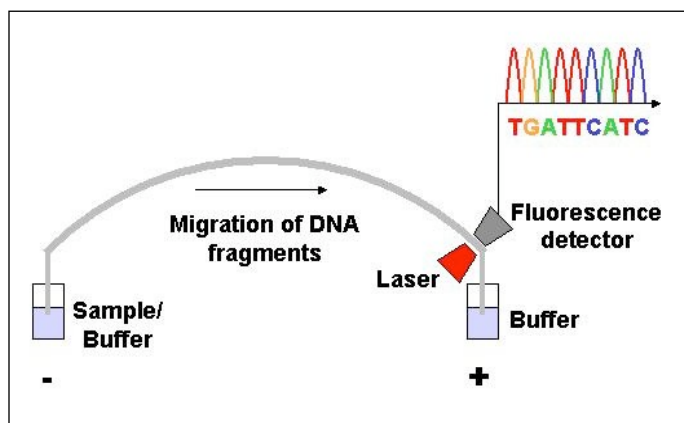
**Figure 2.** DNA is efficiently sequenced by including dideoxynucleotides among the nucleotides used to copy a DNA segment. (a) In this example, a labeled primer (designed from the flanking vector sequence) is used to initiate DNA synthesis. The addition of four different dideoxynucleotides (ddATP is 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)

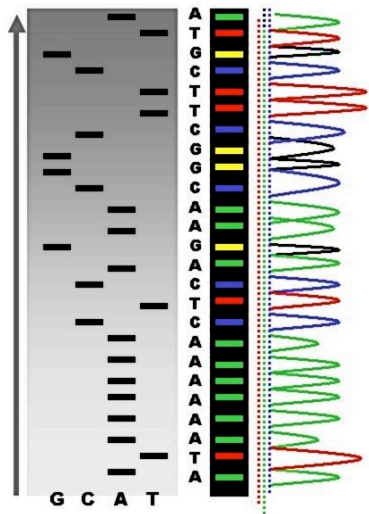
An alternative to the labeling of the primer is to label the dideoxy nucleotides instead, commonly called 'dye terminator sequencing'. The major advantage of this approach is the complete sequencing set can be performed in a single reaction, rather than the four needed with the labeled-primer approach.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



**Figure 3:** 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.

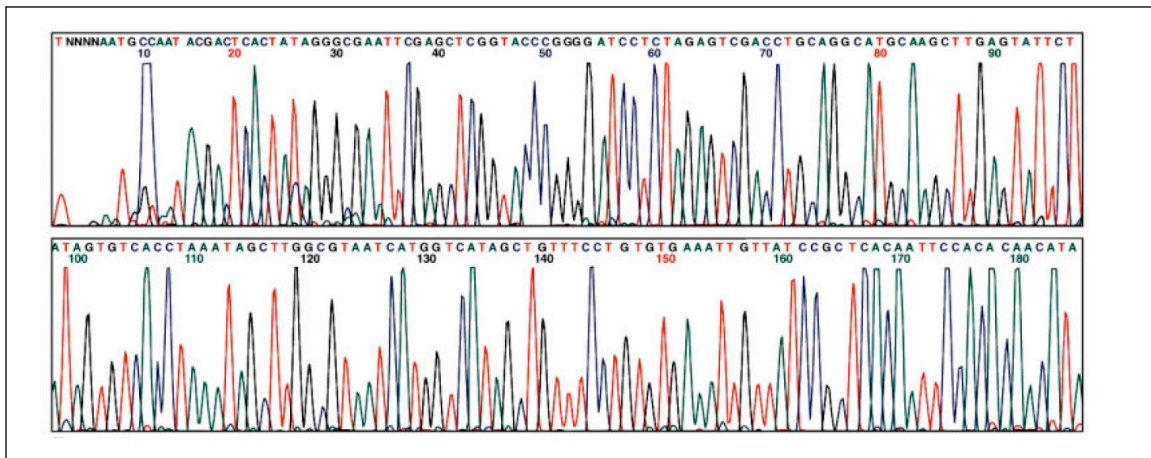


**Figure 4.** Modern automated DNA sequencing instruments (DNA sequencers) can sequence up to 384 fluorescently labelled samples in a single batch (run) and perform as many as 24 runs a day. However, automated DNA sequencers carry out only DNA size separation by capillary electrophoresis, detection and recording of dye fluorescence, and data output as fluorescent peak trace chromatograms (see Fig. 16, 17).



**Figure 5:** 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 6.** 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.

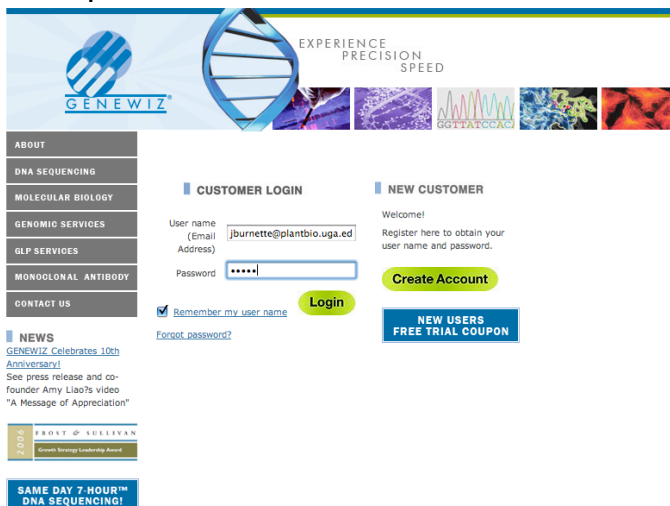
## Analyzing DNA sequencing results

Now that your DNA samples have been sequenced, you need to analyze the data. First you need to assess the quality of the data and then determine what the data mean.

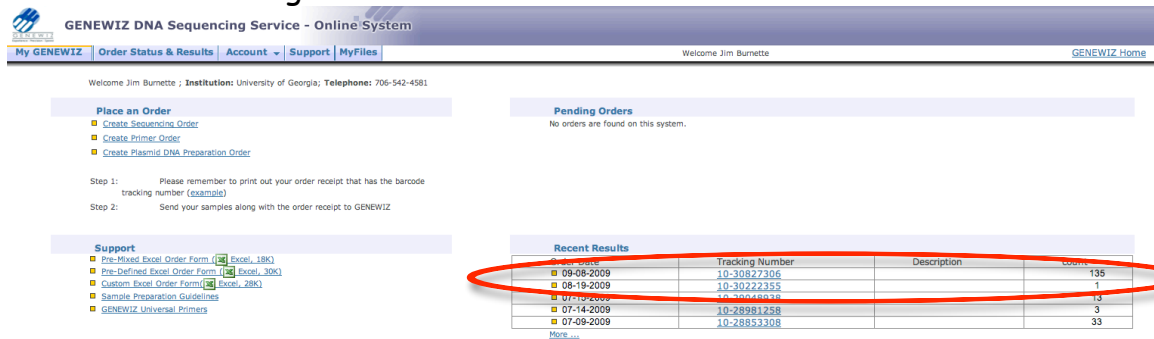
As you work on a sequence you must take notes about each step. Use a Google Docs to do this.

Steps to assess the quality of the data:

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 10-30827306.







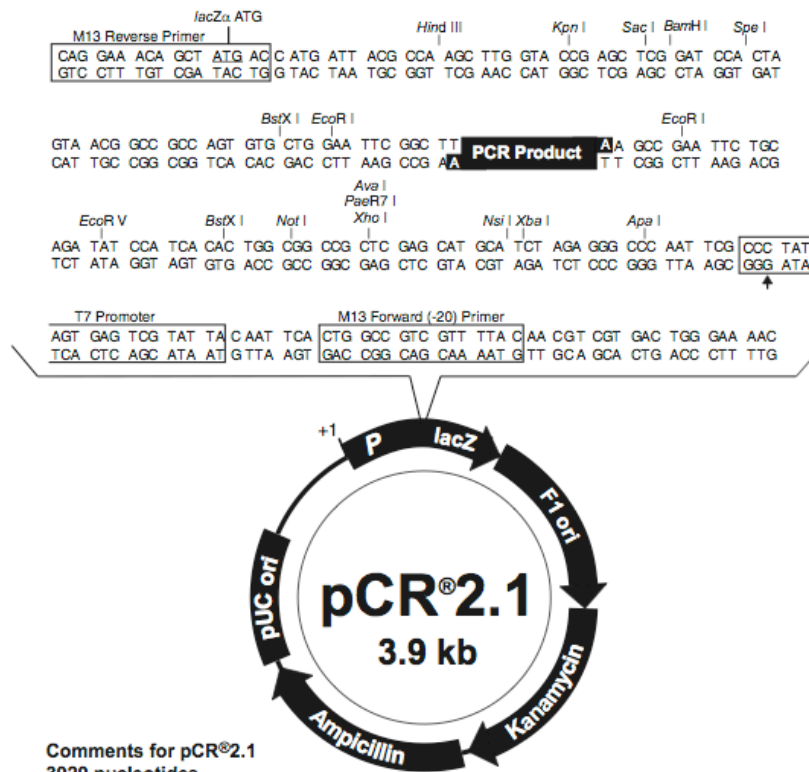


7. Using the Trace View you will need to determine the quality of each of your sequence results.

Before you do any analysis of your sequence, you must remove plasmid sequence.

## Map and Features of pCR<sup>®</sup>2.1

**Map of pCR<sup>®</sup>2.1** The map of the linearized vector, pCR<sup>®</sup>2.1, is shown below. The arrow indicates the start of transcription for the T7 RNA polymerase. The complete sequence of pCR<sup>®</sup>2.1 is available from our Web site ([www.invitrogen.com](http://www.invitrogen.com)) or by contacting Technical Service (page 18).



**Comments for pCR<sup>®</sup>2.1**  
3929 nucleotides

- LacZ $\alpha$  gene: bases 1-545
- M13 Reverse priming site: bases 205-221
- T7 promoter: bases 362-381
- M13 (-20) Forward priming site: bases 389-404
- f1 origin: bases 546-983
- Kanamycin resistance ORF: bases 1317-2111
- Ampicillin resistance ORF: bases 2129-2989
- pUC origin: bases 3134-3807

*continued on next page*

The PCR products were cloned between the EcoRI restriction sites and the plasmid was sequenced using the M13 Reverse primer. So, there will be

vector sequence on either side of the PCR insert. This sequence will interfere with analysis.

There is a modified BLAST search that will compare your sequence to a database of plasmid vectors. This will show you what part of your sequence is vector and what part is interesting sequence.

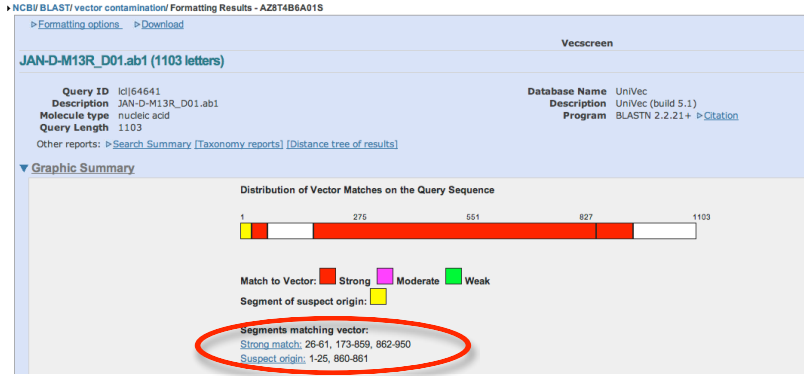
1. Open the NCBI VecScreen web page.

<http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html>

2. Copy-and-Paste a sequence in the window and click "Run VecScreen."

3. Click "View Report."

4. The results graphic shows plasmid vector sequence in red or yellow and insert sequence in white.



5. In this case the insert sequence starts at 62 and ends at 172.

6. To extract this sequence open the web page:

[http://target.iplantcollaborative.org/fasta\\_formatter.html](http://target.iplantcollaborative.org/fasta_formatter.html)

7. Copy-and-Paste the complete sequence in the text field and enter the start and stop numbers in the text boxes. Click "Format."

Fasta Formatter

Enter sequence

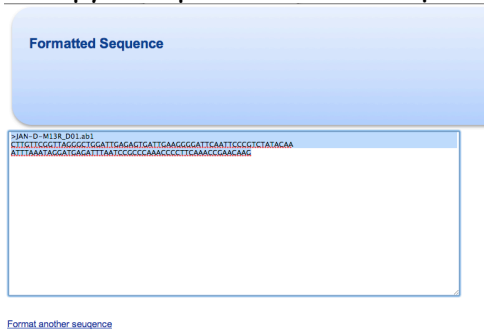
```
>JAN-D-M13R_D01.ab1
NNNNNNNNNNNCTCGGNCCNCTAGTAACGGCCGCCAGTG
TGCTGGAATCGCCCTTCTGTTGGTTAGGCTGG
ATTGAGAGTATTGAAGGGGATCAATCCCGCTATACAAATTA
AATAGGATGAATTTAATCCGCCAAACCGCTC
AAACCGAACAGAAGCGCGAATTCTGCAGATATCCATCACACTGG
CGCCCGCTNAGCATGCATCTAGAGGCCCAATTC
GCCCTATAGTCANTGNATNANNNTCACTGGCGCTGTTTACA
ACCTCGTGACTGGAAAACCTGGCGTTACCCAAAC
TTAATCGCCTTGACGACATCCCGCTTCCCGAGCTGGCGTAATAG
CGAAGAGCCCGCCGATCGCCCTTCCCAACAG
TTGCGGAGCTGAATGGGGAATGGAGCGCGCGCTAGCGCGCAT
TAACGGCGCGGGTGTGTTGTTACGGCAGCGTG
ACCCCTACACTGCCAGCCCTAGCGCCGCTCTTTGCTTTCT
TCGCTTCTTTCTCGGCACTGCGCGCTTCC
CCGTAAGCTCTAATCGGGGCTCCTTTAGGTTCCGATTAGT
GCTTACGGCACCTCGACCCAAAACCTGATT
AGCGTGATGTTACAGTACTGGCCATCGCCCTGATAGACGGNTT
TCGCCCTTGAGCTTGGAGTCCAGCTTTAAT
AGTGGACTTGTTCAAACTGGAACAACNCTAACCCATCTCGG
```

Enter a start and stop positions if you want to extract a subsequence.

Start position: 62 Stop position: 172

Format

## 8. Copy and paste the subsequence into your notes.



## 9. Remove the PCR primer sequence from the sequence.

```
>WaxyB2
GGGCTTGTTTCGGTTAGGGCTGGATTGAGGGGATTGGAGTGGATTAAATCCCCCTTCTATAC
AAATTTAAATAGGAGGGGATTTAATCCCCTCCAATCCCCTCTCAAACCCCTTCAAACCGAACAAGCCC
```

The PCR priming sites are underlined.

Now let's take an initial look at the sequence. How would you find out what the sequence is?

Record the results in your Google Docs.

A new way to analyze sequence is using a program called TARGeT. This program will search a whole genome with your query sequence and produce a phylogenetic tree. This tree will show the relationships between the sequences.


Using TARGeT to analyze sequence.

1. Open the TARGeT web page: <http://target.iplantcollaborative.org/>

## 2. Click on the link "TARGeT using blastn with DNA query."

**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 manuscript




**Choose your TARGeT start page:**

- TARGeT using blastn with DNA query
- TARGeT using blastn with DNA query

For best results use the latest Firefox browser and the latest Java. Javascript must be enabled. Safari 3.0 and 4.0 beta work well also. TARGeT is not tested on IE at all.

If you use TARGeT in your research please cite the Nucleic Acids Research Article



**Tools based on TARGeT**

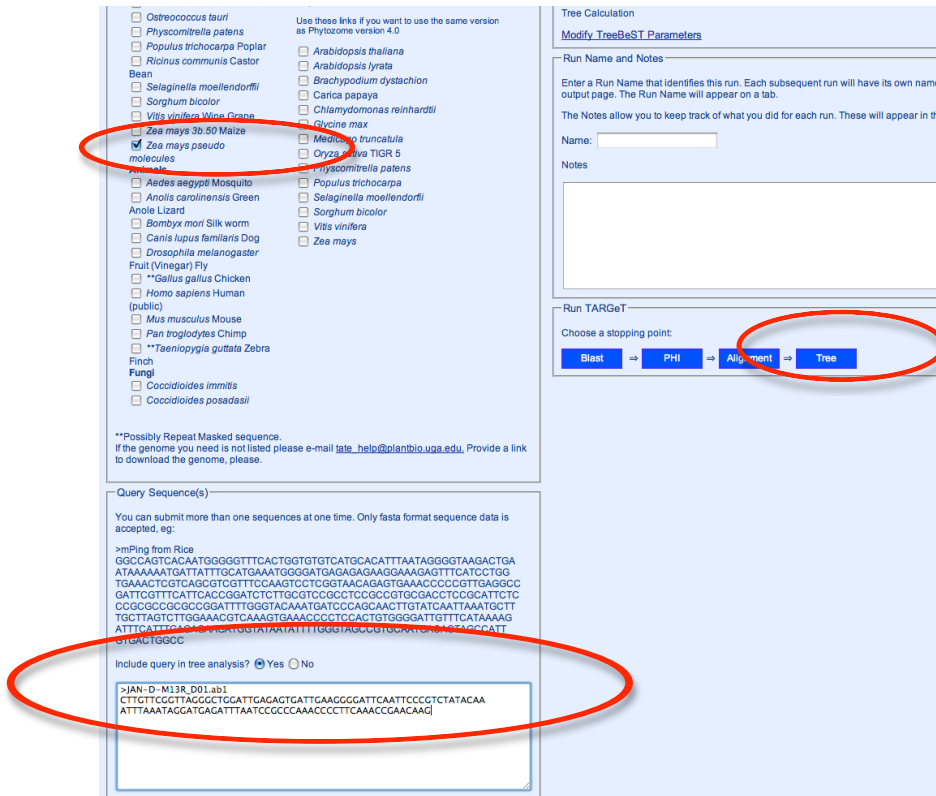
**Identify conserved regions of a multiple sequence alignment.**  
This tool can help you to choose the sequences from the conserved region for a TARGeT search or other uses.

**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

## 3. On the data entry page:

- Select "Zea mays pseudo" This is the Maize genome sequence organized in chromosomes.
- Paste the sequence into the Query Text Field.
- Click "Tree."



*Ostreococcus tauri*  
 *Physcomitrella patens*  
 *Populus trichocarpa* Poplar  
 *Ricinus communis* Castor

Bean  
 *Selaginella moellendorffii*  
 *Sorghum bicolor*  
 *Vitis vinifera* Wine Grape  
 *Zea mays* 3b.50 Maize  
 *Zea mays pseudo* molecules

Use these links if you want to use the same version as Phytozome version 4.0

*Arabidopsis thaliana*  
 *Arabidopsis lyrata*  
 *Brachypodium distachyon*  
 *Carica papaya*  
 *Chlamydomonas reinhardtii*  
 *Glycine max*  
 *Medicago truncatula*  
 *Oryza sativa* TIGR 5  
 *Physcomitrella patens*

*Aedes aegypti* Mosquito  
 *Anolis carolinensis* Green Anole Lizard  
 *Bombyx mori* Silk worm  
 *Canis lupus familiaris* Dog  
 *Drosophila melanogaster*  
Fruit (Vinegar) Fly  
 *Gallus gallus* Chicken  
 *Homo sapiens* Human (public)  
 *Mus musculus* Mouse  
 *Pan troglodytes* Chimp  
 *Taeniopygia guttata* Zebra Finch  
Fungi  
 *Coccidioides immitis*  
 *Coccidioides posadasii*

\*\*Possibly Repeat Masked sequence.  
If the genome you need is not listed please e-mail [help@plantbio.uga.edu](mailto:help@plantbio.uga.edu). Provide a link to download the genome, please.

Query Sequence(s)  
You can submit more than one sequences at one time. Only fasta format sequence data is accepted, eg:  
>mPing from Rice  
GGCCAGTCACAAATGGGGGTTCACTGGTGTGTCTGCACACATTAATAGGGGTAAGACTGA  
ATAAAAAATGATTTATTCATGAAATGGGGATGAGAGAGAAAGAAAGTTTCATCTGG  
TGAACCTCGTCAGCGTCTTTCCAAATGCTCGGTACACAGAGTGAAGCCCGCTTGAGGCC  
GATTCGTTTATTCAACGGATCTCTTGGTGGCCCTCCGGCTGGGACCTCGGATTTCT  
CCGGCCCGCCGGGATTTGGGTCAAATGATCCAGCACTTGTATCAATTAATGCTT  
TGCTAGCTTTGGAAACGTCAAAGTGAACCCCTCCACTGTGGGGATTGTTTCATAAAG  
ATTTGATTTGAAATGCTATTTATTTGGGTAGGCTTGGGCTGGGCGGATTTGAGCGATT  
CTGACTGGCC

Include query in tree analysis?  Yes  No

```
>JAN-D-M13R_D01.ab1
CTTGTCTGCTAGCGCTGGATTGACAGTGATTGAAGGGGATTC AATTCCGCTCTATACA
ATTTAAATAGGATGAGATTTAATCCGCCCAACCCCTTCAAAACCGAACAAG
```

Tree Calculation  
[Modify TreeBeST Parameters](#)

Run Name and Notes  
Enter a Run Name that identifies this run. Each subsequent run will have its own name as output page. The Run Name will appear on a tab.  
The Notes allow you to keep track of what you did for each run. These will appear in the Name:  
Notes

Run TARGeT  
Choose a stopping point:  
Blast => PHI => Alignment => Tree

## 4. The results will be explained in class.

Links to Bookmark:

<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>

<http://www.ebi.ac.uk/Tools/muscle/index.html>

<http://target.iplantcollaborative.org/>

You must take notes while you are analyzing your sequences. [See this example](#) for ideas on how to organize your notes:


1. **Compare multiple elements at one time.** Using a multiple alignment we can quickly see which of your elements are unique and should be investigated further. To do a multiple alignment you need the trimmed sequence of all the small elements.

- Change "Output Format" to "ClustalW2 (Strict)."
- Paste the sequences.
- Click "Run."

EBI > Tools > Sequence Analysis

### MUSCLE

MUSCLE stands for **M**ultiple **S**equence **C**omparison by **L**og-**E**xpectation. MUSCLE is claimed to achieve both better average accuracy and better speed than [ClustalW2](#) or [T-Coffee](#), depending on the chosen options.

 [Download Software](#)

**MUSCLE**  
multiple sequence  
alignment method  
with reduced  
time and  
space  
complexity

<b>RESULTS</b> interactive	<b>SEARCH TITLE</b> Sequence	<b>YOUR EMAIL</b> 
<b>OUTPUT FORMAT</b> ClustalW2 (Strict)	<b>OUTPUT TREE</b> none	<b>OUTPUT ORDER</b> aligned

**a**

Enter or Paste a set of Sequences in any supported format: Help

```

>KGT-C-M13R_R_C02.ab1
CTTGTTCGGTTTGAAGGGGTTTGAAGAAGGATTGTAGGAGATTAATCCCCTCTTATTTAA
ATTTGTATAGAAGGGGATTAATCCCCTCTATCCCCTCAATCCAGCCCTAACCGAACAA
CA
>KGT-B-M13R_R_B02.ab1
CTTGTTCGGTTTGAAGGGGTTTGAAGGGGTTAGAGTGGATTAATCCCCTCTTATTTAA
ATTTGTATAGAAGAAGATTAATCTCCTCCAATCCCCTCAATCCAGCCCTAACCGAACAA
AG
>JAN-D
CTTGTTCGGTTAGGGCTGGATTGAGAGTGATTGAAGGGGATTCAATCCCCTCTATACAA
  
```

Upload a file:  no file selected

**b**

**c**

A multiple alignment is a very useful way to compare more than two sequences at the same time. When all nucleotides in a column are the same an "\*" is placed below the column. From the output you can easily see that one sequence is very different from the other two and that two of the



2. **Small elements: Determine if there are similar elements in the maize genome.** Blast the element using the NCBI nucleotide blast against the *Zea mays*.

- a. Paste the sequence.
- b. Select "Others" and Type in *Zea*. Select "*Zea mays* L. (taxid:4577)" from the pop-up menu.
- c. Select "More dissimilar sequences."
- d. Click "Blast."

The screenshot shows the NCBI BLAST search interface. The 'Enter Query Sequence' section contains a text box with the sequence: >JAN-D  
CTTGTTCCGGTTAGCGCTGGATTGAGAGTGATTGAAGGGGATTCAATCCCGTCTATACAAAT  
TTAAATAG  
GATGAGATTTAATCCGCCCAAACCCCTTCAAACCGAACAAG. The 'Job Title' is 'JAN-D'. The 'Choose Search Set' section has 'Database' set to 'Nucleotide collection (nr/nt)' and 'Organism' set to 'Zea mays L. (taxid:4577)'. The 'Program Selection' section has 'Optimize for' set to 'More dissimilar sequences (discontiguous megablast)'. The 'BLAST' button is highlighted in blue. Red arrows point to the 'Others (nr etc.)' database selection, the 'Zea mays L. (taxid:4577)' organism selection, the 'More dissimilar sequences (discontiguous megablast)' program selection, and the 'BLAST' button.

- e. Record any interesting hits in your notes.
- f. Using a multiple alignment compare your sequences to wx-B2 and to other students' sequences. Record your findings in your notebook.



### 3. Analyze larger PCR inserts for other TEs.

- Run a nucleotide blast against maize using the NCBI website following steps a-d.
- Do you find any previously annotated elements in your element? How close is the match?
- Use Blast2Seq to find inverted repeats. In this case you will blast the element against itself.
- Open the NCBI Blast page and choose nucleotide blast.
- Check "Align two or more sequences" and paste the element in both text boxes.
- Choose "Somewhat similar sequences."

NCBI/BLAST/blastn suite

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide subjects using

**Enter Query Sequence**

Enter accession number, gi, or FASTA sequence [Clear](#)

>JBD-K-M13R\_E05.ab1  
 CTTGTTCCGTTTGAAGGGGTTTGAGGCCTTGTTTCGTTTATGCCGGATTGGTGGGTCGGAA  
 CGATTACTAACCGGATTGTTTCTAATTTATATAAACTTTGATTAGCTGAAACGATTCC  
 GGGTGCAATCCGACAGAAAAGAACAAGCCCTAAGGGCCCGTTTCGTATCGACCGGATTGG  
 C

Query subrange

From

To

Or, upload file  no file selected

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](#)

>JBD-K-M13R\_E05.ab1  
 CTTGTTCCGTTTGAAGGGGTTTGAGGCCTTGTTTCGTTTATGCCGGATTGGTGGGTCG  
 CAA  
 CGATTACTAACCGGATTGTTTCTAATTTATATAAACTTTGATTAGCTGAAACGATT  
 CC

Subject subrange

From

To

Or, upload file  no file selected

**Program Selection**

Optimize for

Highly similar sequences (megablast)

More dissimilar sequences (discontiguous megablast)

**Somewhat similar sequences (blastn)**

Choose a BLAST algorithm

**BLAST** Search nucleotide sequence using Blastn (Optimize for somewhat similar sequences)

Show results in a new window

- Scroll down and click the "Algorithm parameters" link.
- Change the "Word Size" to 7.
- Click BLAST.

g

**Algorithm parameters** Note: Parameter values that differ from the default are shown in yellow.

**General Parameters**

Max target sequences: 100  
Select the maximum number of aligned sequences to display

Short queries:  Automatically adjust parameters for short input sequences

Expect threshold: 10

Word size: 7 (dropdown menu shows 7, 11, 15)

**Scoring Parameters**

Match/Mismatch Scores: 2,-3

Gap Costs: Existence: 5 Extension: 2

**Filters and Masking**

Filter:  Low complexity regions  
 Species-specific repeats for: Human

Mask:  Mask for lookup table only  
 Mask lower case letters

**BLAST** Search nucleotide sequence using Blastn (Optimize for somewhat similar sequences)  
 Show results in a new window

h

i

j. The results will be described in class. We will also work together to annotate one element.

k. Once potential internal elements are found, you need to look for Target Site Duplications. Again we will do this in class.

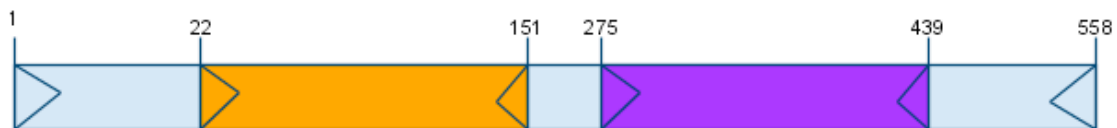
After you have annotated your element you should draw a diagram showing where the internal elements and any TSDs. I did this in Google Docs. (Note this diagram is not complete.)

>AY-S

```

CTTGTTTCGGTTTGAAGGGGTTTGAGGCCTTGTTTCGTTTATGCCGGATTGGTGGGTCCGAA
CGATTACTAACCGGATTGTTTCTCTAATTTATATAAACTTTGATTAGCTGAAACGATTCC
GGGTGCAATCCGACAGAAAAGAACAAGGCCATAAGAGCCCGTTCGTATCGACCGGATTGGG
CTGTTTCCAGGCTCCTTCCAGGTGGAATAAATTTGGTCCGGTTTGAATTCAGTTGGGGCC
ATTTCAACCGTTCGTTTAGCTTTGAACCGGAAGCAAAACAGGCTCGGTTTAACTTTCTC
CTCCCTCCATGGCCCGGAACAAAGTCTTGTTGTCTCCTCTGGATAGAATCCTGGCTAT
GGTAATACAAATTCACCAGTTTGTAGTGAACCAGCTGGATCCACTCTATTTGGGCCAGGA
ATAAAACAGGCCCTGTTTAAACCGTAACAAACGAACAGGCCCTAAGGGGGATTAGAGGG
GATTAATCCCCTCTATTTAAATTTGTATAGAAGGGGATTGAATCCCCTTCAATCCCCTC
TCAATCCAGCCCTAACCGAACAAAG

```



4. Find related elements in the maize genome using TARGeT. You can use your new elements to find other related ones in the maize genome using the maize browser. The drawback is that the results are not organized in a way to easily see how related various elements are. A program called TARGeT was written to help find related sequences but also display the relationships using a phylogenetic tree. First we need to learn about phylogenetic trees.

### What are phylogenetic trees?

Here we have a graphical representation of a phylogenetic tree. Notice the terms and what they refer to.

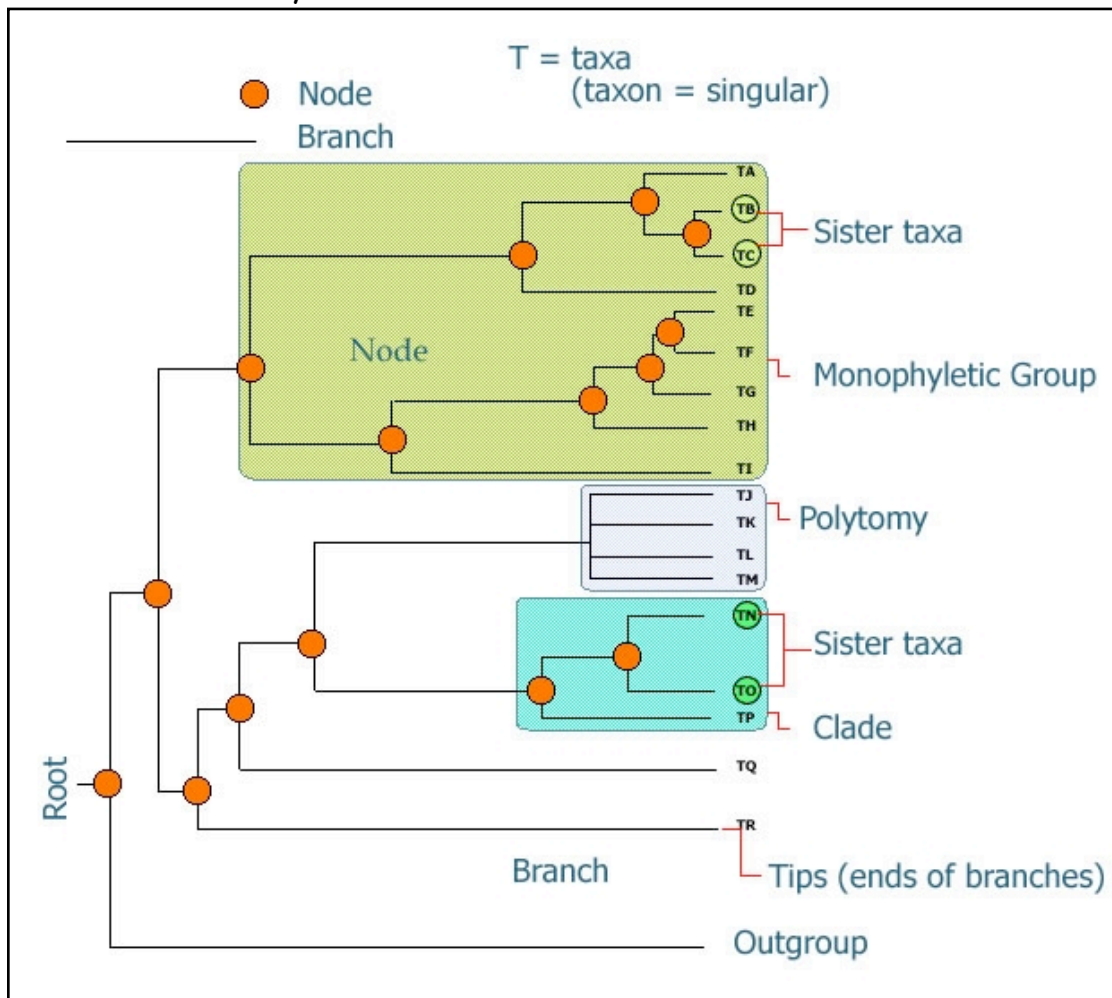


Figure 1. This figure shows a graphical representation of a phylogeny. The important features of the phylogenetic tree are

Before you can interpret a tree you need to understand some terminology. The tips of a tree represent the sampled individuals. These units are called

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

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

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

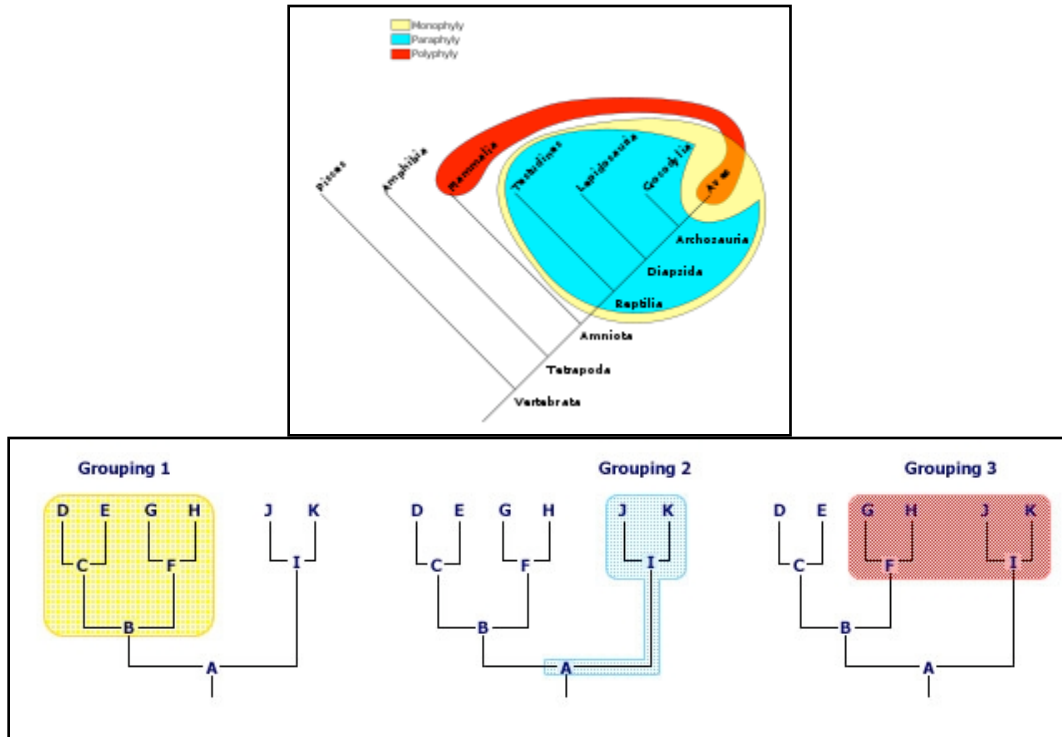


Figure 2. Monophyletic groups are highlighted in yellow, paraphyletic groups are highlighted in blue, and polyphyletic groups are highlighted in red. The tree of the vertebrates gives us an example of a monophyletic group-the sauropsids, a paraphyletic group-the reptiles, and a polyphyletic group-the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

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

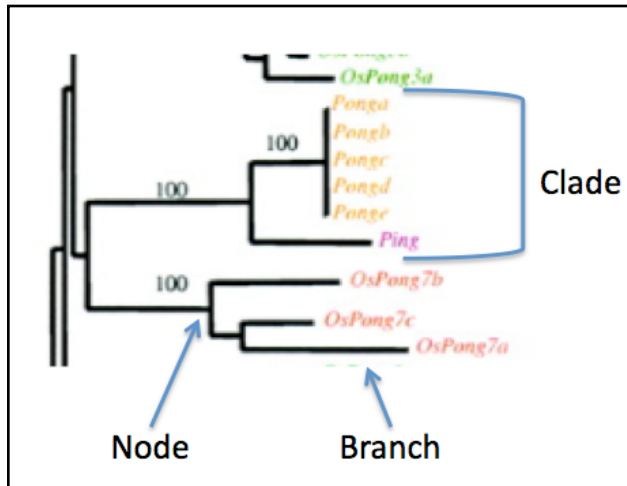


Figure 3. A clade that includes the elements most closely related to the Ping element in the rice genome.

### How are phylogenetic trees constructed?

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

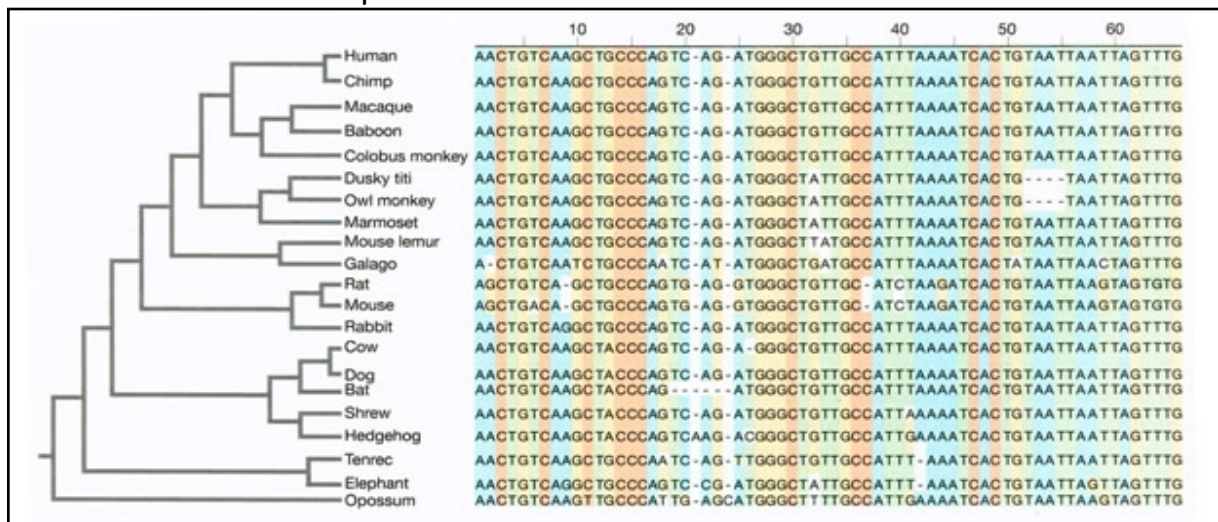


Figure 4. Sequence alignments show the relationships between the mammals. The nucleotide sequences alignment visually demonstrates nucleotide similarities and differences while also showing the presence of gaps in the sequences. The level of similarity between the sequences is used to reconstruct the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

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

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

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



that requires the fewest changes is the most likely answer. There are many different approaches to generate the best tree. The most common methods include: neighbor-joining, Bayesian, and maximum likelihood methods. If you are interested, we can go into more detail about the specifics of these methods.

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

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



## Using TARGET to infer phylogenetic trees

Now that you have an understanding of what a phylogenetic tree is you need to learn how to construct them. We will use the TARGET pipeline to construct trees. There are four steps to constructing a tree:

1. Obtain the sequences you want to compare. This is usually done with a BLAST search.
2. Reformat the Blast results into a machine friendly format called FASTA. Put together 'split hits' in the blast result.
3. Generate a multiple alignment with the sequences.
4. Construct the tree.

TARGET combines all of these steps into one web page. TARGET really shines at step 2 and saves you hours of busy work. The whole process using TARGET requires 2-3 minutes. The results of each step are presented to you in a unified web page.

### 1. Open the TARGET web page:

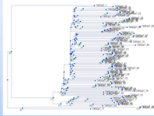
<http://TARGET.iplantcollaborative.org/>

### 2. Select the second link "TARGET using blastn with DNA query."

**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 manuscript](#)



**Choose your TARGET start page:**

- **TARGET using blastn with DNA query**
- TARGET using tblastn with Protein query

*For best results use the latest Firefox browser and the latest Java. Javascript must be enabled. Safari 3.0 and 4.0beta work well also. TARGET is not tested on IE at all.*

If you use TARGET in your research please cite the [Nucleic Acids Research Article](#)



**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 255 visitors.

Search	Number	Total
blastn	4.2x10 <sup>4</sup>	1.3x10 <sup>7</sup> nt
tblastn	2.4x10 <sup>4</sup>	1.0x10 <sup>6</sup> aa

Contact TARGET Help with questions and comments.  
Hosting and sysadmin provided by the iPlant Collaborative.



Visitor locations  
Click to see



3. Select the "Zea mays pseudo" database and paste your query sequence into the "Query Sequences" text box.

This database is the most recent version of the maize genome sequence.

**TARGeT: Using a Nucleotide Query (blastn)**

Choose a Database and enter a Query.

Database(s):

You can choose more than one database.

<p><b>Complete Genomes</b></p> <p><b>Plants</b></p> <p><input type="checkbox"/> Arabidopsis lyrata</p> <p><input type="checkbox"/> Arabidopsis thaliana*</p> <p><input type="checkbox"/> Chlamydomonas reinhardtii</p> <p><input type="checkbox"/> Glycine max Soybean</p> <p><input type="checkbox"/> Lotus japonicus</p> <p><input type="checkbox"/> Medicago truncatula</p> <p><input type="checkbox"/> Oryza sativa sub indica</p> <p>Rice, indica</p> <p><input type="checkbox"/> Oryza sativa sub japonica</p> <p>Rice, japonica</p> <p><input type="checkbox"/> Ostreococcus lucimarinus</p> <p><input type="checkbox"/> Ostreococcus tauri</p> <p><input type="checkbox"/> Physcomitrella patens</p> <p><input type="checkbox"/> Populus trichocarpa Poplar</p> <p><input type="checkbox"/> Ricinus communis Castor</p> <p>Bean</p> <p><input type="checkbox"/> Selaginella moellendorffii</p> <p><input type="checkbox"/> Sorghum bicolor</p> <p><input type="checkbox"/> Vitis vinifera Wine Grape</p> <p><input checked="" type="checkbox"/> Zea mays 3b.201 Maize</p> <p><input checked="" type="checkbox"/> Zea mays pseudo molecules</p> <p><b>Animals</b></p> <p><input type="checkbox"/> Aedes aegypti Mosquito</p> <p><input type="checkbox"/> Anolis carolinensis Green</p> <p><input type="checkbox"/> Anolis lizard</p>	<p><b>Partial Genomes</b></p> <p><input type="checkbox"/> Musa sp. Banana and Plantain</p> <p><input type="checkbox"/> Solanum lycopersicum Tomato</p> <p><input type="checkbox"/> Oryza rufipogon</p> <p><input type="checkbox"/> Dropped Sequence</p> <p><input type="checkbox"/> Mg</p> <p><input type="checkbox"/> Pv</p> <p><b>TE Databases</b></p> <p><input type="checkbox"/> Oryza MITEs from TIGR</p> <p><input type="checkbox"/> Zea MITEs from TIGR</p> <p><b>EST Libraries</b></p> <p><input type="checkbox"/> Zea maysEST Library</p> <p><b>Phytozome Version 4 Genomes</b></p> <p>Use these links if you want to use the same version as Phytozome version 4.0</p> <p><input type="checkbox"/> Arabidopsis thaliana</p> <p><input type="checkbox"/> Arabidopsis lyrata</p> <p><input type="checkbox"/> Brachypodium distachyon</p> <p><input type="checkbox"/> Carica papaya</p> <p><input type="checkbox"/> Chlamydomonas reinhardtii</p> <p><input type="checkbox"/> Glycine max</p> <p><input type="checkbox"/> Medicago truncatula</p> <p><input type="checkbox"/> Oryza sativa TIGR 5</p> <p><input type="checkbox"/> Physcomitrella patens</p> <p><input type="checkbox"/> Populus trichocarpa</p> <p><input type="checkbox"/> Selaginella moellendorffii</p> <p><input type="checkbox"/> Sorghum bicolor</p>
---	---

Scroll down for the Query input box.

Query Sequence(s)

You can submit more than one sequences at one time. Only fasta format sequence data is accepted, eg:

```
>mPing from Rice
GGCCAGTCACAAATGGGGGTTTCAC TGGTGTGTCATGCACATTTAATAGGGGTAAGACTGA
ATAAAAAATGATTATTTCATGAAATGGGGATGAGAGAGAAGAAAGATTTTCATCTGG
TGAACCTGTCAGCGTGTTC CAAAGTCCCTGGTAACAGAGTGAACCCCTGAGGCGC
GATTCGTTTATTACC CGGATCTCTTGGTCCGCCCTCGGACCTCGGACCTCGGATTCTC
CCGCCCGCGCCGGATTTTGGGTACAAATGATCCAGCAACTTGATCAATTAATAGCTT
TGCTTAGTCTTGGAAACGTCAAAGTGAACCCCTCCACTGTGGGGATTGTTTCATAAAAG
ATTTCAATTTGAGAGAAGATGATAATATTTGGGTAGCCGTGCAATGACACTAGCCATT
GTGACTGGCC
```

Include query in tree analysis?  Yes  No

```
>JAN-D
CTTGTTCGGTTAGGCGTGCATTGACAGTGATTGAAGGGGATTCATTCCTCTATAAA
ATTTAAATAGGATGAGATTTAATCCGCCCAAACCCCTCAAAACCGAACAAG
```

4. In the right-hand you should fill out the Run Name and Notes textboxes.

This will help you document your trees later.

Steps of TARGeT and optional settings

1. Blast:

blastn

[Modify Blast Parameters](#)

2. Pli:

Re-formatting of the Blast Results.

[Modify Pli Parameters](#)

3. Multiple Alignment:

Multiple Alignment of the Blast Results

[Modify MUSCLE Parameters](#)

4. Phylogenetic Tree:

Tree Calculation:

[Modify TreeBuild Parameters](#)

Run Name and Notes

Enter a Run Name that identifies this run. Each subsequent run will have its own name so you can easily identify it on the output page. The Run Name will appear on a tab.

The Notes allow you to keep track of what you did for each run. These will appear in the output window.

Name:

This is an example of 82-like element

Run TARGeT

Choose a stepping point

[Blast](#) [Pli](#) [Alignment](#) [Tree](#)

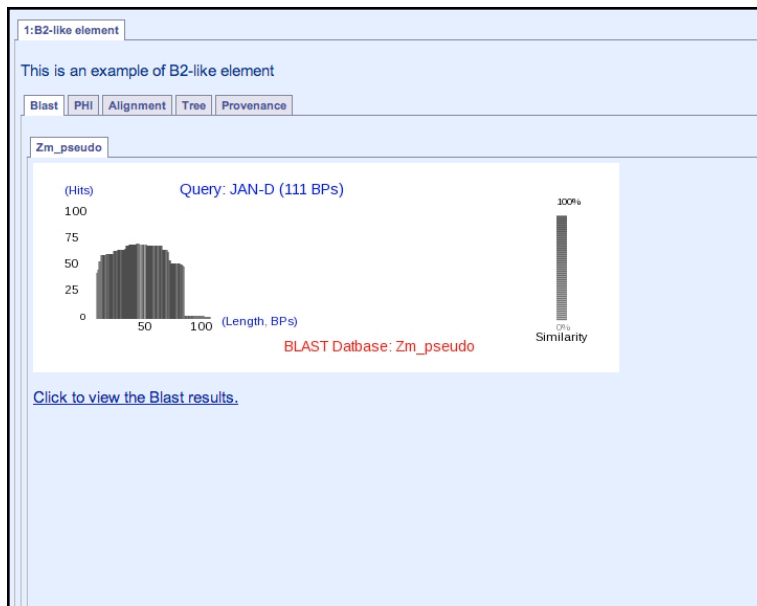
The other sections in this column allow you to modify the programs that perform each step of TARGeT. For now the defaults are fine. Later we will discuss when you need to modify these parameters.

5. Click the **Tree** Button.

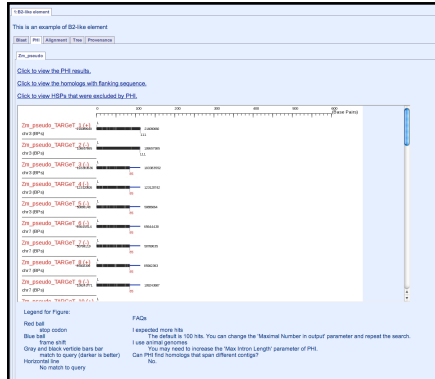
Your query will be submitted. Be patient it will take a couple of minutes.

6. The results of the TARGeT run are organized in a set of tabs. Each step gets a tab. There is a fifth tab called Provenance. You can download the files generated by TARGeT on this tab. If you included a Run Name and Note you will also see them.

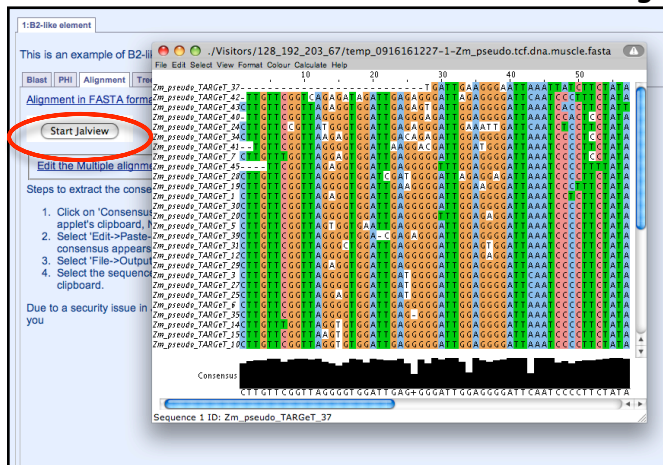
8. You will find the BLAST results on the first tab.



9. PHI output tab. This tab shows the results after the Blast sequence is re-formatted and introns (if present) are computationally 'removed'. The left hand window shows the sequence identified by TARGET as significant matches to the query. The format of the sequence is called FASTA.



10. Multiple Alignment tab. The multiple alignment program (called Muscle) uses the FASTA file generated by PHI. To see the alignment you need to click on the "Start Jalview" button. The alignment will open in a new window.



11. You will find the tree constructed from the multiple alignment on the tree tab. The tree was constructed with a program called TreeBest.

1: B2-like element

This is an example of B2-like element

Blast | PH | Alignment | Tree | Provenance

There are three options for viewing the tree.

1. With Jalview you can open the tree and the multiple alignment. By selecting one or more tip you can obtain the
  - 1. Click on the button to launch the tree in Jalview.
  - 2. Uncheck "Fit to Window" in the "View" Menu. Expand the window to view your tree.
  - 3. Click branch labels to select them.
  - 4. Click on the Button to Pasted selected sequences in the window.

2. With PhyloWidget you can change the tree layout and have the tips colored by species. PhyloWidget also has n
 

1. Open the Newick File. Copy it to the clipboard.
2. Open PhyloWidget.
3. Select "File->Load Tree->Manual" Input from the PhyloWidget menu bar.
4. Paste Newick file in window.
5. For help with PhyloWidget read the extensive documentation.

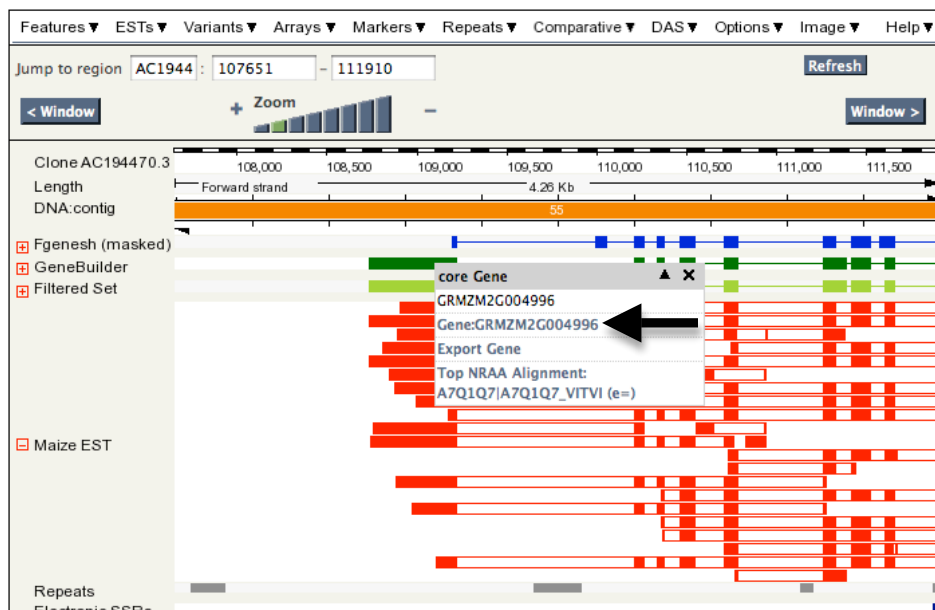
3. Download the Newick file and use your preferred tree rendering program. Right click or Apple Click on the link to
 [Newick Format Tree File.](#)

Pick branches from tree to use in a new Blast.

Annotating your MITE.

Refer to pages 65-81 for using the maize genome browser.

1. Find the genomic location of the MITE in the B73 genome using pages 65-72.
2. From the Contig view, find the location of the MITE. A good candidate for study will meet these criteria:
  - A. In an intron.
  - B. In a region with no other TEs.
  - C. Strong EST evidence for flanking exons.
  - D. Intron not too big (1000 bp or less).
3. To learn more about the gene, click on the green or blue gene model to get a contextual menu. In that menu, APPLE + click on Gene: GRMZMXXXXXXX



4. A new page will open with information about the gene. Read through this page.



For your notes you should have:

- The gene name.
- The location of the MITE.
- The sketch of the gene showing exon locations and the MITE location.
- A summary about the gene function from the *Gene View* page.
- Anything else you think is relevant for studying this MITE.

6. Following the instructions on page 73 get the sequence surrounding the MITE. In this case be sure to enter a number for the flanking nucleotides that will include the exons that flank the MITE. For this example I should ask for 1000 on each side. For a smaller intron, you may not need that much sequence.

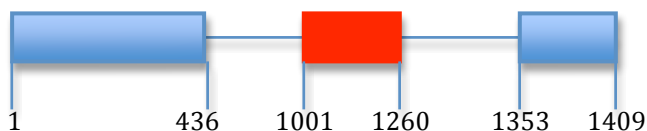
If your gene is on the bottom strand you need to make one change. Select "Forward relative to coordinate system" in the Orientation drop-down menu.

Query location : GRMZM2G004996\_1 1 to 260 (-)  
 Database location : clone:AC194470.3:1-183544 109651 to 109910 (+)  
 Genomic location : 1 293798132 to 293798391 (+)

Alignment score : 260  
 E-value : 5.6e-146  
 Alignment length : 260  
 Percentage identity: 100.00

5' Flanking sequence  (bp)  
 3' Flanking sequence  (bp)  
 Coordinate system   
 Orientation  ←  
 Alignment markup    
 Feature markup    
 Line numbering

7. Make a sketch of just the exon-intron-exon group with your MITE.



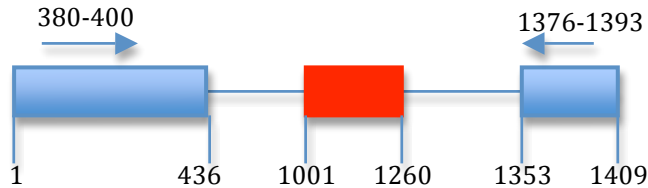


>Genomic +/- 1000

```
ATAAATCTCCTTTAATTTTTCTTCCACCAAACCAGCCCCTGAGTACGATGGACACTGGCT
CCCATACGGTGAACATACCCAAAACAAATGACGATGTCCCAATGTTATGTGCGAAATGCC
AGAAAGGCTGAAAGCAATCAATTC AAGGAAAGCCTAGGCACTTAAAAAATGACGTGCGG
CGTTACTGTGAGTTC CACAAGGATAAAAAACGCCGTCATGTAACGAACGACCAAAAAC
AAAAGGCTGTCATGTAAAAATAATCAAACCTTCTTTACAAGACTATGGAATCCCTTGCTCT
GAATCATGTGCGTAATAAGGCAGCACAGCCCCATCTTATGCGTGAGTGTGTACGAACAGC
AGCAACATCGATTGCCCCAAAACCTTGGGGGTTTGCATCATCTTCCCTTTTGATACATACA
AATCACAGTGACTTTGCGTGTGGAATGAAGATTTGGAATTAAGGTACACGGCAGGCTTGC
TATGTGCATAGAGGAGAGGGTTGTAACCTGATCTTATGCATTTTTTAGAGCAGCCAGTCT
CTTCTCTAGTTCCCTCAACTCCGAACTGCAGATGAGAGAAGCAAAGCATTTACTTTTATT
TTAACCCAAAGGGACCACAATTTT CAGTAATGTGGCAAGGGTGTGGCCTTCGCTAGCAG
CAATTCATGTTTTGTACGCAGCGAGGAATCGTTGGTATACAGCTTGTATGTATGCTTAGTC
TAGTCTTATTATTAGTTAGAAAGTACAAAAGATCATTAATAATATAGAAAACCAATGGTAAG
GTAATCAGATGCATAAAAACGACATAGAAACGTTTGAAGCTGTCATTTTTGTGCGTGAA
AGTTGCCTAGATATCTCAGCACGTCACCAGTAATGCTCGAAAGTCAATACAAGTAAAGAA
TGTTAAACCTTATACCAGACCCGTCCTGCTATTTCCAAGAAATGCAGACAACAGCTGTG
TTTCTTCAAACCATTCCAAACGTCTCAGATGAAAGCTACTCCCTACGGTTTCTATTA
GTTGTGCTTTTAGGACAACGACACGGTCTCCAAAATATAACTTTGACCAATGTTTTTTGT
TAAAATACAAATGAACCTTTAATACATTTATACTTTTATAAAAAGTACTTTTTATGACAAA
TTGGTGCATATAAATATTAGGTTCCAAAAC TAAATAACAAAATAGTTATTTGTAGTCAAA
ATTTTATAAGTTTGACTCGAACCTTATCCAAAACGACAAC TAATAGGAAACCGGAGGGAG
TAAGCAATCAAAGCAATAGCCGTGTGACTAATTAATCTAATTTTGGGCATCCAACGGTTT
CGCACAAACTGGGCAACTCACTGTGTAACACCCCTAAATACAGGGGATAGAGCCTTAGAA
CTATCTTAGTTATTT CAGTAAACCTAAAAAATAAGTCAGGGAGCTTAAATGTTATCCTG
TCCTTTTGGTACTGATTT CAGGTGGAAACAATAATACAGCAACCATATGACTGAGAGGAT
GGAAGTTGAACAACAATAAACAATGGTAGAATTGCCGGTGAAGTATAGAGCAAACCTTCC
ATCAACCTGAACCTTCTTCCAGCAATTTTTCTTTGGGAGCCGAGGACAAC TACAAGGG
GAAAGATGAAAAAGGTTTATATTAAGGAATATATATATGGAATACAAGAATCAGCATCGT
CAACTACCTGTGATGCAATGTCTACACCAATCTCATCCAGAACCTAAAAGACAATGGCAA
GAAACAGAATGAATAATCAAAGCAAAGTGC GGGGCTTATGGAAGAACAAATGGAAATAAA
AACAGTGACTGACCTGATTAGCAAGTTCTT CAGTTTCTTCTCGGCCTGGTCATCGTCTA
AGACATCATCGATTGAATCAGACATCATCTCATCTGCACAATGCATAAGATTAATCAGA
ACCCTCACATGCATCATTTATATATTAATGCGTAAGCACACACCACATTCATAAGAACTT
CCCTAGAACAACAAAAGATAAAAACATCTATAATTGCAAGAAATCATTCTGCTAAGCTTT
GGAGAAGTTTGCATACTGT CATATCCATTTGTGCTGACTGCTTTTGGAAATCTTGCATTA
TTTTCATCTGCTTGGTAGGTTCCATTTGCTGCAAAAATAGATAAGTACACATTCATAAAAA
CATATGTTTTAGGGGGGAAACACATGTTTGATTATTA AAAAATGTACTCTGAGTAATAG
AAGTTACAAGTTATATACACCCGTAGTTCTTGCATAAA
```

8. Refer to pages 78-81 to pick primers with the following modifications:
- The Targets should be the end number of the left exon and the second number should be the length of the intron, so I used 463, 917.
  - Change the Product size ranges to a range that will fit your intron. I needed 750-1000.

9. Before you order your primers, draw them on your sketch to make sure they are in the correct place.



In your notes you should include the new sketch and screen shot of your primer results.

Fill out the form on the Data web page to order your primers.