PREFACE

This manual contains nine chapters covering a few basic methods and procedures used in plant molecular biology. These methods are diverse and include measurements of the size of plant genomes, DNA and RNA extraction and analysis, bioinformatics, the polymerase chain reaction, cloning, DNA sequencing, measurements of gene expression, and transient transformation. A set of “Learning Objectives” and “Lab Objectives” is included in each chapter. The first set lists the topics students are expected to learn and is a useful study guide. The second list names the main lab activities. Some labs will be completed in one session, while, others will be completed over more than one session. Furthermore, there will be sessions in which students will work on more than one activity. Some experiments will require short and intermittent interventions (every 5 to 10 minutes), and will give the appearance of a lot of down time; there is simply no way to compress the procedures to shorten incubation or reaction times. This is the nature of molecular biology techniques. In real world situations, researchers carry out several procedures in parallel.

In addition to the objectives, each chapter provides some background information on both the biological system that will be addressed, and the principles in which the selected lab methods are based. These two general topics underscore the teaching philosophy of this course. Methods are just the means and tools used by scientists to carry out experiments and test hypotheses. The primary concern of a scientist is identifying a problem and then determining what questions should be asked to the system to solve the problem. The questions are formulated in the form of an experiment. For the experiment to be meaningful, and more specifically for the results of the experiment, the experimental methods have to be reliable and be applied in a judicious manner. To accomplish this goal, the researcher has to be aware of the biological and physical-chemical principles on which the methods are based. This is important not only for interpretation of results, but also for troubleshooting and introducing modifications suited to systems for which the methods were not originally designed. Biological systems do not change with time (the life time of a researcher that is), but methods are constantly changing or becoming obsolete as new ones are developed. For this reason, what should endure is a critical attitude towards the use of methods and procedures.
CONTENTS

Chapter 1 – Flow Cytometry and Genome Size
Chapter 2 – Plant DNA Extraction and Analysis
Chapter 3 – Bioinformatics
Chapter 4 – Polymerase Chain Reaction
Chapter 5 – Cloning
Chapter 6 – DNA Sequencing
Chapter 7 – Southern Blotting and Hybridization
Chapter 8 – Analysis of Gene Expression
Chapter 9 – Plant Transformation
MEASURING THE SIZE OF PLANT GENOMES

Learning Objectives:
After completion of this unit, students will be able to prepare plant nuclear samples for analysis in a flow cytometer. In addition, students will be able to analyze and interpret flow cytometer output data to estimate the genome size of a plant species. Interpretation of data will be carried out in the context of the cell cycle and the genetic constitution of the organism. Furthermore, students will be able to apply the knowledge about genome size data in the construction of genomic libraries, genome sequencing projects, and many other procedures described later in this manual.

Laboratory Objective:
Measure the genome size of a plant species. This will be accomplished by measuring the DNA content of isolated nuclei in a flow cytometer. Nuclei will be physically isolated from plant tissue, stained with a fluorogenic compound that has both a high affinity for DNA and an ability to significantly increase its fluorescence yield upon binding DNA. The fluorescence emitted by stained nuclei will be measured in a flow cytometer. The emitted fluorescence of each nucleus will be proportional to its DNA content. The size of the genome will be calculated using a species as genome size standard.

Background
The genome of an organism comprises the full complement of its genetic information. Plants possess three genomes: nuclear, mitochondrial and chloroplast. The presence of the third of these genomes distinguishes plants from other eukaryotic organisms. The cytoplasmic genomes are relatively very small [chloroplast 120-160Kb (Palmer, 1985); mitochondrion 200-2,500Kb (Lang et al., 1999)] in comparison to nuclear genomes which contain between 63 and 122,000 Mb (Bennett & Leitch, 1995). While a diploid leaf cell in G1 will contain two copies of the nuclear genome, it may contain multiple copies of the chloroplast and mitochondrial genomes. In this lab we will be concerned exclusively with the nuclear genome.

The nuclear genome is defined as the total amount of unreplicated nuclear DNA found in one gamete. The size of the genome is also known as the C-value, a quantity which is expressed in either picograms (pg) or megabase pairs (Mbp). This definition implies that typical somatic cells contain two genome copies. Thus, if we try to measure the size of a nuclear genome by measuring the DNA content of nuclei extracted from typical somatic cells, then we need to take into account the presence of two copies. Such measurement can get slightly complicated when the nuclei under analysis are derived from growing tissues. In such tissues, a significant fraction of the cells are going through the cell cycle. This is the process in which cells replicate their DNA and then divide into two daughter cells; this process is known as the cell cycle and it is separated into four phases:

\[
\text{Gap 1 (G1) } 2n \rightarrow \text{Synthesis (S)} \rightarrow \text{Gap 2 (G2) } 4n \rightarrow \text{Mitosis (M)} \rightarrow \text{Gap 1 (G1) } 2n ...
\]
Cells in growing tissues can be seen fluctuating between interphase and mitosis. Interphase comprises the first three stages of the cell cycle: Gap 1 (G1) is the stage that precedes the period of DNA synthesis (S), which is followed by the Gap 2 (G2) stage. During G2 cells prepare for mitosis (M). Analysis of plant nuclei via flow cytometry reveals, in most cases, two groups of nuclei differing in DNA content by a factor of 2. These groups correspond to nuclei which are in different stages of the cell cycle. Unlike animal nuclei, plant nuclei undergoing mitosis can’t be detected by flow cytometry because they lose their nuclear membrane during mitosis and are eliminated during isolation.

Nuclei with DNA content equivalent to a power of 2 series (2, 4, 8, 16, 64...) of the basic disomic nuclei can be detected in some species. These nuclei are the products of endopolyploidy – a phenomenon in which cells go through the cell cycle without entering mitosis. This phenomenon increases the ploidy level of specific somatic tissues, and appears to be developmentally regulated. For instance, it can be detected in the pods and cotyledons of legumes (Lagunes-Espinoza et al., 2000), and in *Arabidopsis* plants as a developmental gradient (Galbraith et al., 1991).

Eukaryotic organisms, plants in particular, display a wide range of genome sizes. The smallest higher plant genome is that of *Genlisea margaretae* with 63 Mbp (Leushkin et al., 2013), while the largest belongs to *Fritillaria assyriaca* with 122,000 Mbp (Bennett et al., 1995) – a 2,000-fold difference! The variation in genome size must be viewed in the context of evolutionary changes, and as such, these can be explained by two major phenomena: a) changes in ploidy level, and b) changes in the amount of repetitive DNA.

Changes in ploidy level occur when a single ancestral species doubles its chromosome number and generates an autopolyploid species. For instance, *Solanum tuberosum*, the potato, is an example of an autotetraploid. Another route for a change in ploidy level is interspecific hybridization combined with chromosome doubling resulting in the creation of an allopolyploid species such as *Triticum aestivum*. This species contains the genomes of three ancestral species. In cytogenetics the chromosome number of the basic genome is referred to as *x*, the gametic chromosome number is referred to as *n*, and the somatic chromosome number as 2n. For instance, in the diploid *Arabidopsis thaliana* 2n = 2x = 10, in the autotetraploid potato 2n = 4x = 48, and in the allohexaploid wheat, 2n = 6x = 42 (AABBDD).

Alterations in the amount of repetitive DNA are generally associated with the activity of transposable elements, mainly retrotransposons. These elements use the “copy and paste” mode of propagation. Retrotransposons are transcribed into RNA, which is “reverse transcribed” into DNA and then inserted elsewhere in the genome. In *Zea mays* these elements constitute up to 75% of the genome.

A final observation about the variation in genome size is the C-value paradox, a term that refers to the lack of correlation between the complexity of an organism and the size of its
genome. For instance, the mango tree, *Mangifera indica*, and the melon vine, *Cucumis melo* have genomes of approximately the same size, 439 and 454 Mbp, respectively; these species differ dramatically in plant size and life histories. A more striking comparison is that of the human and maize genomes which contain approximately 3,000 Mbp.

**Application**

Why is the size of the genome important in molecular biology? As you will see throughout the semester, genome size plays an important role in the selection of suitable strategies for genome analysis. Among these are the construction of a genomic library, Southern blotting and hybridization, the polymerase chain reaction (PCR), and whole genome sequencing! There are many strategies used in the isolation and characterization of a gene. Some of these strategies rely in the construction of a genomic library, a collection of clones which are supposed to represent the entire genome. Thus, to be useful, this library should have at least one copy of every gene. Ideally, we would like to isolate a single nucleus and grab one chromosome at a time, clip and clone or sequence contiguous segments of manageable size from one end to the other. Unfortunately, this approach is entirely unrealistic with current technologies. What it is done instead is to physically isolate nuclei, purify their DNA, and fractionate this DNA into suitable size fragments which can be ligated to a cloning vector of choice, or to primers designed for specific sequencing approaches. These procedures can generate from thousands to millions of clones or sequences. The question is: how many clones or fragments should a library contain to have a high probability of detecting any single gene or sequence? Intuitively we know that the larger the library the higher the probability of success, but knowing the number of clones or sequences that we would need to screen, to have a high probability of success, would go a long way!

The solution to this problem comes from statistics, more specifically from the binomial distribution which can be used to analyze the frequency of certain events. This sounds and appears complicated and intimidating, but it is neither. Let’s start by presenting the problem with a sheet of paper divided into 20 numbered squares (1 to 20). Next, let’s make a very, very large number of photocopies of this sheet, cut the squares out from all sheets, and finally place the squares in a box and mix them well.

**Question:**

How many squares do you have to pull out to have a 99% probability that at least one of the squares is going to have the number 13? (or any number between 1 and 20). Some may ask, wouldn’t it be easier to just stay with the first sheet of paper and cut out square 13? Yes, but this approach is the equivalent of isolating a single nucleus and grabbing one chromosome at a time to select a single gene. We know that this approach is not feasible. We have to consider an operation in which we go with multiple sheets cut into small squares. This operation resembles very much the way genomic libraries are constructed. We can use the binomial distribution to figure out the number of squares we need to pull out to have a given probability of finding a
particular square, or the ideal number of clones in a library to have a given probability of finding a particular gene.

The binomial distribution is used to calculate the frequency of successive independent events in which there are only two possible outcomes, each with a given probability. For instance, the number of times we need to flip a coin to have a 99% chance of getting at least 1 head (the probability of success is ½ and probability of failure is ½), or the number of squares to pull out to have a 99.99% probability of getting at least one #13 square (the probability of success is 1/20 and probability of failure is 19/20). The probability of success when screening for a clone from a genomic library is given by the fraction of the genome represented by the clone. For instance, if the average size of a clone is 100 kilobases (100 Kb) and the size of the genome is 400 megabases (400 Mb), then the probability of success for a clone is 100/400,000, or 0.00025; the probability of failure would be 0.99975. Thus, the size of the clone and the size of the genome will have a significant effect on the size of a useful library.

Different cloning vectors have different holding capacities for DNA inserts. Similarly, different sequencing technologies produce a wide range of sequence reads (from 75 bases to several Kb). As far as libraries are concerned, at the short end of the spectrum are the plasmid vectors which are used for routine cloning procedures involving inserts of up to 5 Kb. However, the typical vectors used in the construction of genomic libraries are lambda phages (10-20 Kb), cosmids (30-40 Kb), fosmids (30-40 Kb), bacteriophage P1 (100 Kb), and bacterial artificial chromosomes (100-300 Kb). From these numbers and the previous calculations we can infer that the bigger the insert, the fewer the number of clones that will be required, and on the other hand, the bigger the genome, the bigger the library will have to be. Let’s now examine the binomial distribution.

The binomial distribution has the following expression:

\[(p + q)^n = 1\]  \hspace{1cm} (1)

where, p is the probability of success, q is the probability of failure, \((p + q = 1)\), and n is the number of trials. In the case of library construction, \(p\) is the fraction of the genome (G) represented by the insert (the clone, or sequence) with an average insert size (g), or \(g/G\). To illustrate this point we will need to expand expression (1), but not before showing the mathematical function that describes the probability of events:

\[(p + q)^n = \sum_{r=0}^{n} \binom{n}{r} p^r q^{n-r}\]  \hspace{1cm} (2)

where \(\binom{n}{r}\) is the number of ways in which we can have \(r\) number of successes in \(n\) trials. The expansion of this expression is written as follows:
\[(p + q) = \binom{n}{0} p^n + \binom{n}{1} p^{n-1} q + \cdots + \binom{n}{n-1} p q^{n-1} + \binom{n}{n} q^n \quad (3)\]

The first term of the expansion represents the probability of sampling \(n\) times and being successful every time. In other words, the probability of screening \(n\) clones and obtaining the desired clone every time (this term will be very small!). The second term represents the probability of sampling \(n\) times and being successful every time, except once, and so forth and so on, until we get to the penultimate term which is the probability of sampling \(n\) times and being successful only once. The last term is the probability of failing every time. With the exception of the last term, the sum of the preceding terms represents the probability \(P\) of being successful at least once! At last we can reduce the expression to a simple and less intimidating form:

\((p + q)^n = P + q^n \quad (4)\)

Now we can ask the question, if we construct a genomic library with an average insert size of \(g\) from a genome of size \(G\), then how many clones do we have to screen to have a \(P\) probability (usually between 0.99 and 0.9999) of detecting at least one clone of a specific sequence?

Solving for \(n\) and remembering that \(p + q = 1\), we have the expression:

\[n = \frac{\log (1-P)}{\log(1-p)} \quad (5)\]

We can now calculate \(n\) by replacing \(p\) with \(g/G\).

The **Poisson distribution** can also be used to assess the occurrences of rare events. This distribution approximates the binomial distribution when \(p\) is small and \(n\) is large. In the Poisson distribution the probability of \(r\) successes \(P(r)\) tends to the following expression:

\[P(r) = \frac{\mu^r}{r!} e^{-\mu} \quad r = 0, 1, 2, 3, 4, \ldots \quad (6)\]

where the mean \(\mu = np\), and \(e\) is the base of natural logarithms (2.71828). Thus, the first terms of the Poisson distribution are written as follows:

\[P(0) = \frac{\mu^0}{0!} e^{-\mu}, \quad P(1) = \frac{\mu^1}{1!} e^{-\mu}, \quad P(0) = \frac{\mu^2}{2!} e^{-\mu} \quad (7)\]

The first term is the probability of sampling \(n\) times and failing every time. Computationally, this is the best term to calculate \(n\). Accordingly, \(\mu^0 = 1\) and by definition \(0! = 1\) as well. The probability of failure is then reduced to \(P(0) = e^{-\mu}\), and since \(\mu = np\), the probability term is expressed as:

\[P(0) = e^{np} \quad (8)\]
Solving for $n$ leads us to the expression:

$$n = - \frac{\ln P(0)}{p} \quad (9)$$

In other words, the number of clones we need to screen to have a chance of finding at least one target clone is directly proportional to the negative natural logarithm of the probability of failure, and inversely proportional to the probability of success $p = (\text{insert size/Genome size})$. If we want to find a clone with a probability of 0.999, then $\ln P(0) = \ln(1 - 0.999)$.

Flow Cytometry

Four basic methods have been used to measure the C-value of higher organisms: chemical analysis, reassociation kinetics, Feulgen microdensitometry, and flow cytometry. The latter is the current method of choice due to its simplicity, reliability, accuracy and expediency. In this lab session, we are going to use flow cytometry to estimate the genome size of several plant species. This task will be accomplished by measuring the DNA content of isolated nuclei from these species relative to the DNA content of nuclei from some standards (Aramuganathan & Earle, 1991; Galbraith et al., 1983). The strategy is to physically isolate intact nuclei, stain them with a fluorogenic compound which has a high affinity for DNA, and measure the amount of fluorescence emitted by single nuclei. The fact that the amount of fluorogenic compound bound to a nucleus is proportional to its DNA content will allow us to estimate the genome size.

Samples will be prepared by chopping young leaf tissue in chopping buffer with a brand new razor blade to release the nuclei. RNAse is included in this buffer to hydrolyze nuclear RNA which has the capacity to bind some of the fluorogenic compound. The nuclei will be isolated from the homogenate by successive filtration and centrifugation steps. Filtration through 50 and 20 $\mu$m nylon meshes will eliminate large pieces of tissue. Low speed centrifugation will pellet the high density nuclei and leave behind soluble and low density cellular components (proteins, sugars, mRNA, ribosomes, plastid membranes). After two consecutive centrifugation steps, the nuclei will be resuspended in staining buffer containing propidium iodide (PI), a fluorogenic compound with high affinity for double stranded DNA. Propidium iodide is a planar molecule that intercalates between base pairs with an average of 1 molecule every 4 to 5 base pairs; Electronic interactions between PI and adjacent base pairs enhance its fluorescence properties. PI absorbs short wavelength light (blue)
and emits it as a longer wavelength light (red). Flow cytometers are equipped to measure the fluorescence intensity emitted by each nuclei.

The flow cytometer is an instrument that uses microfluidics to separate small particles (cells, nuclei, chromosomes) into single liquid droplets (Figure 1). This task is accomplished by injecting the sample into the middle of a sheath flow in the flow chamber (hydrofocusing). A tunable transducer connected to the exit nozzle breaks up the sheath into small droplets, each containing a single particle. Some physical characteristics of the particles can be measured using a laser beam focused on the droplet (some models have up to four lasers).

A set of optical devices (lenses, filters, and photomultipliers) and a computer are used to measure and process the optical signals (Figure 2). The relative volume of a particle can be measured using small-angle forward light scattering because this angle is proportional to the size of the particle. In addition, if the particle carries a fluorophore, then the intensity of the fluorescence emitted by the particle can be measured as well. For instance, a two-parameter histogram (gating) can be constructed to identify subpopulations of particles of similar size and fluorescence intensities. More than one laser and band pass filters can be used to measure a variety of signal emitted.

In this lab however, we will use one-parameter histograms to measure the relative frequency (y-axis) of subpopulation of particles (nuclei) according to the mean value (x-axis) of their fluorescence intensity (Figure 3). These histograms are constructed as follows: the fluorescence intensity sensor is divided into multiple channels, each with a very small range. When the fluorescence of a particle is detected by the sensor, an “event” is recorded by the appropriate channel. As “events” accumulate in adjacent channels, peaks representing
particles with similar fluorescent intensities become apparent. These peaks can be viewed as populations of nuclei. Two peaks of nuclei are likely to be detected in samples derived from growing tissues, each peak corresponding to nuclei in G1 and G2 phases of the cell cycle. Thus, the mean fluorescence of the first peak (G1) should be approximately half the fluorescence of the second peak because the G2 nuclei will have twice the amount of DNA as G1 nuclei. The relative height of these peaks will depend on the age of the tissue from which the nuclei are extracted. Old tissues will likely have most of the nuclei in G1 and will display a single peak. In contrast, young tissue will be actively growing and a significant number of cells will be going through the cell cycle. Consequently, a relatively high proportion of these cells will have nuclei that have undergone DNA replication, and these G2 nuclei will have twice as much DNA as those in G1. G2 nuclei from young tissues will form a significant peak.

Some species, like Arabidopsis thaliana, display endopolyploidy, a phenomenon in which the nuclei undergo DNA replication, but fail to divide creating polyploid nuclei. In these species this phenomenon is associated with older tissues. Young tissues including those of reproductive organs remain diploid eliminating the possibility that increased ploidy is transmitted to the next generation. When these species are analyzed by flow cytometry they display multiple peaks with fluorescence means that are proportional to a power of two series (21, 22, 23, 24, …). This observation also means that peaks with a mean fluorescence that doesn’t correspond to the series likely represent artifacts. For instance, a peak representing pairs of G1 and G2 nuclei physically stuck together, or peaks representing cellular debris.

An accurate estimate of nuclear DNA content depends on both the quality and quantity of the extracted nuclei. High quality nuclei are those which are intact and free of cellular debris. In contrast, nuclei of poor quality include those that have been fractured (less fluorescence), those associated with cellular debris capable of capturing additional fluorogenic material (more fluorescence), and those with both problems. A quantitative measure of variation in a population (in this case, the population of nuclei) is given by the coefficient of variation (CV). The CV is basically the standard deviation expressed as a percentage of the mean \[ CV = \left( \frac{\sigma}{\mu} \right) \times 100 \]. This parameter makes it possible to compare variation of peaks of different fluorescence intensities. In summary, narrow peaks will have low CV values (see Figure3) and wide peaks will have large CV values. In summary, a good estimate of the nuclear DNA content can be obtained by measuring a relatively large number of high quality nuclei. The flow cytometer we will use in this class has the capacity to capture 20,000 events per second!

The amount of fluorescence is proportional to the amount of DNA in the nucleus. Thus, the DNA content of nuclei in a peak can be calculated on the basis of the DNA content of a species used as standard (Johnston et al., 1999) according to the following formula:

\[
\text{Sample DNA content} = \frac{\text{Relative Fluorescence of Sample}}{\text{Relative Fluorescence of Standard}} \times \text{Standard DNA content} \quad (10)
\]
The genome sizes of the standards were originally calculated using chicken red blood cells (CRBC) as a standard. The rationale for using CRBC as a standard is based on the fact that these cells can be counted in a hemocytometer, and the total cell number can be calculated for a given volume. The size of the chicken genome can be calculated by extracting DNA from a fixed volume of CRBC, and quantifying the extracted DNA. The DNA content of each cell is calculated by dividing the total amount of DNA extracted over the estimated number of cells in the original volume of chicken blood. Since extraction and purification are never 100% efficient, the efficiency of extraction is calculated by adding a known amount of radiolabeled DNA to the CRBC. The fraction of the recovered labeled DNA is used to correct the calculated DNA content of the CRBC.

References
Leushkin EV, RA Sutormin, ER Nabieva, AA Penin, AS Kondrashov, MD Logacheva. 2013. The miniature genome of a carnivorous plant Genlisea aurea contains a low number of genes and short non-coding sequences. BMC Genomics 14, 476.

Web Pages of Interest
Flow Cytometry Facility at UF http://www.biotech.ufl.edu/cores-by-division/cellomics
MATERIALS AND SUPPLIES

A. BIOLOGICAL MATERIAL

<table>
<thead>
<tr>
<th>Size Standards</th>
<th>Mbp/1C</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Vigna radiata</em></td>
<td>675</td>
</tr>
<tr>
<td><em>Sorghum bicolor</em> Pioneer 8595</td>
<td>748</td>
</tr>
<tr>
<td><em>Pisum sativum</em> cv. Minerva Maple</td>
<td>4,531</td>
</tr>
<tr>
<td><em>Hordeum vulgare</em> cv. Sultan</td>
<td>5,365</td>
</tr>
<tr>
<td><em>Vicia faba</em> cv. GS011</td>
<td>12,863</td>
</tr>
<tr>
<td><em>Allium cepa</em> cv. Ailsa Craig</td>
<td>15,908</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Plant Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Arabidopsis thaliana</em> (2n = 10)</td>
</tr>
<tr>
<td><em>Arachis hypogaea</em> (2n = 4x = 40)</td>
</tr>
<tr>
<td><em>Capsicum annuum</em> (2n = 12)</td>
</tr>
<tr>
<td><em>Glycine max</em> (2n = 4x = 40)</td>
</tr>
<tr>
<td><em>Lens culinaris</em> (2n = 14)</td>
</tr>
<tr>
<td><em>Oryza sativa</em> (2n = 24)</td>
</tr>
<tr>
<td><em>Phaseolus acutifolius</em> (2n = 22)</td>
</tr>
</tbody>
</table>

B. REAGENTS

Magnesium Chloride, MgCl₂.6H₂O
Mix-bed resin AG 501-X8 (Bio-Rad)
MOPS
Triton X-100
Propidium Iodide
RNase A (Sigma R6513)
Sodium Citrate
Triton X-100

C. SOLUTIONS

Stock Solutions

**Magnesium Chloride, 0.5 M**

\[
\text{MgCl}_2 \cdot 6\text{H}_2\text{O} \quad 10.15 \text{ g} \\
\text{Water, adjust to} \quad 100.00 \text{ ml}
\]

**Sodium Citrate, 0.5 M**

\[
\text{Na}_3\text{Citrate} \quad 12.90 \text{ g} \\
\text{Water, adjust to} \quad 100.00 \text{ ml}
\]
MOPS, 0.5 M - pH 7.2

MOPS 10.47 g
Water, adjust to 100.00 ml

NOTE: Dissolve in 70 ml of water. Adjust the pH to 7.2 with NaOH 4 N, and bring to final volume with water. Store at 4°C.

Propidium Iodide, 5 mg/ml

Propidium Iodide 25.00 mg
Water, adjust to 5.00 ml

NOTE: Store at 4°C in the dark.

RNase, 10 mg/ml

RNase A (cloned) 10.0 mg
Buffer 1.0 ml

NOTE: RNase buffer is Tris.HCl 10 mM (pH 7.5), NaCl 15 mM. Store solution at -20°C. Crude preps should be boiled for 15 min and then allow them to slowly reach room temperature before use. Chromatographically purified enzyme (Sigma R6513 or R4642) does not need to be boiled before use.

Triton X-100, 10%

Triton X-100 10.0 ml
Water 90.0 ml

NOTE: Transfer solution to a 250 ml flask, add 2 g of mixed-bed resin, and stir with a magnetic stirrer for about 1 h. Eliminate the beads by filtration and store at 4°C.

Working Solutions

Chopping Buffer

MOPS, 0.5 M, pH 7.2 4.0 ml
MgCl₂, 0.5M 9.0 ml
Na₃Citrate, 0.5 M 6.0 ml
Triton X-100, 10% 1.0 ml
RNase, 10 mg/ml 0.1 ml
Water, adjust to 100.0 ml

NOTE: Prepare fresh and keep in the dark.

Staining Buffer

Chopping buffer 50.0 ml
Propidium I 5 mg/ml 0.5 ml
D. LABWARE AND SUPPLIES

Nylon mesh 20 and 50 μm
Pipets and pipet tips
Microcentrifuge tubes, 1.5 ml
Razor blades
Spatulas
Weighing boats, small and medium size

E. EQUIPMENT

Microcentrifuge
Flow cytometer

NOTE: We will use a BD LASER II SYSTEM (BD Bioscience), a bench-top flow cytometer with four lasers operating at a wavelength of 355-nm, 405-nm, 488-nm and 635-nm. Signal is detected through a set of interchangeable bandpass filters by a high-performance, high dynamic range photomultiplier
PROCEDURE FOR ESTIMATING NUCLEAR DNA CONTENT

1. Place tissue (.25 to 1 g.) in a medium size weighing boat.
2. Add 1.5 ml of ice-cold chopping buffer and slice tissue every ~ 0.5 mm with a razor blade. Repeat slicing at a 90° angle, but do not overdo it or the nuclei will be damaged.
   **NOTE:** Wear disposable gloves. The chopping buffer has RNAse, and the staining buffer has propidium iodide which has a high affinity for DNA, including yours, and therefore is also a powerful mutagen!

3. Filter homogenate first through a 50 μm nylon mesh into a clean small size weighing boat, and then through a 15-20 μm nylon mesh into another small size weighing boat.
   **NOTE:** Transfer the homogenate by placing the 50 μm filter over a small weighing boat and then hold the boat containing the sample from the corners bending the boat and tilting it towards the filter. You can use a spatula to help pushing all the sample to the filter. Either pour the filtrate into the second filter, or use a pipet for the transfer.

4. Transfer filtrate to a 1.5 ml microcentrifuge tube.
5. Pellet nuclei at 1,000 rpm for 3 minutes, and then discard supernatant.
6. Resuspend pellet in 500 μl of staining buffer, and incubate for 5 min at room temperature.
7. Pellet nuclei at 1,000 rpm for 3 minutes, and then discard supernatant.
8. Resuspend pellet in 500 μl of staining buffer.
9. Analyze sample in the flow cytometer.

Flow Cytometer Output

The flow cytometer will be set up to collect Relative Fluorescence Intensity (PE-A). The relative fluorescence channels will be in a logarithmic scale. A one-parameter histogram as depicted in Figure 3 will be generated for each sample and the standards. The companion table will identify each of the populations and will list the corresponding statistics: frequency, % of total signal, median, geometric mean, mean, coefficient of variation, maximum and minimum values, standard deviation and the mode.

The species listed as standards will be run through the flow cytometer first. The samples will be run next, and the size of the genome of the different species should be determined according to the relationship shown below. Be mindful of what happens to the DNA content of the nucleus during the cell cycle and the number of copies of the genome in a nucleus.

\[
\text{Sample DNA content} = \frac{\text{Relative Fluorescence of Sample}}{\text{Relative Fluorescence of Standard}} \times \text{Standard DNA content}
\]
Study Questions.

1. What is the genome?
2. How many types of genomes do plant cells have?
3. How many copies of each genome do plants cells have?
4. What are the phases of the cell cycle?
5. How does the DNA content of the cell change during the cell cycle?
6. What are the main reasons plants display a 2000-fold variation in genome size?
7. How do you prepare fairly purified nuclei for flow cytometry?
8. How do you measure DNA content of nuclei in flow cytometry?
9. Why do you treat your nuclei preparation with RNAse?
10. What factors affect the correct estimation of nuclear DNA content?
11. What quantitative criterion would you use to describe the quality of a FC nuclear prep?
12. How would you identify an artifactual peak in a FC histogram?
13. What would distinguish the FC profile of an old vs. that of a young tissue?
14. Why were chicken red blood cells used to estimate the genome size of std. plant genomes?
15. What would be the FC signature of a species with endopolyploidy?
16. Could you use FC to tell whether a species is auto or allopolyploid?
17. Can you use FC to identify ploidy variants in one species?
18. Give an example of why the genome size of species is of value in molecular biology.
19. How much do you have to increase the size of a library if you want to decrease the chances of missing a gene 100-fold?
20. Challenge question: You have made a genomic library from a hybrid plant which was generated from a cross between a line that carries a single resistance gene and a line that possesses the susceptible allele for that gene. If you know the genome size of the species, then what will the nature of the genomic library source affect the size of the library you need to screen to pull out a clone of the resistance allele?

Part I
Summary (One Page Maximum. 25 points)
a) the main objective of the experiment,
b) the most important findings,
c) the significance of the results, and
d) major conclusions

Part II – Results (20 points)

Summary of the results obtained in the lab. Include all the values recorded for all species, including those used as standards and calculate the genome size for each one. Also include a printout of the histogram of the sample processed by your group. Show the calculations for the sample your group analyzed, use the space below the table.

<table>
<thead>
<tr>
<th>Species</th>
<th>1st peak</th>
<th>2nd peak</th>
<th>3rd peak</th>
<th>C-Value (Mbp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Mungbean, <em>Vigna radiata</em></td>
<td></td>
<td></td>
<td></td>
<td>675</td>
</tr>
</tbody>
</table>

**Standards**

**Samples**

1 Bean C, *Phaseolus vulgaris* (Leaves)
2 Bean C, *Phaseolus vulgaris* (Cotyledons)
3 Bean J, *Phaseolus vulgaris* (Leaves)
4 Bean wild, *Phaseolus vulgaris* (Leaves)
5 Pea, *Pisum sativum*
6 Peach, *Prunus persica*
7 Pepper, *Capsicum annuum*
8 Pepper, *Capsicum frutescens*
9 Rice, *Oryza sativa*
10 Soybean, *Glycine max*
11 Tomato (garden), *Solanum lycopersicum*
12 Tomato (wild), *Solanum pennellii*
13 Maize (W22), *Zea mays*  

Part III – Answer the following questions:

1) (10) Most of the samples yielded more than one peak. What are the biological characteristics and physical principles that explain the presence of these peaks?
2. Select the species, or one of the species, with which you worked directly and then calculate the size of a genomic library which will give you a 99% and a 99.99% probability of detecting a genomic clone bearing a specific sequence.
   a) (10) In a plasmid vector (Average insert size = 2 Kb)
   b) (10) In a BAC vector (Average insert size = 200 Kb)
   c) (10) Compare the results obtained for 99 and 99.99 % probability and explain the similarity in ratios of the two values for a and
CHAPTER 2

PLANT DNA EXTRACTION AND ANALYSIS

DNA EXTRACTION AND PURIFICATION

Learning Objectives............................. 2.1
Lab Objectives................................. 2.1
Background ....................................... 2.1
Extraction Procedures......................... 2.2
Evaluation Criteria......................... 2.5
The Protocols................................. 2.8
References .................................. 2.9
Materials and Supplies....................... 2.10
Nuclear Fraction Protocol............... 2.14
Evaluation Procedures .................. 2.16
**DNA EXTRACTION AND PURIFICATION**

**Learning Objectives**

After the completion of this unit students will be able to purify high quality nuclear plant DNA from a variety of sources, and be capacitated to evaluate the quality of extracted DNA using a variety of chemical and physical procedures. Students will be able to identify factors that affect DNA quality during extraction, and the appropriate steps needed to control them.

**Lab Objective**

Extract high quality DNA - high molecular weight and contaminant-free. DNA will be extracted from young leaf tissue. Grinding in liquid nitrogen will mechanically disrupt tissue. Nuclei from the powdered tissue will be removed from the soluble cellular fraction by differential centrifugation, and the DNA in the nuclear pellet will be purified via successive chemical and physical fractionation steps. DNA quality will be assessed spectrophotometrically, fluorometrically, and via agarose gel electrophoresis.

**Background**

High quality DNA is required in many molecular biology procedures such as the construction of genomic libraries, and Southern blotting and hybridization of restriction enzyme digests among other procedures. Let’s first examine the structures and chemical environment associated with genomic DNA before we consider its extraction and purification. Nuclear DNA is found in chromatin, a highly structured association of DNA with histones, high mobility group proteins, phosphoproteins, transcription factor proteins, RNA and phospholipids. Chromatin, along with a host of enzymes, RNA, and various small molecules are encased by the nuclear envelope. The nucleus in turn is surrounded by the cytoplasm where one can find several organelles and the vacuole, and a myriad of proteins, ribosomes, mRNA, and small solutes (amino acids, organic acids, sugars, phenolic compounds). A lipid bi-layer membrane encases all the cell contents including nucleus, organelles, vacuole, and the cytoplasm. The organelles also contain their own genomic DNA. Finally, unlike animal cells, plant cells are surrounded by a tough cell wall, which is made up of highly organized and complex polymeric structures containing cellulose, pectin, protein, and aromatic compounds.

Purification of DNA from such complex structures and varied chemical environment presents five major challenges and concerns:

1) Tissue disruption (Cell wall)
2) Suppression of active compounds (Degrading enzymes, phenolics)
3) Elimination of contaminants (Pigments, polysaccharides, protein, and RNA)
4) Minimization of DNA breakage (Shearing)
5) Maximization of yield (μg/g of tissue)

**Extraction procedures**

Extraction of DNA from plant cells is complicated by the presence of cell walls, which require mechanical disruption. This task can be accomplished by grinding fresh tissue in a blender or in a mortar in the presence of an extraction buffer. However, liquid nitrogen (77.2° K, or -196° C) is commonly used to grind tissue because frozen tissue is brittle and can be easily and efficiently broken up in a mortar or in any other grinding device. The effectiveness of this step has a significant bearing on yield, the total amount of DNA extracted from the sample. Although DNA can be extracted from different tissues or organs, leaf samples are most commonly collected, frozen in liquid nitrogen, and stored at -80° C until extraction.

The purpose of mechanical disruption is to break up the cell wall to release the nuclei and make them immediately accessible to the extraction buffer. However, disrupting cellular and sub-cellular structures brings about several problems and complications. These mainly include the release and activation of nucleases, enzymes that degrade DNA, and exposure of DNA to phenolic compounds and phenol oxidases. Nucleases require magnesium ions for activity, and for this reason EDTA, a Mg-chelating agent, is included in the extraction buffer to ensure the inactivation of these enzymes. A more challenging problem with most plant tissues is the release of phenolic compounds. These compounds can oxidize either spontaneously or enzymatically by various phenol oxidases; many of these enzymes are activated during tissue disruption. Oxidation of phenolic compounds yields reactive quinones or semiquinones which form adducts with both nucleic acids and proteins, and polyphenols can cross-link these macromolecules because they have more than one reactive group. Oxidation can be controlled by the addition of reducing agents to the extraction buffer. The most commonly reducing agents are dithiothreitol (DTT), reduced glutathion, and β-mercaptoethanol. Another strategy to keep oxidation in check is to lower the pH of the extraction solution by using a buffer with a low pKa.

Proteins present a problem either by their undesirable enzymatic activities (nucleases and oxidases) as well as by their structural association with DNA in chromatin. Enzymes can be inactivated and structural proteins stripped off the DNA by detergents or chaotropic agents. Detergents are amphipathic molecules that contain both a non-polar (aliphatic or aromatic) and a polar domain (ionic or non-ionic). The non-polar domain allows these molecules to associate with other non-polar domains such as the hydrophobic pockets of proteins and membrane lipids, while the polar domain confers them solubility in water. Detergents used in DNA extraction buffers disrupt, denature, and solubilize membranes and proteins including those found in the nuclear compartment. Examples of commonly used detergents are sodium dodecyl sulphate (SDS), lauroyl sarcosinate (Sarkosyl), and hexadecyl trimethylammonium bromide (CTAB). Some chaotropic agents are also used in the extraction of DNA. These compounds are effective
DNA Extraction

at high concentration (~6-8 M), and were originally thought to destroy water structure, but recent experiments suggest that they directly interact with the hydration shell of macromolecules disrupting hydrogen bonding as well as hydrophobic interactions of macromolecules. Urea, guanidinium chloride, and guanidinium isothiocyanate are examples of the most commonly used chaotropes. Sodium chloride is usually added to the extraction buffer because it helps in the solubilization of a group of proteins called globins. In addition, the presence of sodium chloride is required in the DNA precipitation step.

Following mechanical disruption of the tissues, inactivation of reactive molecules and solubilization of cellular structures leaves a complex homogenate that includes cellular and extracellular components in addition to those from the extraction solution. The challenge at this point is to isolate the DNA from this milieu. One issue to consider at this point is that stripping proteins off chromatin leaves DNA, a very long linear molecule, susceptible to mechanical damage by shearing forces. Thus, if the goal is to extract fairly large DNA fragments, then shaking, vortexing, and fast pipetting should be avoided during DNA isolation.

DNA can be purified from the homogenate by a series of physical and chemical fractionation steps. An organic extraction followed by a centrifugation is the most commonly used method of fractionation. Organic extractions are usually performed with phenol or a mixture of phenol:chloroform followed by chloroform, or just chloroform alone. Either isoamyl alcohol or octanol is normally added to the chloroform to avoid foaming during extraction. When phenol is prepared for DNA extraction, it is equilibrated with an aqueous buffer solution at a pH above 7.5 (Tris.HCl, pH 8.0). The reason for this equilibration is that if the extraction is carried out with acidified phenol (water equilibration), DNA will partition to the phenol phase along with the protein fraction. We will take advantage of this property later on when we try to eliminate DNA during RNA isolation. Although water and phenol are apparently immiscible, water-saturated phenol contains approximately 28 % of water, whereas phenol-saturated water contains 7% phenol. In contrast, the partition coefficient of chloroform into water is much smaller than that of phenol, whereas that of phenol into chloroform is very high. Thus, a chloroform extraction after a phenol extraction is used mostly to remove the phenol from the aqueous phase.

During organic extraction the aqueous homogenate is thoroughly mixed with the organic solvent to increase the effectiveness of the extraction. Thorough mixing results in an emulsion, the formation of tiny droplets of organic solvent surrounded by the aqueous phase. Thus, emulsification increases the area of aqueous-organic interphase to facilitate the extraction procedure. During extraction, molecules partition into the phase for which they have the highest affinity. A centrifugation step is used to separate the aqueous and organic phases. The aqueous phase or upper layer contains small water-soluble compounds (sugars, organic acids, flavonoids, etc.), some proteins, and nucleic acids (DNA and RNA). If this phase is cloudy after centrifugation, then a second organic extraction is recommended to eliminate excess proteins and polysaccharides. The organic or lower phase contains lipids and other hydrophobic
substances such as pigments (chlorophyll, carotenoids), detergent that has been stripped off the proteins, proteins that are soluble in the organic phase. Precipitated proteins, cell walls, and insoluble polysaccharides accumulate in the interphase after centrifugation. If the aqueous phase has a relatively high salt content, it may not separate very well from the phenol phase. In this case, a phenol:chloroform mixture is used to remedy this problem because chloroform is a much denser solvent ($d_{20}=1.486$).

Nucleic acids in the upper phase can be separated from the other compounds by alcohol precipitation. This step also concentrates the nucleic acids. Alcohol precipitation is a rather simple procedure, which requires the presence of a monovalent cation and the addition of either isopropanol or ethanol. The strategy of precipitating nucleic acids is based on the elimination of the net negative charge provided by the phosphate groups. DNA is water soluble because it interacts with water, mainly through the negatively charged phosphate groups. Cations at low concentration in a DNA solution act as free ions with a hydration shell. Addition of an excess amount of cations ($\text{Li}^+$, $\text{Na}^+$, $\text{K}^+$, $\text{NH}_4^+$) creates ion pairs with the phosphate group neutralizing the charge of DNA, which in turn makes it less polar and therefore less soluble in water. However, interaction between ions of opposite charge is weakened in aqueous solutions due to the presence of their hydration shell and the high dielectric constant of water. This interaction can be better visualized by examining the force of attraction as described by Coulombs law:

$$F = \frac{q_1 \times q_2}{d^2 \times \epsilon}$$

where $F$ is the force of attraction between ions of opposite charge, $q_1$ & $q_2$ are the charges of the ions, $d^2$ is the square of the distance between the charges, and $\epsilon$ is the dielectric constant of the medium. The dielectric constant is an expression of the extent to which a material can insulate charges from each other. The dielectric constant of water at room temperature is 80, of ethanol is 24.3, and that of isopropanol is 18.3. These alcohols are usually added to the DNA solution after the addition of monovalent cations to lower the dielectric constant of the medium, which results in a more efficient neutralization of the negative charges of DNA. In the absence of a negative charge, the DNA is no longer soluble in water and precipitates out. The dielectric constant is inversely proportional to temperature. Thus, at higher temperatures the dielectric constant of the medium decreases. However, many protocols oftentimes recommended that ice-cold ethanol or isopropanol be added to the DNA, or that after addition of alcohol the mixture be placed in the cold. This step is mostly recommended with dealing with low concentrations of DNA. If the treatment does in fact increase precipitation, then the most likely explanation is that the low temperature decreases the solubility of DNA, and this effect is greater than the change in the dielectric constant. In summary, addition of one of these alcohols strengthens the ion pairs making the DNA salt much less polar, and therefore less soluble, resulting in its precipitation. The DNA precipitate can be air-dried and dissolved in a small volume of TE buffer (Tris.HCl & EDTA). At this point the extracted DNA can be evaluated.
**Evaluation Criteria**

The two major criteria used to evaluate extracted DNA are quality and quantity (yield). The quality criterion is based on its chemical purity, the average fragment size of the DNA, and its suitability for biochemical manipulation. The quantity criterion is concerned with the yield, or the amount of DNA obtained from a given amount of tissue.

Some aspects of chemical purity of the extracted DNA can be evaluated by spectrophotometry. Absorption spectrophotometry measures the amount of light absorbed by a substance at a specific wavelength. DNA absorbs electromagnetic radiation in the UV range with a peak of absorption at approximately 260 nm. Thus, this property is used to measure DNA concentrations. However, an A$_{260}$ reading does not always provide a good estimate of the DNA concentration because other molecules, which can be present in the DNA solution, can also absorb light in the UV range. RNA has spectrophotometric characteristics that are very similar to those of DNA, so it can’t be distinguished from DNA by spectrophotometry. For this reason, contaminating RNA can lead to an overestimation of the DNA concentration. The presence of RNA in the DNA solution can be detected by agarose gel electrophoresis (see below). If RNA is present in DNA preparation, then it can be removed by treating the sample with RNAses; this enzyme and the released ribonucleotides can be removed by a second organic extraction and DNA precipitation.

Phenolics and proteins (phenyl alanine, tyrosine, tryptophan) also absorb in the UV range with a peak of absorption at 280 nm. The presence of these contaminants can be detected by calculating the A$_{260}$/A$_{280}$ ratio. The ratio of highly pure DNA is 1.8, and that for RNA is 2. Thus, lower ratios indicate the presence of contaminants that absorb at 280 nm. If the low ratios are due to contaminating proteins, then this problem can be corrected by an organic extraction. But if the low ratios are due to the presence of phenolic compounds, then it may be too late to correct the problem because it is very likely that the phenolic compounds are present as DNA adducts, or elements crosslinking nucleic acids and proteins. A telltale sign of a phenolics problem is the visible brownish tinge of the DNA solution. As mentioned earlier, oxidation of phenolic substances can be suppressed by including a reducing agent in the extraction buffer, and also by using a low pH buffer in the extraction solution.

Contaminating polysaccharides, pectins in particular, present a major problem because they inhibit enzymes used in downstream manipulations of DNA. The absorption spectrum of pure DNA shows a trough at about 240 nm, but the presence of polysaccharides/pectins reduces the depth of this trough because they absorb light of this wavelength. An excess of pectins is indicated sometimes by a cloudy or milky appearance of the DNA solution, and in most cases they can be eliminated by a high-speed centrifugation; however, in the worse cases the DNA solution adopt a gel-like consistency. Excess polysaccharide/pectins can be eliminated through an efficient organic extraction step.
The amount of pectins and phenolic compounds in leaf samples varies tremendously among species. For instance, citrus and peach leaves have a relatively high pectin content, while Solanaceous species like peppers and tomatoes are rich in phenolics, and grape leaves are rich in both. These disparities require the adaptation and optimization of extraction protocols to meet the demands of specific tissues.

The physical integrity of the extracted DNA is another quality criterion. Chromosomes are very long DNA molecules. During extraction the DNA-associated proteins are removed and mild shearing forces invariably break the large naked chromosomal DNA molecules. Nucleases are also another source of DNA degradation. The physical integrity of the DNA depends on the extent to which breakage occurs and how breakage is controlled. How this factor is managed depends also on the planned use for the DNA. At one end of the spectrum, routine PCR amplification of small targets does not require high molecular DNA, but at the opposite end is the construction of BAC libraries that requires high molecular weight DNA. In the latter case, nuclei are isolated and embedded in agarose blocks before DNA is purified. Embedding in agarose restricts movement of DNA during extraction and this restriction prevents excessive breakage. In any case, agarose gel electrophoresis is normally used to assess the average DNA size. This technique will be described in greater detail later on. But briefly, the mobility of DNA in agarose is inversely proportional to its molecular weight. Thus, large fragments do not move very far from the origin, while the small fragments have greater mobility. In practical terms, extracted DNA of large molecular weight appears as a single band near the origin, whereas degraded DNA is detected as a "smear" distributed along the gel. RNA, regardless of its molecular weight, moves very close to the buffer front because it forms intra-molecular bonds and acquires a compact 3D configuration that allows it to move fast.

The third quality criterion used to evaluate extracted DNA is whether it is amenable to biochemical manipulation. In other words, can the extracted DNA be used as a substrate by standard nucleic acid modifying enzymes like restriction enzymes or DNA polymerases? Thus, a functional assay like a restriction enzyme digestion, or a PCR-amplification is required to answer this question. In either case, agarose gel electrophoresis is used to evaluate the results of these procedures.

The quantity of DNA extracted from a given amount of tissue will depend mainly on the average cell size of the tissue, and the efficiencies of the tissue-grinding method and subsequent fractionation steps. Relatively high yields can be obtained by selecting young tissue because it has both small cells (more cells/unit of weight) and less cell wall material with a minimum level of lignification.

DNA concentration can be determined spectrophotometrically as indicated above. Spectrophotometers measure the fraction of monochromatic light absorbed by the sample. Thus, this measurement is independent of the intensity of the incident light, but dependent on the path length (cuvettes are usually 1 cm), the concentration of the analyte, and the absorption properties of the analyte. To facilitate this measurement, *Escherichia coli* DNA has been
established as the "standard" with a relationship of 1 unit of absorbance at 260 nm (path = 1 cm) corresponding to a DNA concentration 50 μg/ml. The potential sources of error for this method have been explained above.

Spectrophotometry is traditionally carried out in spectrometers that require cuvettes to hold both the sample and the reference solution. This approach demands high volume samples, and it is practically useless in cases in which the sample volume is very small. The newly developed NanoDrop ND-1000 spectrophotometer overcomes this problem because it can measure the absorbance of 1 μl samples without the need of a cuvette. The NanoDrop uses surface tension to hold the sample between two optical surfaces connected to optical fibers. These fibers are in turn connected to a light source and a light detector to measure absorbance. The distance between these surfaces can be varied between 0.2 and 1 mm and the absorbance can be measured at those distances increasing the dynamic range of the instrument. In summary, spectrometry can be used to measure DNA concentrations and to detect the presence of phenolic and protein contaminants, but not by RNA contamination.

Fluorometry on the other hand can solve the problem of measuring the DNA concentration in the presence of contaminants. This technique takes advantage of the properties of certain dyes which have a very high affinity for double stranded DNA, and not for single stranded DNA or RNA, and increase their fluorescence several-fold upon binding double stranded DNA. Measurement of DNA concentration is based on a standard curve of fluorescence established with a reference solution of pure DNA. Unlike absorbance measurements, fluorescent measurements depend on the intensity of the light source, and several factors associated with the sample; these include fluorophore concentration, sample pH among others. In this particular procedure, the fluorescent signal does not come directly from DNA, but from a dye that specifically binds DNA and upon binding becomes fluorescent. These dyes have a very high quantum yield which makes it possible to quantify DNA in the ng/ml range. Examples of these dyes are Hoechst 33258, and PicoGreen. The Hoechst 33258 dye binds to the minor groove of DNA and has a greater affinity for AT-rich regions.

Some of you may be asking why is it that spectrophotometry doesn’t require a standard curve, but fluorometry does. The answer is that spectrophotometry measures the fraction light of a particular wavelength that is absorbed by the target molecule (an intrinsic property of nucleic acids) along a path of a given length regardless of the intensity of the light source. Whereas the outcome of fluorometry measurements is indirect, and depends on the intensity of the light source, the geometry of the instrument, and the amount of dye bound to the DNA; this amount will depend to some extent on the concentration of the dye in the solution, and the time that it is allowed to reach equilibrium with the nucleic acid. For these reasons, fluorometric measurements always require a standard curve. However, both measurements provide a relative estimate of the nucleic acid concentration. In the case of spectrophotometry, the estimate is based on E. coli DNA, whereas in fluorometry it is based on the DNA that is being used as a standard. You should keep in mind that either method can assess the physical integrity of DNA.
The Protocols

The most widely used protocols for extracting genomic plant DNA are based on either the CTAB procedure (Murray and Thompson, 1980), or the Dellaporta procedure (Dellaporta et al., 1983). These procedures aim to isolate DNA from an unfractionated homogenate of leaf tissue (other organs have been used too). The basic strategy is to disrupt the tissues mechanically by grinding in liquid nitrogen, and then to disrupt all cellular, sub-cellular, and supra molecular structures by means of a detergent (CTAB or SDS).

The CTAB procedure relies on the strong detergent action of CTAB to dissolve membranes, and denature proteins including those associated with DNA. There are two ways in which DNA can be purified from the CTAB homogenate: by organic extraction and isopropanol precipitation, or by CTAB precipitation. In the first procedure, the homogenate is extracted with an organic solvent, usually chloroform, or a succession of phenol and chloroform. After this extraction the mixture is centrifuged to yield three phases: an organic phase at the bottom containing detergent and lipid compounds that include chlorophyll, carotenoids, terpenoids, etc.; a hard inter-phase containing starch, cell walls, insoluble polysaccharides and denatured proteins; and an aqueous upper phase that contains DNA, RNA, soluble monosaccharides, organic acids, salts and water-soluble pigments. Nucleic acids are precipitated from this aqueous phase by addition of isopropanol. However, in many species, a variety of polysaccharides tend to co-extract with the DNA. To eliminate these contaminants DNA can be precipitated with CTAB. DNA is insoluble in 1% CTAB, but it can be kept in solution by the presence of sodium chloride (NaCl) at a minimum concentration of 0.7 M. This property is used to separate DNA from polysaccharides by differentially precipitating the DNA by simply reducing the NaCl concentration to 0.35M. Several other steps are sometimes required to eliminate all contaminants.

The Dellaporta protocol is an adaptation of the miniprep protocol used for the isolation of plasmid DNA. The detergent sodium dodecyl sulfate (SDS) is used in this protocol to disrupt cellular structures during extraction. SDS is soluble at room temperature, and insoluble in the presence of high concentrations of salt and at low temperatures. In contrast, potassium dodecyl sulfate (KDS) is simply insoluble. Thus, to remove SDS after extraction, an excess of potassium ions (K acetate) is added to the homogenate. The potassium will displace sodium and in turn ensure the precipitation of the newly formed KDS along with the associated denatured proteins. This process is enhanced by the relatively high salt concentration that results from the addition of the K Acetate. Chilling the homogenate induces the precipitation of any remaining SDS molecules. After removal of the precipitate by centrifugation, the DNA is precipitated with isopropanol.

More recently, a new flexible plant DNA extraction protocol has been developed which can be easily adapted to meet the needs of different tissues (Vallejos, 2007). The first step in
this protocol is essentially the same as in the other protocols – mechanical disruption of the tissue with liquid nitrogen. In contrast to the other protocols however, the main strategy of this protocol is to first separate the nuclear fraction from the bulk of soluble contaminants by differential centrifugation. After this step, the densest cellular components (nuclei, starch and cell walls) are pelleted at the bottom of the tube, while soluble pectins, ribosomes, organelles, phenolic substance and other soluble components remain in the supernatant, which is discarded. RNA associated with the nuclear fraction is removed by treatment with RNAse. The CTAB procedure, essentially as described by Murray and Thompson (1983), is used to extract the DNA following a digestion with RNAse. One variant of this protocol is the use of an acidic resuspension buffer for samples known to be rich in phenolic compounds. The spontaneous oxidation of these compounds is inhibited at low pH. This protocol yields high molecular weight DNA of high purity. The class will use this protocol to extract DNA from different various samples.

Several DNA extraction kits are now available in the market, and some of them have been designed for automation. The main objective of these kits is to extract DNA that is suitable for PCR amplification. Purification of DNA by these kits is based on a solid-phase extraction procedure. This procedure exploits the ability of silica particles to adsorb DNA in the presence of a high concentration of mild chaotropic agents (perchlorate or guanidinium salts) at a pH below 7.0. Chaotropic agents destroy the hydration shell around both DNA and the silica particles. In addition, the low pH (~5.0) keeps the silanol groups (SiOH) protonated removing the negative charges on the surface of the particles; both of these factors facilitate the adsorption of DNA by silica particles. Impurities can be separated from Silica-adsorbed DNA by washing with a solution of the chaotropic agent. After repeated washes, the DNA can be eluted from the silica particles by a solution of low ionic strength at pH 8.0. Future experiments in this class will use a silica-based kit for the purification of plasmid DNA.

References


MATERIALS AND SUPPLIES

A. BIOLOGICAL MATERIALS

<table>
<thead>
<tr>
<th>Seedlings</th>
<th>Planting time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capsicum annuum, pepper</td>
<td>4 weeks</td>
</tr>
<tr>
<td>Phaseolus vulgaris, common bean</td>
<td>2 weeks</td>
</tr>
<tr>
<td>Pisum sativum, garden pea</td>
<td>2 weeks</td>
</tr>
<tr>
<td>Solanum lycopersicum, tomato</td>
<td>4 weeks</td>
</tr>
<tr>
<td>Zea mays, corn</td>
<td>2 weeks</td>
</tr>
</tbody>
</table>

B. REAGENTS

- Acetic Acid
- Ammonium Acetate
- Chloroform
- CTAB, Hexadecyltrimethylammonium Br
- Ethanol 100%
- Hydrochloric acid
- Isopropanol
- β-Mercaptoethanol
- Nitrogen, Liquid
- Octanol
- Potassium Acetate
- RNAse A
- Sodium Acetate
- Sodium Chloride
- Sodium Dodecyl Sulphate (SDS)
- Sodium EDTA
- Tris Base
- Bio-Rad Versafluor Kit

C. SOLUTIONS

Stock Solutions

**CTAB, 10 % (Store at Room Temperature)**

<table>
<thead>
<tr>
<th>CTAB</th>
<th>50.0 g</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₂O add to</td>
<td>500.0 ml</td>
</tr>
</tbody>
</table>

**NOTE 1:** Wear a dust mask to handle CTAB.

**NOTE 2:** Dissolve in a beaker by stirring with a magnetic bar in 80% of final volume. Either stirring over a hot plate (low heat) or brief warming in a microwave oven while avoiding foam can speed up this process. Carefully pour solution into a graduated cylinder, rinse beaker with water and pour into cylinder until reaching final volume.
Cover cylinder with parafilm, or any other sealing film, and mix before pouring into storage bottle.

**EDTA, 0.5 M, pH 8.0** (Store at 4º C)

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Na₂EDTA.2 H₂O</td>
<td>186.1  g</td>
</tr>
<tr>
<td>NaOH pellets</td>
<td>19.0  g</td>
</tr>
<tr>
<td>H₂O add to</td>
<td>1000.0 ml</td>
</tr>
</tbody>
</table>

**NOTE**: Dissolve salts in 800 ml of water and stir vigorously with a magnetic bar. Adjust pH to 8.0 with 4N NaOH. Bring to final volume and autoclave.

**RNAse Stock** (Store at -20º C)

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNAse A</td>
<td>100.0 mg</td>
</tr>
<tr>
<td>Tris. HCl (1 M, pH 7.8)</td>
<td>0.1 ml</td>
</tr>
<tr>
<td>NaCl (5 M)</td>
<td>0.3 ml</td>
</tr>
<tr>
<td>H₂O add to</td>
<td>10.0 ml</td>
</tr>
</tbody>
</table>

**NOTE 1**: Aliquot 1 ml into microcentrifuge tubes & store at -20º C.

**NOTE 2**: Alternatively, a pure and cloned RNase can be purchased that doesn’t require boiling before use.

**Sodium Acetate, 2 M, pH 5.2** (Store at Room Temperature)

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH₃COONa.3 H₂O</td>
<td>206.8 g</td>
</tr>
<tr>
<td>Acetic Acid</td>
<td>27.6 ml</td>
</tr>
<tr>
<td>H₂O add to</td>
<td>1000.0 ml</td>
</tr>
</tbody>
</table>

**NOTE**: Dissolve salt and acid in 700 ml of water; adjust pH with either acetic acid or 4N NaOH. Bring to final volume.

**Sodium Acetate, 3 M** (Store at Room Temperature)

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH₃COONa.3 H₂O</td>
<td>408.2 g</td>
</tr>
<tr>
<td>H₂O add to</td>
<td>1000.0 ml</td>
</tr>
</tbody>
</table>

**Sodium Chloride, 5 M** (Store at Room Temperature)

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl</td>
<td>292.0 g</td>
</tr>
<tr>
<td>H₂O add to</td>
<td>1000.0 ml</td>
</tr>
</tbody>
</table>

**Tris.HCl, 1M, pH 8.0** (Store at 4º C)

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris base</td>
<td>121.1 g</td>
</tr>
<tr>
<td>HCl</td>
<td>42.0 ml</td>
</tr>
<tr>
<td>H₂O add to</td>
<td>1000.0 ml</td>
</tr>
</tbody>
</table>

**ETHANOL**

**ISOPROPRANOL**
DNA Extraction

NITROGEN (Liquid)

Chloroform (Store at Room Temperature)
- Chloroform: 960.0 ml
- Octanol: 40.0 ml

Ethanol/Na Acetate (Store at 4° C)
- EtOH (95%): 800.0 ml, 76.0 %
- NaAcetate (3 M): 66.7 ml, 0.2 M
- H2O: 133.3 ml

TE Buffer (Store at 4° C)
- Tris.HCl, 1 M, pH 8.0: 1.0 ml
- NaEDTA, 0.5 M: 0.2 ml
- H2O add to: 100.0 ml

Lysis Buffer 2X (Store at Room Temperature)
- Tris.HCl (1 M, pH 7.8): 10.0 ml, 100 mM
- EDTA.Na (0.5 M, pH 8.0): 1.0 ml, 5 mM
- NaCl (5 M): 28.0 ml, 1.4 M
- CTAB (10 %): 20.0 ml, 2 %
- H2O: 41.0 ml
- Total volume: 100.0 ml

Nuclei Resuspension Buffer (Store at 4° C)

<table>
<thead>
<tr>
<th>Buffer Type</th>
<th>NRB-A</th>
<th>NRB-B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris.HCl (1 M, pH 8.0)</td>
<td>90.0 ml</td>
<td>10.0 ml</td>
</tr>
<tr>
<td>EDTA.Na (0.5 M, pH 8.0)</td>
<td>1.0 ml</td>
<td>1.0 ml</td>
</tr>
<tr>
<td>H2O add to</td>
<td>100.0 ml</td>
<td>100.0 ml</td>
</tr>
</tbody>
</table>

NOTE: Use B for routine extractions and A when phenols are a problem.

RNAsse (Prepare fresh)
- RNase A stock (10mg/ml): As needed
  
  NOTE: Pipet amount needed (+10%) into a 1.5 ml tube. Lock the tube and boil for 15 min. Allow tube to cool to RT° before adding it to the sample.

Sample Resuspension Buffer (Store at 4° C)

<table>
<thead>
<tr>
<th>Buffer Type</th>
<th>SRB-A</th>
<th>SRB-B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Na Acetate 2 M, pH 5.2</td>
<td>5.0 ml</td>
<td>---</td>
</tr>
<tr>
<td>Tris.HCl 1M, pH 8.0</td>
<td>---</td>
<td>10.0 ml</td>
</tr>
<tr>
<td>EDTA Na 0.5 M, pH 8.0</td>
<td>1.0 ml</td>
<td>1.0 ml</td>
</tr>
<tr>
<td>NaCl 5 M</td>
<td>4.0 ml</td>
<td>4.0 ml</td>
</tr>
<tr>
<td>H2O add to</td>
<td>100.0 ml</td>
<td>100.0 ml</td>
</tr>
</tbody>
</table>

NOTE 1: Add β-mercaptoethanol before use to a concentration of 1.0 %
NOTE 2: Use B for routine extractions and A when phenols are a problem.

Bio-Rad DNA Quantitation Kit

Stocks

Hoechst 33258 1 mg/ml
- Hoechst Reagent 10mg/ml 100.0 μl
- Water 900.0 μl

DNA Std 100 μg/ml
- Calf thymus DNA 1mg/ml 100.0 μl
- 10X TEN buffer 100.0 μl
- Water 800.0 μl

DNA Std 10 μg/ml
- Calf thymus DNA 1mg/ml 100.0 μl
- 10X TEN buffer 100.0 μl
- Water 800.0 μl

Working Solutions
Hoechst 33258 1 μg/ml (in a 50 ml tube)
- 10X TEN buffer 4.0 ml
- Water 36.0 ml
- Hoechst 33258 1 mg/ml 40.0 μl

D. LABWARE AND SUPPLIES
- Conical tubes, Polypropylene, 50 ml
- Dewar Flask
- Glass pipets
- Microcentrifuge tubes
- Micropipet tips
- Oak Ridge tubes, Polypropylene, 50 ml
- Pasteur pipets, long

E. EQUIPMENT
- Centrifuge, clinical or high-speed with T°-control
- Freezer,-80°C
- Fluorimeter
- Micropipets
- Spectrophotometer.
NUCLEAR FRACTION PROTOCOL

1. Collect 4 g of young leaf tissue. Wrap sample in aluminum foil. Write sample name and date on the foil, and immediately freeze in N\textsubscript{2}(l), and store the sample in -80\degree C freezer. Alternatively, the sample can be processed immediately after harvest.

**NOTE:** Wrap sample in a compact package. Loosely packaged samples will melt easily during subsequent handling. Temporary melting before extraction will result in oxidation of polyphenols and browning of DNA.

2. Add approximately 30-40 ml of ice-cold *sample resuspension buffer* (SRB) to a 50 ml conical polypropylene centrifuge tube, and place it on ice.

**NOTE:** Use SRB-A for Solanaceae, and SRB-B for the others.

3. Add N\textsubscript{2}(l) into a mortar with the pestle in it. Fill it to the rim. Put in the frozen leaf sample (or the fresh sample) and wait for the N to evaporate. Immediately grind plant tissue into a fine powder & transfer it to the 50 ml polypropylene tube with the aid of a spoon/spatula. Close tube, and immediately shake or vortex, to ensure thorough wetting and suspension of the powder. Add more SRB to the 50 ml mark, mix and place on ice.

**NOTE:** Dip the spoon/spatula in N\textsubscript{2}(l) prior to use. They must be cold to avoid melting the powder on contact.

4. Spin tubes at 1500 RPM for 10 min, and at 4\degree C.

5. Discard supernatant. Eliminate debris from tube wall by rotating tube while pouring off super.

6. Add NRB to the 10 ml mark. Add 50 \mu l of freshly boiled RN\textsubscript{ase} (10 mg/ml), and immediately vortex tube to resuspend pellet, and incubate at room temperature for 15 min.

**NOTE:** Use NRB-B for SRB-B extractions and NRB-A for SRB-A extractions.

7. Add 10 ml of 2\texttimes Lysis Buffer, mix gently but thoroughly, and incubate at 65\degree C for 1 hour. Mix by inverting the tube gently every 5 min.

8. Add 15 ml of Chloroform/octanol, and extract by inverting the tube gently 20-25 times.

9. Allow tube to reach room temperature. Mix one more time, and separate phases in a clinical centrifuge at 4,500 rpm for 10 min at 15\degree C. Use a wide-bore pipet, or an inverted 10-25 ml glass pipet, to transfer the upper phase to a 50 ml polypropylene Oak Ridge tube, and centrifuge samples at 10,000 RPM for 15 min.

10. Pour supernatant to a new 50 ml conical polypropylene tube and precipitate DNA by the addition of 2/3 volume of Isopropanol. Mix thoroughly by inverting tube gently. Allow DNA to precipitate for 10 minutes to 1 hour.

11. Discard liquid and wash DNA ppt with 20 ml EtOH (76%)/NaAcet (0.2M). Let DNA ppt. stand in wash solution for at least 30 min. DNA can be kept in this solution overnight, or for weeks or months. DNA precipitate can be handled with a glass hook.
12. Wash DNA pellet in **EtOH (100%)** briefly. Discard liquid and transfer pellet to a 1.5 ml microcentrifuge tube. Air-dry pellet for 10-20 min.

**NOTE:** If the pellet is relatively large, then it can be placed on a clean Kim-wipe that has been laid over a paper towel. Allow pellet to drain, but do not allow it to dry onto the Kim-wipe. Transfer DNA to microcentrifuge tube after most of the liquid has been removed.

13. Add 200 to 500 μl of **TE Buffer**. Allow DNA to hydrate and dissolve for at least 1 h or overnight. Afterwards, incubate the sample at 65° C for 20 min to ensure DNA is completely dissolved. Store DNA at 4°C.

**NOTE:** Do not freeze genomic DNA of high quality. Freezing will result in decrease of the average fragment size.
EVALUATION OF DNA - YIELD AND QUALITY

A. DNA Concentration - Spectrophotometry.

1. Use TE buffer to make a 1 in 50 dilution of your DNA preparation. Use a 1.5 ml micro centrifuge tube.

2. Take a spectrophotometric reading at 260 and 280 nm.

3. Calculate the DNA concentration using the following relationship: \(1A_{260} = 50 \mu g/ml\). Remember to include the dilution factor in your calculations.

   NOTE: If the readings at 260 nm are >2.0, then it means you have a very high concentration of DNA (or of contaminants). Dilute the DNA further. You may choose to dilute the dilution, perhaps just 1/10. Conversely, if the readings are too low <0.05, then it means your DNA prep is too dilute and you need to make a new dilution, perhaps 1/10 or 1/20.

4. Calculate the yield of DNA (\(\mu g\) per gram of tissue). For this calculation you will need the DNA concentration, the total volume in which the DNA was originally dissolved, and the weight of the tissue used for the DNA extraction.

5. Assess the chemical purity of your DNA. \(A_{260}/A_{280}\) ratios is 1.8 or more, then the DNA is of good quality with negligible amounts of protein or phenolic contamination. Ratios below 1.8 indicate contamination and an overestimation of the DNA concentration. Since the optical properties of RNA are similar to those of RNA, it is possible to have a good \(A_{260}/A_{280}\) ratio and still have a significant amount of contaminating RNA. In this case, the \(A_{260}\) reading will also lead to an overestimation of the DNA concentration.

B. DNA Concentration - Fluorometry.

Preparation of a Standard Curve.

1. Turn on the VersaFluor fluorometer (Bio-Rad) at least 15 min before the first reading. Make sure that the 360 nm Excitation and 460 nm Emission filters are in place

2. Dispense 2 ml aliquots of the Hoechst 33258 reagent (1 \(\mu g/ml\)) into seven 2 ml tubes.

3. Dispense 10, 5 and 2 \(\mu l\) of 1mg/ml “standard” DNA solution in tubes 1, 2 and 3, respectively, and 10, 5 and 2 \(\mu l\) of 0.1 mg/ml “standard” DNA solution in tubes 4, 5 and 6, respectively. Tube 7 will be used to zero the fluorometer. Mix well and incubate in the dark for at least 10 min, but not more than 60 min.

4. Press the SETUP button and select MEDIUM gain

5. Zero instrument. Press the RANGE button and adjust to 00000. Press EXIT.

6. Transfer content of tube 7 (baseline) to a cuvette, place it in the chamber and press the SET ZERO button. After the blinking stops, the display should read 0 ± 5. If not, repeat step.
7. Set the range by placing contents of tube 1 in the instrument, wait 10 seconds, then press \textit{RANGE} and adjust reading to 10000. Press \textit{EXIT}.

8. Take readings of all other standard points.

\textbf{Measurements}

1. Add \(5 \mu l\) of the DNA sample directly into a 2 ml tube containing 2 ml of the Hoechst 33258 reagent (1 \(\mu g/ml\)). Close tube, mix well and incubate in the dark for 10 minutes before taking a reading.

2. Calculate DNA concentration. If the standard curve is straight, then approximately each relative fluorescent unit (RFU) is equivalent to 1ng of DNA. Remember the number of \(\mu l\) of the DNA sample that were added to the cuvette. As before, if the reading exceeds 10000 RFUs, then a smaller aliquot should be measured. Conversely, if the reading is too low (<100), then greater volume should be measured, but it should not exceed 10 \(\mu l\).

3. Calculate yield as before.

\textbf{C. Electrophoretic Analysis}

1. Prepare an agarose minigel.

2. Prepare sample. Add the equivalent of 1 \(\mu g\) of the DNA sample to a 1.5 ml micro centrifuge tube. Add water to bring up the volume to a total of 30 \(\mu l\), and 6 \(\mu l\) of 6X DNA loading buffer. Mix well and spin for 10 seconds. Remember to balance the rotor.

3. Load the gel. Load one well with a molecular weight standard, and then another with 18 \(\mu l\) of the sample from step 2.

4. Set up the power supply to 100 Volts and time to 20 minutes, and run the gel.

5. Photograph the gel.

\textbf{D. Biochemical quality.}

1. Test whether the DNA can be digested with a restriction enzyme.

2. Add the equivalent of 1.5 \(\mu g\) of DNA to a 1.5 ml tube, and then add water to bring the volume up to 25 \(\mu l\), mix well and spin down for 5 seconds. Add 16 \(\mu l\) of this DNA sample to a tube containing 2 \(\mu l\) of reaction buffer and 2 \(\mu l\) of EcoRI (10 units/\(\mu l\)). Mix well, spin for 30 seconds and incubate for 1 h at 37\(^\circ\) C.

3. Add 4 \(\mu l\) of 6X loading buffer, mix well and spin for 30 seconds.

4. Analyze the digested DNA via agarose gel electrophoresis along the undigested DNA.
## DNA Extraction

### Spectrophotometric Readings

<table>
<thead>
<tr>
<th>Sample</th>
<th>Dilution</th>
<th>A_{260}</th>
<th>A_{280}</th>
<th>A_{260}/A_{280}</th>
<th>µg/ml</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### NanoDrop

<table>
<thead>
<tr>
<th>Sample</th>
<th>Dilution</th>
<th>A_{260}</th>
<th>A_{280}</th>
<th>A_{260}/A_{280}</th>
<th>µg/ml</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Fluorimetric Readings

<table>
<thead>
<tr>
<th>Std</th>
<th>DNA Stock</th>
<th>µl</th>
<th>DNA, ng</th>
<th>RF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 mg/ml</td>
<td>10</td>
<td>10,000</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1 mg/ml</td>
<td>5</td>
<td>5,000</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1 mg/ml</td>
<td>2</td>
<td>2,000</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.1 mg/ml</td>
<td>10</td>
<td>1,000</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.1 mg/ml</td>
<td>5</td>
<td>500</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.1 mg/ml</td>
<td>2</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample</th>
<th>µl</th>
<th>RF</th>
<th>ng/µl</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample</th>
<th>µl</th>
<th>RF</th>
<th>ng/µl</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
RESTRICTION ENDONUCLEASES

Learning Objectives
Upon completion of this module, students will be able to use the appropriate Restriction-Modification systems for the structural analysis and manipulation of DNA.

Lab Objectives
Analyze DNA isolated in a previous lab session with different restriction enzymes and identify specific restriction patterns of genomic and plasmid DNAs.

Background
Prokaryotic organisms possess a defense mechanism conferring the ability to distinguish self from non-self and to eliminate non-self – similar to the immune mechanism of animals. This defense mechanism is based on the enzymatic Restriction - Modification (R-M) system, and it is unique to bacteria. The R-M system not only works against phage infection, but it is equally effective against foreign plasmid, episomal, and bacterial DNAs. The R-M system has been turned into one of the most powerful tools in molecular biology. It has been, and continues to be, a key element in the construction of cDNA and genomic libraries, characterization of DNA sequences, and assembly of complex gene constructs used in plant transformation among multiple other applications.

Before we address the use of restriction enzymes, we shall examine the characteristics and properties of the different R-M systems. The R-M system comprises two distinct, but complementary enzymatic activities. The **Restriction Activity** is carried out by the restriction enzyme after it binds the specific cognate DNA sequence, while the **Modification Activity** is carried out by the methylase enzyme that selectively methylates the sequence recognized by the **Restriction Activity**. In nature, the methylase protects the restriction sites present in the DNA of the host, while the restriction enzyme works on foreign DNA bearing restriction sites.

R-M systems have been classified into four major groups according to the combination of enzyme subunit composition, co-factor requirements, binding site, and reaction product characteristics as described in Table 1.

Types I, III and IV carry both restriction and modification activities in the same enzyme complex, while Type II systems are typically binary - one enzyme carries the cleavage activity and another the methylase activity. All groups require magnesium (Mg$^{2+}$) as co-factor, but only Types I and III require ATP. Type I R-M enzymes bind to a specific recognition asymmetric sequence (e.g.: GAANNNNNNRTCG) and cleave randomly at distant sites that come in contact with the endonuclease domain of the enzyme. Type II R-M enzymes recognize specific palindromic (two-fold symmetry) sequences that are 4 to 8-nucleotides long and cut either
within or at the borders of the recognition site. Type III cleaves at specific distances away from the recognition site; enzymes of this group that recognize contiguous sequences cleave on one side, while those that recognize interrupted sequences cut on both sides. Type IV enzymes recognize specific DNA sequences that contain specific A or C methylation.

Table 1. Characteristics of the major groups of Restriction-Modifications (R-M) systems.

<table>
<thead>
<tr>
<th>Type</th>
<th>Structural Feature</th>
<th>Cofactor</th>
<th>Site</th>
<th>Cleavage</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Heteropentameric, bifunctional</td>
<td>Mg$^{2+}$, ATP</td>
<td>Asymmetric</td>
<td>Outside, random</td>
</tr>
<tr>
<td>II</td>
<td>Homodimeric, monofunctional</td>
<td>Mg$^{2+}$</td>
<td>Symmetric</td>
<td>Inside, or at border</td>
</tr>
<tr>
<td>III</td>
<td>Heterotetrameric, bifunctional</td>
<td>Mg$^{2+}$, ATP</td>
<td>Symmetric</td>
<td>Outside, to one side; or both in interrupted seqs.</td>
</tr>
<tr>
<td>IV</td>
<td>Mono &amp; dimeric</td>
<td>Mg$^{2+}$</td>
<td>Mod. A or C</td>
<td>Outside, to one side</td>
</tr>
</tbody>
</table>

Out of the four types, Type II enzymes are the most widely used in molecular biology because they are relatively easy to handled, have a highly predictable pattern of activity, and as a group they present a great diversity of specificities. Restriction enzymes are named after the species in which they are found. The first letter is capitalized and corresponds to the initial of the genus, and the second and third letters are derived from the species name. Originally, the nomenclature rules called for the use of italics in the first three letters, but that rule has been dropped in recent years due to the difficulty of transmitting certain fonts, including italics, in some electronic media. The name of the enzyme also carries a strain indicator, and a roman numeral reflecting the order in which the enzyme was found in the species as indicated in the example below:

**EcoRV**: *Escherichia coli*, Strain RY13, fifth (V) enzyme identified

As indicated above, Type II recognition sites are marked by palindromes, sequences that read the same in both directions. Examples of palindromic words are: LEVEL, RADAR, and MADAM. When considering palindromes in DNA, the two strands need to be taken into account. Take for instance the recognition sequence for the enzyme EcoRI below:

The sequence in the upper strand reads GAATTC, and if the DNA is flipped 180°, the

![Figure 1](image.png)

**Figure 1.** This figure depicts a dimeric restriction enzyme scanning a DNA sequence for its recognition site. Upon detection of the site, this enzyme will cut at the intervals marked by the vertical arrows.
sequence will still read the same. Notice that for this phenomenon to occur, the second half of the sequence must be an inverted repeat of the first half. Starting from the center of the sequence, one has an A on the left and a T on the right, which are complementary bases in DNA. Moving one place over, the next base on the left is another A on the left and a T on the right, and finally at the third place there is a G on the left and a C on the right. Computer programs can be trained to recognize Type II restriction enzymes. However, these sites can also be recognized manually. You can scan a DNA sequence for AT, TA, CG, or GC di-nucleotides. If one of these is detected, then you could check whether the flanking bases are complementary (A - - T, T - - A, C - - G, or G - - C). In this fashion it is possible to recognize 4-, 6-, or 8-base Type II restriction sites. It follows from this analysis that the recognition site for a restriction enzyme could be defined by just half of the sequence. For example, for EcoRI one can just state that the sequence is GAA, the TTC trinucleotide that follows is predetermined by the first trinucleotide.

How do these enzymes recognize the same sequence in both strands? The answer comes from the observation that these enzymes act as homodimers (crudely portrayed in Figure 1). Typically, dimers of these enzymes bind DNA randomly and move along until a specific recognition sequence is detected. Upon site recognition, the two strands are cut at precisely the same site as indicated by the vertical arrows in Figure 1. Table 2 below has a list of a few restriction enzymes with their respective recognition sequences and the cleavage sites marked by a slash (/). These enzymes can generate three types of termini: 5' overhangs (BamHI), 3' overhangs (PstI), or blunt ends (DraI). We can now deal with the more practical aspects of restriction enzymes as we consider the nature of these overhangs as they are very relevant in labeling and cloning procedures. The generation of complementary ends, also known as “sticky ends,” has facilitated cloning. A cloning vector cut with the same restriction enzyme as the DNA targeted for cloning will have complementary “sticky ends” which can be easily ligated to create a recombinant molecule.

A single restriction enzyme is not always required for the creation of recombinant DNA molecules. Close examination of the termini generated by restriction enzymes reveals that certain pairs of enzymes with different recognition sequences can produce compatible ends. For example BamHI recognizes and cuts G/GATCC, while BglII recognizes and cuts A/GATCT. Nevertheless, these enzymes generate compatible “sticky ends” which could be ligated together.

<table>
<thead>
<tr>
<th>Restriction Enzyme</th>
<th>5' Sequence</th>
<th>3' Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>BamHI</td>
<td>5'- AAAGTTTACG</td>
<td>GATCCTTTAGCATG - 3'</td>
</tr>
<tr>
<td></td>
<td>3'- TTTCAAATGCTTAG</td>
<td>GAAATCGTAC - 5'</td>
</tr>
<tr>
<td>BglII</td>
<td>5'- TCGGAAACCAG</td>
<td>GATCTCCATATACC - 3'</td>
</tr>
<tr>
<td></td>
<td>3'- AGCCTTTTGGAAGTCAG</td>
<td>AGGTATATGGA - 5'</td>
</tr>
<tr>
<td>Recombinant DNA</td>
<td>5' - AAAGTTTACGGATCTCCATATACC - 3'</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3' - TTTCAAATGCCAGGTATATGG - 5'</td>
<td></td>
</tr>
</tbody>
</table>
Notice that after ligation neither enzyme can be used to cleave the ligation site again, but MboI or Sau3AI could do it (see Table 2). At the other end of the flexibility scale are the enzymes that yield blunt ends. All blunt ends are compatible with each other for ligation and cloning. Thus, as far as ligation and cloning are concerned, restriction enzymes that recognize different sequences but yield blunt ends produce “compatible ends.”

**Table 2.** List of a few Type II restriction enzymes and their corresponding recognition sites.

<table>
<thead>
<tr>
<th>Name</th>
<th>Recognition Site</th>
<th>Termini</th>
<th>Name</th>
<th>Recognition Site</th>
<th>Termini</th>
</tr>
</thead>
<tbody>
<tr>
<td>AluI</td>
<td>AG/*CT</td>
<td>blunt</td>
<td>HpaII</td>
<td>*C/*CGG</td>
<td>5'</td>
</tr>
<tr>
<td>BamHI</td>
<td>G/GATC*C</td>
<td>5'</td>
<td>MboI</td>
<td>/GATC</td>
<td>5'</td>
</tr>
<tr>
<td>BglII</td>
<td>A/GATCT</td>
<td>5'</td>
<td>MspI</td>
<td>*C/CGG</td>
<td>5'</td>
</tr>
<tr>
<td>DraI</td>
<td>TTT/AAA</td>
<td>blunt</td>
<td>NotI</td>
<td>GC/GG<em>C</em>CGC</td>
<td>5'</td>
</tr>
<tr>
<td>DraII</td>
<td>RG/GNCCY</td>
<td>5'</td>
<td>PstI</td>
<td>*CTGCA/G</td>
<td>3'</td>
</tr>
<tr>
<td>EcoRI</td>
<td>G/AATT*C</td>
<td>5'</td>
<td>SacI</td>
<td>GAG*CT/C</td>
<td>3'</td>
</tr>
<tr>
<td>EcoRV</td>
<td>GAT/ATC</td>
<td>blunt</td>
<td>Sau3AI</td>
<td>/GATC</td>
<td>5'</td>
</tr>
<tr>
<td>HindIII</td>
<td>A/AGCTT</td>
<td>5'</td>
<td>XhoI</td>
<td>C/T*CGAG</td>
<td>5'</td>
</tr>
</tbody>
</table>

*C = 5methyl cytosine that renders the sequence resistant to the restriction enzyme.

/ = Cleavage site.

There are some exceptions to the recognition patterns exhibited by Type II enzymes. For example, DraII is one of the few that can recognize discontiguous sequences (Table 2); an N, which means that any of the four bases could be present at this position, interrupts the palindrome. Several bases may interrupt the palindrome in some cases. This enzyme also shows that sometimes the recognition specificity could be relaxed somewhat. Notice the bases at each end of the recognition site. The R and Y at the ends stand for purine and pyrimidine, respectively.

Methylation is the most common form of base modification in biological systems. As we have seen at the beginning of this section, the key to the functionality of the R-M system is methylation. The absence of an R-M system in eukaryotes makes it unlikely that prokaryotic methylation patterns be found in eukaryotes. In fact, methylation patterns in prokaryotes are very different from those in eukaryotes, and plants in particular. Some plant genomes, especially those that are large, have a significant proportion of the genome methylated. Silenced transposable elements and different types of repetitive DNA usually occupy these methylated regions. Plant methylases act on CG, CHG, or CHH sites. Thus, certain restriction enzymes are sensitive to plant methylation patterns as indicated in Table 2. Asterisks preceding
some C residues indicate that when that particular C is methylated, the enzyme will not cleave that site. This property of restriction enzymes is an important one in the analysis of plant genomic DNA.

There are restriction enzymes from different bacterial species that recognize and cleave the same sequence; these enzymes are known as isoschizomers. For instance, both MboI and Sau3AI recognize the sequence /GATC and cut precisely at the same site. Some isoschizomers differ in their sensitivity to methylation patterns and are used to detect differences in the methylation pattern of DNA samples. HpaII and Mspl recognize the sequence C/CGG and cut at the same location. However, neither enzyme will cut if the external C is methylated, but only Mspl will cut if the internal C is methylated.

Methylases like EcoRI methylase are used in some cDNA library construction protocols. In this scheme, newly synthesized cDNA is treated with EcoRI methylase to methylate all EcoRI sites in the cDNA before blunt end ligation of EcoRI adapters (other similar procedures are also used). After ligation, the cDNA is digested with EcoRI to generate EcoRI “sticky ends” which can then be easily ligated to a vector with EcoRI “sticky ends.” Internal EcoRI sites and the integrity of the cDNA are protected by methylation.

Given the different size of the recognition sequences, one could infer that, on average, the frequency of sites is inversely proportional the number of nucleotides recognized by the enzyme. Thus, a 4-cutter like MboI will have on average one site every 256 bases \( (0.25^4) \), a 6-cutter like HindIII one site every 4,096 bases \( (0.25^6) \), and an 8-cutter like NotI one every 65,536 bases \( (0.25^8) \). If the bases were distributed at random in a genome, then restriction digests of a eukaryotic genome would yield DNA fragments with a distribution centered on the preceding figures, depending on the enzyme used in the experiment. However, these are not the observed results. It is true that 4-cutters generate more fragments than 8-cutters, but there is clear variation in the distribution of DNA fragments produced by different enzymes that recognize sites of the same length. The major reason for these discrepancies is the fact that bases are not randomly distributed in the genome. For instance, coding regions tend to have a higher GC content than introns and intergenic regions. The enzyme DraI (TTT/AAA) is likely to cut in introns and intergenic regions, in contrast to NotI (GC/GG*C*CGC) which not only cuts infrequently due to the length of its recognition sequence, but also due to the fact that it is inhibited by C methylation.

**Technical Aspects**

Let’s now go over some of the most practical aspects of storing, handling and using restriction enzymes in general. Several companies offer a relatively large selection of restriction enzymes. These enzymes, like many other nucleic acid modifying enzymes, are sold in a solution containing 50% glycerol, and shipped on dry ice. Once they arrive in the lab, they are stored at -20°C. This temperature is low enough to prevent degradation of the enzyme, but
not too low to freeze-solid a 50% solution of glycerol. Freezing these enzymes will cause irreversible denaturation as freezing weakens and destroys the tertiary structure of proteins that depend on hydrophobic interactions. For this reason, storage at temperatures below -20°C is not recommended as this will denature the enzymes. Similarly, restriction enzymes will be inactivated if they are frozen in a glycerol-free (<5%) aqueous solution. Keep these enzymes on ice at all times when you are working with them.

The activity of restriction enzymes is expressed in units. A unit is defined as the amount of enzyme capable of digesting 1 μg of a DNA substrate in 1 hour under a specified set of conditions defined by temperature and the contents of an accompanying buffer. The units of activity are not comparable between restriction enzymes because not all restriction enzymes are evaluated with the same DNA substrate. This substrate could be bacteriophage lambda, Adenovirus-2, or a specific plasmid. The choice depends on whether the substrate has a site for the enzyme, but after that is determined, the number of restriction sites in each substrates varies between 1 and >10. This is why it is recommended that digestions of genomic DNA be carried out with at least 5 units/μg of DNA. The presence of contaminating inhibitors in the DNA prep may require higher concentrations of restriction enzymes. Most commercial restriction enzyme preparations are stable for 1 hour at 37°C, and a few for 2 hours or more. In general, incubation beyond those times does not increase the extent of digestion.

Most enzymes are optimized to work at 37°C, but those isolated from thermophilic bacteria have optimum temperatures between 50 and 65°C, and the recommended temperature is indicated in the accompanying technical information sheet. The manufacturer commonly supplies buffers in which reactions are to take place; most buffers are supplied at 10 times the desired concentration. Reaction buffers vary in the nature of the buffer, its concentration and pH, and the presence and concentration of sodium and/or potassium ions. Some companies have up to 10 different buffers whereas others only have a few. In general, enzymes have sub-optimal performance in buffers that are not recommended for the enzyme. There are occasions in which a DNA sample needs to be cut with two different restriction enzymes. The best-case scenario is the one in which the two enzymes have the same reaction buffer. Alternatively, the enzymes may have different buffer requirements. In this case, we can explore different possibilities. The DNA could be digested in the same reaction buffer with two different restriction enzymes that have different buffer requirements. Restriction enzymes may have from 0 to 90% activity in suboptimal reaction buffers. Thus, a double digestion might be OK if the activity of the second enzyme is not significantly affected in the sub-optimal buffer. If the activity is severely restricted for the second enzyme, then there are two alternatives. One is to start with the buffer that has the lowest salt concentration, and perform the digestion with the first enzyme in its optimum reaction buffer. After the appropriate time, the first enzyme can be heat-inactivated at 72°C, and then the salt concentration of the buffer can be adjusted to optimum conditions for the second enzyme. Not all restriction enzyme can be heat-inactivated. When DNA quantity is not a limiting factor, the most practical path might be to purify the DNA
(organic extraction, or cleaning kits) after the first digestion, and perform the second digestion with the reaction buffer that is optimum for the second enzyme.

Another practical aspect of restriction enzyme digestions is related to the DNA concentration, the enzyme concentration, and the final volume required for the experiment. Let’s consider a situation in which we are going to detect single copy sequences from barley using Southern blot hybridizations. To do this, we first need to digest barely DNA with a restriction enzyme (EcoRI) and then separate the restriction fragments by agarose gel electrophoresis. Let’s just say that the sample wells in the gel have a capacity for 60 μl. We are going to have to digest and run in each lane 17 μg of DNA if we want to detect single copy sequences. Comparatively, only 2 μg of common bean DNA would be required because the bean genome is about 1/8 the size of the barely genome. We will discuss this in greater detail when we do Southern hybridizations. Returning to the barley samples, we have measured our DNA preparations and determined that the DNA concentration is 250 μg/ml. At this concentration, 68 μl of DNA will be required per sample. The majority of restriction enzymes are sold at a concentration of 10 units/μl. According to the recommendation of using 5 units per μg of genomic DNA, we are going to need 85 units of EcoRI, or 8.5 μl. It turns out that glycerol concentrations in excess of 5% can alter the specificity of some restriction enzymes, and EcoRI in particular. This altered specificity is called “star activity” and the name is derived from the fact that an asterisk has been used in tables to warn users about this phenomenon. The reaction has to be carried out in a minimum volume of 85 μl so glycerol will not be in excess of 5%. To accommodate addition of 10X reaction buffer we will carry this reaction in total volume of 90 μl (DNA 68 μl, EcoRI 8.5 μl, 10X reaction buffer 9 μl, water 4.5 μl). We will now mix the components thoroughly and incubate the reaction at 37°C for 1 hour. At the end of the incubation we need to add 1/5 vol of 6X loading buffer to stop the reaction and facilitate loading in the gel. However, a total volume of 108 μl will not fit the 60 μl sample well. What can we do? There are two possible solutions. One is to precipitate the digested DNA with ethanol and resuspend it in a smaller volume factoring in the loading buffer so that the final volume does not exceed 60 μl. The other solution is to start with a DNA sample of a higher concentration and a restriction enzyme of high concentration as well. A third possibility is possible with a few restriction enzymes, which are available in high concentration formulations.

References


Useful Links:
http://rebase.neb.com/rebase/rebase.html
www.thelabrat.com/restriction/index.shtml
MATERIALS AND SUPPLIES

A. BIOLOGICAL MATERIAL
   Plant genomic DNA

B. REAGENTS AND SOLUTIONS
   Restriction Enzymes: \textit{DraI}, \textit{EcoRI} and \textit{PstI}
   Restriction Enzyme Buffers
   Reaction Stop Buffer = Loading Buffer (See next Section)

C. LABWARE AND SUPPLIES
   Microcentrifuge tubes
   Pipet tips

D. EQUIPMENT
   Microcentrifuge
   Micropipets
   Water Bath with temperature control
PROCEDURE FOR RESTRICTION ENZYME DIGESTION

In this lab we are going to test the effect of different restriction enzymes on plant DNA, specifically on the DNA isolated in the previous lab. Digested DNA will be analyzed via agarose gel electrophoresis. A sample of undigested DNA will be included as control. A set of positive controls will be provided to the class.

1. Set up the following reactions:

<table>
<thead>
<tr>
<th>RE Buffer, 10 X</th>
<th>D</th>
<th>E</th>
<th>P</th>
<th>Cont</th>
<th>LD</th>
<th>LE</th>
<th>LP</th>
</tr>
</thead>
<tbody>
<tr>
<td>RE Buffer, 10 X</td>
<td>2 μl</td>
<td>2 μl</td>
<td>2 μl</td>
<td>-</td>
<td>2 μl</td>
<td>2 μl</td>
<td>2 μl</td>
</tr>
<tr>
<td>DraI</td>
<td>2 μl</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2 μl</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>EcoRI</td>
<td>-</td>
<td>2 μl</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2 μl</td>
<td>-</td>
</tr>
<tr>
<td>PstI</td>
<td>-</td>
<td>-</td>
<td>2 μl</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2 μl</td>
</tr>
<tr>
<td>H2O</td>
<td>x μl*</td>
<td>x μl</td>
<td>x μl</td>
<td>n μl</td>
<td>11 μl</td>
<td>11 μl</td>
<td>11 μl</td>
</tr>
<tr>
<td>DNA, 1 μg</td>
<td>y μl</td>
<td>y μl</td>
<td>y μl</td>
<td>y μl</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>λDNA 0.5 μg</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5 μl</td>
<td>5 μl</td>
<td>5 μl</td>
</tr>
</tbody>
</table>

*Total Volume 20 μl 20 μl 20 μl 20 μl 20 μl 20 μl 20 μl

*x + y = 16 μl; m + y = 20 μl

2. Incubate reactions at 37° C for 1 hour.

3. Stop reaction by addition of 1/5 volume of 6X loading buffer.

**NOTE:** The “loading buffer” for agarose gel electrophoresis stops the reaction, adds a tracking dye, and increases the density of the sample to facilitate sample loading. Samples are usually loaded in the gel by slowly ejecting the sample while the pipet tip is submerged in the buffer filling the sample-well. The glycerol added with the loading buffer increases the density of the sample and causes it to go to the bottom of the well.
AGAROSE GEL ELECTROPHORESIS

**Learning Objectives**

After completion of this module, students will be able to use agarose gel electrophoresis to separate DNA fragments according to size. Thus, students will be able to use this technique to assess the physical integrity of extracted DNA. Or measure the size of specific DNA fragments, in particular those produced by restriction enzymes. Knowledge about the physical-chemical properties of DNA, buffers, and agarose will enable students to make the appropriate decisions on the selection of agarose concentration, electrophoresis buffer, and power settings for optimal resolution of DNA fragments.

**Lab Objectives**

Analyze the size distribution of DNA samples using agarose gel electrophoresis. An agarose slab, with sample-wells, will be formed in a gel apparatus after pouring molten agarose suspended in a buffer solution. DNA samples will be loaded in the wells and forced to migrate through the agarose gel matrix by applying an electric field. DNA will be visualized by staining with a fluorescent dye with high affinity for double stranded nucleic acids, and viewed and photographed under UV light.

**Background**

Agarose gel electrophoresis is the standard method for size fractionation of nucleic acids. This procedure is based on the fact that double stranded DNA is negatively charged in alkaline buffers ($7 \leq \text{pH} \leq 8.5$). The negative charge is due to the phosphate backbone of DNA. Thus, when DNA is placed in a buffered solution (pH ~8) and an electric field is applied, it will move towards the ANODE (+). Furthermore, if the DNA molecules are forced to move through a restrictive porous matrix, then these molecules will separate according to size. The extent of separation, and the range of fragment sizes that can be separated, will depend on the average pore size. The components of the system are an agarose gel, an electrophoresis buffer, an apparatus to hold the gel and buffer, and electrodes to connect to a power supply. These components and their properties are described below.

Agarose is a linear galactan hydrocolloid polymer of the disaccharide agarobiose. This polysaccharide is isolated from marine algae. The commercial products are mixtures of polymers containing varying degrees of modifications by ester sulfates, methoxyl groups, ketal pyruvates and carboxyl groups. The average length of the polymer, its chemical composition, and its net charge, all have dramatic effects on the physical-chemical properties of agarose gels. These properties include: gel strength, melting and gelling temperatures, and pore size among others.
Gel strength is one of the properties that has made agarose a very popular analytical tool. Low concentrations (0.5-2 %) of this polymer can form strong gels which are easy to handle. Gel strength depends on polymer length, temperature, and pH (see buffer below). Strength comes from the ability to form a gel through a gelation process which entails the formation of both double helices by segments of two polymers and hydrogen bonds. Polymer length is positively correlated with strength. Low temperatures (above freezing) also strengthen the gel as they strengthen hydrogen bonds.

The melting and gelling temperatures depend in great part on the chemical compositions of the modifying moieties. For instance, addition of hydroxyethyl groups will lower these temperatures. Low-temperature melting agaroses have many uses in molecular biology, particularly when a DNA fragment or DNA fragments with a specified size range need further manipulation. Following size separation of DNA by electrophoresis in one of these agaroses, and identification of the position of the DNA, the section containing the desired DNA fragment(s) is cut out and transferred to a test tube where the agarose block can be melted at a temperature that is not too high to denature the DNA. Once the DNA is free from the gel structure, it can be enzymatically manipulated by different enzymes such as restriction enzymes, kinases, phosphatases, ligases, and many others. Some low-temperature gelling agarose preparations are used at “high concentrations” (4-10 %) for the separation of small DNA fragments (10-500 bp).

The electrophoresis buffers most commonly used for the separation of double stranded DNA are TAE (Tris.Acetate, EDTA), TPE (Tris.Phosphate, EDTA), and TBE (Tris.Borate, EDTA). The ionic strength of the buffer plays a role in DNA mobility. Low ionic strength buffers like TAE or TPE produces sharp bands, but can get exhausted very quickly. DNA moves through the gel more slowly in high ionic strength buffers than in those of low ionic strength because at high ionic strength there are more anions, other than DNA, to carry the charge. This phenomenon causes the DNA to move more slowly and allows it to diffuse more, but separation of fragments can proceed for a longer period of time.

Buffers are exhausted when most of the ions migrate out of the center of the field. Some of the consequences of buffer exhaustion should be viewed in light of Ohm’s Law (V = IR). When the ion concentration drops, so does the current, resulting in a proportional increase of the electrical resistance, which in turn heats up the gel. In the worst cases, the gel can heat up to the melting point causing DNA denaturation (and warping of the acrylic gel bed). For this reason, low ionic strength buffers should be used for short runs, or alternatively they could be circulated by means of a pump from one electrode tank to the other. TBE buffer is commonly used for long runs.

Let’s examine further the changes that occur at each buffer reservoir. Water is electrolyzed at each electrode (see equations below). At the cathode (-), protons are reduced by electrons supplied by the cathode producing hydrogen gas and hydroxyls which tend to raise the pH, a trend that is suppressed by the presence of the buffer. In contrast, at the anode (+),
electrolysis of water is driven by the oxidation that produces protons and oxygen gas. This acidifying effect is suppressed by the presence of the buffer. Thus, when the buffer is exhausted, the cathode becomes alkaline and the anode acidic. Buffer exhaustion can lead to melting of the agarose gel, and to agarose degradation at the cathode end as the heat and alkaline conditions can hydrolyze the agarose polymer. Notice in the equations below that on an electron basis, there are twice as many hydrogen molecules produced at the cathode as oxygen molecules at the anode. In practical terms, this phenomenon can be used to diagnose whether the leads to the power supply were connected correctly as the cathode should produce twice as many bubbles as the anode.

**Cathode**

\[ 2e^- + 2H_2O \rightarrow H_2 + 2OH^- \]
\[ HA + OH^- \rightarrow H_2O + A^- \]

**Anode**

\[ H_2O \rightarrow 2H^+ + \frac{1}{2}O_2 + 2e^- \]
\[ H^+ + A^- \rightarrow HA \]

The electrical field that is normally used in the separation of nucleic acids varies between 2 and 10 volts/cm. A high field-strength is used for short runs in a minigel, while a low field strength is used for long runs (overnight). In general, high field strengths lead to a high resolution of small fragments, while low field strengths lead to high resolution of large DNA fragments. There are many types of power supplies used with electrophoresis with different levels of sophistication. At one end are those that just provide a constant voltage drop, while more advanced designs have the options of setting constant voltage (V), current (I), or power (P = VI), and in addition they have a timer that allows the operator to set the length of the run. There are two issues related to safety. One is the need to place a cover over the electrophoresis tank to avoid the possibility of an electrical shock. The other is rather an explosive issue. It is recommended that at the end of the electrophoresis run the leads be disconnected first from the power supply rather than from the gel apparatus. The reason is that hydrogen and oxygen gases can accumulate under the lid and any electrical sparks produced while disconnecting the leads may trigger an explosion.

DNA fragments resolved in agarose gels can be visualized by staining them with a fluorogenic substance that has high affinity for DNA. Ethidium bromide (EtBr) is the most commonly used DNA stain, and it is normally used at a concentration of 0.5 μg/ml. EtBr is a planar molecule (similar to propidium iodide) that intercalates into the DNA (whether it is in the gel, or in your hands!). Its high affinity for DNA makes it a powerful mutagen and a
cancer, and for these reasons you MUST WEAR GLOVES when handling this chemical. Intercalation of EtBr into DNA reduces the fluorescent quenching effect exerted by the solvent (water), resulting in an increase in fluorescence yield. EtBr has a peak of absorption in the UV range and an emission peak at approximately 610 nm in the red zone of the spectrum. The distribution of fluorescent DNA zones in the gel can be captured by photography (Fig. 2). A typical setup comprises a light-shielded compartment in which the UV light is projected from under the gel, and a camera over the gel can capture the positions of the fluorescent zones (DNA bands) in the gel. A filter is used to block the UV-light from entering the camera allowing only red light through. Photographic images can be obtained either with film or electronically. Modern equipment has many safety features, but you should be aware of the potential dangers of UV-light (blindness and skin cancer) and take all the precautions to protect your eyes and skin from UV sources. It should be pointed out that there are now several fluorescent dyes for sensitive detection of nucleic acids like SYBR Gold, SYBR Green, and the family of cyanine dimers also known as the TOTO family of dyes, which are much more sensitive than EtBr, and more expensive as well.

There are two ways to stain DNA separated by electrophoresis. One way is to first run the gels, and then stain them by incubation in a solution of EtBr. The other is to incorporate EtBr into the gel. In this procedure, the DNA will capture the dye during electrophoresis. The dye is actually positively charged and migrates in opposite direction to the DNA. An issue to consider here is that DNA containing EtBr will migrate more slowly in the gel because upon intercalation the dye unwinds the DNA molecule a few degrees. The unwinding increases the volume of the molecule and also reduces its flexibility. The distance traveled by a DNA fragment in an agarose gel during electrophoresis is inversely proportional to the logarithm of its molecular weight (Log MW). This relationship is linear within a certain range of MW, and this range is inversely proportional to the agarose concentration. In turn, the average pore size depends on the agarose concentration. The linear relationship between DNA mobility and the log of its MW makes it possible to determine the size of any DNA fragment by comparing its migration relative to the migration of molecular weight standards. Conventional agarose gel electrophoresis can be used to resolve DNA fragments between 0.2 and 25 Kb. Smaller fragments can be resolved by specialty agarose preparations, but single base resolution can be obtained by polyacrylamide gel electrophoresis (PAGE). On the other hand, high MW fragments (20 Kb to 5 Mb) can be resolved by clamped-homogeneous electric field (CHEF) electrophoresis. The latter procedure is used to separate entire chromosomes of fungi.
Agarose gels are prepared by first suspending the agarose powder in an aqueous buffer, and then melting it at high temperature (boiling). The molten agarose is poured into a mold (gel bed) where it is allowed to harden (gel) and form a porous lattice. Well-forming combs are used as templates to shape rectangular receptacles (also called wells) for the DNA samples. Wells are formed near the CATHODE (-). Samples are prepared by mixing the DNA with a “gel loading buffer” that contains glycerol and some tracking dyes. The glycerol increases the density of the sample and facilitates underlayering the sample in a well that it is already filled with electrophoresis buffer. The tracking dye helps keep track of the “loading” process, and because it is negatively charged, it is also used to monitor progress of the electrophoretic separation. After loading the wells, an electric field is applied to the gel and the negatively charged DNA moves towards the ANODE (+). The rate of DNA migration in the gel depends on the average pore size of the gel, the buffer (salt concentration and pH), and the size of the DNA fragment.

References
MATERIALS AND SUPPLIES

A. BIOLOGICAL MATERIAL

DNA

Unprocessed (Genomic, plasmid, PCR product)
Processed (Digested by restriction enzyme, ligated, etc.)

B. REAGENTS

<table>
<thead>
<tr>
<th>Acetic Acid</th>
<th>EDTA Na₂</th>
<th>Sodium Acetate, CH₃COONa.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agarose, Type I</td>
<td>Ethidium Bromide</td>
<td>SDS</td>
</tr>
<tr>
<td>Boric Acid</td>
<td>Glycerol</td>
<td>Tris base</td>
</tr>
<tr>
<td>Bromo Phenol Blue</td>
<td>Phosphoric Acid 85%</td>
<td>Xylene Cyanole FF</td>
</tr>
</tbody>
</table>

C. SOLUTIONS

Stock Solutions

EDTA, 0.5 M, pH 8.0

Na₂EDTA.2 H₂O  186.1 g
NaOH pellets  19.0 g
H₂O add to  1000.0 ml

NOTE: Dissolve salts in 800 ml of water and stir vigorously with a magnetic bar. Adjust pH to 8.0 with 4N NaOH. Bring to final volume and autoclave.

Sodium Dodecyl Sulphate, 20 %

SDS  40. g
H₂O add to  200.0 ml

NOTE 1: Wear a dust mask to handle SDS.

NOTE 2: Dissolve in 140 ml of water while stirring with a magnetic bar in a beaker. Speed up the process by either stirring over a hot plate (low heat) or brief warming in a microwave oven. Avoid foaming. Carefully pour solution into a graduated cylinder, rinse beaker with water and pour into cylinder until reaching final volume. Cover cylinder with parafilm, or any other sealing film, and mix before pouring into storage bottle.
Working Solutions

TAE Electrophoresis Buffer, 50X
Tris base 242.0 g
Acetic acid 57.1 ml
EDTA.Na 0.5 M 100.0 ml
H2O add to 1000.0 ml

TBE Electrophoresis Buffer, 5X
Tris base 54.0 g
Boric acid 27.5 ml
EDTA.Na 0.5 M 20.0 ml
H2O add to 1000.0 ml

TPE Electrophoresis Buffer, 10X
Tris base 108.0 g
Phosphoric acid, 85% 15.5 ml
EDTA.Na 0.5 M 40.0 ml
H2O add to 1000.0 ml

Loading Buffer - Blue Juice 6X

<table>
<thead>
<tr>
<th></th>
<th>[final]</th>
</tr>
</thead>
<tbody>
<tr>
<td>EP buffer 10X</td>
<td>0.25 ml 0.5 ml</td>
</tr>
<tr>
<td>NaEDTA (0.5 M)</td>
<td>0.20 ml 0.4 ml</td>
</tr>
<tr>
<td>SDS (20 %)</td>
<td>0.05 ml 0.1 ml</td>
</tr>
<tr>
<td>BrPheBlue (25 mg/ml)</td>
<td>0.10 ml 0.2 ml</td>
</tr>
<tr>
<td>Xylene Cyanole FF (25 mg/ml)</td>
<td>0.10 ml 0.2 ml</td>
</tr>
<tr>
<td>Glycerol</td>
<td>3.50 ml 7.0 ml</td>
</tr>
<tr>
<td>H2O</td>
<td>0.80 ml 1.6 ml</td>
</tr>
<tr>
<td>Final vol</td>
<td>5.00 ml 10.0 ml</td>
</tr>
</tbody>
</table>

Ethidium Bromide Stock, 10 mg/ml
NOTE 1: Store in dark bottle and in the dark.
NOTE 2: Wear gloves to handle EtBr. It is both a MUTAGEN and a CARCINOGEN!

D. LABWARE AND SUPPLIES
Micropipet tips
Staining Trays
X-ray film (used)

E. EQUIPMENT
Agarose Gel Apparatuses
Balance
Imaging System
Micropipets
Microwave Oven
Orbital shaker
Power Supplies
UV Transilluminator
AGAROSE GEL ELECTROPHORESIS

A. Gel preparation

1. Prepare minigel apparatus:
   - Place gel bed in electrophoresis tray, and place dams at both ends.
   - Level apparatus with the gel tray in place.

2. Pour gel. Molten agarose (1%) will be readily available for the class. Pour the molten agarose onto the gel bed until it just barely reaches the reams on both sides. Immediately place the comb in the comb slots closest to the cathode (black terminal). Wait 15 to 30 minutes for the agarose to gel.

B. Electrophoresis Run

1. Carefully remove the comb with a left-right rocking motion first, and then back and forth rocking motion, and without tearing the gel. Remove the dams too.

2. Pour electrophoresis buffer to fill the tanks on both sides of the gel and enough to fill the wells and barely cover the gel. Eliminate any bubbles that may be trapped in the wells.

3. Load the samples into the wells with a micropipet in the following order:
   1. Molecular Weight Markers
   2. Undigested plant DNA
   3. DraI digested plant DNA
   4. EcoRI digested plant DNA
   5. PstI digested plant DNA
   6. DraI digested λ DNA
   7. EcoRI digested λ DNA
   8. PstI digested λ DNA

   **NOTE 1:** All samples should contain loading buffer.
   **NOTE 2:** Make sure the volume of the sample does not exceed volume of the well.
   **NOTE 3:** Make sure a MW standard is loaded in one of the wells. If a lambda DNA digest is used as MW std., then it should be heated at 65°C for 3 min prior to loading.

4. Connect the leads to the gel apparatus and then to the power supply. Set voltage to 100 and run time to 20 min. Run samples to the desired length based on migration of tracking dye.

   **NOTE 1:** The negatively charged DNA will migrate to the ANODE (+)!
   **NOTE 2:** Make sure the lid is on the gel rig to prevent an accidental electrical shock!

5. Turn power off, disconnect leads, and transfer gel to a staining tray.
Visualization of DNA Fragments.

1. **Stain gel for 15 min in ethidium bromide 0.5 µg. ml.**

2. Discard staining solution into appropriate container.
   
   **NOTE:** Do not pour down the sink!

3. Rinse in tap water.

4. Photograph gel with the video imaging system. Keep the output: electronic and hard copy.
   
   **NOTE 1:** Make sure the filter used to block off UV light is in place.
   
   **NOTE 2:** Take special care not to scratch the surface of the trans-illuminator.

Appendix

If you were to prepare the molten agarose these are the steps that you would have to follow:

1. Weigh the correct amount of agarose powder (0.7-2.0%) and transfer to a flask:
   - Large gel 300 ml
   - Medium gel 80 ml
   - Mini gel 20 ml
   
   **NOTE:** Thickness of gel varies between 4 and 10 mm.

2. Add the correct vol. of Electrophoresis Buffer to the flask and resuspend the powder.

3. Melt agarose.
   
   Microwave: weight flask and contents. Heat in the microwave until agarose melts completely. Remove flask and shake slowly to mix in water condensed on the sides of the flask. Point flask away from your face (and those of other’s as well) as molten agarose can overheat and upon shaking boil over. Place on a balance and add water to the original weight. Immediately swirl to mix in the water.
   
   Magnetic stirrer/hot plate: Cover the flask with aluminum foil and bring to a boil with constant stirring. Heat until agarose is completely dissolved.

4. If the gel bed is made of relatively thin acrylic (~3/16”), then cool agarose to 55-65° C (or incubate in water bath at ~60”), pour gel, and put the comb in place.
   
   **NOTE 1:** The molten agarose can be cooled by placing the agarose flask, with a magnetic bar, on a tray containing an ice-water mix. Monitor temperature drop with a thermometer.
   
   **NOTE 2:** Minigels may not require cooling the molten agarose because the gel beds are too thick to be warped by heat.

5. Let agarose set for 15-30 min before loading samples.
   
   **NOTE:** Cover gel w/saran wrap or buffer if it is not going to be used immediately.
Study Questions

1. In contrast to animal tissues, what barriers do you need to overcome to disrupt plant tissues for nucleic acid extraction? How do you overcome this barrier?

2. Why is EDTA included in DNA extraction buffers and solutions?

3. Proteins represent a barrier to DNA extraction, either as structural components of chromatin or as DNA degrading enzymes. What steps and reagents are used to eliminated them or avoid their action?

4. Phenolic compounds are present in most plant tissues, and some species have them in unusual high amounts. What kind of problems do they create? What measures can be taken to control phenolic-related problems?

5. Why is RNA a usual contaminant in DNA extracts? What steps were taken in the procedure you used to eliminate RNA?

6. What DNA property is used in the spectrophotometric determination of concentration?


8. What are the advantages of absorption over fluorescent measurements, and vice versa?

9. Why is that you do not need a standard curve when you use a spectrophotometer to measure DNA concentrations?

10. What qualities of a DNA sample can you assess via agarose gel electrophoresis?

11. Why do you think pectins are usually contaminants in DNA preparations? How do you eliminate these carbohydrates?

12. Why does RNAse retain activity after being boiled?

13. What are the components of the R-M system? What roles do they play in the organism that carries them?

14. How would you recognize a sequence that is recognized by a Type II restriction enzyme.

15. What is an isoschizomer?

16. Why do restriction enzymes are sold in a solution containing 50 % glycerol? Can glycerol affect the activity of the restriction enzyme? If it does, then how do you manage it?

17. You digest genomic DNA with two different restriction enzymes that recognize different 6-base sequences. After resolving the resulting DNA restriction fragments you find out that instead of discreet bands you have smears, and that the distributions of the signal in these smears are very different. Explain the presence of smears and the difference in the distribution of DNA fragments?

18. Would you be able to separate DNA molecules in a dense medium containing a high concentration of sucrose? Explain.

19. What chemical reactions take place at the cathode? What is the cause of buffer exhaustion?

20. How is DNA visualized after separation by agarose gel electrophoresis?
Lab Report: DNA Extraction and Analysis.

Part I.
Summary (One Page Maximum, 20 points)
a) the main objective of the experiment,
b) the most important findings,
c) the significance of the results, and
d) major conclusions

Part II
Results (30 points)
Summary of the results obtained in the lab. Include the following:
- Species:
- Tissue:
- Sample weight:
- Extraction Method:
- DNA resuspension volume
- DNA concentration

Spectrophotometry
a) Dilution
b) $A_{260}$ (dilution)
c) $A_{260/280}$:
d) $A_{260/230}$:
e) DNA concentration = ng/μl
f) Total amount of DNA extracted =
g) DNA/g of tissue

Fluorimetry

<table>
<thead>
<tr>
<th>Std</th>
<th>DNA Stock</th>
<th>μl</th>
<th>DNA, ng</th>
<th>RF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 mg/ml</td>
<td>10</td>
<td>10,000</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1 mg/ml</td>
<td>5</td>
<td>5,000</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1 mg/ml</td>
<td>2</td>
<td>2,000</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.1 mg/ml</td>
<td>10</td>
<td>1,000</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.1 mg/ml</td>
<td>5</td>
<td>500</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.1 mg/ml</td>
<td>2</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample</th>
<th>μl</th>
<th>RFU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a) Calculate the parameters of the standards (intercept and slope)
b) Use the linear relationship to calculate your DNA concentration, total amount of DNA, and the amount of DNA per g of tissue.
Agarose gel electrophoresis

%- agarose:
Power (Volts)
Electrical Field (V/cm)
Run time:
Include Gel picture
What can you conclude about your sample when you compare it to the control?

Part III.

1. (20) Are there any differences between the two DNA concentration estimates that you obtained for your sample? Explain.

2. (20) Answer the following questions regarding the use of restriction enzymes:
   a) Are there any differences in the distribution of DNA fragments from the different restriction enzyme digestions including the uncut control? Explain.
   b) Considering that restriction enzymes used in this experiment were all six-cutters, one would expect the average fragment size to be 4,096 base pairs long with an almost equal distribution of fragments above and below that size. Why do the patterns differ from the expected pattern?

3. (10) Standard cloning procedures involve cutting open the cloning vector and the target DNA with the same restriction enzyme to generate compatible ends. However, there are some circumstances in which molecular biologists generate compatible ends cutting open the vector with one restriction enzyme, and the target DNA with a different restriction enzyme. Moreover, in some cases, the ligated DNAs can be cut at the same site, but in other cases they can’t. Indicated whether the following pairs of restriction enzymes would be compatible in a cloning experiment, and what ligations would allow re-cutting, even when you have to use a different restriction enzyme (Hint: you can use catalogs or the web to find the answers).

<table>
<thead>
<tr>
<th>Vector</th>
<th>Insert</th>
<th>Compatible ends? Y/N</th>
<th>Re-cut? Y/N</th>
<th>What Rest. Enzyme?</th>
</tr>
</thead>
<tbody>
<tr>
<td>EcoRV</td>
<td>Dral</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EcoRV</td>
<td>HaeIII</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BamHI</td>
<td>Sau3AI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BamHI</td>
<td>BglII</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EcoRI</td>
<td>Apol</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 3

BIOINFORMATICS

Learning Objectives ............................. 3.1
Lab Objectives ................................. 3.1
Background ..................................... 3.1
Bioinformatics ................................. 3.4
Entrez ............................................. 3.4
Databases ....................................... 3.5
BLAST ............................................ 3.6
Sequence Retrieval ......................... 3.9
Structural Analysis .......................... 3.12
Identification of Exon Coord. ............ 3.13
GeneSequer .................................... 3.13
References .................................... 3.14
Study Questions .............................. 3.15
Lab Report ..................................... 3.16
BIOINFORMATICS

Learning Objectives

After the completion of this unit, students will be able to use the National Center for Biotechnology Information data retrieval system to search diverse databases and retrieve appropriate information related to specific DNA and protein sequences from diverse organisms. Finally, students will be capacitated to use this and similar resources to perform *in silico* structural analysis of DNA sequences.

Lab Objectives

Find and download the assigned sequence from NCBI databases, and perform a structural analysis of the sequence.

Background

Before proceeding with the bioinformatics section *per se*, we should first review the most basic steps in the flow of genetic information. These steps were originally described by Francis Crick in the late 1950's and constitute the “Central Dogma” of molecular biology (Figure 1). According to this model, the information contained in DNA is preserved by DNA replication, and it is transferred to RNA by transcription. Finally, RNA sequences are translated into proteins. The “Central Dogma” is now viewed as an oversimplification of the system as more has been learned about how genetic information flows among these three biopolymers (Shapiro, 2009).

The commitment to DNA replication takes place during the G1 phase of the cell cycle, while the actual synthesis occurs during the S phase; this is followed by G2 and mitosis ensuring the perpetuation of the genetic information. Transcription on the other is the process by which the information encoded in DNA is passed on to an intermediary carrier, the mRNA (Figure 1). The role of the *promoter* region of the gene is to recruit transcription factors which in turn recruit RNA polymerase and position this enzyme at the *start site* of...
transcription. RNA polymerase synthesizes the transcript by “reading” the template DNA strand (Negative or anti-sense) and elongates the transcript until it reaches the terminator sequences. Following termination, the RNA transcript is cleaved upstream of the termination point at the polyadenylation site, and a poly A tail is added by poly(A) polymerase. The primary transcript (pre-mRNA), with the exception of the poly(A) tail, has the same sequence as the coding strand and is complementary to the template strand.

The coding sequences of most eukaryotic genes are interrupted by non-coding DNA stretches of variable length called introns, while the interrupted stretches of coding sequences are called exons. Following transcription, introns are removed from the primary transcript during the RNA splicing step to yield the mature mRNA. One characteristic feature of introns is the presence of specific dinucleotides at the 5' and 3' ends, GT and AG, respectively. Following splicing, exons remain in the same linear order as they are found in the genomic DNA sequence. Notice in Figure 1 that the first and last exons contain non-coding and coding sectors (Figure 1). The non-coding sectors are known as the 5' and 3' untranslated regions (5' UTR and 3' UTR). These sequences contain important information pertaining to regulation of translation as well as transcript stability. Comparative analysis of gene structure between plants (Arabidopsis) and animals (humans) has revealed significant differences. On average, an Arabidopsis gene is 2.4 Kb long, has five exons with an average length of 172 bp, and four introns 173 bp long; in contrast, the average human gene is 28 Kb long, it has 8.8 exons 130 bp long, and 7.8 introns 3Kb long.

Finally, the genetic information encoded in mRNAs is translated into proteins sequences by ribosomes according to the established genetic code. In this code, the identity of an amino acid is dictated by three-base codons. Translation starts at the AUG start codon and proceeds until it reaches a termination codon (UGA, UAA, or UAG). Given that there are only four bases, a total of 64 codons are possible. Three of these encode a stop signal and the other 61 specific amino acids. Considering that there are only 20 amino acids used in protein synthesis, these numbers clearly underscore the redundancy that exists in the genetic code. With the exception of Methionine which is encoded by a single codon (AUG) all other amino acids are encoded by 2 (Cysteine) to 6 (Leucine, Serine) codons. This redundancy is also viewed as a degeneracy of the code. The third base of a codon is the most variable among redundant codons, followed by the second and first bases. A consequence of this redundancy is that it makes it possible to have a single protein sequence encoded by different DNA sequences! This fact will be relevant when searching for similarities between sequences from distantly related taxa.

Recent advances in DNA sequencing technologies have resulted in a rapid expansion of databases containing nucleic acid and protein sequence information. However, the fact that nucleic acid and protein sequences are obtained via a diverse array of strategies and procedures makes it important to clearly identify the origin of the sequences. DNA sequences in data banks arise from two sources: genomic DNA and cDNA. Genomic DNA sequences are derived from chromosomal DNA by way of either cloning, or PCR amplification from genomic DNA.
Typically, genomic libraries are constructed using a variety of vectors that include bacteriophage lambda, cosmids, fosmids, P1, and bacterial artificial chromosomes (BACs). These genomic libraries are constructed for the purpose of characterizing a specific region of the genome (a gene), or the entire genome. Databases may contain the complete genomic sequence of a gene, the full sequence of a clone (20-300 Kb), or the entire sequence of a genome. It may also be possible to find simply the termini of genomic clones (~500 bp). Genomic sequences derived via PCR are also common in databases. These sequences are obtained from either random or specific sequences. Recently, there has been an explosion of genomic sequences obtained with second generation sequencing instruments that use massive parallel sequencing of PCR-based vector-free libraries. We will visit these approaches in the DNA sequencing section.

In contrast to genomic DNA sequences, cDNA (complementary DNA) sequences are obtained in vitro via reverse transcription of mature messenger RNA (mRNA). In essence, cDNAs are DNA versions of mRNA sequences. We should bear in mind that full cDNA sequences differ from the genomic sequence from which they were transcribed by the absence of introns, and the presence of a poly A tail. We should also bear in mind that although the majority of higher eukaryotic genes possess introns, a small fraction of them lack these sequences, which presumably were lost during evolution.

The synthesis of cDNAs, or cDNA libraries, is carried out by a retroviral enzyme called “reverse transcriptase” (RT). Retroviruses are parasitic RNA viruses that encode a RT enzyme which reverse transcribes the viral genomic RNA into double stranded DNA. This DNA is then inserted into the genomic DNA of the host by the virus-encoded integrase enzyme. Afterwards, the virus is propagated by the host transcription machinery. Although there are many RT enzymes that are commercially available, the most commonly used RT enzymes are those derived from the Moloney Murine Leukemia Virus (M-MLV), and the Avian Myeloblastosis Virus (AMV).

cDNA libraries are constructed from mRNA isolated from tissues at specific stages of development (e.g.: 10-day old seedlings), or from tissues after a particular treatment (heat, cold, hormone, pathogen, etc). The purpose of these libraries is to capture and identify the set of genes that are expressed under specific conditions. Libraries vary in the percentage of full-length cDNA clones, the higher their frequency, the better the quality of the library is. Alignment of complete cDNA sequences with the corresponding genomic sequences can reveal the location of exon-intron junctions and consequently exon and intron sequences. Furthermore, alignments of a large number of cDNA -genomic pairs in a species have been used to develop software capable identifying exon-intron junctions in genomic DNA sequences. These software tools are used to analyze sequences of entire genomes to identify and predict gene structures.

Improvements in sequencing technologies, including automation, have facilitated sequencing of large cDNA libraries to generate collections of “Expressed Sequence Tags” (EST). These collections are made up of short single pass sequences of varying quality. However, the sheer volume of data that are generated with this approach can provide useful
information about the expression profiles of tissues from which mRNA was extracted.

Protein sequences are generally obtained from conceptual translations of cDNA sequences. However, many protein records include additional information derived from direct experimentation. This information may include enzymatic activity, post-translational modifications, particular amino acid motifs, and physical-chemical characteristics.

**Bioinformatics**

The current and overwhelming amount of sequence information has created an enormous challenge for biologists. This challenge has led to the birth of the bioinformatics field. Bioinformatics uses the principles and methods of computer sciences and systems engineering to store, manage, analyze and interpret complex arrays of biological data, and to help in the formulation of biological relevant questions. A number of bioinformatics resources have been developed for the research community. Among the most important are those provided by the National Center for Biotechnology Information (NCBI), which is part of the National Library of Medicine (NLM) at the National Institutes of Health (NIH; [http://www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/)). In the early 1990s NCBI identified the need to archive and retrieve various types of biological information. To fulfill these objectives, NCBI developed **Entrez**, a data storage and retrieval system that provides access to a number of integrated databases ([http://www.ncbi.nlm.nih.gov/books/NBK3836/](http://www.ncbi.nlm.nih.gov/books/NBK3836/)). These databases encompass a diverse array of information including scientific literature, nucleotide and protein sequences, 3-D structure of biomolecules, and genomic data among many others. Entrez can be used effectively as a tool for discovery because it can extract information from different databases, which together can be used to formulate hypotheses and plan experiments.

**Entrez**

“**Entrez** is NCBI’s primary text search and retrieval system that integrates the PubMed database of biomedical literature with 39 other literature and molecular databases including DNA and protein sequence, structure, gene, genome, genetic variation and gene expression” (NCBI, Entrez Help Manual). The different databases have different degrees of interconnectivity. For instance, fields in one database record may have links to a number of other databases in the system. Thus, a protein sequence may have links to PubMed, a DNA database, a 3-D structure database, and a gene expression profile database. This interconnectivity among databases can turn **Entrez** into a discovery tool. **Entrez** can be used to locate a specific DNA sequence that has not been characterized to any detail, and with the aid of some NCBI tools, like BLAST (Basic Local Alignment Search Tool), it is possible to retrieve similar sequences and make inferences about structure, function and evolution of those sequences. For more detailed information about the NCBI databases, Entrez and other tools, students are referred to the NCBI Handbook available at the NCBI web site ([http://www.ncbi.nlm.nih.gov/books/NBK21101/](http://www.ncbi.nlm.nih.gov/books/NBK21101/)). Some
databases comprise several smaller databases. For instance, the Nucleotide Database is an umbrella database that comprises the EST, GSS, and CoreNucleotide Databases. Each record in a database represents an objective observation of data, and has a unique identifier (UID). The integration of the databases in the system permits Entrez to retrieve records with links to other records within the database, or to records in other databases. The sources used by a database could vary depending on its nature, for example, all publications come from the PubMed database, but protein sequences may originate from a variety of sources that include conceptually translated DNA sequences from GenBank, SWISS-PROT managed by the Swiss Institute of Bioinformatics and the European Bioinformatics Institute, the Protein Information Resource (PIR), etc.

An effective utilization of *Entrez* requires knowledge about the databases included in the system. However, it is necessary to recognize first that all the sequence information maintained by NCBI is shared with its partners in the International Nucleotide Sequence Database Collaboration (INSDC) comprised by: the European Bioinformatics Institute (EBI) of the European Molecular Biology Laboratory (EMBL), and the DNA Data Bank of Japan (DDBJ). The three partners share the contents of their own databases exchanging information on a daily basis.

**Databases**

The Core Nucleotide database contains DNA sequences included in *GenBank*, the *RefSeq* database (Reference Sequence), the *TPA* database (Third Party Annotation), and *PDB* (the Protein database). *GenBank* is an annotated collection of primary nucleotide sequences and their corresponding protein translations stored in PDB. *GenBank* contains primary sequence data generated by research labs. Only the labs or contributors of the sequences in GenBank have editing control. *GenBank* can contain multiple entries for the same sequence because submissions can originate from different labs. The *RefSeq* database is a collection of unique sequences derived from GenBank. These sequences are selected from model organisms and curated by NCBI staff. The *TPA* database contains sequences submitted by non-NCBI staff. These sequences are derived or assembled from sequences that are already in GenBank, and contain new annotations. This work may or may not be published in the scientific literature.

To simplify searches, NCBI has developed an umbrella Nucleotide Database that includes three databases: The Core Nucleotide database, the *dbEST* (expressed sequence tags database), and the *dbGSS* (genome survey sequences database). ESTs are usually short (<1 kb) single-pass sequences of cDNA clones obtained from a particular tissue and/or developmental stage. Included in this database are also sequences generated by differential display and Rapid Amplification of cDNA Ends (RACE) experiments. Usually, very little is known about these sequences, which tend to be of low quality. The extent of annotation for these sequences varies from no information at all, to annotation indicating similarity to known genes. By definition,
ESTs lack intronic sequences because they are derived from mRNA. These sequences are generated by large-scale sequencing projects and are submitted in batch. GSSs are also short sequences (<1Kb) about which little is known. In contrast to ESTs, these sequences are derived from genomic DNA and therefore they may or may not contain non-coding DNA sequences. GSSs include, but are not limited to, single-pass sequences of genomic DNA, BAC-end sequences, exon-trapped genomic sequences, and genomic sequences amplified by PCR.

One of the questions the reader may be asking by now is how does one establish a connection between a DNA sequence and the protein it encodes, and by extension the activity or function of this protein. In the early days of molecular biology, this link was established for genes that were expressed at very high levels in specialized tissues (e.g.: phaseolin, the storage protein of the common bean; the globin protein from rabbit reticulocytes, etc.). In developing bean seeds or rabbit reticulocytes, the mRNA for those proteins was the most abundant and easy to isolate and purify. cDNAs from these mRNA were first cloned, and these clones were used to probe genomic libraries to isolate genomic clones of those genes. A variety of more sophisticated methods were later used to clone genes expressed at lower levels. Among these was the purification of a protein with a known enzyme activity followed by the production of an antibody against the enzyme. This antibody would then be used to “fish out” the nascent peptide chain still attached to the ribosomes “translating” the mRNA encoding the enzyme. The associated mRNA was then converted to cDNA and cloned. Other cloning schemes involve complementation of bacterial or yeast mutants, and map-based cloning methods that also required complementation assays. In the meantime, cloned genes were painstakingly sequenced, and the sequence data begun to accumulate in data banks. In view of the richness of these data banks, identification of new genes has relied heavily on the detection of sequence similarities across taxonomic groups.

**BLAST**

Comparative analysis of biological units led Charles Darwin to develop the Theory of Evolution. He took structural and morphological similarities among species as evidence of relatedness and common ancestry. More recently, it has been determined that inferences of relatedness based on morphology correlate well with similar inferences made on the basis of DNA and protein similarities. Large-scale methodical and analytical sequence comparisons have been made possible through the development of sequence alignment algorithms. These algorithms are designed to carry out a quantitative evaluation of the degree of similarity between sequences. This measurement can be factored in when making inferences about the homology of the sequences – a qualitative term that is used to indicate whether two sequences share common ancestry. Thus, sequence comparisons are made in the context of evolution.

The first alignment methods aimed to obtain the best global alignment of two sequences, a type of alignment that engages the entire length of the sequences. This alignment works well
with closely related sequences of approximately equal lengths, but has serious limitations when comparing a cDNA sequence with its genomic template because it forces pairing of exonic bases from the cDNA with intronic bases from the genomic sequence. Similar problems are also encountered comparing distantly related sequences in which substitutions, insertions and/or deletions (indels), and rearrangements may have occurred. Furthermore, not all segments of the genome evolve at the same rate. Exon sequences evolve at a much slower rate than intron sequences because they are under stronger selection pressure. In fact, exon sequences tend to be highly conserved, particularly at the protein level. On the other hand, introns have a faster rate of evolution and usually vary in sequence and sequence lengths between closely related species. Effective sequence comparisons across distant taxonomic groups have been facilitated by the development of local alignment methods. One of these is the Basic Local Alignment Search Tool, also known as BLAST (Altschul et al., 1990), which is one of the analytical tools of Entrez.

BLAST attempts to produce an alignment between a sequence of interest, the **query**, and sequences found in a database, the **subject**. There are several versions of BLAST from which a researcher can select:

<table>
<thead>
<tr>
<th>Program</th>
<th>Query</th>
<th>Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>BlastN</td>
<td>Nucleotide sequence</td>
<td>Nucleotide</td>
</tr>
<tr>
<td>BlastP</td>
<td>Amino acid sequence</td>
<td>Protein</td>
</tr>
<tr>
<td>BlastX</td>
<td>6-frame conceptual translation of a DNA sequence</td>
<td>Protein</td>
</tr>
<tr>
<td>tBlastN</td>
<td>Amino acid sequence</td>
<td>6-frame conceptual translation of a DNA sequence.</td>
</tr>
<tr>
<td>tBlastX</td>
<td>6-frame conceptual translation of a DNA sequence.</td>
<td>6-frame conceptual translation of a DNA sequence.</td>
</tr>
</tbody>
</table>

The query sequences should be presented to BLAST in the FASTA format. In this format a sequence has two components: the first line preceded by the “>” sign followed by a single descriptive title, and the next lines with the nucleotide or amino acid sequence. Upon submission of the **query** sequence, BLAST breaks up both query and database sequences into "words" (11 bases or 3 amino acids), and constructs a table with these words and their "neighboring words." For nucleotide alignments BLAST keeps a **Score** adding 2 points for each identically aligned pair, and subtracting 3 points for a mismatch. After a successful alignment, BLAST proceeds with the flanking neighboring “words.” When no matches are detected, a **gap** is created after subtracting a penalty from the Total Score, extending the gap has a penalty of a lesser value. Extension of the alignment continues until the score decreases by a critical amount. In this fashion, it is possible to obtain segmental alignments of a cDNA with the genomic
sequences from which it was derived, or between only the exons of two genomic sequences. Amino acid alignments use a substitution matrix that adds points for perfect matches (Leu for Leu = 4) or likely substitutions (Leu for Ile = 2) and subtracts points for those that are unlikely (Cys for Cys = 9; Cys for Glu = -4). Along with the Score, BLAST also calculates the “Expect Value” ($E$), a statistical parameter representing the number of equal or better matches that would be detected by chance given the size of the database. Thus, the smaller the $E$ value the greater the similarity of the sequences. The user can adjust the threshold value for $E$.

Mutations occur in the DNA, but selection acts, in the vast majority of cases, on the protein. Thus, coding regions are under higher selection pressure and tend to be more conserved, particularly at the protein level. For this reason, it is possible to compare two sequences at the DNA level and detect no significant similarities, but a significant one when the comparison is performed at the protein level. To facilitate these types of comparisons, the BLASTx version of BLAST allows the comparison of a DNA sequence query against the protein database. DNA sequences are normally presented as single stranded and written from the 5' end to the 3' end; however, there is no information about whether the sequence corresponds to the sense, or the anti-sense strand. For this reason, BLASTx generates conceptual translations of the six possible frames. The first frame divides the DNA sequence in 3-base codons starting with the first base, while the second and third frames start with the second and third bases, respectively. The first codon of the 4th, 5th and 6th frame starts with the 1st, 2nd and 3rd base of the complementary strand starting at its 5' end. The first base of the complementary strand is the one that pairs with the last base of the sequence in the FASTA file. The reverse comparison can be made with the tBLASTn, or a comparison of two DNA sequences at the protein level with tBLASTx. This comparison demands a lot more computer time because it carries out 36 comparisons for each pair of sequences that it analyzes.
RETRIEVAL OF SPECIFIC DNA SEQUENCES

Use Entrez to find and retrieve a specific sequence from the nucleotide database. The retrieved sequence will be used immediately to design primers suitable for PCR amplification. Primers designed by students will be ordered and later used for PCR amplification from genomic DNA. The amplified DNA will be cloned and sequenced at the ICBR DNA Sequencing Core Lab. Clones will be used as probes to detect DNA restriction fragments via Southern hybridizations. In addition, the primers will be used to assess expression levels of the corresponding sequences via reverse transcriptase PCR.

**Using Entrez**

- Direct your browser to the NCBI web page at [http://www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov). This is a valuable resource and students are encouraged to explore the site. However, it will be wise to start with the Training & Tutorials link under the heading NCBI Home on the left panel. After clicking the T & T hyperlink, you will see the NCBI Education Page, the NCBI Handbook, and the NCBI Help Manual. The first has many links to PowerPoint presentations and video tutorials, while the second and third have downloadable PDF files for chapters and manuals. It is also possible to get started exploring the hyperlinks under the “Get Started” subheading.

- We are now going to look for a sequence encoding Actin 2 in *Phaseolus vulgaris*. Go back to the previous page ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). Type the query *Actin2* (no spaces) in the window between the drop down menu window that says All Databases, and the Search button, and click on the Search button. This will take you to the Entrez home page. Notice the numbers to the left of the database names. These numbers indicate the number of matches for the query in each of the databases.

- Click on the Nucleotide database hyperlink appearing under the Genomes subheading.

- The new page shows across the top that the Nucleotide database was searched with the query “actin2." The first line has a drop down menu for Display Settings. The second line indicates that a total of 115 nucleotide sequences have been detected, 43 of them in the Nucleotide collection and these are displayed below in groups of 20. There are also 72 sequences in the EST collection, and these can be displayed by clicking on the hyperlinked number.

  Over to the right side of the screen, under the Filter your results heading, you will see the All (43) link is highlighted. No bacterial sequences were detected. Out of the 43 sequences, 35 were found in GenBank and the INSDC (International Nucleotide Sequence Database Collaboration), 14 of these are mRNA, and there are also 8 records in the RefSeq database.

- The small number of records found in this search is somewhat surprising because a substantial amount of work has been done with the family of actin proteins from different species. To
investigate this apparent anomaly, type Actin 2 (leave space before number) in the query box and hit the Search button. Notice that you now have over 300,000 hits distributed among the three major nucleotide databases, 199,000 in Bacteria, and close to 166,000 in GenBank. When two words are typed next to each other, Entrez automatically inserts the Boolean operator AND between the words, Actin and 2 in this case, and searches and selects all records where both “actin” and “2” appeared somewhere. This explains the relatively high number of hits. These results illustrate the impact of query syntax on the accuracy of results.

- In cases where results are not as expected, it is a good idea to investigate the possible problems by inspecting the details of the search. This can be done by scrolling down to the Search details window on the right hand of your screen. This window will display the order in which the database was searched, and the limits, if any, that may have been used in the search.

- Entrez can also look for sentences, words that appear in a specific order. A sentence can be formed by enclosing the query in double quotation marks. If you search for “Actin 2"", you will see that the hits in GenBank will drop to about 735, and that there are ZERO hits in the Bacteria category. Actin is a eukaryotic protein!

- You can increase the specificity of the search by combining terms in the query using Boolean operators (AND, OR, NOT). For example, search for “Actin 2" AND “Phaseolus vulgaris". This search will yield 14 sequences, 13 in the Nucleotide collection and 1 in the EST collection. Seven records in the Nucleotide collection belong to Arabidopsis thaliana and only one to Phaseolus vulgaris. This is because the search was unrestricted. Some of the Arabidopsis records were for entire chromosomes, and it is very likely that the annotation for some of the genes in the Arabidopsis chromosomes made reference to Phaseolus vulgaris and elsewhere in the record there is a reference to Actin 2.

- The specificity of the search can be further increased by restricting the search using the Advanced or the Limits hyperlinks that appears under the search box. Click on the Advanced hyperlink, and type “Phaseolus vulgaris” in the first box; click on the dropdown menu and change All Fields to Organism. Type “Acting 2” in the second search box and press the SEARCH button. Voilà! The output appears more targeted, doesn’t it?

- Searches in the Advanced window are recorded in a history list you can see under the search boxes. This means that you can do individual searches selecting a particular field for each search, and later combine these individual searches that are added to the list using Boolean operators. You can left click the Search #number and select a Boolean operator. The item will appear in a search box on top, and the chosen Boolean operator will appear in front of the next search box. You can now make a second selection to narrow down your search.

**Download Target Sequence**

Entrez lists records in the Summary format after every search. These consist of a hyperlinked descriptive title that includes the name of the organism, the name of the sequence,
and the type of sequence (genomic DNA, chromosome, or mRNA- full or partial). The format of
the record can be changed with the drop down menu of the **Display Settings** that appears at the
top of the list. Clicking on the title hyperlink opens the record in the GenBank format. This
format shows the title and then the GenBank identifier which is a permanent identifier that may
have one or more versions depending on the number of revisions. The next fields are the
Keywords, Source Organisms, and Name of the Authors and Publication. Finally, you will find
the **Features of the sequence** including the type of sequence, and the sequence itself. The
GenBank format carries a lot of relevant information and can be saved by your browser.
However, most applications require the FASTA format. The most practical way to download the
sequence in this format is to first change the display to FASTA (text) using the **Display Settings**
drop down menu, and then copy and paste the sequence onto a word processor file (MS Word,
WordPad, or Notepad).

**Problem - No Sequence Was Detected.**

What if the sequence is not found? So far, we have only searched the Core Nucleotide
collection. If no sequences are found, then it may be possible to still find the target sequence in
either the EST collection, or the GSS collection. There are several ways in which those
collections can be accessed. One is at the very beginning in the Entrez page, the second is by
clicking on the number hyperlinked in the results page after the first search, and finally, by
selecting the appropriate database from the drop down menu at the top of the search screen. A
word of caution, previous searches, and Limits imposed on the searches do not carry when
changing databases.

There is always a chance that the sequence is not detected via a name search. As
mentioned in the discussion of databases, the EST and GSS collections are large and the entries
are not always fully annotated or have no annotation at all. Thus, an attempt can be made to find
the target sequence through a similarity search using BLAST. For this approach, you can start
with the known sequence from a better characterized organism (Soybean, Arabidopsis, poplar,
rice), and use the sequence to search the EST database for sequences from the target organism.
One last ditch approach is to download similar sequences from various organisms and align them
with a software like CLUSTAL Primers can be designed from conserved regions, and in the most
extreme cases, one can design nested degenerate primers. The last ditch approach is beyond the
scope of this lab.
STRUCTURAL ANALYSIS OF DNA SEQUENCES

Sequences obtained in the previous steps will be used to retrieve the full genomic sequence from the Phytozome database (www.phytozome.net). The *Phaseolus vulgaris* genome has been sequenced (Schmutz et al., 2014), and the genome sequence is archived in the Phytozome database where many annotation features are available. The objective of this section is to retrieve both genomic and CDS sequence, and annotate the genomic sequence indicating the coordinates of each exon identified in the sequence. This task can be performed via alignment of the genomic and transcript sequences using the Specialized BLAST tool.

This information will be used to design primers suitable for the amplification of genomic DNA using the polymerase chain reaction (PCR). PCR primer design is a fully automated task performed with specialized software. However, the user must provide critical information to the software in order to identify the suitable primers. With this objective in mind, let’s examine the source and structural features of the sequences in question.

In general, two types of sequences can be retrieved from the Nucleotide Collection: genomic DNA, and mRNA-derived cDNA. Genomic sequences may contain exons, introns, or both, while mRNA-derived sequences will only contain exons. If a genomic sequence is retrieved, then primers designed from anywhere in the sequences will be suitable for amplification of genomic DNA. However, the same primers may not be suitable for reverse transcriptase-PCR amplification (RT-PCR) if at least one of them resides partially or completely in an intron. The absence of intron sequences in mRNA will render the primer completely useless for the task. Conversely, primers designed with a cDNA sequence may be useful for RT-PCR, but may not be suitable for amplification of genomic DNA if one of the primers spans an exon-exon junction in the cDNA. Primers that anneal to exon-exon junctions will not be able to prime genomic DNA because the corresponding exons in the genomic DNA will be separated by one intron. The challenge here is to design primers that will be suitable for both PCR-amplification of genomic DNA and RT-PCR.

Let’s review some additional considerations. Primer pairs derived each the same exon, or from single different exons each will be useful in RT-PCR and PCR of genomic DNA; however, in the second case primers annealing to different exons would be useful as long as they are located within a reasonable distance. Although much less common in plants than in animals, introns with lengths in excess of 5 Kb can hinder amplification via conventional PCR procedures. If primers are designed from exons flanking a single intron, then the amplification product that uses genomic DNA as template will be bigger than the amplification product derived from RT-PCR. This is because the genomic DNA template will have the intronic sequences, but the template mRNA would have had the introns spliced out during RNA processing after transcription. In fact, the size difference in amplification products can be used as a way to determine whether an RNA preparation is contaminated with DNA. In summary, the key for the successful design of primers that can work in both the amplification of genomic DNA and mRNA will depend on the ability to properly identify exon and intron regions.
The identity of exon and intron sequences, as well as the position of their junctions, can be determined from alignments of the genomic and cDNA sequences. The conservation of exon sequences across taxonomic groups allows the performance of similar alignments with heterologous sequences (tomato genomic DNA vs. pepper cDNA). Thus, while alignments with native sequences can provide strong evidence for the location of exon-intron and exon-exon junctions, alignment with homologous sequences can only be used to make inferences. The **BLAST** tool of Entrez is among the many software tools that can be used for DNA sequence alignment.

**Identification of Exon Coordinates in Genomic Sequences**

The exact mechanism for the recognition of exon-intron junctions is not fully understood, but a few structural features have been identified and appear to be conserved throughout evolution. Namely, a **GU** dinucleotide at the 5' end of the intron, also called the **donor** site, and an **AG** dinucleotide at the 3' **acceptor** site. In addition, plant introns have a polypyrimidine track near the 3' acceptor site and a CURAY (Cytosine, Uridine, puRine, Adenine, pYrimidine) motif near the 3' end. In addition, putative exons are analyzed for three-based periodicity and the presence/absence of stop codons. Several computer programs have been written to scan DNA sequences and recognize exon-intron junctions. Recognition is aided by the correct identification of the structural features mentioned above. However, these features need to be in the proper context because the presence of additional exonic and intronic signals, which have not been fully identified, is required for proper splicing. For this reason, most of these programs are “trained” with a substantial set of cDNA-genomic DNA sequence pairs from a particular species to be able to reliably identify exon-intron junctions in that species. **GeneSeqer** is a Tool of the Plant Genome Database (**PlantGDB**) portal. This tool uses both the **SplicePredictor** program to identify potential Acceptor-Donor pairs and alignments with of genomic DNA with native or homologous EST, cDNA, or protein sequences to generate a plant gene structure model.

**GeneSeqer – Detection of Exon-Intron Junctions**

- Click on the Tools drop-down menu and select GeneSeqer.
- **Step 1**: Select-species-specific splice site model. Select the Medicago model because the query is going to be another legume species.
- **Step 2**: Input genomic DNA sequence. Copy and paste the genomic sequence in FASTA format in the indicated window, but not before replacing the phvul number with the name of your sequence. Click on the radio button for the **FASTA** format. Do not need to type the sequence name if you entered the FASTA format with the sequence name following the > sign. Select the **Original** radio button.
- **Step 3: Select or input EST/cDNA sequences.** It is possible to either enter a particular species, or select entire taxonomic groups. Instead of clicking on one of the four choices at the top of this section (All Plants, All Grasses, ...), begin typing *Phaseolus* in the **Species name** box and a list of *Phaseolus* species will appear in the window below. After clicking on *Phaseolus vulgaris*, a line with the name of the species will appear under the **Organism** window. Paste the transcript sequence in FASTA format in the appropriate window.

- **Step 4: Select Alignment Stringency Level.** Select the **Strict** alignment.

- **Step 5: Submit Job.** Click on the **Submit** button. This will take you to a new screen showing partial output, alignments of the query with mRNA-derived sequences. The introns appear as stretches of unpaired DNA sequence. Notice that these stretches start with a GT and end with AG. The partial output screen will update every minute.

- The Output page has a Navigation panel on the left and two output panels on the right. The upper panel has a color-coded graphical representation of the output. Place the pointer over the **Help** hyperlink to decipher the color codes.

- Save the entire file. You can get to the most important part of this text by clicking on the **Scroll down to “Predicted gene location”** hyperlink. This section has the coordinates of each exon and intron in the sequence.

**References**


Study Questions
1. What happens to an mRNA molecule after it is transcribed from the DNA template?
2. The NCBI Protein database has a vast number of protein records, but many of those are not the product of protein sequencing. Explain why.
3. Are all exons translated into proteins?
4. There are specialized software programs that have the ability to detect introns in open reading frames. What kind of “signatures” do most introns have in common?
5. A total of 64 codons are available. Three of them do not code for specific amino acids, but the remaining 61 code for 20 different amino acids. Explain.
6. Why is it possible to find similarities in the protein sequences of distantly related organisms, but not in the DNA sequences that encode those proteins?
7. What kind of records is contained in the EST database?
8. What kind of records is contained in the GSS database?
9. Many records in GenBank indicate that they are from mRNA, yet the sequences for those records do not contain any uracils. Explain why.
10. What steps are performed by the appropriate BLAST program to compare a DNA sequence to entries in the protein database?
11. You were looking for a malate dehydrogenase sequence from *Hevea brasiliensis*, but you couldn’t find an entry for that sequence in GenBank. However, you know that two groups have generated a vast *H. brasiliensis* collection of ESTs, but these collections are not annotated; that is, EST sequence entries do not have information about their identity. How would you go about in your search?
12. What does the “Expect” value of a BLAST search indicate?
13. You have used a genomic DNA sequence to search the Ref_mRNA database and obtain several hits. Almost all the hits show three interrupted alignments all with high scores and extremely low E-values. What do those results mean?
14. You have conducted a BLASTn search with a sequence from a new plant species found in an exploration trip to Borneo. This search yielded no matches at all. There is a possibility that this sequence is very unique to this newly found species. However, there is the possibility that the sequence encodes a gene that evolves at a much faster rate than other genes. How do you go about finding a match which may give a clue about the identity of the sequence?
15. What is a FASTA file?
Lab Report: Bioinformatics
A specific sequence will be assigned to each group

Part I.
Summary (One Page Maximum. 25 points)
a) The main objective of the experiment,
b) The most important findings,
c) The significance of the results, and
d) Major conclusions

Part II
Results
Summary of the results obtained in the lab. Include the following:
a) (20) The Sequences identified in the search in FASTA format. Include the sequence name in the title line.
b) (25) The *Phaseolus vulgaris* genomic sequence with exon annotation:
   Organism Name:
   Sequence name:
   Sequence type:
   CDS:
   Exon 1
   Intron 1:
   Exon 2:
   Intron 2:
   Sequence

Part III.
1. Find the complete mRNA sequence of the ‘telomerase reverse transcriptase’ (tert) from a plant and from the unicellular organism *Tetrahymena thermophila*.
a) (5) Provide their sequences in FASTA format including the GenBank identifier.
b) (25) Align their DNA and protein sequences using the BLAST tools. Include the alignments in the report and describe the results you observe in those comparisons. To what major conclusion do you arrive?
PCR AMPLIFICATION

Learning Objectives

At the completion of this unit students will be capacitated to design PCR experiments for the amplification of specific nucleic acid sequences. This skill includes the design of suitable primers, setting up an amplification reaction mixture, designing a thermocycler protocol, and evaluating the amplification products.

Lab Objectives

Design a suitable primer pair for the amplification of a genomic sequence identified in the previous lab. Amplify target genomic DNA sequence and evaluate the amplicon via agarose gel electrophoresis.

Background

The Polymerase Chain Reaction (PCR) is an ingenious method devised by Kary Mullis to selectively amplify specific DNA sequences from practically any source of DNA. He developed this technique while he was working for the Cetus Corporation, and was awarded the Nobel Prize in Chemistry in 1993. This method is based on the simple idea that if the sequences flanking a target sequence are known, then one can design and synthesize short complementary DNA oligomers which can be used in combination with a DNA polymerase, and the four nucleotides (A, C, G and T) to synthesize DNA in vitro. The implementation of this idea requires the means by which the temperature of the reaction can be rapidly changed and controlled and with precision. This can be accomplished with a set of water baths, or with a programmable thermocycler. This instrument is used to take the reaction through its three basic temperature steps:

a) Denaturation. Separation of DNA strands at high temperature (94-95°C)
b) Annealing. Base-pairing between the oligo primers and the priming sites flanking the target sequence. It takes place at temperatures ranging from 35 to 72°C. The annealing temperature depends on the thermodynamic properties of the primers which are dictated by the base composition and the length of the annealing sequence.
c) Extension. This is the step where the DNA polymerase synthesizes new DNA. The enzyme adds nucleotides to the 3' end of the primers. A thermostable polymerase with an optimum temperature of 72°C carries out this step.

An exponential amplification of the target sequences can be attained if the three step cycle is repeated multiple times. During the development of PCR, fresh DNA polymerase from
a mesothermic organism (*Escherichia coli*) had to be added after every cycle because the enzyme was denatured during the DNA denaturation step. This problem was overcome by the adoption of DNA polymerases from thermophilic organisms like *Thermophilus aquaticus*.

The *E. coli* DNA polymerase I was the first polymerase to be characterized; other prokaryotic enzymes were later found to have several functional similarities to Pol I. In general terms, DNA polymerases require a template, a primer annealed to the template, a 3'OH end on the primer, the nucleotides to be added to the new DNA chain, and Mg\(^{2+}\) ions. All polymerases synthesize DNA in the 5' to 3' direction. DNA Pol I also has a 3' to 5' exonuclease activity (proofreading activity), and a 5' to 3' exonuclease activity all in the same peptide. The Danish biochemist Hans Klenow discovered that the 5' to 3' exonuclease activity resided near the amino terminus of the protein, and that this fragment could be removed proteolytically; the remaining peptide is called the Klenow fragment in his honor, and it is widely used in many protocols. It has also been determined that the remaining polymerase and exonuclease activities reside in different domains. In fact, there are mutants that lack the 3' to 5' exonuclease activity and are more error prone than the wild type enzyme. Taq DNA polymerase, one of the most widely used thermostable polymerases, is one with the lowest fidelities because it lacks the 3' to 5' proofreading exonuclease activity.

An interesting characteristic of Taq DNA polymerase is its ability to add an unpaired adenine residue at the 3' end of the newly synthesized strand. Other nucleotides can be added too, but at a much lower frequency. This is an important characteristic to consider before cloning PCR products. The common expectation is that PCR amplification products are blunt ended and therefore amenable to cloning into vectors opened with a restriction enzyme that leaves blunt ends (e.g.: EcoRV). However, ligations of PCR products amplified with Taq DNA polymerase to blunt end vectors is extremely low or nil. The reason of course is that the amplification product is not blunt ended. Two approaches have been used to overcome this cloning problem. One of these is the T/A cloning strategy which requires the generation of plasmid vector with an unpaired T at the 3' end, a feat that is accomplished by incubating the blunt ended vector with Taq DNA Pol and dTTP alone at 72°C. The plasmid’s unpaired T could pair up with the 3' A overhang in the amplicon. The second strategy is to use Pfu DNA Polymerase instead, because this enzyme has a 3' to 5' exonuclease activity and leaves blunt ends in the amplification products. This strategy will be used in the next cloning section.

The outcome of a PCR amplification reaction depends on the characteristics of the reagents and the thermocycling program. An examination of these factors is presented below, and these should be considered when designing a PCR experiment. PCR is a very useful technique, but mishandling of the components and the thermocycling program can easily lead to a number of PCR artifacts, or to complete failure.
PCR AMPLIFICATION REAGENTS:

PCR Buffer: This buffer is usually supplied as a 10 X. The buffer used for Taq DNA polymerase usually contains Tris.HCl, pH 8.3, and 50 mM KCl. The Pfu polymerase buffer is similar, but has a relatively higher optimum pH (8.8).

Magnesium Chloride. Mg is a required co-factor of DNA polymerases. The standard concentration is 1.5 mM, but certain amplification reactions require a different Mg\(^{2+}\) concentration. High Mg\(^{2+}\) concentrations can decrease the specificity of the amplification. Some commercial PCR Reaction Buffers already include MgCl2.

Deoxynucleotides (dATP, dCTP, dGTP and dTTP). It is essential to use nucleotides of the highest quality. They are normally used at concentrations between 100 to 400 \(\mu\)M. They are usually prepared as a mixture of the four bases at 10X of the desired concentration. Nucleotides decompose at a relatively high rate and should be stored at -20\(^{\circ}\) C, and kept on ice during preparation of PCR reactions. It is also a good idea to prepare stock solutions and dispense them in small aliquots to avoid repeated freeze/thaw cycles.

Primers. Primer specificity is the most important issue in PCR. Specificity depends on several interrelated factors that include the physical chemical characteristics of the primer itself, the genome size, and the physical characteristics in which the primer will anneal to the template. Among the first characteristics of the primer to be considered is its length which has to be considered in the context of genome size. The longer the primer is, the fewer the potential annealing sites there are for a given genome size. Thus, for a given primer length, the larger the genome, the more annealing sites there will be. To calculate the frequency of annealing sites for a primer, we can use the same approach as the one used for calculating the frequency of restriction sites. If the frequency of a single base is \((1/4)\), and the frequency of an \(n\)-base sequence is \((1/4)^n\), then the frequency of the \(n\)-base sequence in a genome of size \(G\) is \(4^n/G\). Try to calculate how many annealing sites would be expected for a 10-mer (ten-base oligomer), a 15-mer, and a 20-mer in the Arabidopsis (140 Mbp) and onion (15,290 Mbp) genomes.

Another important characteristics of oligo primers is the annealing temperature (Tm), also referred to as the melting temperature. This term is derived from reassociation kinetics studies of nucleic acids. It marks the temperature for the midpoint of the reassociation curve. Several methods have been used to calculate the Tm. The most rudimentary method is the empiric method based on the Watson and Crick DNA model in which 2\(^{\circ}\) C are added for every A:T pair and 4\(^{\circ}\) C for every G:C pair. To increase the accuracy of this simplistic calculations researches adopted the empirical formula used for DNA:DNA hybridization studies. This formula takes into account the molarity of monovalent cations ([Na\(^{+}\)] and [K\(^{+}\)], the relative GC content, and the length of the annealing fragment (N): Tm = 81.5\(^{\circ}\)C + 16.6\(^{\circ}\)C (log ([Na\(^{+}\] + [K\(^{+}\]]) + 0.41\(^{\circ}\)C (%GC) – 675/N

However, thermodynamic analysis can provide more accurate estimates of the Tm. An
analysis of the annealing behavior of oligonucleotides in thermodynamic terms has revealed the existence of three domains, and has also addressed the stability of each of them (Owczarzy et al., 1997; Santa-Lucia et al., 1996). The principal domain corresponds to the entire length of the oligonucleotide, and its stability is given by the length of the sequence and the base composition. The second domain is the hexamer at the 3' end. The reason for focusing on a hexamer is that DNA polymerases (see below) need only six base pairs to bind DNA before proceeding with extension. This is perhaps the most important domain determining specificity and yield of the amplification reaction. A highly stable 3'-end is counterproductive because it can lead to non-specific annealing and extension activities. Primer annealing is a dynamic process in which segments of the primers partially anneal to non-specific sites of the complex DNA template (genomic DNA) for very brief periods of time. In contrast, annealing to the specific site is stable. If the six contiguous bases (hexamer) at the 3'-end form a stable duplex at a non-specific site, then there may be enough time for the polymerase to bind and extend at that site, turning the non-specific site into a competitor of the target site. This is why it is advisable to design primers with a stable 5'-end domain, the third domain of the oligo.

Thermodynamic analysis has also revealed that a nucleotide can play a substantial role in the annealing behavior of the nearest-neighbor nucleotide. To illustrate this point a table has been constructed for the free energy (ΔG) of transition of dinucleotides between their annealed and free states.

**Table 1.** Free Energy (ΔG) values for contiguous dinucleotides (nearest-neighbors) measured at 25°C.

<table>
<thead>
<tr>
<th>1st dN</th>
<th>2nd dN</th>
<th>dA</th>
<th>dC</th>
<th>dG</th>
<th>dT</th>
</tr>
</thead>
<tbody>
<tr>
<td>dA</td>
<td></td>
<td>-1.9</td>
<td>-1.3</td>
<td>-1.6</td>
<td>-1.5</td>
</tr>
<tr>
<td>dC</td>
<td></td>
<td>-1.9</td>
<td>-3.1</td>
<td>-3.6</td>
<td>-1.6</td>
</tr>
<tr>
<td>dG</td>
<td></td>
<td>-1.6</td>
<td>-3.1</td>
<td>-3.1</td>
<td>-1.3</td>
</tr>
<tr>
<td>dT</td>
<td></td>
<td>-1.0</td>
<td>-1.6</td>
<td>-1.9</td>
<td>-1.9</td>
</tr>
</tbody>
</table>

The table above can be used to calculate the free energy of transition of hexameric domains in the 20-mer 5'-GGCGGAAACAAGGGTCACAA-3'. Starting from the 5'-end we can take adjacent dinucleotides and calculate the free energy of transition for the first hexamer: GGCGGA. The free energy of transition of this hexamer is -15 Kcal/mol. This calculation can be repeated for all fifteen hexamers of the oligonucleotide as seen in Figure 1. Each mark in the plot corresponds to the free energy of a hexamer (each line at the bottom). This
plot indicates that the most stable domain is located at the 5'-end, and the second most stable domain is in the center. It can also be seen that the 3'-end has a moderate level of stability. The negative value of the $\Delta G$ term indicates the amount of energy that needs to be added to the system to disrupt the annealing. The free energy of transition for the entire oligomer is calculated by adding the free energies of all 19 overlapping dinucleotides.

Implicit from the previous considerations are the formation of secondary structure and/or primer-dimers. Secondary structures in the form of hairpin loops occur when one segment of an oligo is complementary to another segment of the same oligo. This structure will interfere with annealing to the template. Primer dimers form when the 3' end of one primer is complementary to the 3' end of the second primer. During PCR amplification these primers will anneal to each other forming dimers. If annealing is 100%, then the major effect will be a highly unbalanced competition with the template targets because primer concentration is always several order of magnitudes over those of the template, and this will inhibit amplification. On the other hand, if the primers show only overlapping complementarity, then there will be unbalanced completion as described, but in addition, the polymerase will extend the non-overlapping regions producing an enormous amounts of very small ds DNA.

There are several on-line freeware, and commercial computer programs for primer design. These programs implement, to different degrees, the principles described above. They do not necessarily use the same thermodynamic table. In this lab we are going to use Primer3, a program designed at the Whitehead Institute (Rozen & Skaletsky, 2000; Untergrasser et al., 2012). This software will allow you to set the limits for several parameters and also the stretch of sequence to be amplified.

One final point about primers is their concentration in the reaction mixture. Primers are normally used at a concentration of 100 nM, and stock solutions are usually prepared at a concentration of 1 $\mu$M or 10X. The molecular weight of the primers can be calculated by adding the products of the number of the nucleotide molecules by its molecular weight as...
indicated in the formula shown below. However, oligos are normally synthesized without the 5' phosphate group. For this reason, the weight of the phosphate group is subtracted from the sum of the products. The missing 5' phosphate in the synthesized oligo will be of particular concern when considering cloning PCR products by ligation.

\[
MW = (329.2 \times \# G's) + (313.2 \times \# A's) + (304.2 \times \# T's) + (289.2 \times \# C's) - 78
\]

To gain an insight into the amplification reaction, students are encouraged to calculate the number of primer molecules in a 20 μl PCR-reaction. This number should be compared to the number of template molecules in the reaction mixture.

**Additives.** Bovine serum albumin, or BSA, has been used as an adjuvant in PCR reactions at concentrations of up to 400 ng/μl. This protein appears to work as a scavenger of several inhibitors including detergents (SDS & Triton X-100) and phenolic compounds. Gelatin at a concentration of 0.01% has also been used in PCR reactions as a stabilizer of Taq DNA polymerase. Manufacturers offer “Molecular Biology” grade preparations for these proteins, which mainly means they are free of contaminating nucleases.

Ammonium sulphate, glycerol, DMSO, and betaine are sometimes added to the amplification reaction to increase specificity and/or yield. These compounds can interfere with the formation of hydrogen bonds. They are added at concentrations that have a marginal effect on the target, but a substantial effect on interfering structures.

For instance, preventing the formation of secondary structure of single stranded DNA during amplification, or weakening unspecific annealing of the primer. The first effect may increase yield and the second may increase specificity.

**DNA Template.** The quality and quantity of the DNA template will affect the specificity and yield of the PCR reaction. As described in Chapter 2, DNA quality has two main components: purity and length. As far as PCR is concerned, there are two purity issues that are extremely relevant: the presence of extraneous DNA from unknown origin, and the presence of chemicals that can interfere with the amplification reaction. Contaminating DNA can yield misleading results and should be avoided at all costs. This type of contamination can occur during sample collection, extraction, and/or at the time when PCR reactions are set up. Micropipets can create a mist when pipetting in a sample. Part of the mist can accumulate in the shaft of the pipet and become the source of contamination in subsequent reactions. This problem can be eliminated with pipet tips that have aerosol barriers. The second type of contaminants in the sample comprises those that have an inhibitory effect in the amplification reaction. Some of these contaminants may be the results of poor extraction procedures (e.g., polysaccharides). The effect of this type of contaminants can be overcome sometimes by either increasing the amount of enzyme, or by diluting the sample which has the effect of diluting the contaminant. Other contaminants can have an origin in the lab and be the result of inadvertent additions to the sample. For instance, a relatively high concentration of EDTA in the DNA sample can reduce the
effective concentration of Mg in the reaction, a cation that is an essential co-factor of the DNA polymerase.

Size, the second component of DNA quality, is not a major concern in most PCR applications because the amplification targets are usually relatively small segments. However, size is extremely important in long range PCR where amplification targets range between 10 and 20 Kb.

The amount of template is also an important factor in yield and specificity. For routine application it is recommended that 30,000 to 40,000 copies of the template be included in a 20 μl reaction. This is one instance in which it is very useful to know the size of the genome. Take the sizes of the Arabidopsis (128 Mb) and the maize (3000 Kb) genomes and calculate the amount of DNA that is needed for a PCR reaction of these species. Also to consider at this point is the amount of template needed when the sample is a 3Kb plasmid carrying a 1 Kb insert. Compare the number of template copies with the number of primer molecules in the reaction and determine the molar ratio. If you are in a calculating mood, then try to calculate the total amount of DNA (in ng) that can be synthesized with the nucleotides added to the reaction mix in a 20 μl reaction.

Water. This is normally added to round up the final volume of the reaction. Water should be sterile and of the highest quality.

DNA polymerase. Taq DNA polymerase is the most commonly used enzyme for PCR. This enzyme is extracted from Thermus aquaticus, an obligate thermophilic bacterium found in the thermal pools of Yellowstone National Park where it grows at temperatures just below the boiling point of water. The enzyme is stable at 95°C. DNA polymerases with similar properties have been isolated from other thermophilic bacterial species such as T. thermophilus (Tth polymerase), T. flavus (Tfl polymerase), Thermococcus litoralis (Tli polymerase or Vent polymerase), Pyrococcus furiosus, (Pfu polymerase), and from various Thermus and Pyrococcus species with polymerases that are sold as Deep Vent polymerase. The standard PCR reaction contains 0.5 units of Taq polymerase in a 20 μl reaction.

Three characteristics of these enzymes are relevant for PCR amplification:

**Processivity**: This is the number of nucleotides that can be extended in one catalytic reaction by one DNA polymerase molecule. Taq has low processivity as compared to Pfu polymerase. For this reason, typical amplifications with Taq DNA polymerase are limited to targets between 3 to 5 Kb, whereas Pfu polymerase is used for Long-Range PCR because it can process longer targets (up to 20-40 Kb targets). Processivity is sometimes confused with the rate of nucleotide incorporation which is higher in Taq (~3500 bp per minute) than in Pfu (~700 bp/min). It is generally recommended that in amplifications with Taq polymerase the extension time should be programmed to 1 min./Kb of target, and 2 min./Kb when using Pfu. Commercial preparations for Long-Range PCR contain mixtures of enzymes to complement rate and processivity.
**Fidelity**: This is the accuracy with which the enzyme incorporates bases that are complementary to those of the template DNA. *Taq* polymerase has the lowest fidelity among the thermostable polymerases with about $8 \times 10^{-6}$ mutation/bp/duplication, whereas *Pfu* has only $1.3 \times 10^{-6}$ and new *Pfu* mutants are even lower. The low fidelity of *Thermus* DNA polymerases is mainly due their lack of 3' to 5' exonuclease activity. This feature also contributes to the propensity of adding an unpaired nucleotide at the 3' and of the amplicon. This issue is a great relevance when attempting to clone PCR products.

**Thermal stability.** The half life of *Taq* polymerase at 95°C is 90 minutes, whereas that of Deep Vent Polymerases is around one day. The last two properties make *Pfu* and other Vent Polymerases more suitable for “Long-Range PCR” because extension times are much longer than those used in conventional PCR.

**Thermocycling Program.** There are usually four sections in the program.

1. **Initial Denaturation.** Carried out for 2 min at 92-95°C. In this step the reaction is heated to the point where the hydrogen bonds of the sample’s double helix are broken. This results in the separation of the DNA strands. In some PCR kits the polymerase is complexed with an antibody to prevent the polymerase from accidental and artifactual extension. To completely denature the antibody and free up the polymerase, a denaturation step of up to 15 minutes is recommend. The use of the Ab-Taq complex is called “hot start.”

2. **Multiple Amplification Cycles** (25 to 35). Usually carried out in three steps; however, in cases where the annealing temperature is close to or coincides with the “extension” temperature, the PCR can be turned into a 2-step PCR (Denaturation and annealing-extension):
   a) **Denaturation** step at 95°C for 30 seconds.
   b) **Annealing** is the second step in the cycle where the sample is cooled down to a temperature where the primers can anneal to their cognate sites on the template. Sometimes, this temperature has to be established empirically, a process that is facilitated by thermocyclers in which temperature gradients can be programmed.
   c) **Primer Extension** is normally set at 72°C, the optimum temperature for *Taq* and other DNA polymerases. The length of this step depends on the size of the target sequence. As a rule of thumb, 1 min per Kb is normally used for *Taq* and 2 min/Kb for *Pfu*.

3. **Finishing Step.** This is carried out to ensure that all amplicons have been fully finished. Not all protocols include this step, which is basically a 5 to 10’ extension.

4. **Final Holding Step.** Thermocycler is programmed to hold the samples at 4 to 15°C for an indeterminate length of time. This phase ends when the samples are removed from the thermocycler. It is advisable to set the temperature at 15°C in places of high humidity because at lower temperatures an excess amount of water can condense around the
thermocycler block, and this may be detrimental to the instrument. DNA is very stable at this temperature.

PCR amplification is expected to proceed in an exponential manner. For example, consider the amplification of 500 bp target from a genome that is 1000 Mbp long. A 25 ng DNA sample of this genome would be expected to yield $429 \mu g$ of amplification product after 35 cycles. In reality, PCR yields are two to three orders of magnitude lower than the expected amounts. PCR amplification profiles approximate a sigmoidal curve with a final plateau. The reason for this plateau is not the exhaustion of reagents, but a feedback inhibition caused by increased product competition. In the later stages of PCR a relatively large number of amplification products are present in the reaction volume. During the annealing step two types of duplexes are formed: a) primer:single-stranded-amplicon, and b) single-stranded-amplicon:single-stranded-amplicon. The second becomes a competitive inhibitor for the first leading to a plateau in the amplification profile.

**Troubleshooting**

Agarose gel electrophoresis is the preferred method used for evaluating the outcome of a PCR reaction. There are several possible outcomes. The best one of course is when the amplicon of the expected size is obtained in abundant quantities. However, let’s now consider some undesirable outcomes, the most likely explanations for their occurrence, and possible. This is by no means an exhaustive treatment of the subject.

- PCR amplicon is of an unexpected size. If the primers are exactly as selected, then one possibility to consider is the presence of an insertion/deletion (indel) between the sequence used for primer design and the sequence that is used as a template for amplification. An indel can be identified by including a reaction with the genotype used for primer design.

- Several amplicons are detected. There could be many reasons for this outcome. One occurs when the amplification target belongs to a multigene family or repetitive DNA, and the primers were designed in conserved regions. It is possible that the annealing sites for at least one of the primers are not all identical. For instance, on some occasions, multiple bands of low intensity could be observed, but all except one disappear when the temperature is increased to a discriminating point. If the outcome doesn’t change, then new primers need to be designed. Another possibility could be due to the amplification of “inverted repeats.” In this case, amplification could be directed by a single primer that anneals to inverted repeats present within short distance of each other. This case can be tested by running PCR reactions with a single primer at a time, along with the two primer amplification.

- There is a smear in the gel. This pattern could be due to the amplification of a truly repetitive DNA family. However, assuming that the primers were designed from a single copy sequence, smears are commonly caused by indiscriminate amplification, which in most cases is likely due to an excess of one of the reagents in the mix (template, dNTPs, Mg, DNA polymerase). Check
the concentration of the PCR components, and the temperature settings of the thermocycler program. A lower than desired annealing temperature may be the cause too.

- No amplification at all. This is the most challenging problem because any factor, including the thermocycler program could be at fault. Trouble shooting this problem requires carrying out a number of tests with discipline and organization. The first test should be ensuring the reliability of the sequence used for primer design -- the quality of the primers is as good as the quality of the sequence that was used for their design. However, if the amplification has already been successful in previous runs, then the template, reagents and thermocycler program should be examined. Test the quality of the template by carrying out a reaction with a template that has previously worked. If the reagents are suspect, then a set of reactions should be set up in which only one component at a time is replaced with a new one. Often times the culprit is the dNTP stock. Multiple freeze/thaw cycles result in their excessive degradation. This problem can be averted by minimizing the number of freeze/thaw cycles, by dispensing the mix in small aliquots.

In summary, the most effective way to troubleshoot PCR problems is to systematically test each component of the PCR amplification reaction. Polymerases are really stable and are usually not the problem, but the other components might be. Magnesium and buffer salts are usually kept at -20° C. Sometimes not all the salts go into solution upon thawing, and this may be the cause of the problem. Lack of, or poor amplification can result from using a template with a relatively high EDTA content. Excess EDTA will chelate Mg making it unavailable as a cofactor for the polymerase. Assuming that everything is in place at the correct concentration, the lack of amplification can be corrected by simply increasing the annealing temperature. The explanation is that perhaps at least one of the primers partially anneal with an excess of non-specific sites creating a competitive inhibition; increasing the annealing temperature will eliminate the inhibition by eliminating the excess partial annealing.

References


Saiki RK., DH Gelfand, S Stoffel, R Higuchi, G Horn, KB Mullis, HA Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science


PCR-PRIMER DESIGN

Primer3 will be used for primers design. There are three versions of this program as shown below. This program has been adopted by NCBI.

Primer3  http://frodo.wi.mit.edu/primer3/
Primer3Plus  http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi
Primer3web  http://bioinfo.ut.ee/primer3/

Primer3 has several input parameters that can be controlled to select suitable PCR primers from a particular sequence. The web page has three sections: the first section has a box where the sequence can be pasted. The second section has subdivision for the left and right primers, and a hybridization primer. Each subdivision has a small box that can be clicked on to add a check mark requesting the program to select and design that primer. In addition, primer sequences can be pasted in the boxes below so the program can test whether they are suitable as PCR primers. Finally, the third section has a number of sub-sections that allow the user to set a series of parameters under which Primer3 will select and design primers.

- **Sequence ID:** If the sequence was entered in FASTA format, then there is no need to fill this box. However, if there is a name associated with the FASTA format, and a name is also entered in this box, then the program will include a warning in the output to the effect that two names were provided for this sequence.

- **Targets:** There are two ways in which the target of amplification can be selected. First, enter the position of the first base followed immediately by a comma and the number of bases in the target. For example, entering “151,200” will indicate that the target for amplification is from base 151 to base 350). Alternatively, square brackets can be used in the sequence to enclose the [target]. Primers will be selected from the regions flanking the target area.

- **Excluded regions:** Multiple regions from which primers should not be selected can be listed in this box. The notation is similar to that used for Targets. Let’s consider an example of a 1000-base mRNA sequence where the first exon encompasses bases 1 to 400, the second 401 to 700, and the third 701 to 1000. The objective is to design primers for genomic amplification of the second intron, but significant sections of the flanking exons should also be included. One could enter in the Excluded Regions box “1,400 601,200”. This entry will indicate that no primers should be considered within the first 400 bases, and for the region spanning from base 601 to base 800. Alternatively these regions can be enclosed by the “smaller-than” and “greater-than” signs (<ACG…CTGG>).

- **Product Size Ranges:** Primer3 will first look for primers for the first size range in the box, if it can’t find one in that range, then it will move on to the next one. The first range in the default is 150-250. Taking into account the excluded region specifications entered above and the default Product Size Range, Primer3 could generate primers in the region between base 400 and 600 or...
between bases 800 and 1000 in the example considered above. To avoid this possibility, the Product Size Range can be altered to suit the user’s objectives. For instance, entering “210-600” can exclude the possibility of designing primers for amplification within the second or third exons alone. Notice that the minimum of 210 exceeds the span of the exclusion zone between bases 600 and 800. And the upper limit encompasses exons 2 and 3. Remember that in this example the primers are designed with a mRNA sequence, but the intended template is genomic DNA. Thus, the amplification product will be bigger than the one calculated by Primer3. How much bigger? This is practically impossible to predict. Although plant introns are generally not as big as animal introns, some can extend from 2 to over 5 Kb. Thus, it is within the realm of possibilities to consider an intron that is so large that it would hinder the amplification of the target genomic sequence by conventional PCR. Several strategies can be implemented to select primers at desirable regions.

Several conditions can be adjusted by the user. Among these are primer size, primer Tm, product Tm, and maximum self-complementarity in the primers. Each one of the listed factors is hyperlinked to a page containing definitions and explanations of the terms. After selection of the desired restrictions, press the Pick Primers button to select the primers. The program will bring up a new page with the results. A table with the sequence of the forward and reverse primer including all their physical properties and the locations in the sequence, the size of the sequence that was analyzed, and the size of the expected amplification product. In addition, the entire sequence will be printed indicating the physical location of the primers, the target sequence, and any excluded regions. Finally, the output includes a set of additional primer pairs that meet the pre-set requirements.
MATERIALS AND SUPPLIES

A. REAGENTS

DNA template
Pfu DNA polymerase
10X Pfu buffer
dATP
dCTP
dGTP
dTTP
Primers

B. SOLUTIONS

dNTPs, 2 mM
  dATP, 100 mM   10 μl
  dCTP, 100 mM   10 μl
  dGTP, 100 mM   10 μl
  dTTP, 100 mM   10 μl
  Water         460 μl

Electrophoresis Buffer

Molten Agarose

C. EQUIPMENT

Thermocycler
PCR AMPLIFICATION

1. Program the thermocycler as follows:

<table>
<thead>
<tr>
<th>Step</th>
<th>Temp</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial denaturation</td>
<td>95°C</td>
<td>45 sec</td>
</tr>
<tr>
<td>Amplification, 35 cycles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denaturation</td>
<td>95°C</td>
<td>45 sec</td>
</tr>
<tr>
<td>Annealing</td>
<td>60°C</td>
<td>45 sec</td>
</tr>
<tr>
<td>Extension</td>
<td>72°C</td>
<td>2 min</td>
</tr>
<tr>
<td>Finishing</td>
<td>72°C</td>
<td>5 min</td>
</tr>
<tr>
<td>Hold</td>
<td>15°C</td>
<td>Forever</td>
</tr>
</tbody>
</table>

2. Allow the thermocycler to warm up for at least 15 min before you run your samples.

   NOTE: This can take place while you set up your reactions.

3. Set up a PCR Reaction in a 200 μl PCR tube by adding the following reagents in order:

   - PCR Buffer, 10X: 5 μl
   - dNTPs, 2 mM: 5 μl
   - H2O: 27 μl
   - Pfu DNA Pol. 0.25 u/μl: 1 μl
   - F Primer, 1 μM: 5 μl
   - R Primer, 1 μM: 5 μl
   - Template DNA 20 ng/μl: 2 μl
   - TOTAL vol: 50 μl

   **NOTE 1:** Keep all reagents and the reaction mix on ice.

   **NOTE 2:** A cocktail will be made with all the reagents, except primers and template. The total volume of the cocktail will be 38 μl.

   **NOTE 3:** Once everything has been added to the tube, spin the tube for a few seconds to bring down all droplets to the bottom of the tube. Mix and spin again. Place tubes in the thermocycler, close the lid and run the sample. Make sure the lid of the thermocycler’s heater is on to avoid condensation of water at the top of the tube.

4. Analyze the PCR product by agarose gel electrophoresis. Pipet a 5 μl aliquot of the PCR reaction mix into a new tube. Add 20 μl of 1.25 X loading buffer, mix by vortexing and spin tube for a few seconds to bring liquid down to the bottom. Load 12.5 μl in the gel.

   **NOTE:** Make sure to include a molecular weight standard when running the agarose gel. This will be needed to determine whether the amplification product is of the
expected size and to “guesstimate” the concentration of the amplicon.
5. If the amplicon is of the expected size, and the total amount is estimated to be in excess of 100 ng, then proceed with the TopoCloning step.
Study Questions
1. What is the basis of PCR amplification?
2. What is the function of the denaturation step?
3. How do you determine the temperature of the annealing step?
4. What factors determine the duration of the extension step?
5. Why would it be extremely difficult to prime an amplification reaction with primers shorter than six nucleotides?
6. Could you carry out a PCR amplification reaction with E. coli DNA Polymerase I? If possible, what would you do differently?
7. Why does Taq DNA polymerase have a higher error rate than Pfu polymerase??
8. Long range PCR is used for the amplification of long stretches of DNA, 10 to 20 kb although up to 40 Kb have been reported. What characteristic of Pfu makes it a key component of long-range PCR amplification reactions?
9. What characteristic of Taq DNA Polymerase makes it unsuitable for the production of amplicons to be cloned through blunt-end ligations?
10. Name one factor that affects primer specificity. Explain.
11. What are the general characteristics of a good PCR primer?
12. Why do PCR amplification products reach a plateau? This appears to occur in spite of the fact that PCR amplification is supposed to increase amplicons in an exponential fashion. Explain.
13. You downloaded a sequence from an EST collection residing in GenBank and used this sequence to design primers for amplification. However, several attempts to amplify the target sequences have failed. Indicate the possible reasons for failure (there is more than one) and how would you go about addressing those reasons.
14. You have designed primers from a high quality sequence, and these primers yield two amplification products differing in size by 350 bases. Upon cloning, sequencing, and BLASTing the sequences of these amplicons you have learned that both have a match to one gene. Give a possible explanation for these results.
15. Give the most likely explanation for a situation in which the amplification product is larger than expected.
Lab Report: PCR amplification.

Part I.
Summary (One Page Maximum. 20 points)

a) The main objective of the experiment,
b) The most important findings,
c) The significance of the results, and
d) Major conclusions

Part II
Results (20 points)
Output for primer design
Clearly explain how the sequence was presented to the primer selection program.

Part III.
1. (15) Consider the genomes listed on the table below and calculate the frequency of a 10-base, 15-base and a 20 base oligo. How is this information useful in primer design?

<table>
<thead>
<tr>
<th>Species</th>
<th>C-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Prunus persica</em>, peach</td>
<td>265 Mbp</td>
</tr>
<tr>
<td><em>Solanum lycopersicum</em>, tomato</td>
<td>907 Mbp</td>
</tr>
<tr>
<td><em>Allium ampeloprasum</em>, garlic</td>
<td>24,255 Mbp</td>
</tr>
</tbody>
</table>

2. Find the genomic sequence of the small sub-unit of NADH:Cytochrome b5 reductase 1 from *Arabidopsis thaliana* and provide the following:

a) (5) The GenBank sequence identifier and the sequence in FASTA format.

NOTE: To get full credit, include the entire sequence when designing the primers; this way the nucleotide positions will be consistent from one section to the next.

b) (10) A primer pair suitable for the amplification of the first intron and some flanking exonic sequences (>50 nucleotides). Show the criteria you used for primer selection.

c) (15) A set of primers for the amplification of the fifth and sixth introns with the inclusion of some exonic sequences from exons five and seven. Include a graphical representation of how primers were selected.

d) (15) What will be the size of the amplicons generated with primers designed in “b” and “c” when the template is genomic DNA, and when the template is a full length cDNA clone.
Chapter 5

CLONING

Learning Objectives ....................... 5.1
Lab Objectives ................................ 5.1
Background ................................... 5.1
References ................................... 5.7
Extraction & Analysis of pDNA .... 5.8
Materials and Supplies ................. 5.10
Cloning PCR amplicons ............... 5.14
Qiagen Miniprep Procedure .......... 5.16
Lab Report ................................... 5.17
Study Questions ............................. 5.18
CLONING IN PLASMID VECTORS

Learning Objectives

After completion of this unit students will be able to design experiments aimed at cloning PCR amplified DNA using plasmid vectors. More specifically, this task will entail the selection of the appropriate plasmid vector for a given cloning target, and the colony screening procedure best suited for the selected plasmid. Finally students will be able to isolate plasmid DNA from bacterial cultures.

Lab Objectives

Clone PCR fragment obtained in a previous lab session, screen for bacterial colonies that are likely to contain recombinant DNA, confirm insertions, and purify recombinant plasmid DNA suitable for sequencing.

Background

Plasmids are extra-chromosomal self-replicating DNA molecules found in all bacterial species. This ability to replicate has turned them into the workhorse vector molecular biologists use for the propagation and manipulation of DNA sequences. Plasmids are also used for the construction and delivery of transgenes, and the production of recombinant proteins. Plasmid vectors have been constructed by removing unnecessary sequences from native plasmids, while keeping those sequences which are essential for replication. In addition, engineering of plasmid cloning vectors also entailed manipulation of genes involved in copy number control in bacterial cells, and incorporation of selectable-marker(s) to maintain the plasmid in the host, and a cloning site for the creation of recombinant molecules. Protocols have been devised for the introduction of plasmids into bacterial cells (transformation), and for their extraction and purification.

The most important plasmid component is the replicator unit, which contains the origin of replication (ori) and elements involved in the control of initiation of replication. The ori sequence marks the start of replication recruiting the replication machinery, and neighboring sequences encoding primer and countertranscript RNA regulate copy number and compatibility with other plasmids. pBR322 was one of the first plasmid vectors designed for cloning. This vector was known as the pMB1 replicon with a copy number between 15 and 20 plasmids per cell. A mutation in this replicon “relaxed” the plasmid and allowed cells to accumulate about 500 to 700 copies. This mutation has been incorporated in many other plasmids like pUC18 and pBlueScript. An example of low copy-number is the F factor of Escherichia coli, which is under strict control (1 – 2 copies /cell). The bacterial artificial chromosome (BAC) vector was
developed from the $F$ factor; this vector can carry inserts as large as 300 Kb and it is used for the construction of genomic libraries (Shizuya et al., 1992).

Plasmids that have the same ori sequence belong to the same incompatibility group and can’t co-exist in the same bacterial cell. Of interest in plant molecular biology are the binary vectors used for Agrobacterium-mediated transformation. These plasmids carry two ori sequences. These sequences allow these plasmids to be propagated in $E. coli$ before they are introduced into Agrobacterium for delivery to a plant. They also carry an antibiotic resistance gene that is expressed in bacterial cells, and a different antibiotic resistance gene that is expressed in the transgenic plant.

A variety of selectable markers have been engineered into plasmid vectors to ensure that once they are introduced into $E. coli$ by transformation, they can be maintained continuously. These markers usually encode antibiotic resistance. Addition of the appropriate antibiotic to the media guarantees that only cells carrying the plasmid will grow. Immediately after transformation however, the cells are usually incubated for a short period of time in liquid medium in the absence of antibiotic. This incubation period gives the transformed cells time to express the antibiotic-resistance gene before they are spread on selective (+ antibiotic) plates.

Plasmid vectors have been engineered with a multiple cloning site (MCS), also called the polylinker region, to facilitate the introduction of foreign DNA in the form of restriction fragments. The MCS consists of a series of unique restriction sites in the plasmid. The uniqueness of these sites allows the restriction enzyme to cut open the plasmid generating a single linear molecule. The termini of the plasmid can then be chemically attached to a foreign DNA fragment (the clone) with compatible ends, generated by the same or compatible restriction enzyme (Chapter 2), to re-circularize the plasmid. The MCS is usually flanked by sequences that can be used for priming sequencing reactions or PCR amplification.

One potential problem during cloning is the presence of background colonies. Those are colonies that contain a plasmid that was re-circularized without an insert. These colonies grow because the plasmid carries the antibiotic resistance, but they are useless without the insert. For this reason, a second selectable marker is included in cloning plasmids in order to distinguish colonies that carry a recombinant plasmid from those that don’t. One of the most popular selectable markers of this kind is the $lac-Z$ gene encoding the $\beta$-galactosidase enzyme. This enzyme works as a tetramer. Each monomer

![Figure 1. Basic structure of a plasmid vector.](image)
Cloning

has two independent domains: alpha (α) and omega (ω). Deletion of the segment encoding the alpha domain in chromosomal DNA leads to the production of the ω-peptide, which has no enzymatic activity. However, if the alpha peptide is produced independently, it can re-store the activity of the omega fragment. This restoration of activity is called alpha complementation in trans.

Most E. coli strains used as hosts in cloning experiments carry a deletion of the alpha fragment in lac-Z gene; these strains produce the inactive omega fragment. Cloning plasmids have the sequence of the alpha fragment under the control of the lacZ promoter. However, the MCS is embedded in the coding sequence of the alpha-fragment. Thus, a functional β-galactosidase enzyme can be detected when the bacterial host carries the plasmid with an uninterrupted sequence of the α-fragment – α-complementation. The expression of this enzyme can be induced by the addition of lactose, or more commonly by the lactose analog isopropyl-β-D-thiogalactopyranoside (IPTG). β-galactosidase activity can be detected through the hydrolysis of the lactose-like substrate 5-bromo-4-chloro-3-indolyl- β-D-galactopyranoside (better known as X-Gal). Upon enzymatic cleavage of this artificial substrate, the released indols oxidize spontaneously and form a blue indigo precipitate in colonies carrying the intact plasmid. Colonies carrying a plasmid in which the α sequence is interrupted by an insert (the clone) are white because the insertion disrupts the coding sequence, and the α peptide is no longer functional.

Other selectable markers have used to replace the blue/white colony selection procedure. The ccdB gene (control of cell death), derived from the F’ plasmid, encodes a cytotoxic protein in E. coli where it inactivates gyrase, an essential topoisomerase II. The sccB gene of Bacillus subtilis encodes a levansucrase enzyme that catalyzes the transfer of fructose to levan, a fructose polymer. This enzyme is expressed in the periplasm of E. coli where accumulation of levan is lethal. The use of these genes reduces the background level (blue colonies) during cloning. The general strategy is to embed the MCS within these genes in a given cloning plasmid, and to plate cells transformed with these recombinant plasmids on media that favors the expression of these genes. Thus, an insertion into the MCS of these plasmids will inactivate the selectable marker (ccdB or sacB) and will render cells carrying a recombinant plasmid viable. In contrast, cells with the intact plasmid will not be able to survive in the selectable media.

The traditional method for generating recombinant DNA molecules (vector + insert) is through the ligation of compatible ends. In this procedure, the plasmid is opened up with a restriction enzyme and the cloning target is produced with a restriction enzyme that produces ends that are compatible with the ends generated in the cloning vector. The most commonly used enzyme for joining DNA molecules is Bacteriophage DNA ligase. This enzyme catalyzes the formation of a phosphodiester bond between the adjacent 5’ phosphate and 3’ hydroxyl termini of DNA strands brought together by compatible termini, or by chance in blunt ended DNA ligations. The presence of the 5’ phosphate is essential for ligation. Thus, to avoid re-
circularization of the plasmid without an insert, the plasmid is usually de-phosphorylated with calf intestinal alkaline phosphatase, or shrimp alkaline phosphatase. In this case, ligation takes place only in one strand at each ligation point. Ligation of blunt ended DNA fragments is much less efficient than that of cohesive termini. One way to increase the efficiency of blunt end ligations is through the addition of polyethylene glycol (PEG) to the reaction mix. PEG and other similar polymers are called volume excluders. These polymers interact with a significant amount of water in the solution, making it unavailable for solvation of DNA. In this way, the concentration of the reactants is artificially increased in the reaction volume.

Cloning of PCR products requires special considerations. First of all, PCR primers are usually not phosphorylated at the 5' end. Thus, if a blunt-end PCR product is ligated to a blunt-end vector, then the ligation will take place at two points instead of four. The ligase will establish a phosphodiester bond between the phosphorylated 5' end of the plasmid generated by the restriction enzyme and the 3' end of the PCR product, on both sides of the insert. However, if the plasmid is de-phosphorylated, ligation will be impossible.

The ligation of two DNA termini is not different from any bimolecular reaction in which the rate depends on the concentration of substrates. However, special consideration should be given to the ligation of a DNA fragment to a linearized vector. Let’s consider the ligation between the two termini of the same vector molecule. In a solution containing only the linearized vector, at very low vector concentrations, the concentration of one terminus in the immediate space occupied by the other terminus will be higher than the concentration of the termini of the other vector molecules in the solution. It follows then, that at very low concentrations circularization of DNA molecules is going to be the prevalent outcome. At the other extreme, a high vector concentration will result in concatemerization of the vector.

A theoretical treatment of ligation outcomes has been provided by Dugaiczyk et al.
Cloning (1975). This treatment addresses the ratios of the local concentration of one terminus in the neighborhood of the other on the same DNA molecule \((j)\), and the total concentration of DNA termini \((i)\) in the ligation reaction. When \(j > i\) circularization occurs, when \(j < i\) concatemerization prevails, and when \(j = i\) and equal number of concatemerization and circularization occurs. Thus, to maximize the number of monomeric recombinant molecules in a ligation (a circle with one insert and one vector molecule), it is necessary to obtain the optimum concentration of termini and an optimal relative proportion of insert and vector termini. Through theoretical calculations and practical considerations, the standard ligations are carried out with 50 ng of plasmid DNA (~3Kb) and a 3 molar excess of insert DNA. Commercial preparations vary in terms of the required temperature and duration of the ligation. In general, ligations of blunt ends are much less efficient than those involving compatible “sticky ends.” Blunt end ligations require the timey convergence of two blunt ends in the correct orientation and a ligase. For this reason, blunt end ligations are carried out at low temperatures (4 to 16°C) for long periods of time (8-16 h).

Another consideration in cloning PCR products is the polymerase that was used for amplification. In the previous chapter you read that Taq DNA polymerase usually leaves an unpaired ‘A’ overhang at the 3’ end of the amplicon, although other bases can also be added instead. A very small percentage of the amplicons have one or two blunt ends. This phenomenon creates a problem for blunt-end ligation approaches which are inherently less efficient than ligation with complementary ends. One solution is to treat a vector, which has been cut with a restriction enzyme that leaves blunt ends, with Taq DNA polymerase and an excess of dTTP. This treatment will add an unpaired ‘T’ at the 3’ ends of the vector. This type of ligation is called “T/A cloning.” This approach to cloning is usually inefficient.

A new generation of plasmid vectors has been constructed to facilitate the cloning of PCR products. In these vectors, the cloning target is introduced by a site specific recombination step catalyzed in vitro by an enzyme preparation. The most commonly used enzymes are the Vaccinia virus Topoisomerase IB, the well-characterized att Lambda recombination system (Gateway® technology of Invitrogen), and the Cre-recombinase of the Cre/Lox system of bacteriophage P1 (Dupont patent). These systems require the presence of specific sequences at the termini of the sequences that are targeted for cloning.

In this lab we are going to use the StrataClone Blunt PCR Cloning Kit from Stratagene (La Jolla, CA). This kit uses the activities of both the Vaccinia virus topoisomerase and the Cre/lox system from bacteriophage P1. Topoisomerases are essential enzymes that help overcome topological barriers to the flow of genetic information – from replication to transcription. These enzymes introduce single- or double-strand breaks to relax supercoiling of the DNA. Upon cleavage the enzyme forms a covalent phosphodiester bond between a tyrosine and one of the DNA termini. Type I topoisomerases cleave only one strand, the IA sub-class forms a covalent bond with a 5’ phosphate, whereas the IB sub-class forms a bond with a 3’
phosphate. The Type II superfamily generates staggered double-strand breaks. The *Vaccinia* virus topoisomerase belongs to the IB subclass and cleaves the phosphodiester backbone of the DNA after the sequence 5'-CCCTT forming a covalent 3' phosphodiester bond with a tyrosine residue. This bond stores energy which is used for the re-ligation step to an acceptor strand - the PCR amplicon in this case.

The StrataClone Blunt PCR cloning vector contains two blunt-ended DNA arms, each charged with a topoisomerase I at one end. A short incubation of these arms, with the PCR product, yields a mixture of recombinant DNA molecules. The desired recombinant molecule will have the PCR product flanked by the two different arms of the vector. Each of these arms carries at the other end the *lox* sequence (locus of crossingover) of the P1 bacteriophage. Upon entering the host cell (*E. coli*, StrataClone SoloPack strain), the *lox* sequences will be ligated by the *Cre* recombinase enzyme, which is expressed in the host cell, creating a circular molecule.

Recombinant plasmid generated *in vitro* can be introduced into transformation “competent” *E. coli* cells. Two types of competent cells can be prepared for transformation. “Chemically competent cells” (ccc) can be prepared by treating mid-log cells with a solution of CaCl$_2$ (Mandel and Higa, 1970; Cohen et al., 1972). The efficiency of transformation of these cells has been increased significantly through the years with the development of new bacterial strains and treatments with divalent cations, DMSO, hexamine cobalt chloride, and highly pure KCl (Hanahan 1983). The mechanism for the acquisition of competence has not been elucidated. Nevertheless, these treatments can generate “ccc” capable of yielding $10^9$ transformants/µg of supercoiled plasmid DNA. These cells can be prepared in the lab and stored at -80º C, or can be purchased from different suppliers. This lab will use commercially available “ccc” for transformation obtained with the cloning kit. Plasmid DNA can also be introduced into bacterial cells through electroporation, a technique that uses high voltage pulses to generate transient pores in bacterial membranes (Dower et al., 1988). The preparation of “electrocompetent cells”
Cloning is much simpler than the method used for “ccc.” It basically requires washing and concentrating the cells via repeated centrifugation in an ion-free glycerol solution. An electroporator is required for transformation. There are several electroporation devices in the market. These devices can produce a pulse (fraction of a msec) of a large voltage drop (several kVolts) using specialized cuvettes.

References


EXTRACTION AND ANALYSIS OF PLASMID DNA

The most commonly used method for the isolation of plasmid DNA is based on the alkaline lysis mini-prep procedure developed by Birnboim and Doly in 1979. Several modifications have been introduced since, including those adopted by a variety of commercially available kits. Nevertheless, the principles are still the same. In this method, bacterial cells are first concentrated by centrifugation, and then disrupted with a mixture of NaOH and SDS. The detergent disrupts cell structures and denatures proteins, while the alkaline medium denatures DNA. The NaOH-mediated denaturation step has differential effects on the plasmid and chromosomal DNAs. After cell disruption, the large chromosomal DNA is mechanically nicked and breaks up slightly by shearing forces, and under alkaline conditions it fully denatures. In contrast, denaturation of the plasmid is physically prevented by the supercoiled nature of the covalently closed circular molecules. Following the brief treatment with NaOH/SDS, the mix is neutralized by the addition of a high concentration of potassium acetate buffer (pH ~5). Addition of this buffer has several effects. It brings the pH down to near neutrality allowing renaturation of chromosomal DNAs. Renaturation of the chromosomal DNA results in the formation of a large network that is removed in the following step. Care must be taken during previous steps to prevent excessive breakage of chromosomal DNA. If the chromosomal DNA breaks up too much, it will not form a large network and it will partition with the plasmid DNA. Addition of K ions induces the precipitation of protein-SDS complexes because unlike sodium dodecyl sulphate (SDS), potassium dodecyl sulphate (KDS) is insoluble. In addition, the solubility of SDS is decreased at high salt concentrations and low temperatures. Thus, chilling the sample on ice also helps during plasmid purification. The precipitate can be removed by high speed centrifugation, or by filters used in several commercial kits. At this point, the DNA can be precipitated in ethanol or it can be further cleaned by an organic (chloroform/phenol) extraction before it is precipitated.

We are going to use a plasmid purification kit from Qiagen. This kit uses a silica membrane filter to purify the DNA. The principles of this type of purification were described in Chapter 2 of this manual. In general terms, the kit is based on the alkaline lysis procedure up to the neutralization step. After eliminating the precipitated KDS by centrifugation the supernatant containing the plasmid DNA is mixed with a chaotropic agent added to the neutralization solution; this agent will mediate adsorption of the plasmid DNA to the silica. The mixture is filtered through a silica membrane using centrifugal force, and the plasmid DNA will be retained by the silica membrane. The silica-bound plasmid is washed to eliminate impurities, and then it is eluted with elution buffer (Tris.HCl buffer pH ~7.5). Some researchers prefer to elute with plain water to avoid the possibility that Tris or EDTA in the elution buffer could interfere with downstream applications like sequencing. However, often times “pure water” could be slightly acidic due to solubilization of CO₂ from the atmosphere. Elution with this type of water may not be very efficient as a pH above 7 is required for maximum elution of DNA from the silica matrix.
References

MATERIALS AND SUPPLIES

A. Biological Materials

<table>
<thead>
<tr>
<th>Plasmid Vector</th>
<th>DNA Insert</th>
<th>Host</th>
</tr>
</thead>
<tbody>
<tr>
<td>pSC-B-amp-kan</td>
<td>PCR amplicon</td>
<td><em>E. coli</em> Solopack</td>
</tr>
</tbody>
</table>

B. REAGENTS

<table>
<thead>
<tr>
<th>Acetic Acid</th>
<th>Glucose</th>
<th>Sodium Chloride, NaCl</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agar</td>
<td>Hydrochloric Acid, HCl</td>
<td>Sodium Hydroxide, NaOH</td>
</tr>
<tr>
<td>Agarose</td>
<td>Isopropanol</td>
<td>Sodium Dodecyl Sulphate, SDS</td>
</tr>
<tr>
<td>Ampicillin.Na</td>
<td>MgCl₂</td>
<td>Taq DNA polymerase</td>
</tr>
<tr>
<td>Bacto-tryptone</td>
<td>Molecular Weight Markers</td>
<td>10X Taq Buffer</td>
</tr>
<tr>
<td>Blunt PCR Cloning Kit</td>
<td>Octanol</td>
<td>Tris Base</td>
</tr>
<tr>
<td>Chloroform</td>
<td>Polyethylene Glycol, PEG₈₀₀₀₀</td>
<td>Yeast extract</td>
</tr>
<tr>
<td>dNTPs</td>
<td>Potassium Acetate</td>
<td>X-Gal</td>
</tr>
<tr>
<td>EDTA.Na₂</td>
<td>Qiagen Plasmid Kit</td>
<td></td>
</tr>
<tr>
<td>Ethanol</td>
<td>RNAse A (Sigma R6513)</td>
<td></td>
</tr>
</tbody>
</table>

C. MEDIA AND SOLUTIONS

Media and Antibiotics

Luria-Bertani Broth

- Bacto-tryptone: 10 g
- Yeast extract: 5 g
- NaCl: 10 g
- H₂O add to: 1000 ml

**NOTE:** Adjust pH to 7.0 with 4N NaOH and autoclave for 20 min.

Ampicillin Stock

- Ampicillin.Na: 250 mg
- H₂O to final vol: 10 ml

**NOTE:** Sterilize by filtration, and store 1 ml aliquots at -20°C.
X-GAL STOCK, Store at -20°C
X-Gal 200 mg
N'N' DMF 9.8 ml
**NOTE1:** Prepare in PP tube.
**NOTE2:** Transfer 1 ml aliquots to 1.5 ml tubes, seal with parafilm.

AGAR PLATES
Add BactoAgar to media at a concentration of 1.5 %, autoclave, allow it to cool to 55°C before adding antibiotic and Mg stock.

Solutions

Stocks

EDTA, 0.5 M, pH 8.0

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Na$_2$EDTA.2 H$_2$O</td>
<td>186.1 g</td>
</tr>
<tr>
<td>NaOH pellets</td>
<td>19.0 g</td>
</tr>
<tr>
<td>H$_2$O add to</td>
<td>1000.0 ml</td>
</tr>
</tbody>
</table>

**NOTE:** Dissolve salts in 800 ml of water and stir vigorously with a magnetic bar. Adjust pH to 8.0 with 4N NaOH. Bring to final volume and autoclave.

Potassium Acetate, 5 M

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH$_3$COOK</td>
<td>490.8 g</td>
</tr>
<tr>
<td>H$_2$O add to</td>
<td>1000.0 ml</td>
</tr>
</tbody>
</table>

Sodium Chloride, 4 M

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl</td>
<td>23.4 g</td>
</tr>
<tr>
<td>H$_2$O add to</td>
<td>100.0 ml</td>
</tr>
</tbody>
</table>

Sodium Dodecyl Sulphate, 2 %

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDS</td>
<td>4 g</td>
</tr>
<tr>
<td>H$_2$O add to</td>
<td>200 ml</td>
</tr>
</tbody>
</table>

**NOTE 1:** Wear a dust mask to handle SDS.

**NOTE 2:** Dissolve in a beaker, stirring with a magnetic bar, in 80% of final volume. This process can be speeded up by either stirring over a hot plate (low heat) or brief warming in a microwave oven while avoiding foam. Carefully pour solution into a graduated cylinder, rinse beaker with water and pour into cylinder until reaching final volume. Cover cylinder with parafilm, or any other sealing film, and mix before pouring into storage bottle.
**Sodium Hydroxide, 0.4 N**

NaOH 8 g

H₂O add to 1000 ml

**NOTE:** Prepare monthly

**Tris.HCl, 1M, pH 8.0**

Tris base 121.1 g

HCl 42.0 ml

H₂O add to 1000.0 ml

**Working solutions**

**dNTPs, 2 mM**

- dATP, 100 mM 10 µl
- dCTP, 100 mM 10 µl
- dGTP, 100 mM 10 µl
- dTTP, 100 mM 10 µl
- Water 460 µl

**Chloroform**

- Chloroform 960 ml
- Octanol 40 ml

**GTE Buffer (Autoclave)**

[final]

- Tris.HCl (1 M, pH 8.0) 2.5 ml 25 mM
- EDTA.Na (0.5 M) 2.0 ml 10 mM
- Glucose 0.9 g 50 mM

H₂O to final vol 100.0 ml

**NOTE:** Mix stocks and dissolve glucose in water (80-90% of final vol), adjust the pH to 8.0, and then take to final vol with water. Store at 4°C.

**Lysis Solution NaOH/SDS (Prepare fresh)**

- NaOH 0.4 N 1/2 vol
- SDS 2% 1/2 vol

**NOTE:** Add 2% SDS to the 0.4 N NaOH. Keep this mix at RT°.

**Potassium Acetate, 5 M, pH 4.8 (Autoclave)**

- K Acetate (5 M) 52.2 ml
- Acetic Acid (Glacial) 13.7 ml
- H₂O Adjust vol to 100.0 ml
Polyethylene Glycol (MW 8000) 13% (Autoclave)

PEG₈₀₀₀  13 g
Water adjust vol to  100 ml

RNAse A (10 mg/ml)

RNAse A (Sigma R6513)  50 mg
Tris.HCl, 1M, pH 8.0  50 µl
NaCl, 5M  30 µl
H₂O  Adjust vol to  5 ml

NOTE: Dispense 1 ml aliquots and store at -20° C. This is a Mol. Biol. Grade enzyme preparation and does not need to be boiled before use.

TE Buffer (Autoclave)  [final]

Tris.HCl (1 M, pH 8.0)  1.0 ml  10 mM
EDTA.Na (0.5 M)  0.2 ml  1 mM
H₂O to final vol  100.0 ml

C. Labware and Supplies

Falcon tubes, 12-15 ml
Ice buckets
Kim Wipes
Microcentrifuge tubes, 1.5 ml
Micropipets and tips

D. Equipment

Freezer, -20°C
Microcentrifuge
Refrigerator, 4°C
Shaker Incubator
Vortex
CLONING PCR AMPLICONS

Ligation Reaction

1. Set up the ligation by adding the following to a 200 µl PCR tube.
   - StrataClone Blunt Cloning Buffer 3 µl
   - PCR product (5-50 ng) 2 µl
   - StrataClone Blunt Vector Mix 1 µl
   **NOTE:** If the amplification was quite robust, then dilute 1:10 in pure water.

2. Mix gently by repeat pipetting and incubate at room temperature (25°C) for 5 minutes. Place on ice when finished. Incubate at room temperature for 5 minutes.

Transformation of Chemically Competent *Escherichia coli* Cells

1. Thaw one tube of StrataClone SoloPack competent cells on ice. Make sure the cells are not splashed on the wall of the tube.
2. Add 1 µl of the cloning reaction mix to the tube with thawed cells, mix gently, but NOT by pipetting.
   **NOTE:** A positive control transformation should be carried out with pUC18 plasmid.
3. Incubate on ice for 20 minutes.
4. Pre-warm 500 µl LB medium to 42°C.
5. Heat shock transformation mixture at 42°C for EXACTLY 45 seconds.
6. Incubate on ice for 2 minutes.
7. Add 250 µl of LB medium pre-warmed to 42°C.
8. Close the tubes tightly, and incubate for 1 hour at 37°C in an orbital shaker at 200 RPM. Lay the tubes horizontally to improve aeration.
9. Spread 40 µl of 2% X-Gal on two LB-Amp (100 µg/ml) plates.
10. Plate 5 µl and 100 µl of the transformation mixture, and incubate overnight at 37°C.
    **NOTE:** To plate 5 µl, first add the 5 µl to a tube containing 100 ul of LB, then plate the entire volume.
11. Count the number of colonies from the PCR insert and those of the positive control. Estimate the number of cfu/µg of DNA.
**Colony Screen**

**Day 1**
1. Pick 4 white colonies with sterile toothpicks and inoculate 2.5 ml of LB (+Amp) in a Falcon tube.
2. Incubate overnight at 37°C in an orbital shaker set to run at 250 RPM.

**Day 2**
1. Pipette 100 µl of the cell culture into a 1.5 ml microcentrifuge tube.
2. Pellet bacterial cells at 8000 rpm for 2 min.
3. Discard supernatant into biohazard waste container.
4. Add 100 µl of sterile water and resuspend cells by vortexing.
5. Flash-freeze cells in liquid nitrogen.
6. Thaw the samples and keep on ice.
7. Pellet the cell debris in a microcentrifuge for 5 min at maximum speed.
8. Carry out a PCR screen using the supernatants as templates and the PCR primers used in the previous experiments.
   a. Program the thermocycler as follows:
      
      | Step               | Temp | Time |
      |--------------------|------|------|
      | Initial denaturation | 94°C | 2 min |
      | Amplification, 35 cycles |      |      |
      | Denaturation        | 94°C | 30 sec |
      | Annealing           | 60°C | 30 sec |
      | Extension           | 72°C | 60 sec |
      | Finishing           | 72°C | 5 min  |
      | Hold                | 15°C | Forever |

   b. PCR reaction
      
      | Component                  | Volume |
      |----------------------------|--------|
      | PCR Buffer, 10X            | 2 µl   |
      | MgCl₂, 15 mM               | 2 µl   |
      | dNTPs, 2 mM                | 2 µl   |
      | F Primer, 1 µM             | 2 µl   |
      | R Primer, 1 µM             | 2 µl   |
      | H₂O                        | 6 µl   |
      | Taq Pol. 0.25 u/µl         | 2 µl   |
      | Template DNA               | 2 µl   |
      | TOTAL vol                  | 20 µl  |

9. Add 4 µl of 6X electrophoresis loading buffer to the PCR reaction and analyze by agarose gel electrophoresis.
10. Select one of the positive samples and proceed with the plasmid isolation from the overnight cultures.
**QIAGEN MINIPREP PROCEDURE**

1. Pellet 1.5 ml of overnight culture, in a 2 ml tube from the kit, for 1 min at 13,000 rpm.
2. Remove medium with pipette and discard in biohazard container.
3. Add 250 µl of **P1 Buffer** to the bacterial pellet.
4. Resuspend cells thoroughly by vortexing.
5. Add 250 µl of **P2 Buffer** (Lysis buffer), close tube and immediately invert the tube 4-6 times to mix, and incubate at room temperature for no more than 5 minutes.
   - NOTE: It essential to do both mix thoroughly and avoid excessive shaking that can lead to shearing the chromosomal DNA.
6. Add 350 µl of **N3 Buffer** (Neutralization buffer). Mix **thoroughly** by tube inversion 4-6 times.
7. Centrifuge for 10 min at 13,000 rpm in table top centrifuge.
8. Place a **Qiagen Spin Column** in a decapped 2 ml tube. Transfer the supernatant to the column. Close the column lid.
9. Spin for 1 minute and discard the flow-through.
10. Wash column with 750 µl of **PE Buffer** (Ethanol and Na Acetate buffer). Add the buffer and spin for 1 min.
11. Discard the flow-through and place the column back into the 2 ml tube and spin again for 1 min to eliminate all the ethanol.
12. Transfer the column to a new 1.5 ml conical microcentrifuge tube. Add 50 µl of **EB** (Elution buffer, Tris.HCl pH 8.5). Wait for 1 minute before spinning the column for 1 minute.
   - NOTE: It is important to pipette the EB to the middle of the filter without touching it.
13. Use a 5 µl aliquot to quantify the plasmid prep by fluorometry.
14. Determine DNA quantity and quality fluorometry with the Hoescht reagent. Use 2 ul Fluorometric Readings

<table>
<thead>
<tr>
<th>Std</th>
<th>ng/µl</th>
<th>RF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample</th>
<th>µl</th>
<th>RF</th>
<th>ng/µl</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

15. Also, prepare a sample for agarose gel electrophoresis. Pipette into a microcentrifuge tube 13 µl of water, 2 µl of plasmid prep, and 3 µl of loading buffer. A single gel will be run for the entire class. A picture of this gel will be submitted, along with the plasmid samples, to the DNA Sequencing Core Lab.
Lab Report: Cloning

Part I

Summary (One Page Maximum. 25 points)

a) The main objective of the experiment,
b) The most important findings,
c) The significance of the results, and
d) Major conclusions

Part II

Results (15 points)

Summary of the results obtained in the lab.
- The cloning kit components
- The transformation potential of the “ccc” with the control plasmid. Show calculations.
- The number of white clones that tested positive by PCR. Show picture
- The concentration and yield of the purified plasmid.

Part III.

1. (10) Making a phosphodiester bond requires energy. What are the sources of energy for T4 DNA ligase and for Topoisomerase.

2. (15) “Inverse PCR.” is a PCR technique designed to obtain the flanking sequences of a target sequence. In this technique, the PCR product is characterized first for the presence or absence of restriction enzyme sites, and a record is kept of those enzymes that do not cut the target sequence. The next step is to digest the genomic DNA, from which we want to obtain the flanking sequences, with a restriction enzyme that does not cut the target sequence. In the following step, the digested DNA is ligated to generate circles. These circles are formed by ligation of compatible ends located on the flanking sides of the target sequence. What conditions do you need to use in the ligation reaction to generate the DNA circles.

3. (15) How does α-complementation work?

4. (15) Why do we need to incubate the cells for 1 hour at 37°C after transformation before they could be spread on agar plates?

5. (10) What are the effects caused by addition of potassium acetate buffer to the lysed bacterial cells?
Study Questions

1. What is the most essential component of a plasmid?
2. What is the major difference between a typical cloning plasmid vector and the bacterial artificial chromosome (BAC) vector?
3. What is the most logical explanation if Plasmid A can be co-transformed into E. coli with plasmid B, but not with plasmid C?
4. What is the role of antibiotic resistance genes included in plasmid vectors?
5. Cloning plasmids have a polylinker region that is included to facilitate cloning. What is the main characteristic do the restriction sites in the polylinker?
6. What is the principal strategy used to identify false positives in cloning?
7. What situation would explain the presence of an insert that yields blue or bluish colonies?
8. What structural feature of DNA must be present for ligation to take place?
9. How can you increase the efficiency of blunt end ligation? Explain.
10. What ligation conditions are required to produce circular molecules?
11. What ligation conditions are required to produce recombinant molecules?
12. What issue needs to be addressed when cloning PCR products?
13. What are the two main methods for introducing plasmid DNA into bacterial cells?
14. What is the effect of adding NaOH to bacterial cells during plasmid DNA extraction?
15. How is chromosomal-DNA kept separated from plasmid DNA during extraction of the latter?
16. What is the basis of the strategy for isolation plasmid DNA with silica membrane of particles?
Chapter 6

DNA SEQUENCING

Learning Objectives......................... 6.1
Lab Objectives.............................. 6.1
Background .................................... 6.1
Sequence Analysis.......................... 6.8
Lab Report.................................... 6.11
DNA SEQUENCING

Learning Objectives

After completion of this unit, students will be able to prepare plasmid DNA suitable for DNA sequencing reactions, edit, process, and interpret DNA sequencing data. In addition, students will be able to match the needs of a sequencing project with the appropriate sequencing methods.

Laboratory Objectives

Process and edit chromatograms of DNA sequence, align complementary sequences to obtain a consensus sequence, produce a Fasta file for the sequence, and confirm the identity of the sequence via a BLAST search.

Background

DNA sequence information is essential for the study of gene structure and function at one end of the spectrum, and for comparative and evolutionary studies at the other end. DNA sequencing technology has evolved very rapidly in the last few years. The Maxam-Gilbert chemical method and the Sanger dideoxy method represent the pioneer technologies; these are based on electrophoretic analysis of the sequencing reactions. Of these, the Sanger method has been the most enduring, and has been used to obtain the sequences of several prokaryotic and eukaryotic genomes, including those of Escherichia coli, Caenorhabditis elegans, Arabidopsis thaliana, Oryza sativa, Glycine max, and Homo sapiens among several others. Massive parallel sequencing strategies have been developed which increase output by several orders of magnitude over the Sanger technique. These strategies have been dubbed NextGen or second generation sequencing technology and comprise pyrosequencing (454 GS-FLX), sequencing by synthesis (Illumina), sequencing by ligation (ABI SOLiD), and sequencing by pH changes (IonTorrent). A third generation sequencing technology instrument has been introduced recently, and it has the ability to sequence single DNA molecules providing long reads (PacBio).

In this lab we are going to use the sequencing services provided by the DNA Sequencing Core Lab of the Interdisciplinary Center for Biotechnology Research at the University of Florida. The clones obtained by the class will be sent out for sequencing using the Sanger dideoxy method. This method has evolved tremendously since its introduction (Sanger et al., 1977). The method requires:

1) A single stranded DNA template,
2) A primer oligonucleotide,
3) A DNA polymerase,
4) Deoxynucleotides (dNTPs)
5) Dideoxynucleotides (ddNTPs, also known as chain terminators)
6) A label (Radioactive, or fluorescent).
The Sanger method is based on DNA synthesis carried out by a DNA polymerase. In the initial protocol, the template for the sequencing reaction was produced by first cloning the target sequence into the replicating form of the M13 bacteriophage, a single stranded DNA phage, and then inducing phage replication which yielded single stranded DNA. At present, isolation of ss DNA is not required because the reaction is now carried out by Taq DNA polymerase using asymmetric amplification. This means that a single primer that anneals to one of the DNA strands is used to prime the reaction. The sequencing strategy calls for DNA amplification in the presence of a mixture of normal deoxynucleotide triphosphates (dNTP), and a small proportion of dideoxynucleotides (ddNTP). The ddNTPs lack a hydroxyl group at the 3' position of the ribose moiety. Thus, when a ddNTPs is incorporated by the DNA polymerase onto the growing DNA chain, the lack of the hydroxyl group at the 3' position prevents the addition of more nucleotides to the chain, hence the name “chain terminators.” By providing a mixture of dNTPs and ddNTPs, an array of early termination chains of differing lengths can be produced in a sequencing reaction. The original Sanger protocol required four different reactions, each with a single ddNTP and a radiolabeled dNTP. Each of the reactions generated a series of fragments which were terminated by the same ddNTP at distances from the primer that corresponded to the position of the dNTP analog (for instance, incorporation of ddATP instead of dATP). These fragments were then separated according to size by denaturing polyacrylamide gel electrophoresis under conditions that discriminate fragments differing in size by one nucleotide in the range of 10 to 1,000 bases. The products of each of the four reactions were run in different lanes in the gel. The 32P-labeled DNA fragments were visualized by autoradiography. The DNA sequence from these reactions could be deduced by “reading” the autoradiogram from the bottom up. The DNA sequence can be deduced starting at the 5' end by recording the track (A, C, G or T) in which the smallest discernible band appears. The next base in the sequence is given by the identity of the track (A, C, G, or T) in which the next fragment in size is located. Sequences between 200 and 500 bases were originally obtained by reading the four tracks from the bottom to the top of the autoradiogram (See Fig. 1).

A number of modifications have been introduced to the Sanger method to transform it from a manual sequencing method to a fully automated procedure. These improvements include changes in the way the single stranded template is procured, and the introduction of fluorescent labels. The template for the sequencing reaction was originally obtained by denaturation of the double stranded DNA or by inducing the production of filamentous phage (M13) engineered into the cloning vector which contained the target sequence. This method was made obsolete by the development of cycle sequencing which uses double stranded DNA as a template, and a thermostable DNA polymerase (Taq, Vent, etc.). In essence, cycle sequencing amounts to carrying linear PCR with the Sanger reaction. It is linear because a single primer is used. This new approach overcomes the limitations of mesothermic polymerases, and results in longer
fragments, higher yields, and a significant reduction in problems related to the formation of secondary structure.

A significant improvement in the Sanger method was the adoption of fluorescent tags for the chain terminators. The use of a different fluorophore for each of the four chain terminators meant that the labeling reactions could be carried out in a single tube, and the products could all be resolved in a single track of a polyacrylamide gel requiring only 25% of the space used with a radioactive label. Detection of fluorescent labeled DNA fragments is accomplished by means of a laser that scans the polyacrylamide gel near the bottom of the plate. As the DNA fragments move down the gel they are detected by the optical system. The DNA sequence is deduced from the color of the fluorescent label characteristic of each nucleotide (Figure 2). The intensity of the fluorescence of each fragment along with the resolution of its corresponding band (degree of separation from adjacent bands) are taken into account to calculate a quality score, which is used to assess the overall quality of the sequence. Furthermore, fluorescent labeling was improved by the use of complex fluorophores – energy transfer dyes containing a fluorescein as the donor dye linked to a substituted rodhamine molecule. When these fluorescent molecules are covalently linked by a benzo amino acid, they produce a two-peak excitation spectrum, and a one-peak emission spectrum. In these molecules, light energy is transferred from fluorescein to rhodamine resulting in the amplification of the fluorescence signal. Fluorescent labels containing these molecules are known as BigDye chemistry, and provide greater sensitivity in detection. For instance, they have made it possible to directly sequence BAC clones, a task that had previously proven to be difficult due to the size of the molecule and the consequent reduction in the initial number of copies in the reaction. Another innovation in Sanger technology was the introduction of capillaries to replace the polyacrylamide gel. All these innovations have allowed to have fully automated systems.

Overall sequencing strategies have also evolved since the Sanger method was introduced. At first, sequencing of a DNA fragment cloned into a plasmid vector was initiated using a
plasmid-derived primer. To continue sequencing beyond the first 500 bases, the first few bases had to be deleted to bring the following bases closer to the terminus of the plasmid and the sequencing primer (Figure 1A). This was accomplished by opening up the plasmid near the first sequenced bases, digesting the 3' terminus with exonuclease III and the remaining single strand with mung bean nuclease. The blunt end was re-ligated to the plasmid for a second round of sequencing. This type of manipulation is no longer required due to improvements in oligonucleotide synthesis technology. New sequencing primers can be designed from the previously sequenced DNA, and these primers can be ordered on line to be received within one or two days. DNA synthesizers are also available in many laboratories. This method of sequencing is called **primer walking** (Figure 1B). The Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) software can be used to design sequencing primers.

Another sequencing strategy calls for the introduction of transposon-mediated random insertions in the DNA (Figure 1C). There is at least one company that sells a kit that contains both the insert-DNA and the transposase needed to insert it. The insert contains the terminal inverted repeats recognized by the transposase, a selectable marker (antibiotic resistance) and sequences to which sequencing primers (pointing towards the edges of the transposon) can anneal. Thus, instead of deletions, or synthesis of new primers for primer-walking, this method uses random insertion throughout the plasmid as starting points for sequencing. The strategy here is to produce a mini insertion library of the plasmid by inserting the “transposon” *in vitro* prior to transformation. A number (n) of independent insertions are required to have a probability (P) of sequencing the entire plasmid given that a sequencing read is of a given length (L), and the plasmid is of size S. The calculations are similar to those used in Chapter 1. For instance, let’s assume we have a plasmid vector that is 3 Kb carrying an insert that is also 3 Kb, so the total length of the recombinant molecule is 6 Kb. Furthermore, we know that our DNA Sequencing Center can easily obtain 800-base reads of high quality. Thus, a single insertion can be used to obtain 800 bases of sequence to each side of the insertion for a total of 1600 bases. A single insert covers (p) 1600/6000 bases of the recombinant plasmid. How many clones (n) with independent insertions should we sequence to have a 99.99% probability of obtaining the entire sequence of the recombinant plasmid. The answer of course is 30.

Notice that the vector will be sequenced too because there is no way for the transposon to discriminate between vector and insert sequence, unless you carry out the transposon tagging reaction before the insert is ligated to the vector. If only the insert is tagged, then the number of needed clones would be cut down to 12, and if you decide to do this with a probability of only 99%, then the number of clones will be reduced by half. Remember that two sequencing reactions are needed for each clone, one in opposite directions from the insertion. This approach is affordable when sequencing in 96-well format plates. While in the **primer walking** approach we are able to obtain a linear and orderly progression of the sequence, the transposon approach provides the entire sequence simultaneously. A computer program is required to assemble the reads into a single sequence. These programs can be directed to filter out the plasmid sequence.
Larger sequencing projects have been undertaken that rely on computer programs to put together the target DNA sequence. For instance, let’s consider obtaining the sequence of a BAC clone with an insert size of 200 Kb. A primer walking sequencing approach would take a very long time. Instead, a **shotgun sequencing** approach will take a minimum amount of time, but will require the construction of a mini-library in a plasmid vector. Accordingly, the BAC DNA is physically broken up into fragments of approximately 2 Kb, turned into blunt end fragments, which will be cloned into a plasmid vector. Although cutting BAC DNA with a restriction enzyme would seem the easiest way to break up the DNA, this approach has the problem that restriction sites are not uniformly distributed in DNA, and such treatment may lead to a biased mini-library. Instead, physical fractionation with a nebulizer or by sonication can produce an unbiased fragment population. However, the DNA fragments obtained in this fashion have to be enzymatically mended by eliminating 3' or 5' overhangs. Mungbean nuclease has specificity for single stranded DNA and is normally used for this purpose, although other strategies can also be employed. The calculations for the number of sub-clones that are required to have a given probability of obtaining the entire BAC sequence is the same as described previously. All the sub-clones can then be sequenced in 96-well format using different degrees of automation.

There are two main strategies for sequencing entire genomes. One of them is called the clone-by-clone approach, and it relies on the construction of a physical map of the target organism which is accomplished in two steps. First, a genomic library of the organism is constructed in a large insert vector (BAC). The BAC clones are then screened under different strategies to identify overlapping clones leading to the construction of contigs – the assembly of BAC clones that represents the physical organization of the organism’s chromosomes. In the second phase, a **minimum tiling path** is selected for sequencing. The minimum tiling path corresponds to the set of overlapping clones that are physically ordered to represent the chromosomes of the organism, and with the added characteristic that they have minimum redundancy. Sequencing of the minimum tiling path is expected to yield the genome sequence. The second approach is the shotgun approach already described for a single BAC clone, but applied to an entire genome.

Second and third generation sequencing platforms lend themselves to large DNA sequencing projects. The 454 Genome Sequencer FLX System (Roche) is an instrument that incorporates pyrosequencing and microfluidics to carry out massive parallel sequencing reactions capable of yielding 1 million reads with a 600 to 1000 nucleotide length, or a total of 600 million to 1 billion bases in a single 10 or 14-hour run, depending on the platform. This instrument is used for sequencing of small prokaryotic genomes and for gene expression studies. The University of Florida ICBR has two of these instruments in place. The first step for sequencing in this instrument is the construction of a single stranded DNA template library. This is accomplished by first fracturing the DNA into 1 to 2 Kb fragments which are ligated to a pair of adapters after end-repair. The ligated DNA is denatured and captured by beads bearing an oligo complementary to one of the adapters. This is done under conditions in which a single stranded DNA molecule is captured by a single microbead. Subsequently, the bead is populated via emulsion PCR. These beads are then deposited into microwells on a plate. A microfluidics system delivers the reagents and washing solutions to the wells as described below. The third step is the series of pyrosequencing reactions. This sequencing method is based on the coupling
of three different enzymatic activities. As the DNA polymerase incorporates a nucleotide to the growing chain, it releases a pyrophosphate molecule for every nucleotide that is incorporated into the DNA. The enzyme ATP sulfurylase then catalyzes the synthesis of ATP from adenine phosphosulfate and the pyrophosphate produced in the previous reaction. Luciferase, the next enzyme in this chain, uses energy from the newly synthesized ATP to catalyze the oxidation of luciferin to oxyluciferin with the subsequent release of ADP and light (bioluminescence). These reactions take place in a micro cell and are carried out 4 times, each with a different nucleotide for each position in the template. The microcell is completely washed after each reaction. The sequence from each well is deduced from the light emitted by each well following the addition of a specific nucleotide. Also, the amount of light detected from each well is proportional the number of nucleotides that are incorporated into the DNA. Thus, wells where two bases (AA, CC, GG or TT) are incorporated would emit twice as much light as those where a single base (A, C, G, or T) was incorporated. The disadvantage of this strategy is that proportionality is loss above four bases. In summary, one whole plate will yield about 1 million reads 600 to 1000 nt long, or a total of approximately 0.6 to 1 Gbp.

The SOLiD sequencing system (Applied Biosystems) produces a total of 60 Gbp of sequence data, but with shorter reads (35 to 75 nt). Sample preparation is very similar to that of the 454 instrument using beads, but instead of depositing the beads in micro wells, the beads are affixed to the surface of a glass slide. This system uses a “sequencing by ligation” strategy. A primer that anneals to the oligonucleotide ligated to the sample DNA is used as a sequencing primer. A set of four 8-mers each with one of the four bases at the 3’end and the same base at the next position up is released on the surface of the slide. The other six bases are random bases. Each 8-mer has a distinct fluorophore, linked to the 5’ end, which associated with the identity of the second base from the 3’ end. The 8-mers that anneal next to the primer are ligated by a ligase and the color of the fluorophore is recorded for each bead. Unlike other sequencing methods the newly synthesized strand grows in the 3’ to 5’ direction. In the next step the 3 bases at the 5' end carrying the fluorophore are enzymatically cleaved, and the process is repeated 15 more times. The first run provides the identity of a set of 2 bases every 5-bases. To obtain the identity of all the bases the entire process is repeated 4 more times, but each time the reaction is started with a new sequencing primer that anneals one nucleotide downstream from where the previous one annealed. Five different sequencing oligos are used overall. The end-result is that in a 75 nucleotide stretch each base has been queried twice. This instrument can produce 60 gigabases in a single run. The most effective use of this technology is for re-sequencing and for gene expression studies. The DNA sequencing Core of ICBR has one of these instruments.

The Illumina sequencing platform prepares DNA for sequencing using principles similar to those used by the other platforms. DNA is broken into 400 to 600 bp, end-repaired, and ligated to a couple of adapters. The concentration of the DNA is carefully determined. After denaturing the DNA, it is diluted and the solution is spread over a glass surface that has covalently bound oligo primers complementary to the adapters ligated to the DNA. The single stranded DNA, at a low concentration, anneals to those oligos that are used to synthesize the complementary sequence; this sequence will now be covalently attached to the glass plate. The plate contains oligos complementary to both adapters. This arrangement facilitates the synthesis of the complementary strand. The other end of the newly synthesized DNAs can anneal to a plate-
bound complementary oligo and the synthesis cycles are repeated several times to basically PCR-amplify a sequence and form a “cluster” of that particular sequence. A sequencing primer is annealed to the free end of the template to start the sequencing reactions. All four labeled nucleotides are used in the first reaction using a unique fluorophore for each nucleotide. However, each nucleotide has a block at the 3’ end to prevent the addition of a new base. After the fluorescence of each cluster is recorded by the optical system, both the fluorophores and blocks are eliminated to move into the next round of synthesis. Illumina has several platforms that can yield between 30 million to 1 billion reads of 100 to 150 nucleotides long.

The IonTorrent platform sequences DNA on a solid semiconductor chip. The chip has millions of wells that capture a microbead populated with a single sequence produced by emulsion PCR as used by the other platforms. Sequencing is based on synthesis; in this system all the wells are flooded with a single nucleotide. If the nucleotide is incorporated, a proton is released and the corresponding change in pH is sensed and recorded for that well. The error rate in this system is supposed to be minimum as no fluorescently labeled nucleotides are needed or light of certain wavelength needs to be recorded. This platform currently yields up to 80 million reads of about 100 bases long, but a new chip with a capacity to yield up to 300 millions reads has been announced for release in the near future.
SEQUENCE ANALYSIS

The instructions below will take you to the server of the DNA Sequencing Core Lab. You will then be able to download your sequences. From this server you will be able to view the chromatogram trace of the sequence, and also download the chromatogram file to your computer. If requested, the Seq Lab can provide an edited version of your sequence in FASTA format. However, in this lab you will do the editing based on the quality score of the bases.

The instructions below will enable you to download a freeware program from the Geospiza server. This program will allow you to view and edit your sequence. Instructions will also lead you to the UF DNA Sequencing Core server.

1. Download FinchTv to your computer from: http://www.geospiza.com/ftvdlinfor.html. You will be able to select your operating system.

2. Direct your browser to http://dNAlims.dNAtools.com/
3. Click on the Login to dnaLIMS link.
4. Enter the user name and the password, and click on the Submit button.
5. Click on the Download Your DNA Results link under the Sequencing Requests and Results sub-heading on the left of your screen.
6. Select the most recent Order Number first. When you do, the number will appear on the left column. Click on the Submit button at the top of your screen.
7. The new screen will have a table with the sequencing results. The results will be available in three different ways, each with a hyperlink. Left clicking on any of them will bring your sequence. Left clicking the Text link will bring up the unedited sequence in a new window, while right clicking it will bring up a popup menu which will allow you to save the sequence as a file in your computer. Left-clicking the Chromat link will bring up a popup window asking whether you want to open or save the chromatogram. You will be able to see the chromatogram only if you have already installed FinchTv in your computer, otherwise you can right-click on it and save the chromatogram by selecting the “Save Target As ...” choice in the popup menu (highly recommended). Finally, you can have a preview of the sequences by left-clicking on the View link. Each group submitted a single clone for sequencing, and the cloned DNA was sequenced in both directions. Make sure you download the forward and reverse sequence for your clone.

8. Open up the FinchTv program and then open one of your sequences. Alternatively you can go to the directory where the chromatogram files are stored and click on the forward file. FinchTv will display the chromatogram with the color-coded peaks of each DNA fragment along with the quality scores (vertical bars) and a threshold line (Q=20). Bars below this line indicate poor quality for the read. Select the Wrapped View in the drop down menu for View. You can adjust the vertical and horizontal scales with the slide buttons.

9. Edit the sequence by trimming off vector and low quality sequences. The sequencing reaction was primed with an oligo (M13 primer – forward or reverse), which annealed to the vector at a location a few bases upstream (or downstream) from the insertion point. So, you need to locate the point of insertion according to the diagram shown below. The vector used in the
cloning experiment is pSC-B-amp/kan, and the structure of its multiple cloning site and surroundings is depicted in the adjacent diagram:

10. Enter the sequence corresponding to the EcoRI site and the topoisomerase site (GAATTCCGCCCTT) in the Find Sequence box and press enter. This action will take you to the point where the topoisomerase ligated the vector with the amplicon.

11. Block the sequence to the left of the insert, and use the delete command in the Edit menu to delete this part of the sequence. You will notice that the letters and numbers have disappeared in the deleted section, and that the remaining bases have been renumbered.

12. The other end of the sequence should also be edited. One possibility is that the insert was smaller than the sequencing length. If this is the case, then sequencing was carried out over the insert and into the plasmid on the other side of the insertion. To determine whether this is

Figure 4. Structure of the cloning site of plasmid pSC-B-amp/kan. The insert is placed between the topoisomerase recognition sequence “GCCCTT -------AAGGGC” which are flanked by two EcoRI sites.
the case, type in the **Find Sequence** box the complementary sequence of the topoisomerase recognition sequence and the EcoRI sequence (AAGGGCGAATTC), and press enter. If the program locates this sequence, then block it and the rest of the sequence to the end of the chromatogram. Delete as before. The second possibility is that the insert was too long and the other terminus was not sequenced. In this case, scroll down until the quality bars fall short of the quality threshold. Block the part of the sequence with low quality scores to the end of the chromatogram and delete as before.

13. Confirm that this is the sequence that you targeted for cloning. Under the **Edit** menu click on the BLAST link at the bottom of the list and select the correct Blast Program. This will take you to the NCBI BLAST page and your sequence will already be pasted in the sequence window. Make sure that under the **Choose Search Set** panel, the appropriate database is selected.

14. Once you are sure that the sequence is the correct one (Save output as evidence), **Export** the sequence in the FASTA format. Click on the **Export** link under the **File** menu (FinchTv).

17. Check for possible sequencing errors by aligning the forward and reverse sequences using the appropriate BLAST program.
Lab Report: DNA Sequencing.

Part I
Summary (One Page Maximum. 20 points)
a) The main objective of the experiment,
b) The most important findings,
c) The significance of the results, and
d) Major conclusions (How does the Sanger method differ from NextGen technology?)

Part II - Results
1. (15) Print one of the chromatogram traces of your sequence (forward or reverse). Access the Print Options window from the File menu in FinchTv. Select an appropriate scale and make sure to check the last box for Print region with called bases only. Then go to the Print Sub-menu and Print your page. Also, use your word processor (Notepad, Wordpad, MS Word...) to print out the sequence in FASTA format. Indicate the putative location of exons and introns (Sequences are best printed with Courier New font).

2. (15) Provide evidence that the sequence obtained from the DNA Sequencing Core Lab is the sequence that was targeted for amplification and cloning. Matching the PCR primers to the termini of the sequence will give you some points, but more will be required for full credit.

Part III
1. (50) The complete coding sequence of a gene can be obtained through screening and sequencing of selected clones from a genomic, or cDNA libraries. For instance, screening a genomic library using your clone as a probe can lead you to isolate a clone that contains the entire sequence of the gene. On the other hand, screening a cDNA library of full-length clones can also lead to the identification of clones with the complete coding sequence of the gene.

However, in the absence of genomic or cDNA libraries, there are some PCR-based protocols that can get you the same results. Inverse PCR (iPCR) is a technique that can be used to obtain sequences flanking a target sequences. In this procedure, genomic DNA is digested with one or more restriction enzymes that do not have a recognition site within the available sequence. Next, the digested DNA is ligated under conditions that favor circularization. Finally, nested primers pointing in the direction of the immediate flanking sequences are used to amplify the ligated flanking sequences. Similarly, the rapid amplification of cDNA ends, or RACE, is a PCR procedure designed to amplify either the 5' end (5' RACE) or 3' end (3' RACE) of a specific mRNA.

Both iPCR and the RACE procedures require the design of primers that are pointing away from the sequence. Design nested primers that can be used for both iPCR and 5' and 3' RACE using the sequence that was received from the Sequencing Core.
When designing primers you should take into account the following:
- If you are working with a genomic sequence, then a primer located in an intron may work for iPCR, but will not work for RACE.
- If you are working with an EST sequence, remember that if a primer sequence anneals to an exon-exon junction, it will work with RACE but not with iPCR.
- If the external primers are too close to the edge of the sequence, then you will not see any overlapping sequences when you get around to sequencing the new products.
- Finally, the orientation of the sequence is important for RACE because you need to determine which primers you are going to use for 5' and for 3' RACE. This is mentioned here because when cloning fragments with blunt ends, or with compatible ends for a single restriction enzyme site, cloning is not directional, that is, the insert can be ligated to the vector in either orientation. You can establish the orientation of your sequence from the BLASTX alignments.
CHAPTER 7

SOUTHERN BLOTTING AND HYBRIDIZATION

Learning Objectives............................. 7.1
Lab Objectives..................................... 7.1
Background - Southern Blotting.......... 7.1
Background Southern Hybridization ... 7.3
References ........................................... 7.6
Blotting - Materials & Sup ............... 7.7
Blotting - Lab Procedure ................. 7.9
Hybridization - Materials & Sup ...... 7.11
Hybridization - Lab Procedure .......... 7.14
Lab Report......................................... 7.18
SOUTHERN BLOTTING AND HYBRIDIZATION

Learning Objectives

After completion of the two modules in this section, students will be able to carry out Southern hybridizations, and use the information produced by this technique to deduce the structural characteristics of target DNA. Students will be able to select the most appropriate labeling and hybridization conditions to match the probe with the DNA target.

Lab Objectives

Obtain Southern blots of plant DNA digested with a restriction enzyme and detect specific DNA fragments containing a set of specific sequences. This objective will be accomplished by first separating DNA restriction fragments by agarose gel electrophoresis, and then transferring the fragments to a solid support. Specific DNA fragments affixed to the solid support (the Southern blot) will be identified by hybridization to a labeled probe. The hybridization pattern will be visualized by either autoradiography or a phosphoimaging device.

Background - Southern Blotting

Southern blotting and hybridizations is a technique used to detect and identify DNA restriction fragments carrying specific sequences. The Southern blotting procedure was developed by Edward M. Southern at the University of Edinburgh in 1978. In this procedure, DNA fragments are first separated according to size by agarose gel electrophoresis, and then they are transferred (blotted) and affixed onto a solid support – nitrocellulose or nylon membranes – preserving the separation pattern they had in the gel. Thus, the blot constitutes a long-term record of the electrophoretic separation. The immobilized fragments can be analyzed via hybridization with labeled probes to identify specific sequences bearing similarity to the probe. This technique has been used to identify cloned sequences, study genetic variation at the DNA level, and the genetic and physical characterization of genomes of varying complexities, from simple viral genomes to complex eukaryotic genomes. This technique has been adapted for colony and plaque hybridizations, detection of nucleic acids deposited on a small area (dot blots and slot blots), for the detection of RNA in northern blots, and immunodetection of proteins in western blots.

The first solid support used for Southern blots was nitrocellulose (NC). Nucleic acids can be adsorbed onto nitrocellulose membranes at high salt concentrations. However, brittleness, combustibility and a relatively low binding capacity (~100 μg/cm²) for nucleic acids have given way to Nylon membranes. These membranes have a greater tensile strength than NC, and a high nucleic acid binding capacity (>500 μg/cm²). Nylon membranes bind nucleic acids through hydrophobic interactions once the hydration shells have been removed. For this reason, drying
the blot at a high temperature (80°C) after transfer is recommended to fix the DNA onto the membrane. For the same reason, the membrane should not be allowed to dry after hybridization. Doing so will allow the probe to bind to the membrane. Nucleic acids can establish covalent bonds with the nylon membrane after exposing the blot to shortwave UV light. Drying and UV cross-linking not only increases hybridization signal, but it also prolongs the life of the blot and allows multiple hybridizations with the same blot. Some nylon membranes have positive charges on the surface; these charges increase the binding capacity through interactions with the negatively charged phosphate backbone. Finally, nylon membranes should never be handled with bare hands because skin oils leave prints on the surface and produce undesirable background signal – wear gloves or handle membranes with flat tip tweezers.

The main objective of the Southern blotting is to identify a specific DNA fragment via hybridization to a labeled “probe” with a complementary sequence. This hybridization will occur if both target and probe DNA are in single stranded form. For this reason, the DNA in the agarose gel has to be denatured before it is transferred to the solid support. DNA is denatured by treating the gel with a high pH solution (NaOH) to disrupt hydrogen bonding between bases. Increasing the temperature beyond the melting point (T_m) is not a viable option as this temperature melting of the agarose gel.

The most common method of transfer uses mass flow through capillary movement. Under such conditions, the transfer efficiency of small fragments is greater than that of large ones. This is why partial de-purination of DNA in the gel is carried out by treating the DNA with a low concentration of hydrochloric acid (HCl, 0.25 N) for a short period of time (5 to 10 min). Immediately following de-purination, the DNA is treated with a strong base (NaOH, 0.4 N). This treatment accomplishes two tasks: it breaks the phosphodiester bonds of the de-purinated phosphoriboses, and it also denatures the DNA. Thus, after these treatments the DNA transfers to the solid support as relatively small single stranded molecules. The length of the de-purination treatment must be carefully controlled. A long treatment can reduce the size of DNA to a point where it lowers, or significantly reduces, its affinity for the solid support.

Southern transfers have been traditionally carried out with an upward capillary movement. However, this mode of transfer results in an increased accumulation of weight above the gel which in turn increases pressure on the gel, and reduces the transfer efficiency. A more efficient configuration is the downward capillary transfer devised by Chomczynski (1992).

Four modifications to the Southern transfer method have been introduced. One of them is the dry blot procedure in which the movement of DNA from the gel onto the solid support is entirely dependent on the liquid in the gel. A bi-directional transfer can be set up by sandwiching the gel between membranes and filter paper. These procedures are normally used with high copy sequences (plasmid digests or PCR products). Two other procedures that require some specialized equipment include electroblotting and vacuum blotting. The former is more widely used for making western blots, while the latter is preferred for the transfer of nucleic acids.
Background - Southern Hybridization

Detection of nucleic acids immobilized on a solid support is based on their ability to form double stranded molecules between complementary sequences. This idea was first proposed by Denhardt (1966) and it provided the foundation for the method developed later by Southern (1975). This technique, called “Southern hybridization,” or “DNA blot hybridization,” has been widely used for the structural characterization of DNA sequences. The conceptual framework for this type of analysis was derived from renaturation kinetics, a technique extensively used in the 1960's and 1970's to study the complexity of various genomes.

The two main criteria used to measure the effectiveness of Southern hybridization are the sensitivity and specificity of the detection. These criteria depend on a complex set of factors. The sensitivity of detection depends on the total amount of target nucleic acid used in the experiment, the nucleic acid binding capacity of the solid support, the concentration of the nucleic acid probe in the hybridization solution, the length of the hybridization period, and the specific activity of the labeled probe (amount of label/ng of probe). The binding capacity of the solid support was addressed in the previous section. The amount of target DNA is a critical factor that depends on the size of the genome of the organism under consideration. For instance, if one expects to detect a 1 Kb fragment bearing a particular sequence, then it makes a difference whether this sequence comes from an organism with a small genome, or one with a large genome. Let’s compare Arabidopsis (1C = 128 Mbp) with Maize (2700-3000 Mbp). The latter’s genome is about the same size as that of the human genome. If the limit of detection is 1 pg, and the DNA we want to detect is a single-copy sequence 500 bp long, assuming there is 100% binding to the solid support, then we need to load 256 ng of Arabidopsis DNA, and 6,000 ng of Maize DNA. Usually, three to four times those amounts are loaded in the gel. Proportional amounts are loaded for genomes of other sizes. Due to volume limitations, DNA of large genomes has to be first digested, then precipitated with ethanol, and then re-suspended in a volume that is small enough for to fit in the well of a gel.

The extent of a reaction (a hybridization in this case) depends on the reaction rate, which in turn depends on the concentration of the reactants, and the length of time the reaction is allowed to proceed. Hybridizations are normally carried out overnight (~16 hours) for genomic blots. A minimum probe concentration of 10 ng/ml is recommended for these hybridizations, as long as the specific activity of a radioactively labeled probe is above $10^8$ dpm/μg; the specific activity of non-radioactive probes is usually determined empirically via test hybridization. Traditionally, DNA has been labeled with the 32P radioisotope. In this lab however, we will use a non-radioactive labeling procedure. Polymers such as dextran and polyethylene glycol have been used as means of shortening the hybridization time by artificially increasing the effective probe concentration in the hybridization solution. These polymers interact with water very strongly making it unavailable as a solvent for DNA. Removing solvation water from the medium increases the effective DNA concentration. A higher DNA concentration results in a higher rate of hybridization.
The length of the hybridization is another factor that should be taken into consideration. Hybridizations of up to 16 to 24 h are carried out to detect single copy genes of large genomes. Usually, double stranded probes have to be denatured either by treatment with NaOH or by boiling before they are added to the hybridization solution. However, these long hybridization periods may not be productive if they proceed for a very long time as the probe strands can reanneal with each other instead of with the target on the solid support. During the hybridization period both strands can anneal to the target DNA, but as time goes by, the probe strands will also re-anneal with each other to the point that they lose their effectiveness.

The second hybridization criterion is the specificity of the probe for the target DNA. Specificity is determined by the degree of similarity between target and probe, the relative GC content of the hybridizing sequences, the temperature of the hybridization, the concentration of monovalent cations, and the concentration of organic solvents that affect the $T_m$ of double stranded DNA. Hybridizations are normally carried out at relatively high salt concentrations (~0.75 M Na$^+$), a concentration that facilitates base-pairing by stabilizing DNA duplexes. The optimum temperature for hybridization has been empirically determined to be approximately 20 to 25°C below the melting temperature of the target sequence ($T_m$). This temperature can be calculated from the empirical formula developed by Meinkoth and Wahl (1984):

$$T_m = 81.5^\circ C + 16.6(\log M) + 0.41(\% \text{ GC}) - 0.61(\% \text{ formamide}) - 500 / L$$

where $T_m$ is the melting temperature of the sequence, $M$ is the concentration of monovalent cations (Na$^+$) at less than 0.9 M, %GC is the GC content of the sequence, and $L$ is the length of the hybrid in base pairs.

A 250 bp sequence with a 35% GC content and at a 0.8 M Na$^+$ concentration would have a $T_m$ of 91.75°C. Thus, the optimum hybridization temperature would be 25°C below, or 67°C. Hybridizations are routinely carried out at 65-68°C, but more permissive hybridizations can be carried out at lower temperatures, in particular when heterologous probes are used. A Southern blot obtained with DNA from one species can be hybridized with a probe originating from a related species. The sequences are not expected to be identical, and for this reason hybridizations are carried out at lower temperatures when heterologous probes are involved. It has been estimated that on average, 1% decrease in homology between sequences decreases the $T_m$ by 1 to 1.5°C. A similar situation is encountered with multigene families; although members of a family are found on the same organism, the time of the duplication event(s) may have occurred long enough to allow them to diverge to some extent. The specificity of hybridizations can be controlled by both hybridization conditions and the stringency of the washes that are performed after the hybridization period. As the salt concentration of the washing solution is lowered, the stringency and the specificity of the probe are increased. Under the most stringent conditions blots are washed at a concentration of 20 mM Na$^+$. Less stringent conditions are used when hybridizing with a heterologous probe; for instance, using an Arabidopsis (Brassicaceae) probe on a bean (Legume) genomic blot. Hybridization temperature and stringency of hybridization may have to be determined empirically for optimum results.
Another important aspect of hybridization is the method used for labeling the probe. The three most commonly used methods for labeling DNA are nick translation, random primed labeling, and PCR. Nick translation was the first widely used method for labeling. The nick translation reaction contains both DNA polymerase and a small amount of DNase, and it is based on a DNA repair strategy. While the DNase introduces nicks in the DNA strands of the double stranded probe, the DNA polymerase detects the nicks and repairs the DNA using its 5' to 3' exonuclease activity at one end of the nick, and its 5' to 3' polymerase activity at the opposite end. As a consequence of these activities the nick is displaced and at the same time a labeled nucleotide included in the mix is incorporated into the newly synthesized DNA. The amount of labeled probe is limited to the amount of starting material.

Random primed labeling was introduced as a means to produce DNA with very high specific activity (>10^9 dpm/μg of DNA). In this procedure, the double stranded DNA probe is first denatured at high temperatures (95-100°C), and subsequently annealed to random hexanucleotides. In this way, the hexanucleotides prime the DNA for the synthesis of new strands by the Klenow fragment of \textit{E. coli} DNA polymerase (5' →3'polymerase, 3'→5' exonuclease, and no 5'→3' exonuclease activity). \textit{De novo} synthesis of DNA ensures incorporation of a labeled base added to the reaction mixture. The amount of labeled probe produced by this procedure can be doubled or tripled because the Klenow fragment can displace DNA strands positioned on the 3' side of the extension point. Finally, PCR can also be used to introduce a labeled base into DNA. After any of these labeling procedures, the labeled DNA has to be denatured before it is added to the hybridization buffer.

The synthesis of non-isotopic labeled probes has become more popular in recent years. The same methods used for the incorporation of radioactive nucleotides into DNA are used for the incorporation of non-radioactive labeled nucleotides. A common strategy is to label a nucleotide with biotin. For instance, the first type of non-isotopically labeled probes was synthesized with dUTP (to replace dTTP) linked to a biotin molecule via an allylamine linker arm. The biotin labeling strategy is based on the fact that avidin, a tetrameric protein found in egg-white, has an extremely high affinity for biotin exhibiting a very low dissociation constant (K_d = 10^{-15}). Streptavidin is a protein with similar properties and is found in some strains of \textit{Streptomyces} bacteria.

Several strategies have been developed to detect the hybridized probe. One strategy used in the detection of non-isotopically labeled probe requires three steps. In this example, the blot has been hybridized with a dUTP-Biotin labeled DNA probe. In the first step, the hybridized blot is incubated with avidin or streptavidin, and the excess protein is washed off leaving only the biotin-bound avidin. In the second step, the blot is incubated with a biotin-labeled alkaline phosphatase or horseradish peroxidase. Avidin forms a homotetramer and each monomer binds one biotin molecule. Thus, incubating the blot with the biotin-labeled enzyme will result in the binding of biotin linked to the enzyme by the avidin tetramer already bound to the DNA probe on the blot. After washing off excess enzyme, the position of the hybridization zone can be revealed via an enzyme assay with a substrate that an insoluble pigmented product, or one that
produces a chemiluminescence signal. Production of an insoluble precipitate allows for
detection of the DNA target directly on the blot, but limits the blot to a single use. On the other
hand, chemiluminescent reaction products require the use of X-ray film or an image capturing
device like a phosphoimager; this detection approach allows multiple uses for the blot.

Nitrocellulose or nylon membranes are used for this procedure because of their ability to
bind nucleic acids. This property is a double edge sword because during hybridization the
membrane can indiscriminately bind the probe, while the intention is to have the probe
specifically binding the target sequence. This problem is solved by “pre-hybridizing” the blot
with a solution that contains blocking agents such as proteins (bovine serum albumin or casein)
and denatured unrelated DNA (herring sperm DNA). These blocking agents prevent non-
specific binding of the probe and reduce or eliminate background signal. It is recommended that
pre-hybridizations be carried out for 1 to 4 hours, but extending this step doesn’t seem to be
counterproductive.

References

Chomczynski P. 1992. One-hour downward alkaline capillary transfer for blotting of DNA and
RNA. Anal. Biochem. 201, 134-139.
Denhardt D. 1966. A membrane filter technique for the detection of complementary DNA.
Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel
A. Reagents and Solutions

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Concentration</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hydrochloric Acid, HCl</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sodium Chloride, NaCl</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sodium Hydroxide, NaOH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tris Base</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sodium Phosphate, Monobasic, NaH₂PO₄·H₂O</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sodium EDTA, Na₂EDTA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Na₂EDTA·2 H₂O</td>
<td>186.1 g</td>
<td></td>
</tr>
<tr>
<td>NaOH pellets</td>
<td>19.0 g</td>
<td></td>
</tr>
<tr>
<td>H₂O add to</td>
<td>1000.0 ml</td>
<td></td>
</tr>
</tbody>
</table>

**NOTE:** Dissolve salts in 800 ml of water and stir vigorously with a magnetic bar. Adjust pH to 8.0 with 4N NaOH. Bring to final volume and autoclave.

### Stock Solutions

**EDTA, 0.5 M, pH 8.0**

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Na₂EDTA·2 H₂O</td>
<td>186.1 g</td>
</tr>
<tr>
<td>NaOH pellets</td>
<td>19.0 g</td>
</tr>
<tr>
<td>H₂O add to</td>
<td>1000.0 ml</td>
</tr>
</tbody>
</table>

**NOTE:** Dissolve salts in 800 ml of water and stir vigorously with a magnetic bar. Adjust pH to 8.0 with 4N NaOH. Bring to final volume and autoclave.

**Sodium Hydroxide, 4 N**

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaOH</td>
<td>16.0 g</td>
</tr>
<tr>
<td>H₂O add to</td>
<td>100.0 ml</td>
</tr>
</tbody>
</table>

### Working Solutions

**Depurination Solution, 10X**

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCl (conc)</td>
<td>206.0 ml</td>
</tr>
<tr>
<td>H₂O</td>
<td>794.0 ml</td>
</tr>
<tr>
<td>Final volume</td>
<td>1,000.0 ml</td>
</tr>
</tbody>
</table>

**Denaturation Solution, 5X**

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl</td>
<td>175.0 g</td>
</tr>
<tr>
<td>NaOH</td>
<td>80.0 g</td>
</tr>
<tr>
<td>H₂O to final vol</td>
<td>1,000.0 ml</td>
</tr>
</tbody>
</table>

**Neutralization Solution, pH 7.8**

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl</td>
<td>40.0 g</td>
</tr>
<tr>
<td>Tris</td>
<td>61.0 g</td>
</tr>
<tr>
<td>HCl (3 M)</td>
<td>111.0 ml</td>
</tr>
<tr>
<td>H₂O to final vol</td>
<td>1,000.0 ml</td>
</tr>
</tbody>
</table>

**NOTE:** Dissolve solutes in 80% of final vol. Adjust to pH 7.8 with HCl.
Transfer Solution, 5X

<table>
<thead>
<tr>
<th>Component</th>
<th>175.0 g</th>
<th>350.0 g</th>
<th>700.0 g</th>
<th>3 M</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl</td>
<td>4.0 g</td>
<td>8.0 g</td>
<td>16.0 g</td>
<td>0.1 N</td>
</tr>
<tr>
<td>NaOH</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H₂O to final vol</td>
<td>1,000.0 ml</td>
<td>2,000.0 ml</td>
<td>4,000.0 ml</td>
<td></td>
</tr>
</tbody>
</table>

SSPE 20X, pH 7.4

<table>
<thead>
<tr>
<th>Component</th>
<th>175.3 g</th>
<th>350.6 g</th>
<th>701.2 g</th>
<th>3.0 M</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl</td>
<td>27.6 g</td>
<td>55.2 g</td>
<td>110.4 g</td>
<td>0.2 M</td>
</tr>
<tr>
<td>NaH₂PO₄.H₂O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDTA, 0.5 M</td>
<td>40.0 ml</td>
<td>80.0 ml</td>
<td>160.0 ml</td>
<td>20 mM</td>
</tr>
<tr>
<td>NaOH, 4N</td>
<td>50.0 ml</td>
<td>100.0 ml</td>
<td>200.0 ml</td>
<td></td>
</tr>
<tr>
<td>H₂O to 1,000.0 ml</td>
<td>2,000.0 ml</td>
<td>4,000.0 ml</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**NOTE:** Dissolve salts in 80% of final vol. Adjust pH to 7.4 with NaOH (4 N).

SSC 20X, pH 7.0

<table>
<thead>
<tr>
<th>Component</th>
<th>175.0 g</th>
<th>350.0 g</th>
<th>700.0 g</th>
<th>3.0 M</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl</td>
<td>88.0 g</td>
<td>176.0 g</td>
<td>352.0 g</td>
<td>0.3 M</td>
</tr>
<tr>
<td>NaCitrate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H₂O to final vol</td>
<td>1,000.0 ml</td>
<td>2,000.0 ml</td>
<td>4,000.0 ml</td>
<td></td>
</tr>
</tbody>
</table>

**NOTE:** Dissolve solutes in 80% of final vol. Adjust pH to 7.0 with a few drops of NaOH.

B. Labware and Supplies

- Blotting Paper
- Gel Trays
- Nylon Membrane
- Blotting paper
- Ruler
- Scalpel
- Tanks
- Saran Wrap

C. Equipment

- Oven
- UV-crosslinker
SOUTHERN BLOTTING PROCEDURE

I. GEL PREPARATION
1. After photography, place gel in a clean plastic sandwich container and rinse the gel with water one more time.
   **NOTE 1:** WEAR CLEAN GLOVES TO HANDLE THE GEL.
   **NOTE 2:** AVOID SCRATCHING THE TRANSILLUMINATOR! Use Kim Wipes to clean its surface.
2. Add 100 ml of **1X depurination solution** and incubate gel for 5 minutes with gentle agitation, and discard solution.
3. Add 100 ml of **1X denaturation solution** and incubate gel for 15 minutes with gentle agitation. Repeat treatment with **1X denaturation** for another 15 min and discard solution.
   **NOTE 3:** This is a good time to start preparing the membrane (Section II), and the transfer (Section III).
4. Add 100 ml of **transfer solution** to the gel and incubate for 5 min.

II. MEMBRANE PREPARATION
1. Label membrane on lower left corner with a fine-point Sharpie marker (Group#).
   **NOTE:** Handle the membranes with CLEAN tweezers and gloves only, and a clean surface (Kim-Wipe).
2. Wet membrane in dH₂O for at least 20" & soak in **transfer buffer** for at least 15' prior to transfer.

III. CAPILLARY TRANSFER (Wear CLEAN gloves).
1. Place a stack of filter paper on the bench where the blotting is going to take place. Set A filters are cut about 2 cm larger than the gel on each side.
2. Lift gel from its tray and place it on a clean acetate sheet or piece of Saran Wrap.
3. Pour a few drops of **transfer solution** over the gel, and place the membrane on top of the gel. The marking should be on top. Roll a clean glass pipette over the membrane to eliminate bubbles that may have formed between the membrane and the gel. The MW marker should be on the left and the wells should be on the top. The label on the lower left corner of the membrane should be over the lower left corner of the gel.
4. Take one sheet of filter paper from Set A, wet it well in **transfer solution**, and carefully place it over the membrane. Roll pipet over to eliminate bubbles under it again.
5. Turn assembly over and place it on top of the stack of filter paper. Roll pipet over the gel to ensure that there are no bubbles trapped underneath.

**NOTE:** This operation can be done by first lifting the film under the gel. Place both thumbs under the acetate sheet (or Saran Wrap) and the finger tips over the wet filter paper. Lift the assembly, turn it over, and place it over the stack of filter paper.

6. Put a few drops of **transfer solution** over the gel, then, place the pre-wetted filter papers from set B (2 sheets slightly bigger than the gel) on top of the gel. Roll pipet over to eliminate bubbles.

7. Place tank near one side of the stack & pour in **transfer solution** to about 1 cm from the rim.

8. Wet the 2 sheet-wick in **transfer solution** and place it over the stack aligning the corners of one narrow end of the wick with the side corners of the gel that are away from the tank. Make sure that the wick does not touch any material placed beneath the gel; otherwise a short circuit will be formed. This will decrease the transfer efficiency, or eliminate it. The other end of the wick should be submerged in the tank.

9. Top off the tank with **transfer solution**, and cover the wick with an acetate sheet or with Saran Wrap. Allow transfer to proceed for at least 2 h or overnight.

10. After the transfer, remove the wick and the gel.

11. Lift the membrane and place it in a clean sandwich box. Wash it in 100 ml of **5X SSC** for 5 min with gentle agitation.

12. Place membrane over a dry filter paper with the DNA side up and allow the excess liquid to drain, but do not allow the membrane to dry out (about 1-5 min).

13. Affix DNA to nylon membrane by heating it to 80° C for 1 hour.

**NOTE:** This treatment will create covalent bonds between the membrane and the DNA.

14. Dry the membrane.
SOUTHERN HYBRIDIZATION - MATERIALS AND SUPPLIES

A. BIOLOGICAL MATERIALS

DNA probes
Primers
Southern Blots

B. REAGENTS

<table>
<thead>
<tr>
<th>Material</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agarose</td>
<td>Phosphoric acid, 85%</td>
</tr>
<tr>
<td>EDTA Na₂.2 H₂O</td>
<td>Phototope - Star Detection Kit, NEB</td>
</tr>
<tr>
<td>Formamide</td>
<td>Sodium Citrate</td>
</tr>
<tr>
<td>MgCl₂ 50 mM</td>
<td>Sodium Chloride</td>
</tr>
<tr>
<td>N-lauroylsarcosine</td>
<td>Sodium Dodecyl Sulphate</td>
</tr>
<tr>
<td>NaOH</td>
<td>Taq DNA Polymerase, 5 U/μl</td>
</tr>
<tr>
<td>NEBlot Phototope Kit, NEB</td>
<td>Tris base</td>
</tr>
<tr>
<td>PCR buffer 10X</td>
<td></td>
</tr>
</tbody>
</table>

C. SOLUTIONS

Stock Solutions

EDTA, 0.5 M, pH 8.0

Na₂EDTA.2 H₂O 186.1 g
NaOH pellets 19.0 g
H₂O add to1000.0 ml

NOTE: Dissolve salts in 800 ml of water and stir vigorously with a magnetic bar. Adjust pH to 8.0 with 4N NaOH. Bring to final volume and autoclave.

Sodium Dodecyl Sulphate, 20 %

SDS 40.0 g
H₂O add to200.0 ml

NOTE 1: Wear a dust mask to handle SDS.

NOTE 2: Dissolve in a beaker, stirring with a magnetic bar, in 80% of final volume. This process can be speeded up by either stirring over a hot plate (low heat) or brief warming in a microwave oven while avoiding foam. Carefully pour solution into a graduated cylinder, rinse beaker with water and pour into cylinder until reaching final volume. Cover cylinder with parafilm, or any other sealing film, and mix before pouring into storage bottle.
Southern

Sarkosyl 20%
N-lauroylsarcosine 20.0 g
H₂O add to 100.0 ml
NOTE: Same precautions as SDS

SSC 20X, pH 7.0
NaCl 175.0 g 350.0 g 700.0 g 3.0 M
NaCitrate 88.0 g 176.0 g 352.0 g 0.3 M
H₂O to final vol 1,000.0 ml 2,000.0 ml 4,000.0 ml
NOTE: Dissolve solutes in 80% of final vol. Adjust pH to 7.0 with a few drops of NaOH.

TPE Electrophoresis Buffer, 10X
Tris base 108.0 g
Phosphoric acid, 85% 15.5 ml
EDTA.Na 0.5 M 40.0 ml
H₂O add to 1000.0 ml

Working Solutions

Prehybridization solution:
SSC 20X 250 ml 5X SSC
SDS 20% 1 ml 0.02% (w/v)
Sarkosyl 20% 5 ml 0.1% (w/v)
Blocking reagent 10 g 1%
Formamide 500 ml 50% v/v

2X SSC/0.1% SDS
SSC 20X 100 ml
SDS 20% 5 ml
Water 895 ml

0.5X SSC/0.1% SDS
SSC 20X 25 ml
SDS 20% 5 ml
Water 970 ml

D. LABWARE AND SUPPLIES

Filter Paper
Pipet Tips
Plastic Bags
Polypropylene Tubes, 50 ml
Sandwich Boxes
Saran Wrap
Southern

X-ray Film
X-Ray Film Cassettes

E. EQUIPMENT

Agarose Gel Electrophoresis Equipment
Hybridization Oven
Orbital Shaker
Pipets
Power Supplies
SOUTHERN HYBRIDIZATION PROCEDURE

I. PROBE LABELING. The probe will be labeled with a biotinylated base. The extent of biotinylatio will be tested before the probe is used in hybridization. The probe will be prepared by PCR amplifying the insert from the clones made in class. The PCR products will be cleaned in Qiagen columns, the PCR product quantified by fluorometry, and the probe will then be “labeled” with biotinylated bases using a ‘random primer labeling kit.’

A. Probe Synthesis Via PCR

1. Thaw all reagents for a PCR reaction.

2. Set up the following reactions in a 200 μl PCR tube:

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>H&lt;sub&gt;2&lt;/sub&gt;O</td>
<td>20.00 μl</td>
</tr>
<tr>
<td>PCR Buffer, 10X</td>
<td>5.00 μl</td>
</tr>
<tr>
<td>MgCl&lt;sub&gt;2&lt;/sub&gt; 15 mM</td>
<td>5.00 μl</td>
</tr>
<tr>
<td>dNTP stock, 2 mM</td>
<td>5.00 μl</td>
</tr>
<tr>
<td>Primer F/R, 1 μM each</td>
<td>5.00 μl</td>
</tr>
<tr>
<td>Taq DNA Polymerase, 0.25 u/μl</td>
<td>5.00 μl</td>
</tr>
<tr>
<td>Template pDNA (2 pg/μl)</td>
<td>5.00 μl</td>
</tr>
<tr>
<td><strong>Total Volume</strong></td>
<td><strong>50.00 μl</strong></td>
</tr>
</tbody>
</table>

**NOTE:** Mix well and briefly spin tubes before placing them in the thermocycler.

3. Set up the thermocycler to the following program

<table>
<thead>
<tr>
<th>Step</th>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial denaturation</td>
<td>95°C</td>
<td>2 min</td>
</tr>
<tr>
<td>Denaturation</td>
<td>95°C</td>
<td>30 sec</td>
</tr>
<tr>
<td>Annealing</td>
<td>52°C</td>
<td>30 sec</td>
</tr>
<tr>
<td>Primer Extension</td>
<td>72°C</td>
<td>1 min</td>
</tr>
<tr>
<td>Final Extension</td>
<td>72°C</td>
<td>10 min</td>
</tr>
</tbody>
</table>

**NOTE:** make sure to activate the heater for the thermocycler’s lid.

B. Clean PCR-Amplified Probe

1. Transfer the contents of the PCR tube (50 μl) to a 1.5 ml microcentrifuge tube, and add 250 μl of PBI buffer and mix well.

2. Transfer the mix to the Qiagen column mounted in a decapped 2 ml tube.

3. Spin for 60 seconds, discard flow-through, and place column back in the 2 ml tube.

4. Wash the column-bound amplicon by adding 750 μl of PE buffer and spinning for 60 seconds.
5. Discard flow-through, put the column back in the tube and spin one more time for 60 seconds to dry the column.

6. Transfer the column to a new clean 1.5 ml microcentrifuge tube. Apply 20-50 μl of elution buffer directly to the center of the disc, and wait 2 min before spinning the tube for 60 seconds.

C. Quantify the Probe by Fluorimetry
   Proceed as protocol on page 2.1.16, but use 0.1 μg/ml of Hoechst 33258 reagent.

D. Probe Labeling (North2South Random Prime Labeling Kit, Thermo Scientific)
1. Transfer 24 μl of DNA probe (5 ng/μl) to the 1.5 ml tube containing 10 μl of 5X Heptanucleotide Mix (Random Primers).
   NOTE: A control sample will be prepared for the class
2. Denature DNA at 99°C for 5 minutes. Use Lid Lock.
3. Immediately quench on ice-water mix for 5 min.
4. Briefly spin tube to collect all liquid at the bottom.
5. Transfer DNA to the 200 μl PCR tube containing 16 μl of the labeling mix. Mix well by vortexing and spinning down for 5”.

   Labeling Mix
   - dNTP mix 10 μl (dNTPs + Biotin-11-dUTP)
   - Reaction Mix 5 μl
   - Klenow (DNA Pol) 1 μl

6. Incubate in Thermocycler at 37°C for 60 min.
7. Stop reaction by adding 2 μl of EDTA 0.5 M.

E. Clean Labeled Probe
   Proceed as in step B above.

F. Measure Extent of Probe Biotinylation.
1. Hydrate/equilibrate a positively-charged nylon membrane in TE Buffer for at least 10 minutes.
2. Make a set of serial dilutions of the biotinylated probe in TE buffer. Pipette 20μL TE Buffer into a micro centrifuge tube labeled #1 and 10μL into micro centrifuge tubes labeled #2-7 for the control reaction and each experimental sample. Total of 14 tubes.

3. Prepare dilutions of the labeled DNA from both the control DNA and your experimental probe with TE. Dilute the labeled samples 10-fold in TE Buffer by adding 2μl of the stated samples into the #1 tubes and serially diluting the sample into the other tubes using 10μL per dilution.

4. Place the equilibrated membrane onto a clean, dry paper towel. Allow excess buffer to absorb into the membrane, but do not allow the membrane dry out.

5. Spot 2μL of samples and standards onto the hydrated membrane. Allow the samples to absorb into the membrane.

6. Cross-link the DNA with a UV source.

G. Detection
1. Add 20ml of the Blocking Solution, incubate for 30’ with moderate shaking, and discard.

2. Pipet 1 ml of Blocking Solution bathing the blot into a 1.5 ml tube and add 65 μl of Streptavidin-HRP. Mix well and pipet mix back into the bathing solution. Incubate for 15 min and discard.

3. Wash blot 4 times for 5 min each with 20 ml of Washing Solution. Use moderate shaking.

4. Place the membrane in a clean container and incubate for 5 minutes with Substrate Equilibration Buffer. Use moderate shaking.

5. Place membrane on a small piece of Saran Wrap laid flat on the bench. Pipet 5 ml of Substrate Working Solution over the membrane ensuring the entire membrane is covered.

6. Drain membrane and lay flat on another piece of Saran Wrap. Wrap the membrane avoiding wrinkles and bubbles. Use a Kim-Wipe to eliminate any excess liquid that may come out at the edges. Finish folding wrap and expose blot in image capturing device.

7. Add 10ml of the Blocking Solution, incubate for 5’ with moderate shaking, and discard.

II. PRE-HYBRIDIZATION
1. Wet the Southern blot in water, drain and place it inside a 50 ml conical polypropylene centrifuge tube. The DNA side should be facing the center of the tube and the markings towards the outside. Also, the long side of the blot should be parallel to the axis of the tube.

2. Add 5 ml of pre-warmed Hybridization Solution (North2South, Thermo Scientific). Close the tube tightly, place it in the hybridization oven and incubate for 30-60 min 55°C.
III. HYBRIDIZATION

1. Denature the probe for 5 min in a 99° C heating block.
   
   **NOTE:** Start this step after starting the pre-hyb period.

2. Immediately quench the probe on an ice-water mixture for at least 5 min. Spin the tube briefly to ensure all liquid is at the bottom of the tube.

3. Take the hybridization tube out of the oven and pipet the entire contents of the 1.5 ml tube directly into the hybridization buffer. Close the hybridization tube tightly.

4. Place the tube back in the hybridization oven. Allow hybridization to proceed over night (~16 h) at the same temperature as the previous step.
   
   **NOTE 1:** Notice the tube clamps on the rotisserie are placed at a small angle from the horizontal line. This angle will ensure the buffer will move from one end of the tube to the other, in addition to the circular motion, during hybridization and ensure complete coverage of the membrane.

5. Discard the hybridization solution at the end of the hybridization period, but keep the membrane in the tube.

6. Wash membrane three times, for 15 min each time, with 10 ml of **1X Stringency Wash Buffer** (2X SSC/0.1% SDS) at the same hybridization temperature.

7. Remove membrane from tube and proceed with the Detection procedures as described in **Section G**.
Lab Report: Southern Blotting & Hybridization

Part I.

Summary (One Page Maximum. 20 points)
a) The main objective of the experiment,
b) The most important findings,
c) The significance of the results, and
d) Major conclusions

Part II. Results (40 points)
Summary of the results obtained in the lab. Include the following:
a) Electrophoresis:
   Picture of the gel. Label the lanes. Either number the lanes and indicate their identity in a
   legend, or label the lanes with the name of the samples. Indicate the following too:
   - Samples
   - Restriction Enzyme
   - Agarose concentration
   - Power (Volts)
   - Electrical Field (V/cm)
   - Run time
b) Southern Blotting
   Conditions
c) Hybridization
   Probe preparation
   - Amount of probe obtained
   - Amount of probe labeled
   - Enzyme used in labeling reaction
   - Amount of probe used in hybridization
d) Hybridization patterns of the different probes
   - Describe the hybridization pattern for each of the probes
   - Compare the hybridization spectra of each of the probes
   - Do all plants from a single species have the same number of genes?

Part III.

1. Consider 2 genes A and B in the cereals. Gene A produces protein SpepA in Sorghum and 
   MpepA Maize. These proteins have been purified from both species, and have been 
   sequenced. A comparison of their amino acid sequences show 97% similarity. However, 
   only the gene from Sorghum has been cloned. In contrast, no protein information is available 
   from gene B, but it has been cloned in both species and the DNA sequences show 94% 
   sequence similarity. Duplicate Southern blots with restriction enzyme digestions of genomic 
   DNA from the two species have been produced using six different restriction enzymes. 
   These blots have been hybridized with the Sorghum clone of gene A and the Maize clone of 
   gene B. Hybridization with Sorghum clone A shows a very strong one-band signal in all 
   lanes that have Sorghum DNA, but only one band with a faint signal in lanes containing
Maize DNA. In contrast, hybridization with the Maize clone from gene B shows a strong 1-band signal for both species.
a) (20) How do you explain the difference in the hybridization signals obtained with the two probes?
b) (20) Considering the genome size differences, how would explain the fact that both species give the same amount of signal with the probe for gene B?
CHAPTER 8

ANALYSIS OF GENE EXPRESSION

Learning Objectives............................. 8.1
Lab Objectives..................................... 8.1
Background ......................................... 8.1
RNA Isolation - Mat. & Suppl.............. 8.6
RNA Isolation Procedure...................... 8.9
Denaturing Agarose - Mat. & Suppl.. 8.11
Denaturing Agarose - Procedure ....... 8.13
RT- PCR Background......................... 8.15
RT- PCR - Lab Procedure ................. 8.19
Study Questions............................... 8.20
Lab Report....................................... 8.21
ANALYSIS OF GENE EXPRESSION

Learning Objectives

After completion of this unit students will be able to identify the factors that may affect the content of the mRNA population in a tissue at any given time, as well as the factors that affect RNA quality during extraction. Students will also be able to identify the different criteria and methods used in the proper evaluation of plant RNA. Finally, students will also be able to set up experiments to evaluate differential gene expression.

Lab Objectives

Extract total RNA from leaf tissue; evaluate the quantity and quality of the RNA via spectrophotometry and denaturing agarose gel electrophoresis, respectively. Differential levels of expression of a set of genes will also be assessed using real-time reverse transcriptase - polymerase chain reaction (RT²-PCR).

Background

Chapters 2 and 7 described procedures for the isolation of genomic DNA, including methods for measuring quality and quantity, as well as, structural analysis. These methods focused on analysis of genomic DNA – DNA content and structure. However, those methods did not provide information related to gene expression. Analysis of gene expression requires the isolation of pure and intact RNA from tissues in which gene expression is being studied. Isolated RNA can be used to assess expression levels of specific sequences using a variety of methods, including “RNA sequencing,” which is a misnomer because cDNA is what is actually sequenced.

The overriding concern during RNA isolation is the presence of ribonucleases (RNases), and also ways by which they can be suppressed and/or eliminated. RNases are ubiquitous, very stable, and unlike DNAses, which require Mg²⁺ as co-factor, they do not require co-factors for activity. Thus, even if DNAses are not eliminated in the early stages of DNA isolation, their activity can be suppressed by adding a Mg²⁺-chelating agent such as EDTA. In contrast, RNAses have to be suppressed directly and eliminated during extraction. Sources of RNases during RNA extraction are the sample, all glassware and plasticware in the lab, buffers, and human hands. RNases from labware and solutions can be eliminated by chemical treatment with an alkylating agent such as diethyl pyrocarbonate, or with mercury compounds. There are a number of commercially available chemical solutions that can eliminate these enzymes. Baking
at 200° C for at least 4 hours can also eliminate RNases from heat-resistant labware and instrument parts. RNAses from plant tissue can be inactivated and removed during the extraction procedure. The RNA extraction buffer normally contains a reducing agent at a concentration that reduces disulfide bonds, and a detergent that can completely denature RNAses. Proteins are then eliminated via organic extraction with phenol and chloroform. Wearing gloves is essential when working with RNA as human hands are a good source of RNAses. In addition, it is good way to protect the hands from hazardous chemicals.

In general, RNA isolation requires steps similar to those used for DNA extraction. The first is the mechanical disruption of cellular structures. This is most commonly accomplished by grinding the plant tissues in liquid nitrogen. The concern about keeping phenolic compounds from oxidizing and reacting with RNA is the same as before, and it is controlled by the use of reducing agents like β-mercaptoethanol. However, this reducing agent plays an additional role during RNA extraction. Intra-molecular disulfide bonds formed between cysteine residues stabilize the structure of RNases. Inclusion of a reducing agent, like β-mercaptoethanol, in the extraction buffer at high concentrations leads to the reduction of disulfide bonds in the RNases with the concomitant destruction of their 3D structure.

During extraction, a detergent is routinely used to inactivate proteins by denaturation. Two rounds of organic extraction are used to eliminate the denatured proteins. The first round uses phenol, which also removes remove polysaccharides (pectins in particular). Adding chloroform to the phenol increases the efficiency of the extraction. As explained in Chapter 2, nucleic acids normally partition into the aqueous phase during organic extraction. The problem here is now contamination of RNA with DNA. The physical-chemical properties of DNA are such the when the pH of the phenol is low (< 6.0), the DNA partitions into the organic (phenol) phase. Thus, phenol:chloroform (5:1) saturated with an aqueous low pH (<6.) buffer is used in the extraction. The second extraction is carried out with chloroform, and it is used to remove any remnants of proteins, polysaccharides, and phenol as well.

After the two rounds of organic extractions, 2 M lithium chloride is used to differentially precipitate RNA. DNA and other proteins and inhibitors that may still be present are left behind in the supernatant. In the next step RNA is dissolved in TE buffer before it can be precipitated with 2.5 volumes of ethanol in the presence of 0.3 M sodium acetate. RNA isolated in this procedure can be used for northern and RT-PCR analyses.

As with DNA, purified RNA has to be evaluated quantitatively and qualitatively before further analysis can be performed. Quantitation can be performed spectrophotometrically exploiting the ability of nucleic acids to have an absorption peak at 260 nm. One absorbance unit at 260 nm is equivalent to a concentration of 40 μg of RNA per ml. The major problem with spectrophotometry is the low sensitivity (no less than 2-4 μg/ml). This problem can be overcome
by the use of fluorescent dyes like RiboGreen (Molecular Probes, Inc.), which have an affinity for nucleic acids, and can detect RNA at concentrations as low as 1 ng/ml. Unfortunately, neither method of RNA quantitation is RNA specific. However, the presence of DNA could be detected in fluorometric assays by comparing readings before and after DNAse treatment of the RNA sample. The fluorescent dye only binds to nucleic acids, and not to single nucleotides. Thus, a difference in the readings would indicate the presence of DNA. Another approach to detecting DNA in RNA preparations is via PCR without the use of reverse transcriptase.

The physical integrity of RNA can be assessed via denaturing agarose gel electrophoresis. RNA, like DNA, can be separated according to size in agarose gels. RNA is single stranded and has the capacity to form intra-molecular secondary structures using Watson-Crick base pairing over short stretches. These structures prevent reliable size-dependent separation of RNA by agarose gel electrophoresis. To overcome this difficulty, RNA is separated under denaturing conditions using one of two methods: formaldehyde gels, or glyoxal/DMSO treatment. Formaldehyde reacts with amino group of guanine to yield the N2-hydroxymethyl adduct. This adduct prevents the formation of intra-molecular Watson-Crick base pairing and maintains RNA as a single stranded molecule. This adduct is not stable and requires the constant presence of formaldehyde during the electrophoresis run. Formaldehyde is a carcinogen and gel containing this chemical needs to be run in a chemical hood!

The second type of gel requires the denaturation of RNA with glyoxal, a two carbon dialdehyde. In this procedure RNA is incubated at 50°C for one hour in the presence of glyoxal (1 M) and 50 % dimethyl sulfoxide (DMSO). DMSO has the ability to disrupt the hydrogen bonds of Watson-Crick base pairs, while glyoxal works as a bi-functional electrophile. Glyoxal reacts primarily with guanine residues to form a five-membered ring fused between the amine group on C2 and N2 and bearing a pair of vicinal hydroxyl groups. This adduct is stable within the pH range of 6.5 to 7.5. For this reason, electrophoretic separation should avoid a pH outside this range.

On average, only 0.5 to 1% of total RNA extracted from plant tissues represents mRNA, the bulk, or close to 99% corresponds to ribosomal RNA, the 26s and 18s ribosomal RNAs which are approximately 3.4 a 1.8 Kb long. Thus, formaldehyde or glyoxal agarose gel electrophoresis of total RNA would display two prominent bands corresponding to each of the ribosomal RNAs. When RNA has been degraded, these bands are shown associated with a smear below them, or in the most extreme case these bands are completely absent in the gel. Thus, it is possible to have a relatively high reading in the spectrophotometer with highly degraded RNA; single ribonucleosides are still able to absorb light at 260 nm, but they are no longer part of RNA polymers.

Following size fractionation by denaturing agarose gel electrophoresis, RNA can be
transferred to a solid support (nylon membrane) where specific transcripts can be detected by hybridization with labeled probes similar to Southern hybridization. Detection of membrane bound-RNA with labeled DNA probes is called northern hybridization. To restore the RNA’s capacity to base pair with the probe during northern hybridizations, the formaldehyde adducts need to be removed. Since the formaldehyde adduct is unstable, RNA from the formaldehyde is all that is needed. However, in the case of glyoxal-derivatized RNA, the adduct can be removed by treating RNA with a pH 8.0 buffer at 65°C for 10 min; this is carried out after the transfer.

Northern analysis can provide information about transcript size and relative concentration. To determine how the steady state level of a transcript changes in response to internal (developmental) or external (environment) stimuli, it is necessary to compare the amounts of the specific transcript relative a reference standard from the same tissue which is known to be unaffected by the applied stimulus.

Consider two different scenarios in the analysis of a single transcript. In the first one, application of a stimulus increases the expression of a particular transcript, but does not alter the concentration of the other transcripts; in this case, any of the other transcripts could serve as reference. For northern analysis we will load a gel with similar amounts of RNA from the control (before) and the treatment (after). Hybridization of the northern blot with probes for the specific and reference transcripts will show an increase in hybridization signal of the specific transcript in relation to the reference. In the second scenario, the stimulus suppresses a specific transcript only; the reference transcript remains the same as before. In this case, the specific transcript will show a lower abundance with respect to the reference. The use of an “internal standard” will help eliminate variables such as an unequal loading of RNA in the control and treatment. The reference is usually selected among housekeeping gene, or total ribosomal RNA. Many other scenarios could be considered, including one where a stimulus depresses the levels all almost all genes, except a few. If those few are tested, they could appear as transcripts that have been “upregulated.”

Another approach used to measure transcript levels is run on transcription assays. In this assay, intact nuclei are first chilled, isolated, and incubated in vitro with the four ribonucleotides, including a radiolabeled U. The native RNA polymerase is allowed to carry out transcription of nuclear DNA in vitro. Afterwards, RNA is isolated and used as a probe in a reverse northern hybridization of dot blots. These blots are prepared by affixing denatured cDNAs with complementarity to the transcripts of interest. These experiments provide information about whether changes in transcript level are due to changes in transcription. In other words that amount of signal is proportional to the number of transcripts being synthesized, not the amount of transcripts that are already in the cell.

Northern hybridization is very similar to Southern hybridizations, however a few
important issues must be considered when working with RNA blots. RNases should still be considered a threat throughout the life of the northern blot. Another aspect of northern hybridizations is the stability of the RNA-DNA hybrids. An empirical formula for calculating the $T_m$ of these hybrids has been described:

$$T_m = 79.8^\circ C + 18.5(\log M) + 0.58(\%GC) + 11.8(\%GC)^2 - 0.50(\%form) - 820/L$$

where $M$ is the molarity of monovalent cations, $\%GC$ is the percentage of the combined sum of guanines and cytosines in the probe, $\%form$ is the percentage of formamide in the hybridization solution, and $L$ is the length of the hybrid in base pairs. In contrast to stripping the probe off from a Southern blot with boiling SDS solutions or NaOH, probes are stripped from northern blots with 50% formamide at 65$^\circ$C because the harsh treatments used on Southerns can damage the RNA on the northern blot.

Finally, another important aspect of this type of hybridization is the sensitivity of detection. In this regard, in Southern hybridizations the same sensitivity criterion was applied to all single copy genes because there is one copy of any given gene for every copy of the genome loaded in the gel. The same criterion cannot be applied to northern hybridizations because not all genes are expressed at the same level in all tissues. From the Arabidopsis and rice genomes, we know that an organism can encode between 25,000 and 45,000 genes, with each tissue and organ having a unique expression profile at different stages of development. For instance, in a developing seed a few transcripts encoding storage proteins can account for the bulk of the transcripts. These transcripts will be easily detectable by northern analysis. For transcripts expressed at very low levels, the sensitivity of detection could depend on the type of RNA loaded in the agarose gels. One option is to load total RNA which comprises messenger RNA (mRNA), and ribosomal RNA (rRNA). The latter comprises 99% of the total RNA isolated from the tissue. For this reason, up to 10 $\mu$g of total RNA are loaded in a gel. For a more sensitive analysis, mRNA is first purified from total RNA using oligo dT affinity chromatography. In this technique, small beads (cellulose or magnetic latex) carrying an oligomer of deoxythymidine (oligo-dT) are used to trap poly-adenylated mRNA (poly A$^+$ RNA). mRNA eluted from these beads can be quantitated and further analyzed by northern hybridization.
RNA ISOLATION – MATERIALS AND SUPPLIES

A. REAGENTS

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Reagent</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-Mercaptoethanol</td>
<td>Octanol</td>
</tr>
<tr>
<td>Chloroform, CHCl₃</td>
<td>Phenol</td>
</tr>
<tr>
<td>Diethyl Pyrocarbonate, DEP</td>
<td>Sodium Lauroyl Sarkosinate, Sarkosyl</td>
</tr>
<tr>
<td>Ethanol</td>
<td>Sodium EDTA</td>
</tr>
<tr>
<td>Glacial Acetic Acid, CH₃COOH</td>
<td>Sodium Chloride, NaCl</td>
</tr>
<tr>
<td>Hydrochloric Acid, HCl</td>
<td>Sodium Acetate, CH₃COONa</td>
</tr>
<tr>
<td>Lithium Chloride, LiCl</td>
<td>Tris Base</td>
</tr>
<tr>
<td>Nitrogen, liquid</td>
<td></td>
</tr>
</tbody>
</table>

B. SOLUTIONS

Stock solutions

Sarkosyl, 20 %

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sarkosyl</td>
<td>100 g</td>
</tr>
<tr>
<td>H₂O add to</td>
<td>500 ml</td>
</tr>
</tbody>
</table>

NOTE 1: Wear a dust mask to handle Sarkosyl.

NOTE 2: Dissolve in a beaker in 80% of final volume. Stir with a magnetic bar on a stirring plate. This process can be speeded up by turning up the heat in the hot plate (low heat), or by briefly warming in a microwave oven while avoiding foam. Carefully pour solution into a graduated cylinder, rinse beaker with water and pour into cylinder until reaching final volume. Cover cylinder with parafilm, or any other sealing film, and mix before pouring into storage bottle.

EDTA, 0.5 M, pH 8.0

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Na₂EDTA.2 H₂O</td>
<td>186.1 g</td>
</tr>
<tr>
<td>NaOH pellets</td>
<td>19.0 g</td>
</tr>
<tr>
<td>H₂O add to</td>
<td>1000.0 ml</td>
</tr>
</tbody>
</table>

NOTE: Dissolve salts in 800 ml of water and stir vigorously with a magnetic bar. Adjust pH to 8.0 with 4N NaOH. Bring to final volume and autoclave.

Tris.HCl, 1M, pH 7.6

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris base</td>
<td>121.1 g</td>
</tr>
<tr>
<td>HCl</td>
<td>62.6 ml</td>
</tr>
<tr>
<td>H₂O add to</td>
<td>1000.0 ml</td>
</tr>
</tbody>
</table>
**Sodium Chloride, 5 M**

NaCl 292.0 g  
H₂O add to 1000.0 ml

**Working Solutions**

**Extraction Buffer, 1.25 X**

Tris.HCl, 1 M, pH 7.6 12.50 ml  
NaCl, 5 M 10.00 ml  
EDTA.Na, 0.5 M, pH 8.0 1.25 ml  
Sarkosyl, 20% 12.50 ml  
H₂O adjust to 100.00 ml

**NOTE 1:** Add 50.00 ul DEP, mix and incubate overnight. Autoclave for 15 min. Allow solution too cool, then swirl bottle to release CO₂.  
**NOTE 2:** Add β-Mercaptoethanol to 1% prior to use.

**Sodium Acetate 3M, pH 6**

NaAcetate.₃ H₂O 40.80 g  
H₂O to 100.00 ml  

**NOTE 1:** Dissolve salt in 80.0% final volume, adjust pH to 6.0 w/ Glacial AcAcid, and then bring to final vol with H₂O  
**NOTE 2:** Add 50.00 ul DEP, mix and incubate overnight. Autoclave for 15 min. Allow solution too cool, then swirl bottle to release CO₂.

**TE Buffer**

Tris.HCl 1 M, pH 7.6 1.00 ml  
EDTA, Na 0.5 M, pH 8.0 0.20 ml  
H₂O adjust to 100.00 ml  

**NOTE:** Add 50.00 ul DEP, mix and incubate overnight. Autoclave for 15 min. Allow solution too cool, then swirl bottle to release CO₂.

**Lithium Chloride,**

<table>
<thead>
<tr>
<th></th>
<th>2 M</th>
<th>12 M</th>
</tr>
</thead>
<tbody>
<tr>
<td>LiCl</td>
<td>34.0 g</td>
<td>101.7 g</td>
</tr>
<tr>
<td>H₂O add to</td>
<td>400.0 ml</td>
<td>200.0 ml</td>
</tr>
</tbody>
</table>

**NOTE 1:** Filter through Whatman No 1 and then thru 0.45 μm filter.  
**NOTE 2:** Add 100.00 μl DEP, mix and incubate overnight. Autoclave for 15 min. Allow solution too cool, and then swirl bottle to release CO₂.

**Phenol**

Melt crystallized re-distilled phenol in a 65°C water bath.  
Add 8-Hydroxyquinoline to a concentration of 0.1%  
Saturate with DEP-treated H₂O, remove most of the water leaving 1 cm layer on top  
Add 1/5 volume of CHCl₃
Store in light-tight bottle at 4° C for no more than 1 month.

**Chloroform**  
CHCl₃  960.00 ml  
Octanol  40.00 ml

**Ethanol.** Keep at -20° C

**C. LABWARE AND APPLIES**

**Plasticware.**  
Centrifuge bottles (250 ml)  
Centrifuge tubes (50 ml)  
Pipet tips (10, 200, & 1000 ul)  
Micro centrifuge tubes (200, 500, 1500 ul).  
Spoons (disposable, do not autoclave, just let it dry out at RT°)  
**NOTE:** Inactivate contaminating nucleases on the surface of all plasticware by rinsing it in 0.1% DEP, and then autoclaving for 15 min.

**Glassware.**  
Mortar and Pestle  
Graduated Cylinders (100 ml)  
Pipets (5, 10, & 25 ml)  
**NOTE:** All glassware, including pipets and mortar and pestles, must be baked at 200° C for at least 4 h or overnight.

**Pipeting Devices.**  
Pipet bulbs  
Pipet Pumps  
Pipetmans  
**NOTE:** Treat the surface of these devices with RNase AWAY™ Make sure that shaft and tip ejector of Pipetmans is RNase-free.

**Supplies**  
Filter 0.45 μm  
Pipet tips  
Whatman # 1 filter paper

**D. EQUIPMENT**  
Centrifuge, High Speed.  
Dewar Flask  
Freezer, -80° C
PROCEDURE FOR EXTRACTING TOTAL RNA

In the following lab sessions we will carry out an experiment to examine changes in gene expression in *Phaseolus vulgaris*, the common bean, in response to inoculation with *Bean common mosaic necrosis virus*, a potyvirus. The cultivar Calima is of Andean origin and it is susceptible to the virus. In contrast, the Mesoamerican landrace Jamapa displays a hypersensitive reaction (HR). This reaction has been associated with a resistance mechanism in plants. A sample for a wild accession from Argentina, which is susceptible, will also be included in the experiment. The control will comprise plants that have been mock inoculated with just buffer, and the treated plants will have been inoculated with the NL3 strain of BCMNV. A total of six RNA samples will be extracted from these bean lines.

1. Harvest a 4 g leaf sample, quickly place it in a mortar, and immediately pour in N₂ (l).

   **NOTE:** Wear disposable gloves to collect samples and throughout the entire extraction procedure in order to avoid contamination of samples with RNase from the skin. Change gloves if you touch door knobs, refrigerator handles, etc., RNAses are everywhere!

2. Grind sample with a pestle until a fine powder is obtained.

   **NOTE:** Do not pound on the mortar with the pestle. Grind applying a shearing force with circular motion of the pestle.

3. Transfer powder to a pre-labeled centrifuge bottle that contains 4 volumes of *extraction buffer*, and immediately mix the sample in the buffer, thoroughly.

   **NOTE 1:** Use an RNAse-free disposable spoon or spatula to transfer the powder.

   **NOTE 2:** Do not allow the powder to melt before it contacts the extraction buffer. However, the powder should not be added when the powder is cold enough to freeze the buffer. If this happens, the powder will not wet then uniformly. When pockets of the sample thaw in the absence of buffer, polyphenols will oxidize and cross-link nucleic acids and proteins.

4. Add 1 vol of *phenol* to the homogenate and mix well to obtain an emulsion.

6. Separate phases by centrifugation at 4,000 x g and 4°C for 15 min.

7. Transfer the aqueous phase into a new 50 ml centrifuge bottle and re-extract with an equal vol of *CHCl₃*. Separate phases as above.

   **NOTE:** The aqueous phase contains RNA and small water soluble molecules (sugars, amino acids, pigments & some polysaccharides).

8. Transfer aqueous phase to a 35 ml Oak Ridge tube. Centrifuge at 12,000 rpm and 4°C for 20’.

9. Pour the supernatant to a new 35 ml Oak Ridge tube, add 1/5 vol of 12 M *LiCl*, mix well, and incubate in ice-water for 30 min.

10. Pellet RNA at 12,000 rpm and 4°C for 30’. Discard supernatant.
11. Wash the RNA precipitate in 30 ml of ice-cold 2 M LiCl. First disperse the ppt in 2-5 ml, and then bring up to final volume. Pellet RNA for 30 min at 12,000 rpm and 4º C.

**NOTE 1:** If the RNA pellet is still dirty, it will need to be washed again with 2 M LiCl.

**NOTE 2:** After removing the supernatant, spin the tubes one more time for just 1 min to bring down all liquid adhered to the tube walls. Eliminate this liquid with a micropipet.

12. Dissolve RNA pellet in 9 ml of TE buffer, add 1 ml of 3 M NaAcetate, mix well, and then add 25 ml of ice-cold EtOH. Mix well and store at -20º C for at least 1 h or overnight.

13. Pellet RNA at 12,000g for 20 min at 4º C. Discard the supernatant and dry pellet in vacuum, or air dry.

14. Determine RNA quantity and quality

   a) Determine RNA concentration of 1/50 dilution in spectrophotometer. \(1 \text{A}_{260} = 40 \mu \text{g/ml}\). DNA contamination can lead to an overestimation of RNA concentration.

   b) Chemical purity. \(\text{A}_{260}/\text{A}_{280}\) ratio should be 1.8 -2.0 for high quality RNA. Smaller ratios are indicative of protein and/or polyphenol contamination.

   c) Denaturing Agarose Gel Electrophoresis can reveal the intactness of the isolated RNA.

**NOTE:** Ribosomal RNA constitutes approximately 99% of the total RNA. Size fractionation and staining of RNA in agarose gels will reveal two prominent sharp bands corresponding to the 25S and 18S ribosomal RNAs. A light smear similar to that of digested DNA is also visible between these bands and corresponds mostly to mRNA. The lost of sharpness, or an increase in intensity of the “smear” towards the bottom of the gel will indicate the RNA has been degraded.

**Spectrophotometric Readings**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Dilution</th>
<th>(\text{A}_{260})</th>
<th>(\text{A}_{280})</th>
<th>(\text{A}<em>{260}/\text{A}</em>{280})</th>
<th>(\mu \text{g/ml})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
DENATURING AGAROSE GEL ELECTROPHORESIS – MATERIALS AND SUPPLIES

A. REAGENTS AND SOLUTIONS

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agarose</td>
<td>Mix-bed resin AG 501-X8 (Bio-Rad)</td>
</tr>
<tr>
<td>Bromo Phenol Blue</td>
<td>RNase ZAP™ (Ambion®)</td>
</tr>
<tr>
<td>Dimethyl Sulfoxide, DMSO</td>
<td>Sodium Phosphate, Monobasic</td>
</tr>
<tr>
<td>Ethidium Bromide</td>
<td>Sodium Phosphate, Dibasic</td>
</tr>
<tr>
<td>Glycerol Mol. Biol. Grade</td>
<td>Xylene Cyanole</td>
</tr>
<tr>
<td>Glyoxal</td>
<td></td>
</tr>
</tbody>
</table>

B. SOLUTIONS

Stocks

NaH₂PO₄, 1M
NaH₂PO₄. H₂O 13.8 g
H₂O add to 100.0 ml

Na₂HPO₄, 1M
Na₂HPO₄.7H₂O 26.8 g
H₂O add to 100.0 ml

Working Solutions

Sodium Phosphate Buffer 10X, 100 mM, pH 7.0
NaH₂PO₄, 1M 10.0 ml
Na₂HPO₄, 1M 10.0 ml
H₂O add to 200.0 ml

Sodium Phosphate Buffer 1X, 10 mM, pH 7.0
NaPO₄ 10X 100 ml
H₂O 900 ml

Glyoxal, 40%
NOTE: Glyoxal oxidizes upon contact with air. De-ionize immediately before using it. Combine 0.5 g of Bio-Rad Mix-Bed Resin AG 501-X8 and 5 ml of glyoxal in a Falcon tube and vortex for a couple of minutes. Filter through a Bio-Rad Econo column. Check with pH strip the first eluate drops are >5. If pH < 5, then repeat de-ionization procedure.
Gene Expression

**Glyoxal Loading Buffer**
- NaPO₄ 10X, pH 7 1.00 ml
- H₂O (DEPC-treated) 4.00 ml
- Glycerol Mol. Biol. Grade 5.00 ml
- Br. Phenol Blue 25.00 mg
- Xylene Cyanole 25.00 mg

**Total Volume** 10.00 ml

**C. LABWARE AND SUPPLIES**
- Erlenmeyer flask, 250 ml
- Kim Wipes
- Staining tray

**D. EQUIPMENT**
- Electrophoresis Apparatus
- Microcentrifuge
- Micropipet
- Microwave
- Orbital Shaker
- Power Supply
- Water bath, 50°C
PROCEDURE FOR DENATURING AGAROSE GEL ELECTROPHORESIS

The class is going to share the Agarose RNA samples so that a single gel with six samples will be run. The bean leaf samples will be obtained from control and virus inoculated seedling of Calima, Jamapa, and a wild accession.

NOTE: WEAR GLOVES AT ALL TIMES WHEN WORKING WITH RNA.

I. Sample preparation

1. Add to a microcentrifuge tube the following:
   - RNA + H2O: 5.0 ul between 0.2 and 5 ug of Total or polyA+ RNA
   - NaPO4: 100 mM; 2.0 ul
   - DMSO: 10.0 ul
   - Glyoxal 6M: 3.0 ul
   
   **NOTE 1:** Mix well, spin down briefly, and incubate samples at 50o C for 1 hour.
   
   **NOTE 2:** Prepare the gel while the samples are denatured.

2. Cool samples on ice. Add 5 μl of Glyoxal Gel Loading Buffer before loading the samples.

II. Gel Preparation

1. Clean the gel apparatus free of RNases with RNase ZAP™. Spray RNase ZAP™ on the gel bed, gel tank and combs and wait for 5 minutes before scrubbing the surfaces with 2-3 Kim Wipes. Rinse all surfaces 3 times with DEPC-treated water. Dry with additional Kim Wipes, assemble the unit and place it in the chemical fume hood.

2. Add the following to an RNase-free 250 ml Erlenmeyer flask:
   - Agarose: 300 mg
   - PO4 Buffer, 1X: 30 ml

3. Record weight of flask with contents. Heat slurry in microwave until agarose is completely dissolved. Place flask on a balance and bring to the original weight with DEPC-treated water.

4. Pour the molten agarose on the gel bed and allow it to set.

5. Remove the comb by carefully pulling it up with a left-right rocking motion. Fill the electrophoresis tank with 1X PO4 Buffer to just barely cover the gel. Make sure there are no bubbles trapped in the wells.
III. Electrophoresis

1. Load the samples including an RNA molecular weight marker.

2. Run samples at 5 V/cm until front dye reaches 2/3 of the length of the gel. Stop the power supply, disconnect the leads and mix the buffer in the tank every 15 min until done.
   
   **NOTE:** Glyoxal gels use a very low ionic strength electrophoresis buffer and yield sharper RNA bands. However, this buffer gets exhausted very quickly causing local rise in pH, which in turn induces the dissociation of glyoxal and RNA. To prevent this problem buffer should be re-circulated, or mixed at frequent intervals.

3. Transfer the gel to a staining tray and stain with ethidium bromide (0.5 μg/ml) for 10 minutes. De-stain and photograph the gel.

   **NOTE:** Wear gloves while working with ethidium bromide. It is a carcinogen.
RT-PCR

The enzymatic amplification of RNA is another technique used to evaluate the expression of specific sequences. This is a powerful technique because it has the ability to detect relatively small amounts of RNA. For this reason, it is used when total amounts of RNA are limited, or when attempting to detect extremely rare transcripts. Amplification of RNA using the reverse transcriptase-polymerase chain reaction (RT-PCR) is based on coupling the activity of two enzymes: a reverse transcriptase, and a thermostable DNA polymerase. The reactions catalyzed by these enzymes can be carried out in two separate steps; however, the entire procedure can be carried out in a single tube with some commercial kits. One approach is to use enzymes with dual function like the Tth reverse transcriptase/DNA polymerase, and another approach is to mix a reverse transcriptase with a thermostable DNA polymerase that is temporarily inactivated by an antibody. After the RT step at temperatures between 37 and 50°C, the temperature is increased to 95°C to inactivate the RT enzyme and activate the polymerase by denaturing the antibody.

Several protocols have been developed for this type of amplification. Reverse transcriptases from either avian myeloblastosis virus (AMV) or Moloney murine leukemia virus (MMLV) were the first enzymes used for the synthesis of the first cDNA strand. However, these enzymes have proven to be inefficient for reverse transcription of RNAs that form strong secondary structures. These enzymes have been replaced by thermostable enzymes that work well at temperatures in excess of 37°C. One of these enzymes is the DNA polymerase from *Thermus thermophilus* which under certain conditions works as a reverse transcriptase. Other enzymes are the product of recombinant DNA manipulations and have proprietary information issues. Examples of these enzymes are AccuRT which is part of the GeneAmp® AccuRT Hot Start RNA PCR Kit from Applied Biosystems, MonsterScript™ Reverse Transcriptase from Epicentre Biotechnologies, ThermoScript™ RNase H - Reverse Transcriptase from Invitrogen, and Omniscript RT and Sensiscript RT from QIAGEN.

Gene specific primers are commonly used in one-step RT-PCR, while oligo dT primers or random primers are preferred in two-step RT-PCR. In the latter case, a mixture of first strand cDNAs representing the complexity of the RNA population is generated and can be used for several PCR reactions, each with a different pair of gene-specific primers. A major concern in the first step is the efficiency of reverse transcription. If this step is not highly efficient, then the quantitation resulting from the following PCR step will be skewed. This is even more problematic when performing relative quantitation with samples in which the internal standard and the target sequence are reverse transcribed with different efficiencies.

RT-PCR can be used as a qualitative and a quantitative tool to evaluate gene expression. There are basically two modes of measuring RT-PCR results: End-point RT-PCR and Real Time RT-PCR. End-point RT-PCR can’t be used directly to quantify transcripts because the
amplification products reach a plateau during the last cycles, and different samples reach the plateau at different cycles. A laborious quantitative PCR called **competitive RT-PCR** requires a series of amplifications, each with different known amounts of an internal standard. The internal standard must be easily distinguishable from the target sequence. The original concentration of the target sequence can be calculated by plotting the ratio of the standard over the target sequence versus the concentration of the standard. The point at which the ratio is equal to 1 corresponds to the concentration of the target sequence. In addition to being laborious, this approach is prone to some error associated with the choice of standards.

A more reliable quantitative approach is **Real-Time RT-PCR**. In this method, the amount of amplification product is monitored continuously throughout the amplification procedure using a fluorescent reporter molecule. A modified thermocycler capable of measuring the fluorescence of all wells in the array is required. The quantity of the target sequences can be estimated by identifying the **Threshold Cycle** (C_T). This is the cycle in which the PCR product first reaches a **Threshold Level**. This level represents an amount of amplicon that falls within the exponential phase of target accumulation.

There are two main modes of quantitation: absolute and relative. Absolute quantitation aims to determine the exact number of RNA molecules present in a sample. This mode of quantitation requires the use of standards of known concentration. On the other hand, relative quantitation is carried out by measuring the amount of a specific target sequence with respect to a reference sequence, usually a housekeeping gene.

There are two principal methods of measuring the progression of amplification during real time PCR. The first method directly measures the amount of DNA being synthesized by adding to the reaction mix a fluorescent dye which specifically binds double stranded DNA (e.g.: SYBR Green). The amount of double stranded DNA present in the well is directly proportional to the amount of emitted fluorescence. This method is convenient, but has the disadvantage of measuring also non-specific amplification products. However, the presence of secondary or non-target amplicons can be detected in certain instruments through the generation of melting curves. After the final cycle, the amplicons are associated with the dye, and are emitting the maximum amount of fluorescence. At this point, the temperature is raised gradually causing the amplicon to melt releasing the dye with the consequent reduction in fluorescence. The kinetics of melting is unique to specific sequences. A mixture of sequences could yield more than one inflection point in the melting curve. The negative first derivative of the melting curve can identify the temperature at which the inflection point occurs. This is the maximum in the derivative function. The presence of more than one peak is symptomatic of the presence of more than one sequence.

The second method of detection is based on the removal of a specific oligo bound to the target of amplification. In this case, the amplification reaction contains three oligos: two amplification primers and one oligo used as a reporter. During the annealing phase all three anneal to the target. The amplification primers anneal at the flanking sides as in normal PCR, and the reporter oligo anneals somewhere in the middle of the target. *Taq* DNA polymerase
removes the reporter oligo during amplification. The reporter oligo carries some fluorescent dyes, and removal from the target induces a change in fluorescence which is detected by the instrument. The principle of detection is based on fluorescence resonance energy transfer (FRET). Of these methods, Taqman is the most popular, although other reporter types like molecular beacons and “scorpions” have also been used.

In Taqman PCR, the reporter oligo anneals to the target sequence somewhere between the amplification primers. This oligo has a fluorescent molecule linked to the 5' end and a fluorescence quencher attached to the 3' end. Physical proximity of the quencher to the fluorophore prevents the emission of fluorescence. After annealing and during the extension phase, *Taq* DNA polymerase reaches the reporter oligo on one of the strands and it hydrolyzes it using its 5' to 3' exonuclease activity. Upon hydrolysis, the fluorophore and the quencher will be released and fluorescence will increase. The amount of fluorescence is proportional to the amount of amplified product. Molecular beacons use a similar approach, except that in this case the reporter oligo has short inverted repeats at the termini which are also attached to a fluorophore and a quencher. During denaturation and annealing the short inverted repeats of the reporter oligo separate allowing the intervening sequence to anneal to the target. When this occurs the fluorophore is separated from the quencher and it is allowed to fluoresce. Again the amount of fluorescence is proportional to the amount of amplification product. In the case of molecular beacons, the 5' terminus of the reporter oligo is protected from degradation by the *Taq* DNA polymerase. Molecular beacons also contain a fluorophore which fluoresces when the reporter is annealed to the target, but when this reporter is removed by *Taq* polymerase, the fluorophore is brought close to a quencher molecule which reduces the fluorescent signal. The structure of the scorpion oligo is more complicated, but detection of amplification product is based on an increase of fluorescence.

Although the concept behind Real Time PCR is simple, there are many factors that can affect the quantitation estimates. The first is the measurement of the fluorescent signal. To avoid problems related to volume and concentration changes, the signal of the Reporter Dye (SYBR-Green) is expressed in relation to the fluorescence of a reference dye (ROX) added to the mix, also known as Passive Reference. In other words, the SYBR-Green signal is normalized to the fluorescence of ROX. The normalized reporter signal (Rn) is calculated by dividing the fluorescence intensity signal of the reporter dye by the fluorescence intensity of the passive reference dye. The Rn+ value at any point in time is the Rn of the reaction containing all the PCR reagents, whereas the Rn- value is the Rn value of an unreacted sample which can be obtained from the early cycles of amplification when the signal does not change significantly from cycle to cycle, or alternatively, from a reaction mix missing the template. Thus, the corrected fluorescence signal generated by a particular set of PCR conditions is given by the delta Rn (ΔRn) value which is equal to [(Rn+) - (Rn-)]. In summary, the Threshold Cycle (Cₜ) is the cycle in which ΔRn reaches the Threshold Level.

Relative quantitation can be used to measure changes in gene expression of a specific sequence. The Comparative Cₜ Method (ΔΔCₜ) uses the levels of a reference transcript (a
housekeeping gene) to normalize the levels of the target sequence. The measure of the relative changes in target transcript level due to “the treatment” is based on the difference of the ratios of the target \( C_T(\text{Target}) \) over the reference \( C_T(\text{Reference}) \) in the treated sample or \( \Delta C_T(\text{Treatment}) \) and the calibrator tissue or control \( \Delta C_T(\text{Calibrator}) \). Thus, the \( \Delta\Delta C_T \text{ value} = \Delta C_T(\text{Treatment}) - \Delta C_T(\text{Calibrator}) \). Let’s look at the Calibrator, or control first. If the \( C_T \) of the reference is 16, it would mean that it took 16 doubling PCR cycles to reach the threshold value and if the \( C_T \) of the Target is 18, then it means that the target transcript needed two more amplification cycles to reach the threshold level than the Reference. The mathematical expression for the Calibrator is

\[
\Delta C_T(\text{Calibrator}) = C_T(\text{Target}) - C_T(\text{Reference}) = 18 - 16 = 2.
\]

This means that relatively speaking the Target transcripts were at 25% the level of the Reference transcripts; again, it took 2 doubling PCR cycles to the target to reach the same level as the reference. Now, less assume in this example that in the Treatment the \( C_T(\text{Reference}) \) was 17 and the \( C_T(\text{Target}) \) was 15. The roles have reversed. Now it took 2 more cycles to the Reference to reach the same level as the Target. Mathematically this is expressed as

\[
\Delta C_T(\text{Treatment}) = C_T(\text{Target}) - C_T(\text{Reference}) = 15 - 17 = -2.
\]

Thus, the \( \Delta\Delta C_T \text{ value} = \Delta C_T(\text{Treatment}) - \Delta C_T(\text{Calibrator}) = -2 - 2 = -4. \) Thus, the amount of target normalized to an endogenous reference and relative to the amount of target in the Calibrator is given by the expression:

\[
2^{-\Delta\Delta C_T} = 2^{-(-4)} = 16
\]

This means that the target increased 16-fold as a result of the treatment, and this increase was normalized to the level of the endogenous reference.
PROCEDURE FOR REAL-TIME RT-PCR

Applied Biosystems - High Capacity cDNA Reverse Transcription Kit

2X RT Master Mix
RT- Buffer, 10X 2.0 μl
dNTP, 100 mM 0.8 μl
Random Primers, 10X 2.0 μl
RNase Inhibitor 1.0 μl
MultiScribe™ RT 1.0 μl
H₂O 3.2 μl
TOTAL 10 μl

RT Reaction
2X RT Master Mix 10 μl
Total RNA, 1 μg 10 μl
TOTAL 20 μl

Thermocycler Program

Step 1 25°C 10 min
Step 2 37°C 120 min
Step 3 85°C 5 sec
Step 4 15°C Forever

Applied Biosystems - Power SYBR Green PCR Kit

PCR Amplification
Power SYBR Green PCR Master Mix 2X 10.0 μl
Primer F/R, 1 μM 5.0 μl
Template 1/50 dilution 5.0 μl
Total 20.0 μl

Thermocycler Program

Taq Activation 95°C 10 min
Denaturation 95°C 15 sec
Anneal/Extension 60°C 1 min 40 Cycles

Melting Curve
**Study Questions – Gene Expression**

1. What is the major challenge during isolation of RNA?
2. Other than substrate specificity, what are the major differences between DNases and RNases?
3. DNA is one major potential contaminant that must be avoided during RNA isolation. What steps in the protocol used in class helped eliminate DNA? Explain the basis of each of those steps.
4. How is RNA quantified?
5. How would you assess the intactness of RNA? Provide a detailed explanation including methods to eliminate secondary RNA structures.
6. What was the function of DMSO in the preparation of RNA for electrophoresis?
7. What is the approximate percentage of mRNA in total RNA preparations? What kind of RNA is the non-mRNA in the preparation?
8. Name three methods you could use to measure the relative amounts of RNA in one sample, and in two samples. What is required to assess relative changes in the amount of a particular RNA?
9. Assuming all mRNA molecules have an average size of 2000 bp, what would be the mass of a gene that is one of 18,000 genes which represent 10% of the total mRNA population in a preparation of 2 μg of total RNA from root tissue? How many molecules of this gene would be present?
10. What enzymes are involved in RT-RT-PCR?
11. In how many ways can you accomplish the first step in RT-RT-PCR?
12. What is the advantage of RT-PCR over end-point PCR?
13. How is the progress of amplification followed in Real Time PCR?
14. Real Time PCR uses two measurements of quantity for the target. What are those measurements and how are they used?
15. How can you detect secondary amplification product without resolving the amplification product via electrophoresis?
16. How would you determine the relative change in the concentration of a particular amplicon using the comparative Ct method, also known as the ΔΔCt method? Explain it in detail.
17. It is essential to have 100% efficiency of PCR amplification to ensure the reliability of RT-PCR. How would you determine this efficiency?
Lab Report: Gene Expression

Part I.
Summary (One Page Maximum. 20 points)
a) The main objective of the experiment,
b) The most important findings,
c) The significance of the results, and
d) Major conclusions

Part II.
Results

1. (10) Calculate the RNA yield per gram of tissue. Show your calculations.

2. (5) What steps during the RNA extraction reduce/eliminate DNA? Explain.

3. (5) What factors may lead you to overestimate the RNA concentration.

4. (30) Describe the results obtained with RT-PCR and provide an interpretation.

Part III.

1. (10) Consider the expected fraction of mRNA in total RNA, and that a transcript of interest is among those of a class of 10,000 different genes that represent only 0.5% of the mRNA fraction in leaf tissue. Assuming that the average length of a transcript is 2,500 nucleotides, how many molecules of this transcript of interest are present in the total RNA you extracted. Consider the concentration of the RNA and the volume in which it was dissolved.

2. (10) If you wanted to make a cDNA library and have a 99.9% chance of isolating at least one clone of the gene of interest (from the previous question), how many clones would you have to screen?

3. (10) You are interested in the expression of genes induced by 24 h of anoxia (water flooding) in the roots of barley. After extracting RNA from treated roots and controls you carried out a RT-RT-PCR analysis to determine what happens with transcripts of the enzyme alcohol dehydrogenase (ADH) during the treatment. Transcripts of the Krebs Cycle enzyme citrate synthase were used as the internal reference, and the results showed that transcripts for ADH increase 50-fold! However, a colleague of yours points out to you that there have been three articles in well respected journals showing the indeed ADH transcript levels go up after 24 of anoxia, but the increase over the internal references used by the three different labs is only 10-fold. How do you explain your results?
PLANT TRANSFORMATION

Learning Objectives:
After completions of this unit students will be able to identify the different elements involved in Agrobacterium-mediated transformation working knowledge of the Agrobacterium transformation system regarding the structure and organization of transformation vectors and delivery systems.

Lab Objectives:
Carry out transient transformation of leaf tissues via Agroinfiltration, and also obtain stable transformants of Arabidopsis plants via the floral dip method.

Background
There are several examples of transfer of genetic information across taxonomic barriers in different biological systems. Examples in prokaryotes include conjugation and transduction, while in eukaryotes we have the insertion of retroviral genomes into the chromosome of the host. A unique transfer of DNA from bacteria to host cells occurs in the plant kingdom. Agrobacterium tumefaciens bacterium is a pathogen that causes crown gall disease. In essence, this pathogen is capable of delivering some of its DNA into the chromosomes of plant host cells where the bacterium’s DNA integrates into the plant cell DNA. Cells receiving the bacterium’s DNA are in effect “transformed.” The mechanism of DNA transfer, and most of the bacterial and plant components have been characterized to a great extent. In fact, the bacterial components have been modified to allow the transfer of functional genes across taxonomic barriers.

Most of the components required for DNA transfer are located in the tumor-inducing plasmid (Ti plasmid) of Agrobacterium. During natural infection, a section of the Ti plasmid called the T-DNA (Transfer DNA) is transferred to the host cell where it becomes integrated into a chromosome. The T-DNA is delimited by “left border” and “right border” sequences, which are made up of 25 nucleotide-long imperfect direct repeats. In natural systems, the T-DNA carries onc genes that encode enzymes responsible for the synthesis of plant hormones, and ops genes encoding enzymes involved in the synthesis of modified amino acids known as opines (octopine, nopaline, etc.). These “transgenes” are expressed in the infected tissue once the T-DNA is integrated into a plant chromosome. Expression of the onc genes brings about the synthesis of the hormones that promote the growth of a tumor, while the expression of the ops genes mediate the production of opines which represent the major C and N source for Agrobacterium. It has been determined that the only structural feature of the T-DNA required for its transfer from the bacterium to a plant cell is the presence of the “borders,” the genes
encoded in natural T-DNAs are not essential for transfer and have been eliminated in transformation vectors – “disarmed vectors.”

Several bacterial genes encoding virulence factors (vir genes) are required for the transfer of T-DNA. Many of them are located in the Ti plasmid, but a few are chromosomal genes. Genetic analysis of Agrobacterium has revealed the role the vir genes play in T-DNA transfer. The first step takes place when the product of the virA gene, a membrane associated protein, senses certain phenolic compounds in the medium such as acetosyringone. This phenolic compound induces autophosphorylation in the virA protein, and subsequently, trans-phosphorylation of the VirG protein, which in turn leads to the transcription of the other vir genes. The virD2 protein, with the assistance of the virD1 protein, cuts the T-DNA at the “borders” and becomes covalently linked to the 5'end of the T-DNA (right border). T-DNA is transferred as single stranded molecule. The products of the virE2 and virJ genes are single stranded DNA binding proteins. The virJ protein coats the T-DNA in the bacterial cell, while the virE2 protein moves into the plant cell independently of the T-DNA where it coats it replacing the virJ protein. Both the virD2 and virE2 proteins contain nuclear localization signals which direct the protein-T-DNA complex to the nucleus. Transfer of T-DNA into the host cell takes place through the Type IV secretion system (T4SS) of Gram-negative bacteria. This structure is composed of proteins encoded in the virB operon (11 ORFs) and virD4.

Not all the T-DNA that is transferred to the nucleus is integrated into a chromosome. Comparisons of the level of transient and stable expression have shown that the former was significantly higher. It has been demonstrated that the newly imported T-DNA has to be converted to the double stranded form before the trans-gene can be expressed. However, it isn’t known whether single or double stranded DNA is the substrate for integration. Finally, integration appears to be dependent on the products of bacterial and plant genes. Plant encoded proteins have been identified that interact directly with T-DNA bound virD2 and virE2 proteins. Furthermore, it has also been demonstrated that double strand breaks (DBS) in the host DNA are required for integration, and that the non-homologous end join repair mechanism is intimately involved in integration. The evidence for NHEJ involvement is that mutants of the major components of this repair mechanisms can’t be transformed by Agrobacterium.

Natural Ti plasmids are relatively large (~100 Kb) and not amenable to in vitro manipulations. To facilitate molecular manipulation of the system, the Ti plasmid has been divided to create a binary vector system. In this system, the T-DNA has been removed from the Ti plasmid (disarmed vector) and used to construct a second plasmid, which is called the binary vector. In this vector, the onc and ops genes of the T-DNA have been replaced with a multiple cloning site, a plant selectable marker, and reporter genes. Selectable markers confer a selective advantage to transformed cells in the presence of selection pressure. This type of
pressure is exerted by the inclusion of an antibiotic or a herbicide to which the plant is normally susceptible. The selectable marker is usually an antibiotic resistance gene such as neomycin phosphotransferase, (Kan) or hygromycin phosphotransferase (Hyg) which control resistance to kanamycin and hygromycin, respectively. These antibiotics are aminoglycosides capable of inhibiting protein synthesis. Genes encoding resistance to herbicides like glyphosate or basta are also used as selectable markers. To ensure that these markers are expressed during selection, they are placed under the control of constitutive plant promoters. The promoter most commonly used for dicots is the 35S promoter from the *Cauliflower Mosaic Virus*, and for monocots is the Ubiquitin promoter from maize. Visualization of transformed tissue can be accomplished via the inclusion of reporter genes like β-glucuronidase (GUS) from *E. coli*, or the green fluorescent protein (GFP) from the jellyfish *Aequorea victoria*. A terminator of transcription is also included at the 3' end of the marker and reporter genes.

The backbone of the binary vector carries sequences that allow replication in both *E. coli* and *Agrobacterium*. For instance, plasmids may have the origin of vegetative replication (OriV) and the transacting replication function Trf that allow replication in *E. coli*, and the origin of replication pVS1 sequence for stability and propagation in *Agrobacterium*. These plasmids also carry the OriT function which allows conjugal transfer of the plasmid from *E. coli* to *Agrobacterium*. In addition, a selectable marker that is expressed in bacterial cells is also included in the binary vector. This marker is usually a gene encoding antibiotic resistance to help select for bacterial cells carrying the binary vector.

The production of transgenic plants typically requires the combination of two technologies, the *Agrobacterium* delivery system, and a tissue culture and regeneration system. Either one of these technologies can be the limiting factor in the generation of transgenics using *Agrobacterium*-mediated transformation. Not all plant species are susceptible to *Agrobacterium* infection, and not all species possess a reliable protocol for plant regeneration from tissue culture. *Agrobacterium* can readily transfer T-DNA into plant cells, but for stable transformation it is necessary that the newly integrated trans-gene be present in the germ line – the cells that will give rise to the reproductive organs. To accomplish this task, undifferentiated tissue grown in tissue culture is transformed, and based on the totipotency of plant cells, this tissue is induced to regenerate into a plant. Two approaches have been developed for this endeavor: organogenesis and embryogenesis. Organogenesis involves the induction of meristematic tissue capable of generating organized tissues typical of the species, while embryogenesis requires the induction of embryos or embryo-like structures that will lead to the production of a plant.

*Agrobacterium*-mediated *in vitro* transformation is a lengthy and somewhat cumbersome procedure. In contrast to *in vitro* transformations, a couple of simple *in planta* transformation procedures have been developed. The first one was based on the observation that incubation of
Arabidopsis seeds with Agrobacterium can lead to the production of transgenic plants, albeit at an efficiency of 1 in 10,000 seeds. It was later observed that the efficiency of transformation could be increased significantly if the inflorescence was inoculated. This observation led to the development of the so called Floral Dip transformation protocol. This procedure is extremely simple and free of the genetic alterations sometimes associated with regeneration. The efficiency of transformation by floral dip is between 1 and 3%. Examination of these procedures suggested that when seeds were inoculated with Agrobacterium, the bacterial cells remained in the meristem until floral development. More recent evidence indicates that the ovules are the target of transformation. Unfortunately, the floral-dip protocol does work very well only with Arabidopsis, although there are some isolated reports for other crucifers.

Agrobacterium-mediated transient expression can also be used to expedite the functional analysis of trans-genes. Transient expression is also the only alternative in species without a transformation protocol. In this protocol, Agrobacterium cells carrying a transgene are first pre-treated with vir-inducing acetylsyringone, and then are vacuum infiltrated into leaf tissue. The trans-gene is expressed uniformly throughout the tissue at very high levels from two to five days after inoculation.

References


TRANSIENT EXPRESSION – MATERIALS AND SUPPLIES

A. BIOLOGICAL MATERIALS

*Agrobacterium*
Bean Seedlings
*Arabidopsis* plants

B. REAGENTS

<table>
<thead>
<tr>
<th>Chemical Name</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-Br-4-Cl-3-indolyl-β-D-Glucuronide</td>
<td>Potassium ferrocyanide K₄Fe(CN)₆·3 H₂O</td>
</tr>
<tr>
<td>Acetosyringone</td>
<td>Potassium ferricyanide K₃Fe(CN)₆</td>
</tr>
<tr>
<td>Chloramphenicol</td>
<td>Rifampicin</td>
</tr>
<tr>
<td>DMSO</td>
<td>Sodium EDTA</td>
</tr>
<tr>
<td>Ethanol</td>
<td>Sodium Phosphate NaH₂PO₄</td>
</tr>
<tr>
<td>Hygromycin</td>
<td>Sodium Hypochlorite</td>
</tr>
<tr>
<td>Kanamycin</td>
<td>Sucrose</td>
</tr>
<tr>
<td>Magnesium Sulfate MgSO₄·7H₂O</td>
<td>Triton X-100</td>
</tr>
<tr>
<td>MES</td>
<td>Tryptone</td>
</tr>
<tr>
<td>MS salts (Sigma- Aldrich)</td>
<td>Yeast Extract</td>
</tr>
<tr>
<td>Nutrient Broth</td>
<td></td>
</tr>
</tbody>
</table>

C. SOLUTIONS

**Seed Sterilizing Solution** (2.5 % Na Hypochlorite)
- Commercial bleach 100 ml
- H₂O 100 ml

**Ethanol 70%**
- Ethanol 700 ml
- H₂O 300 ml

**MES, 1 M, pH 5.6**
- MES 19.5 g
- H₂O add to 100.0 ml

**NOTE 1:** Dissolve to a volume of about 70 ml, and then adjust pH to 5.6 with NaOH 4N.

**ACETOSYRINGONE, 1 M**
- Acetosyringone 1.96 g
- DMSO add to 10.00 ml
YEB medium

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tryptone</td>
<td>5 g</td>
</tr>
<tr>
<td>Yeast Extract</td>
<td>1 g</td>
</tr>
<tr>
<td>Nutrient Broth</td>
<td>5 g</td>
</tr>
<tr>
<td>Sucrose</td>
<td>5 g</td>
</tr>
<tr>
<td>MgSO₄.7H₂O</td>
<td>490 mg</td>
</tr>
<tr>
<td>H₂O add to</td>
<td>1000 ml</td>
</tr>
</tbody>
</table>

**NOTE 1:** Dissolve in 900 ml of H₂O, adjust the pH to 7.2, and then bring to 1 L.

**NOTE 2:** Autoclave for 20 min.

**NOTE 3:** Add antibiotics before inoculation

---

**vir Gene Induction Medium**

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tryptone</td>
<td>5 g</td>
</tr>
<tr>
<td>Yeast Extract</td>
<td>1 g</td>
</tr>
<tr>
<td>Nutrient Broth</td>
<td>5 g</td>
</tr>
<tr>
<td>Sucrose</td>
<td>5 g</td>
</tr>
<tr>
<td>MgSO₄.7H₂O</td>
<td>490 mg</td>
</tr>
<tr>
<td>MES, 1 M, pH 5.6</td>
<td>10 ml</td>
</tr>
<tr>
<td>Acetosyringone, 1 M</td>
<td>20 μl</td>
</tr>
<tr>
<td>H₂O add to</td>
<td>1000 ml</td>
</tr>
</tbody>
</table>

**NOTE 1:** Dissolve in 900 ml of H₂O, check pH (it should be 5.6), and then bring to 1 L.

**NOTE 2:** Autoclave for 20 min.

---

**MMA medium**

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS salts (Sigma- Aldrich)</td>
<td>1 bag for 1 L</td>
</tr>
<tr>
<td>MES, 1 M, pH 5.6</td>
<td>10 ml</td>
</tr>
<tr>
<td>Sucrose</td>
<td>20 g</td>
</tr>
<tr>
<td>Acetosyringone, 1 M</td>
<td>200 μl</td>
</tr>
<tr>
<td>H₂O add to</td>
<td>1000 ml</td>
</tr>
</tbody>
</table>

**NOTE 1:** Dissolve in 900 ml of H₂O, check pH (it should be 5.6), and then bring to 1 L.

**NOTE 2:** Autoclave for 20 min.

---

**Antibiotics - Stock Solutions**

<table>
<thead>
<tr>
<th>Antibiotic</th>
<th>Stock (mg/ml)</th>
<th>Solvent</th>
<th>Agar (μg/ml)</th>
<th>Liquid (μg/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbenicillin</td>
<td>100</td>
<td>Water</td>
<td>100</td>
<td>30-50</td>
</tr>
<tr>
<td>Chloramphenicol</td>
<td>3</td>
<td>Ethanol</td>
<td>3-100</td>
<td>3-100</td>
</tr>
<tr>
<td>Erythromycin</td>
<td>100</td>
<td>Ethanol</td>
<td>150</td>
<td>100</td>
</tr>
<tr>
<td>Gentamicin</td>
<td>100</td>
<td>Water</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Kanamycin</td>
<td>50</td>
<td>Water</td>
<td>50</td>
<td>10-20</td>
</tr>
<tr>
<td>Rifampicin</td>
<td>10</td>
<td>Methanol</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Spectinomycin</td>
<td>100</td>
<td>Water</td>
<td>100</td>
<td>25-50</td>
</tr>
<tr>
<td>Tetracycline</td>
<td>3</td>
<td>50% Ethanol</td>
<td>3</td>
<td>1.5</td>
</tr>
</tbody>
</table>
X-GLU, 40 mg/ml
5-Bromo-4-Chloro-3-indolyl-β-D-Glucuronide 200 mg
DMSO adjust volume to 5 ml

GUS Staining Solution
Na PO₄, 1 M, pH 7.0 8.0 ml
K₄Fe(CN)₆.3 H₂O, 10 mM 4.0 ml
K₃Fe(CN)₆, 10 mM 4.0 ml
NaEDTA, 0.5 M 1.6 ml
Triton X-100, 10 % 0.5 ml
X-GLU, 40 mg/ml 2.0 ml
H₂O add water to 100.0 ml

NOTE 1: Mix reagents just prior to use.
NOTE 2: Filter sterilize (0.22 μm)

D. LABWARE AND SUPPLIES
Beaker and Flasks
Metromix 300
Pots 4"
Syringe (1 ml) with 26G needle

E. EQUIPMENT
Camera & Filters
Centrifuge
Dry incubator
Magnetic Stirrer and Magnetic bars
Shaker Incubator
UV lamp, 100 Watt
Vacuum Chamber
Vacuum Pump with Vacuum Gauge
TRANSIENT TRANSFORMATION OF COMMON BEAN - VACUUM INFILTRATION

**Plant Preparation** (Start 10 days before transformation experiment)

1. Sterilize common bean seeds in 50% commercial bleach (2.5% Na Hypochlorite) for 3 minutes, and then rinse in sterile water.

2a. Germinate seeds between a germination paper and the side of beaker. Make sure there is water (1 cm-deep at the bottom of the beaker). Cover the beaker with saran wrap, but allow the seedling to come out.

2b. For hypodermic infiltration germinate seeds in 4" pots filled with Metromix 300.

**Inoculum Preparation** (Start 4 days before the transformation experiment)

1. Streak a **YEB** plate (+ proper antibiotics) with the *Agrobacterium* strains, and incubate at 30°C for 36-48 h.

   **NOTE**: Strains should include the control (empty vector) and those with the reporter gene in the binary vector.

2. Pick up a single colony and inoculate a 10 ml **YEB** (+ proper antibiotics) medium, and grow overnight at 30°C and 250 rpm.

3. Inoculate a 200 ml **vir Gene-induction medium**, and grow overnight at 28°C to log phase ($A_{600} = 0.8$).

4. Pellet cells at 4000 rpm for 5 min, discard medium and resuspend in MMA medium to an $A_{600} = 2.4$.

5. Place resuspension in a 250 ml beaker, and incubate resuspension for 1 h at room temperature.

**Transformation - vacuum infiltration**

1. Place a magnetic bar in the beaker containing the inoculum

2. Immerse the leaf/leaves (attached or detached) in the inoculum.

3. Place the beaker in vacuum chamber, and put the vacuum chamber over a magnetic stirrer.

4. Connect the chamber to the vacuum pump and start the magnetic stirrer. Apply vacuum to at least Hg 28" for 5 min, and then break the vacuum rapidly.

5. Rinse the leaves in water and incubate for at least two days at room temperature.

   **NOTE 1**: Place detached leaves up-side up over wet filter paper in a plastic box.

   **NOTE 2**: Placed intact seedlings back in a beaker with the roots sandwiched between the inside of the beaker and a strip of filter paper. Make sure there is water at the bottom, and
that the beaker is covered with Saran Wrap.

**Transformation - hypodermic infiltration**

1. Select a healthy seedling with expanded primary leaves.
2. Remove the needle and load a 1ml syringe with the inoculum to about 0.5 ml. Put the needle back.
3. Turn one of the primary leaves upside down twisting the petiole. Hold one side of the leaf between your thumb and the side of your index. Place the bevel of the needle firmly against the leaf resting on one of your fingers. Inject the inoculum. You should see water soaking in the inoculated area. It should be at least 0.5" in diameter.

**Visualizing the GFP reporter gene**

1. Take the leaves or seedling transformed with the GFP-containing construct a dark room.
2. Turn on the 100 Watt UV-lamp and observed the leaves through a filter.
3. Take a photograph for the report.

**Visualizing the GUS reporter gene**

1. Place the leaf, or a section of the leaf in a petri dish.
2. Pipet 10 ml of GUS staining solution.
3. Place the petri dish in a vacuum chamber and apply vacuum to about 20" of Hg for 1 minute, and then release vacuum rapidly.
4. Incubate for 1 h to overnight at 37°C.
5. Remove GUS staining solution.
6. Wash with 70% ethanol 3 times over a period of 24 h.

**NOTE:** Ethanol will remove chlorophyll and facilitate visualization of the indigo dye.
FLORAL DIP – MATERIALS AND SUPPLIES

A. BIOLOGICAL MATERIAL
   Agrobacterium
   Arabidopsis flowering plants

B. REAGENTS

<table>
<thead>
<tr>
<th>Material</th>
<th>Material</th>
<th>Material</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agar</td>
<td>Hygromycin</td>
<td>Sodium Chloride, NaCl</td>
</tr>
<tr>
<td>Agarose</td>
<td>Kanamycin</td>
<td>Sodium hypochlorite</td>
</tr>
<tr>
<td>Bacto-tryptone</td>
<td>MS salts</td>
<td>Sucrose</td>
</tr>
<tr>
<td>Bacto-yeast extract</td>
<td>Silwet L-77</td>
<td>Tween-20</td>
</tr>
</tbody>
</table>

C. SOLUTIONS

Sterile Water

Seed Sterilization Solution
   Commercial bleach 100 ml
   Tween-20 250 μl
   Water adjust to 500 ml
   **NOTE:** Dilute the bleach before adding detergent.

LB medium
   Bacto-tryptone 10 g
   Bacto-yeast extract 5 g
   NaCl 10 g
   H₂O add to 1000 ml
   **NOTE 1:** Dissolve solutes in 950 ml of water and adjust pH to 7.0, bring to final volume and autoclave for 20 min.
   **NOTE 2:** Add appropriate antibiotics prior to use.

LB plate
   **NOTE 1:** Prepare LB medium as above, but add 15 g of agar before autoclaving.
   **NOTE 2:** Place flask in a 55°C water bath. Add appropriate antibiotics after temperature has equilibrated, and just before pouring plates.

Transformation Solution
   Sucrose 50 g
   Silwet L-77 500 μl
H₂O add to 1000 ml

**NOTE 1:** This solution doesn’t need to be sterilized if it is prepared fresh.

**NOTE 2:** Silwet L-77 should be added immediately before utilization and should be completely mixed in the solution.

**Molten Agarose (0.1% 42°C)**

- Agarose 0.5 g
- H₂O 500 ml

**NOTE:** Autoclave and place flask(s) in 42°C water bath

**Arabidopsis culture plates**

- MS salts 0.5 X
- Agar 0.8 %
- Antibiotics As appropriate

**D. LABWARE AND SUPPLIES**

- Beakers, 500 ml
- Centrifuge tube, 50 ml
- Labels 4"
- Plastic bags 4L
- Pots 4"
- Potting soil
- Saran Wrap
- Sieve or strainer
- Surgical tape
- Watch glass, or small beaker

**E. Equipment**

- Balance
- Centrifuge (clinical or high speed)
- Cold room, 4°C
- Dry Incubator, 28°C
- Growth Cabinet with lights
- Shaker Incubator
- Spectrophotometer
- Vortex
- Water bath (42°C)
Plant Preparation

1. Autoclave potting soil for 30 min
2. Fill 4" pots up to 0.5" below the rim
3. Resuspend a few Arabidopsis seeds in water inside a watch glass, or small beaker.
4. Using a 1 ml pipet tip dialed to 50 ul, pipet a single seed and drop it on top of the soil. Repeat to place about 6 seeds per pot (equally spaced).
5. Cover the pots with Saran Wrap and incubate the pots at 4°C for two days.
6. Transfer pots to cafeteria trays with 0.25" of water, and place in growth cabinet under a 16h photoperiod or continuous illumination.
7. Clip off the first bolt when it appears to induce growth of axillary branches (more flowers).

**NOTE 1:** Plants will be ready at approximately 4-6 day after clipping.

**NOTE 2:** The highest transformation efficiency is obtained with plants that have just a few flowers opened with many racemes with flower buds.

**NOTE 3:** Secondary bolts can be clipped again to delay flowering.

Inoculum Preparation

1. Streak LB plate (+ antibiotics) with the *Agrobacterium* strain(s), and incubate at 28°C for 36-48 h.

   **NOTE:** Include a control (empty vector) along with those carrying the reporter gene in the binary vector.
2. Pick up a single colony and inoculate a 10 ml LB (+ antibiotics) medium, and grow overnight at 28°C and 250 rpm.
3. Inoculate a 500 ml LB-induction medium with 5 ml of overnight stationary phase suspension, and grow overnight at 28°C to mid-log phase, or early-stationary phase.
4. Pellet cells at 4000 rpm for 5 min, discard medium and resuspend in Transformation Solution to an $A_{600} = 0.8$.

   **NOTE 1:** Add Silwet L-77 to a concentration of 0.05% (500 μL/L) to the Transformation solution just prior to resuspending the cells. Make sure it is completely mixed in.

   **NOTE 2:** Place resuspended cells in a 500 ml beaker.
Agrobacterium Dip

1. Dip the entire inflorescence in the cell suspension for 2" to 3" seconds with gentle agitation.

   NOTE 1: When several constructs are targeted for transformation, a viable alternative is to just fill a 1.5 ml microcentrifuge tube with the cell suspension, and dip each raceme independently.

   NOTE 2: Axillary inflorescences that come up after the dip will not produce transformants. Thus they should be clipped off or dipped once they show flowers and flower buds.

2. Place plants in a closed container for 16 to 24 hours to maintain high humidity.

   NOTE: A dome could be used for this purpose, but exposure to direct sunlight should be avoided to prevent overheating. An alternative is to lay plants horizontally in plastic storage boxes.

3. Water and maintain the plants healthy until siliques begin to mature. At that point watering should be stopped.

4. Harvest when all siliques have just turned brown. Place entire inflorescence in a 4L plastic bag and cut it off the plant. Let the contents of the bag dry out before threshing the seeds by shaking the contents of the bag and mildly pressing siliques that had not opened. Remove inflorescence and all large material. Separate seeds from chaff by sieving.

   NOTE: Transformants are usually independent. However, it is a good idea to keep seeds from each plant separated.

Screening for Transformants

1. Weigh 40 mg of seeds and transfer to a 50 ml tube.

2. Add 25 ml of sterilizing solution and vortex for 30 seconds. Sterilize seeds for 5 minutes with occasional mixing. Draw off solution and wash with sterile water 5 times.

3. Resuspend seeds in 4 ml of 0.1% molten agarose (42°C), and plate on a Arabidopsis culture plates. Seal plates with surgical tape and then incubate at 4°C for 2 days.

4. Take plates to grow cabinet under continuous light, and allow seeds to germinate for 1 week.

5. Inspect plates and count the number of transformants.

Part I
Summary (One Page Maximum. 20 points)
a) The main objective of the experiment,
b) The most important findings,
c) The significance of the results, and
d) Major conclusions

Part II - Results
1. (20) Describe the results obtained with the Floral Dip protocol. What are the advantages and disadvantages of this procedure?
2. (20) Describe the results obtained with the Transient Transformation Protocol. What are the advantages and disadvantages of this procedure?

Part III
1. Characterization of the structural features of a gene and its pattern of gene expression don’t always reveal the exact role of the gene in the plant. Functional genomics addresses this issue with several approaches. The generation of gene-knock outs is one of these approaches, and T-DNA insertion mutagenesis is one of the options chosen for knocking out a gene in species for which an efficient transformation protocol exists. In this approach, transgenic plants are generated using a T-DNA construct that has a reporter gene and a plant-selectable marker. If the T-DNA is inserted in the exon of a gene it will most likely inactivate the gene and produce a mutant phenotype. However, when a diploid individual is mutagenized, only one of the chromosomes is affected. In the case of T-DNA insertion mutagenesis, the transformed individual is in a hemizygous state and it is unlikely to show a mutant phenotype. Most of the mutations are recessive. This means that the mutant phenotype can only be detected after the mutagenized plant is selfed. The progeny of such plant will segregate in a 3 normal : 1 mutant ratio.

Problem description: A T-DNA insertion mutagenesis in tomato has generated a mutant line with a recessive lethal mutation. The mutants germinate normally, but the cotyledons are pale yellow and never green up. The seedling dies shortly after germination. This mutant line can be propagated through the heterozygotes. This seems like an interesting gene to investigate. Insert-based primers have been designed to sequence the DNA flanking the insertion, and the sequences on each side of the insert have been obtained. These sequences have been aligned and merged with the entire sequence of the T-DNA (Attachment 1).

a (10). What is the identity of the tomato gene that has been interrupted by the T-DNA.
b (10). What is the identity of the plant-selectable marker in the T-DNA.
c (10). What is the identity of the reporter gene in the T-DNA.
d (20). Provide an annotated form of the sequence in Attachment 1 (see Attachment 2 for details).
e (20). No apparent phenotypic differences exist between normal tomato plants and those that are hemizygous. However, there is the possibility that some differences exist, but these may only be detectable after a biochemical analysis or under specific environmental conditions.
For this reason, it will be essential to distinguish the two genotypes to make the necessary biochemical comparisons. PCR will be the most practical approach to address this problem. One possibility is to design primers that amplify the a segment of the T-DNA. Thus, the presence of the amplification product would indicate that the plant has the insert, and if it doesn’t, then you can assume that the plant is of the normal type and does not have the insert. The problem with this approach is that sometimes more than one insertion occurs in the generation of transgenics. So it is possible to have plants with a T-DNA insertion in a silent region of the genome and another one in the exon of a gene. To get around this problem, you have to design primers to amplify the junction, that is, one primer from the tomato sequence, and the other from the T-DNA. Design such primer pair.

It will also be desirable to design primers that amplify the tomato DNA spanning the region in which the T-DNA was inserted. The amplification product should be short (200-500 bp). If the T-DNA is present no amplification product will be obtained from that chromosome. However, if there is no T-DNA insertion, then an amplification product will be generated. In summary, a normal homozygous tomato plant will produce an amplification product with the tomato DNA-based primers, and no amplification product with the primers designed in the previous section. On the other hand, if the plant is hemizygous, then both sets of primer pairs will generate an amplification product. There is one more problem to solve. There is the suspicion that the interrupted gene has a duplicate, and that designing exon-based primers may result in the amplification of sequences from both genes. Design primers that are expected to specifically amplify a fragment of the interrupted gene.