BIMM 101 Recombinant DNA Techniques Lab Manual

Winter 2021

Authors: Lisa McDonnell, William McGinnis, Mandy Butler, Stephanie Mel, Emily Grossman, Adam Kirby, Thomas Kuret, Christopher Day, Keefe Reuther

Acknowledgements: We would like to sincerely thank Dr. Lorraine Pillus (UCSD Biological Sciences) and her group for providing us with the yeast strain used in our experiments. Thank you to Dr. Scott Rifkin for plasmid samples, and Dr. Jim Berry and Eileen Sylves of the University of Buffalo for providing plasmid samples and inspiration to try this CRISPR experiment as well as some background information on the project. Thank you to Andrew Moore for testing the protocols. Thank you to Dr. Andrew Cooper and Dr. Cindy Gustafson-Brown for feedback on how to improve the protocols.

Contents

About the Contents of the Lab Manual

The lab manual contains two sections:

Background information about techniques and/or parts of a project.

Protocols – these are specific step-by-step instructions that one would use if they were doing the work at the bench (in the lab, sometimes this is called “wet lab” because our type of lab involves working with samples, reagents, etc.)

Instructions for Lab-specific tasks will be provided to you by your instructor.

Keeping a Good Lab (Research) Notebook

The lab notebook is a place to keep track of your thinking, your analysis, and do some scientific writing. IAs and the Instructor can provide you with feedback on your lab notebook entries.

It is very important that you have an entry in your lab notebook for each lab session.

Using the lab notebook

1. Use headings to title new labs and sub-sections of each lab notebook entry. This makes it easier to find information and automatically generate a table of contents (online version).
2. Put the date and the title of the lab at the beginning of the day's entry.
3. Start with the purpose of the lab and an overview of what you will do that day, as well as your predictions about the outcomes (if there are no results for that day, it is still a good idea to summarize the goals)
4. Include a drawing/flowchart of the protocol (for labs where data is being generated from an experiment or a method). If you are making plans, e.g. dilutions, show all your work and thinking.
4. Include the results and your analysis of the results. Write down any relevant observations you make as well, or if anything seems unusual or unexpected.
6. Draw conclusions about the results, with reasoning. Specifically address how the results compare to your predictions. Structure your conclusions in the form of an argument:

    Claim: what can you claim/conclude based on these results?

    Evidence: what data supports your claim(s)?

    Explanation: what is a possible reason for these results? This should be the biology or procedural mechanisms that likely caused these results. Explanations can be hypotheses.

BACKGROUND INFORMATION

Depending on which experiments you are doing or studying, you may need to refer to different parts of the background information. Background information can provide help with practical tasks (like determining how to set up a particular dilution) or understand various lab techniques, such as restriction enzyme digestion. There is also a section specifically on CRISPR-Cas9 editing and an example of how to use it to edit a specific gene in *Saccharomyces cerevisiae*.

Liquid Measurement Units

The abbreviations are as follows:

L = liter
ml = milliliter (10^{-3} L)
µl = microliter (10^{-6} L)
nl = nanoliter (10^{-9} L)
pl = picoliter (10^{-12} L)

Try these conversions:

0.25 L = _____ ml

100 ml = _____ L

0.1 ml = _____ µl

50 µl = _____ ml

100 µl = _____ L

10 µl = _____ pl

Basic Dilutions

Almost all lab work will, at one time or another require dilutions of various liquids. Below is a brief overview of the basic steps involved.

When you have to perform a basic dilution, you have to answer the following questions:

1. What is the final volume needed?
2. What information do I have about starting (stock) and ending (final solution) concentrations?

There are different ways to refer to a specific dilution:

If you are asked to make a **1:10** dilution, to make a **1/10 dilution**, to dilute a sample **10-fold**, or to use a **dilution factor of 10**, this means that you need to dilute your initial sample 10 times. Consequently, you will need the following to prepare the desired dilution:

9 parts diluent (e.g. water or buffer)

1 part sample

Note: **The 1:10 terminology is very confusing**. If you tried to make a 1:10 dilution in terms of ratios, then you would incorrectly have 1 part sample to 10 parts dilutant for a total of 11 parts. However, a 1:10 dilution **really means** 1 part sample to 9 parts diluent.

How do you determine volumes and concentrations to make a dilution?

Below you will find three different equations that are useful when making dilutions. If you know all three equations and you know when to apply a particular equation, then you will be able to solve virtually any dilution problem.

1. **final volume / sample volume = dilution factor**

The **final volume** is the total volume after the dilution. The **sample volume** is the volume of the sample at the initial concentration that will be used for the dilution.

This formula allows you to solve for any of the three parameters if you know the other two.

2. initial concentration / final concentration = dilution factor

The **initial concentration** is the concentration of the sample before the dilution. The **final concentration** is the concentration of the sample after the dilution.

This formula allows you to determine a dilution factor if one is not given. Then it will be possible to use the equation given in A to calculate exactly how (in terms of volumes) you would carry out the dilution.

3. CiVi = Cf Vf (or sometimes written as $C_1V_1 = C_2V_2$)

Ci = initial concentration of undiluted sample
Vi = volume of the undiluted sample that you will add to make the dilution
Cf= final concentration of the sample after dilution
Vf= final total volume of the dilution

Practice problems

Example #1:

Suppose you need 5 ml of a 1:20 dilution of a protein sample. In other words, you want a protein solution diluted 20-fold in a total volume of 5 ml. Using the equation from **1** above (final volume / sample volume = dilution factor), you get the following:

5 ml / sample volume = 20

sample volume = 5/20 = 0.25 ml

So, you need to use 0.25 ml of your protein samples in a total volume of 5 ml.

How much diluent (e.g. water) do you use? Remember that the final volume is the sum of sample volume plus the volume of the diluent.

5 ml final volume - 0.25 ml protein sample = 4.75 ml diluent

Thinking about it another way: You used 1 part sample and 19 parts diluent to prepare a 1:20 dilution. The 19 parts are calculated by dividing 4.75 by 0.25.

Example #2:

Suppose you have a 0.5 M stock of Tris buffer and you want 2 ml of 25 mM Tris buffer. How do you make this dilution?

CiVi = Cf Vf

(500 mM)(Vi) = (25mM)(2ml)

$$Vi = \frac{(25mM)(2ml)}{500\ mM} = 0.1\ ml \text{ or } 100\ \mu l$$

Thus, you need to use 100 µl of your original 0.5M Tris buffer solution. Since the total (final) volume that you want is 2 ml, you will need to add 1.9 mL of diluent (water).

Thinking about it another way: Your starting stock is 0.5 M (or 500 mM) and you want to dilute that down to 25 mM.

Using the equation given in **2** above (Initial concentration/Final concentration = Dilution Factor), you can calculate the dilution factor.

500 mM/25 mM = 20

In other words, you need to dilute your original stock by a factor of 20 to get from

500 mM to 25 mM.

Now that you have a dilution factor, you can solve for sample volume. The final volume that you want is 2 ml. Use the equation given in A) (Final volume/Sample volume = Dilution Factor) to solve for the sample volume.

2 ml/20 = 0.1 ml (or 100 µl)

So, you need to use 100 µl of your 500 mM stock. Because the total volume that you need is 2 ml, you would need to add 1.9 ml of diluent.

Sometimes you need to make **solutions containing more than one chemical**. For example, you may need to make 500 ml of a buffer of 50 mM Tris and 1 mM EDTA. You need to figure out how much **of each** of your stock solutions you are going to use. Remember that the final volume contains BOTH of your solutions. In this case, then, you would NOT make 500 mL of 50 mM Tris and 500 mL of 1 mM EDTA and then pour the 2 together—what would your final concentration be if you did this?

Serial Dilutions

Sometimes it is not feasible to do a dilution in a single step because the original volume you would have to use would be so small that the error in measuring would become significant, or it may be so small that it is impossible to measure! For example, if you wanted to make 1 mL of a 1:5,000 dilution, you would have to measure 1 mL/5000 = 0.2 µl of your original sample. With the tools available in the lab, it is basically impossible to measure a volume that small with accuracy (the error would probably be anywhere from 5 to 10 times the volume that you are trying to measure). The way to get around this problem is to make a *serial dilution*. This means that in order to get to the final desired dilution, you make dilutions of a dilution. The concentration of each succeeding solution is some fraction of the concentration of the preceding solution.

For example

As mentioned above, suppose you want to make a 1:5000 dilution of your sample and you need a final volume of 1 ml. Since you can't accurately measure 1mL/5000, or 0.2 µl, you need to make a serial dilution. So instead of making a 1:5000 dilution, start with just a 1:50 dilution of your original sample as the first step.

How much of the 1:50 dilution do you make? You want to minimize errors due to pipetting so begin with at least 10 µl of your original sample.

10 µl x 50 (for 50 fold dilution) = 500 µl

In other words, you will be making 500 µl of a 1:50 dilution of your original sample.

10 µl + 490 µl of diluent = 500µl total volume

You can now dilute this 1:50 dilution further to make a 1:5000 final dilution. You do this in 2 steps, each step being a 1:10 dilution.

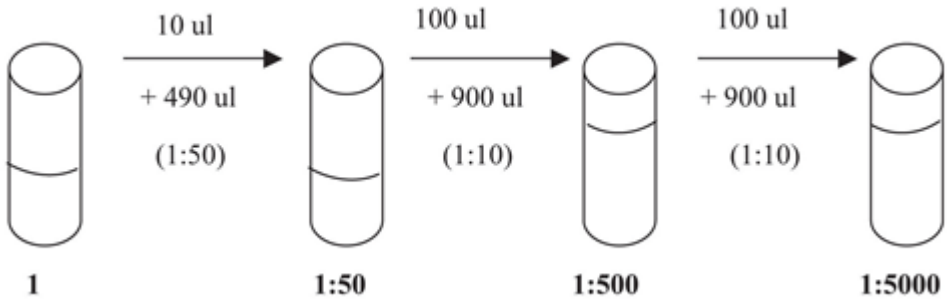
Step 1:

Take 100 µl of your 1:50 dilution and add 900 µl of diluent = 1 mL total (this is now a 1:500 dilution of your original sample).

Step 2:

Take 100 µl of your 1:500 dilution and add 900 µl of diluent = 1 mL total (this is now a 1:5000 dilution of your original sample).

Note: You have to multiply each dilution step to figure out the final dilution (50 x 10 x 10 = 5000)



Another problem

You want 500 µl of 2 mM NaCl. Your stock solution of NaCl is 5M. What do you need to do?

1. Convert your stock solution to units of mM: 5 M = 5000 mM.

2. Determine the dilution factor 5000mM/2mM = 2500

3. Determine the dilution steps.
    - The overall dilution factor has to be 2500.
    - Thus, you could start with a 1:100 dilution (i.e. use 10 µl of your starting solution and add 990 µl of diluent.)
    - You now only need to further dilute by a factor of 25 to achieve a total dilution of 2500.
    - So, dilute your 1:100 dilution again, this time using 20 µl of your 1:100 dilution and 480 µl of diluent.
    - You end up with 500 µl of a 1:2500 dilution.

Adapted from Tricia Laurenson, 1999.

CRISPR-Cas9 Project Overview

The objective of the CRISPR-Cas9 project in this laboratory is to introduce nucleotide changes into a precise location in the genome of yeast. Although discovered only recently, the CRISPR-Cas9 system has been nothing short of revolutionary in allowing biologists to modify and regulate genomes. This could lead to plants that will be more resistant to pests and disease, generate model animals whose genomes have been engineered to process metabolites much like human cells, modify human genomes to correct genetic diseases, generate mosquitos that cannot transmit malaria, just to name a few applications.

The extraordinary power of CRISPR-Cas9 technology raises ethical questions on when and how widely this DNA modification system should be applied to living organisms. For example, there is one report from China claiming that human germlines have already been edited so that the genetically edited humans should be resistant to infection with human immunodeficiency virus (HIV).

There are a variety of CRISPR-Cas systems. We will employ the simplest, Type II, which involves one multifunctional enzyme, Cas9. At one level, CRISPR-Cas9 is just an incredibly precise restriction enzyme, with the precision being supplied by a short guide RNA. In our project, you will design and clone guide RNA sequences into a plasmid vector that can be grown both in bacterial cells and yeast cells. The guide RNA will target the Cas9 endonuclease to a specific yeast gene, and after cleavage, a DNA repair template in the yeast cells will recombine into the cleaved gene, introducing a few edited nucleotides that inactivate the gene products. The inactivated gene results in yeast cells with a visible mutant phenotype. Of course, at each step of this project, we will be performing tests and controls to ensure that the materials needed for the next step are what we want.

Basics of CRISPR-Cas9 Editing

For additional information on CRISPR see the CRISPR 101 e-book:

https://info.addgene.org/download-addgenes-ebook-crispr-101-2nd-edition

Introducing mutations into genomes of a variety of species is critical for research. By identifying phenotypes associated with random mutations, we can begin to elucidate the function of various gene sequences. The ability to introduce targeted, precise genetic changes can be more challenging. Techniques to introduce specific changes, such as TALENS, do exist (for a review see Gupta & Musunuru, 2014), as well as methods to modify RNA molecules, in effect altering gene expression (RNA interference, RNAi). However, CRISPR-Cas9 as a tool to edit DNA sequences has become very popular over the past several years because of the ease, and relatively inexpensive costs, of use.

The CRISPR-Cas9 tool that we are using is based on a system that naturally exists in bacteria. The CRISPR system is a form of immunity found in many bacteria and archaea, against infection by DNA viruses. CRISPR stands for **c**lustered **r**egularly **i**nterspaced **s**hort **p**alindromic

repeats. CRISPR loci are found in the genomes of prokaryotic organisms such as bacteria and archaea. From these loci, short guide RNA sequences are expressed, which bind to the endonuclease called Cas9 (Cas9 name comes from **C**RISPR-**a**ssociated **p**rotein-**9** nuclease, from *Streptococcus pyogenes*). The gRNA-Cas9 complex can target invading DNA molecules, such as DNA viruses, and cleaves them, rendering the virus non-functional and thus preventing infection. CRISPR systems are naturally present in bacteria and are used as an adaptive DNA-based defense against infection by DNA viruses. For more details on the CRISPR locus as a form of adaptive immunity, see the sources below (Dr. Doudna's video lecture, and Ran et al., 2013).

We will not be going over the mechanism of CRISPR immunity in detail here. However, it is important to know the overall concept of immunity, which is that when foreign DNA, for example from a virus, invades a bacteria cell whose ancestors have been previously exposed to the virus, the CRISPR gene locus makes a guide RNA from short sequences that are complementary to sequences in the viral genome. The guide RNA then associates with the Cas9 protein, and cleaves the DNA virus genome, thereby destroying the virus genome and making the cell free of the invading virus. Note that adaptive immunity means that the cell must first be exposed to the pathogen (infectious agent) in order to develop an immunity. Although working by an entirely different mechanism, our antibody-based immune system is also an example of adaptive immunity.

Gupta RM, Musunuru K. Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. J Clin Invest. 2014 Oct;124(10):4154-61. doi: 10.1172/JCI72992. Epub 2014 Oct 1. Review. https://www.jci.org/articles/view/72992

Dr. Jennifer Doudna's lecture on the origins of Crispr-Cas9 technology and how CRISPR-Cas9 editing works: https://youtu.be/SuAxDVBt7kQ

Ran, F., Hsu, P., Wright, J. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8,** 2281–2308 (2013). https://doi.org/10.1038/nprot.2013.143

Use of CRISPR-Cas9 for genetic editing takes advantage of the Cas9 protein. Cas9 is an endonuclease, meaning that it can make double-stranded cuts in DNA. Cas9 needs to be coupled with a small RNA molecule, sometimes called a scaffold-guide RNA (sgRNA). The RNA molecule consists of two parts: a scaffold RNA that binds to Cas9, and a guide RNA (gRNA) that can complementary base pair with the genomic DNA target. Base-pairing between the gRNA and target genomic DNA activates the endonuclease activity of Cas9. CRISPR is gaining widespread use for inducing targeted alterations into genomes of many organisms, including bacteria, yeast (the organism that will be used here), plants, fish, insects, and humans.

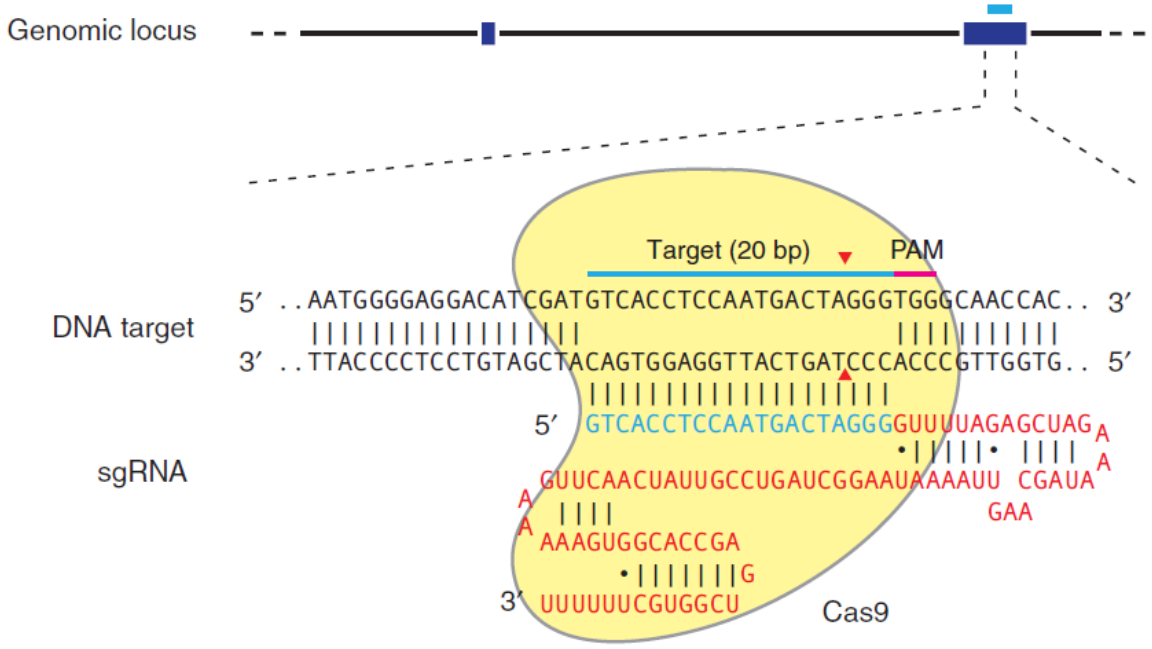


Figure 1 | Schematic of the RNA-guided Cas9 nuclease. The Cas9 nuclease from *S. pyogenes* (in yellow) is targeted to genomic DNA (shown for example is the human *EMX1* locus) by an sgRNA consisting of a 20-nt guide sequence (blue) and a scaffold (red). The guide sequence pairs with the DNA target (blue bar on top strand), directly upstream of a requisite 5′-NGG adjacent motif (PAM; pink). Cas9 mediates a DSB ~3 bp upstream of the PAM (red triangle).

Figure from Ran et al. 2013. Nature protocols,8 (11). **NOTE THERE IS AN ERROR IN THIS FIGURE** – THE target-specific region of the gRNA is indeed an RNA molecule, so it should have uracil instead of thymine (U instead of T). The picture above shows the sequence as a DNA sequence, it should be: 5’ GUCACCUCCAAUGACUAGGG…..

After Cas9 cuts a target DNA sequence the cut is repaired using endogenous processes in the cell. One form of repair is called Non-Homologous End Joining (NHEJ). This form of repair is one that brings together two fragments of DNA, and these two fragments do not have to be homologous to one another to be joined. After the double-stranded break by Cas9, other enzymes bind to the ends of the broken DNA, and modify the ends. The ends are then ligated back together (reforming the broken phosphodiester backbone). The modification and ligation process can be error prone, and it is common for random mutations to be introduced at the repaired break site, such as insertions, deletions, substitutions.

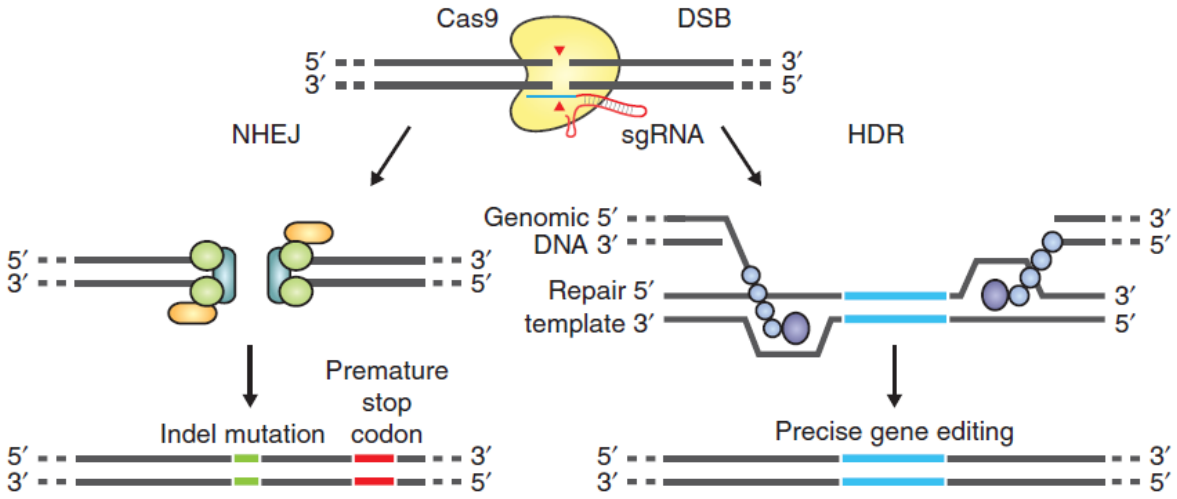


Figure from Ran et al. 2013. Nature protocols,8 (11)

CRISPR terminology:

CRISPR: Clustered Regularly Interspersed Short Palindromic Repeats.

CRISPR is a bacterial adaptive immune system that has been modified for genome engineering. The name comes from repeating segments of DNA that are found in the genomes of prokaryote (bacterial) cells. Each of the repeat sequences is followed by a unique "spacer" DNA sequence, which can encode guide RNAs that target different invading DNA molecules.

Cas9 protein: CRISPR Associated protein 9.

Cas proteins are encoded by genes associated with CRISPR sequences in the bacterial genome. These are nuclease proteins needed to identify and cut the targeted DNA sequence. The first Cas protein to have its activity characterized was Cas9 from *Streptococcus pyogenes*. Cas9 is currently the preferred Cas protein used for CRISPR/Cas gene editing, and is the protein that used for this laboratory.

PAM: Protospacer Adjacent Motif.

Specific DNA sequence that is recognized by Cas9. Cas9 must recognize a PAM sequence, and then this triggers unwinding of some DNA, making the DNA accessible to the guide RNA (gRNA). The target sequence must be 5' to the PAM sequence. Cas9 from *Streptococcus pyogenes*, the Cas9 protein used in this lab, has a PAM sequence of 5' NGG 3'. Novel PAM sequences have been identified in other Cas-like protein systems.

gRNA: Guide RNA.

This is a short RNA molecule that contains the scaffold sequence, necessary for binding of the gRNA to Cas9, and a sequence that is complementary to the target sequence. Sometimes when we talk about the guide RNA, we are only referring to the targeting sequence (but it is important to remember the scaffold sequence is necessary). In bacteria, the transcribed spacer region guides the Cas proteins to foreign DNA sequences (such as invading viruses). The guide RNA that is specific to the target for CRISPR gene editing is usually 20 nucleotides in length.

Target DNA sequence.

Genomic DNA targeted by the CRISPR/Cas9 system by the target-specific guide RNA (gRNA). The target sequence is usually 20 nucleotides long, and is located before (5') the PAM sequence.

Double-stranded (DS) cut, or break: A DS Break made in the targeted gene by the CAS9/guide RNA nuclease.

A break in both strands of the DNA double helix (phosphodiester bond in the DNA backbone of both strands of DNA are broken), caused by the Cas9 protein guided to the target DNA by the gRNA. A DS break elicits DNA repair mechanisms such as Non-Homologous End Joining (NHEJ) or Homology Directed Repair (HDR).

HDR: Homology Directed Repair.

This is a DNA repair mechanism that uses DNA homology to repair a double stranded break (DSB). By designing the HDR template to contain various mutations (insertions, deletions, substitutions) it can be used to introduce specific changes to the target sequence. The region of the HDR sequence containing mutations is flanked by DNA sequence that is homologous to the sequences flanking the Cas9-induced DSB.

HDR template DNA

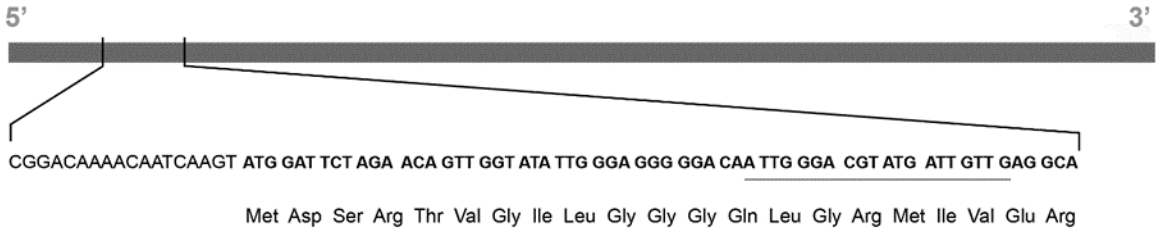
DNA that acts as a template for homology directed repair can be both double or single stranded DNA. In this lab we will use a PCR amplified section of DNA, about 130 bp in length, that is used as a template to repair the dsDNA cut made by the CAS9/guide RNA within the targeted gene.

Example: editing the ADE2 gene in yeast

In this lab we are going to target the yeast gene, *ADE2*, which is involved in *do novo* purine biosynthesis (see below for more details). Mutations that render the gene-product non-functional result in build-up of a precursor in the purine biosynthesis pathway, which turns the yeast cells red in color. In this way, we can easily screen for certain types of mutations. We can also use sequencing of the gene to determine precisely what mutations are introduced after a CRISPR-Cas9 editing experiment.

Below is an example of a gRNA designed to target *ADE2*, and the HDR template designed to introduce nonsense mutations, generating premature stop codons, to render ADE2 non-functional.

The image below shows a portion of the ADE2 gene, specifically near the start of the coding sequence. The thick line represents the double-stranded gene sequence, but then we focus on just the top strand (5' to 3') of the gene sequence, and the corresponding amino acid sequence after the start codon (ATG in the DNA, which would be AUG in the RNA). Note that some 5' non-coding sequence is also shown. The underlined sequence is the gRNA target.



How do you determine the target sequence? This can be done manually, or with the assistance of available programs.

In this lab we will use a program called CRISPR direct (https://crispr.dbcls.jp/). You can enter in a DNA sequence, and the program will identify possible gRNA targets (we will do this together in a subsequent lab).

Notice that the target sequence is followed by a 5' NGG 3' sequence (in the example above, 5'AGG3'). The NGG sequence is the PAM sequence, which must be present immediately downstream (3') to the sequence here the gRNA will bind.

The gRNA sequence to bind to the target region above will have the same sequence as the target:

5′ ATTGGGACGTATGATTGTTG 3′

Once this gRNA is expressed in the yeast cell, it will be an RNA molecule (note that the scaffold sequence is not shown here, but when it is expressed in yeast for a CRISPR-Cas9 experiment the scaffold RNA sequence will be at the 3' end of the molecule):

5′ AUUGGGACGUAUGAUUGUUG 3′

After Cas9 recognizes the PAM, the gRNA can base-pair with the target sequence (note that here we only see the gRNA sequence, and not the scaffold sequence -but in reality the scaffold would be part of the same gRNA transcript, found 3' to the part that binds to the target – see figure above).

```
5′…. ATATTGGGAGGGGGACAATTGGGACGTATGATTGTTGAGGCAAACAGGCTCAAC…. 3′

3′…. TATAACCCTCCCCCTGTTAACCCTGCATACTAACAACTCCGTTTGTCCGAGTTG….5′
                        ||||||||||||||||||||
                    5′ AUUGGGACGUAUGAUUGUUG
```

Cas9 will create a double-stranded cut of the targeted DNA (about 3 bp upstream of the PAM). After the cut, repair mechanisms will be induced. Both NHEJ and HDR can occur, depending on the availability of HDR template and the frequency of HDR and NHEJ (which varies depending on cell type and species).

Homology Directed Repair of ADE2

To induce homology directed repair (HDR) one would need to design an HDR template to be used with Cas9-gRNA to introduce specific mutations into *ADE2*.

The HDR fragment is longer than the gRNA sequence, it typically extends about 50 bp on either side of the target sequence. The outer sequences of the HDR template (sometimes called Homology Arms) are homologous to the wildtype sequence (the same as wildtype), but the middle region of the HDR template can vary from wildtype – including desired sequence changes, such as insertions, deletions, substitutions.

Below is a schematic showing the wildtype *ADE2* sequence, corresponding amino acid sequence, and the HDR template sequence. Note that both the ADE2 genomic sequence and the HDR are actually double-stranded DNA molecules, but here we are just looking at the top strand of each (5' to 3').

Multiple nucleotides in the HDR have been changed compared to the wildtype to result in the introduction of premature stop codons (nonsense mutations). Introduction of these mutations by homology directed repair is predicted to result in a non-functional gene product because translation will terminate early as a result of the nonsense mutations.

The PAM sequence is also mutated. After HDR occurs, the PAM will no longer exist at this site in the gene, so Cas9-gRNA will not target and re-cut the sequence.

The HDR fragment that is produced is much larger than the guide RNA, it should extend at least 50 or so bases to either of the guide RNA target. The repair fragment we will use for the *ADE2* gene is shown in the box above. Note from the diagram that the fragment is edited. Some nucleotides within the gene have been altered to now contain early stop codons. There are three edits to stop codons within the HDR fragment.

(see next page)

```
                        Met Asp Ser Arg Thr Val Gly Ile Leu Gly

5′ CGGACAAAACAATCAAGT ATG GAT TCT AGA ACA GTT GGT ATA TTG GGA GGG GGA

5′ CGGACAAAACAATCAAGT ATG GAT TCT AGA ACA GTT GGT ATA TTG GGA GGG GGA


Gln Leu Gly Arg Met Ile Val Glu Arg Asn Arg Leu Asn Ile Lys Thr Val

CAA TTG GGA CGT ATG ATT GTT GAG GCA AAC AGG CTC AAC ATT AAG ACG GTA

TAA TTG GGA CGT ATG ATT GTT TAG TAA AAC AGG CTC AAC ATT AAG ACG GTA


Ile Leu Asp Arg Glu Ile Ser Pro Ala Lys Gln Ile Ser Asn

ATA CTA GAT GCT GAA AAT TCT CCT GCC AAA CAA ATA AGC AAC 3′

ATA CTA GAT GCT GAA AAT TCT CC 3′
```

The sequence below represents a portion of the ADE2 gene and protein sequence.

Top line: ADE2 protein sequence
Second line: wild-type *ADE2* gene sequence
Third line: HDR template sequence

- The target and guide RNA sequence is indicated shaded.
- The PAM in the original sequence is underlined.
- **Bolded** nucleotides in the HDR show how the HDR sequence differs from the wildtype sequence.

The HDR1 sequence alone (5’ to 3’ strand)

CGGACAAAACAATCAAGTATGGATTCTAGAACAGTTGGTATATTGGGAGGGGGA**TAA**TTGGGACGTATGATTGTT**T**A
G**TA**AAACAGGCTCAACATTAAGACGGTAATACTAGATGCTGAAAATTCTCC

Below is an illustration to summarize various ways that sources of Cas9 and gRNA can be added to cells, and where the molecules go once in the cell. Note that in the example below, the “CRISPR DNA vector” is a plasmid, and the plasmid contains the Cas9 gene and gRNA sequence. If these genes have promoters to drive expression in the specific cell type being transformed, both Cas9 and gRNA can be expressed from the plasmid DNA.

In our experiments we will be transforming yeast with a plasmid containing Cas9 and the gRNA, controlled by yeast-specific promoters. We will also add to the transformation mixture many copies of the HDR template to induce homology directed repair.

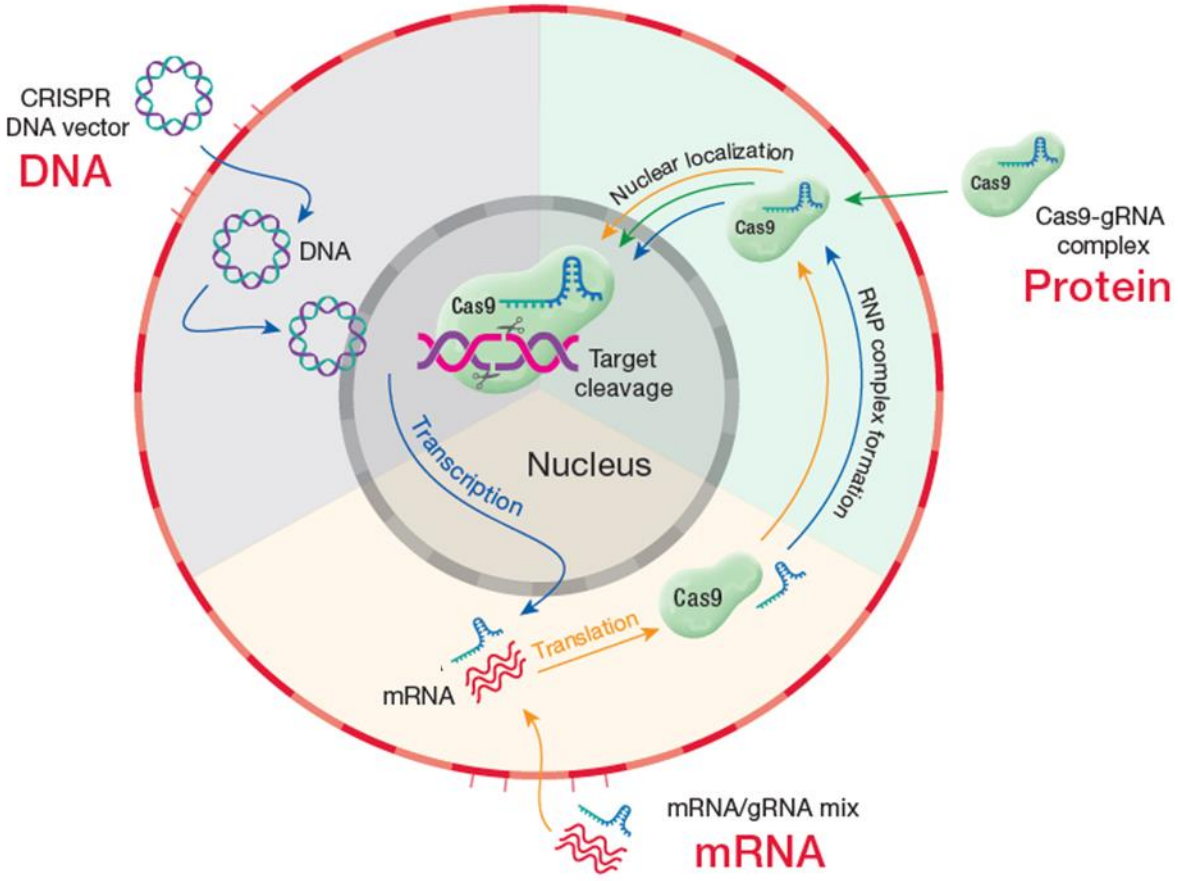
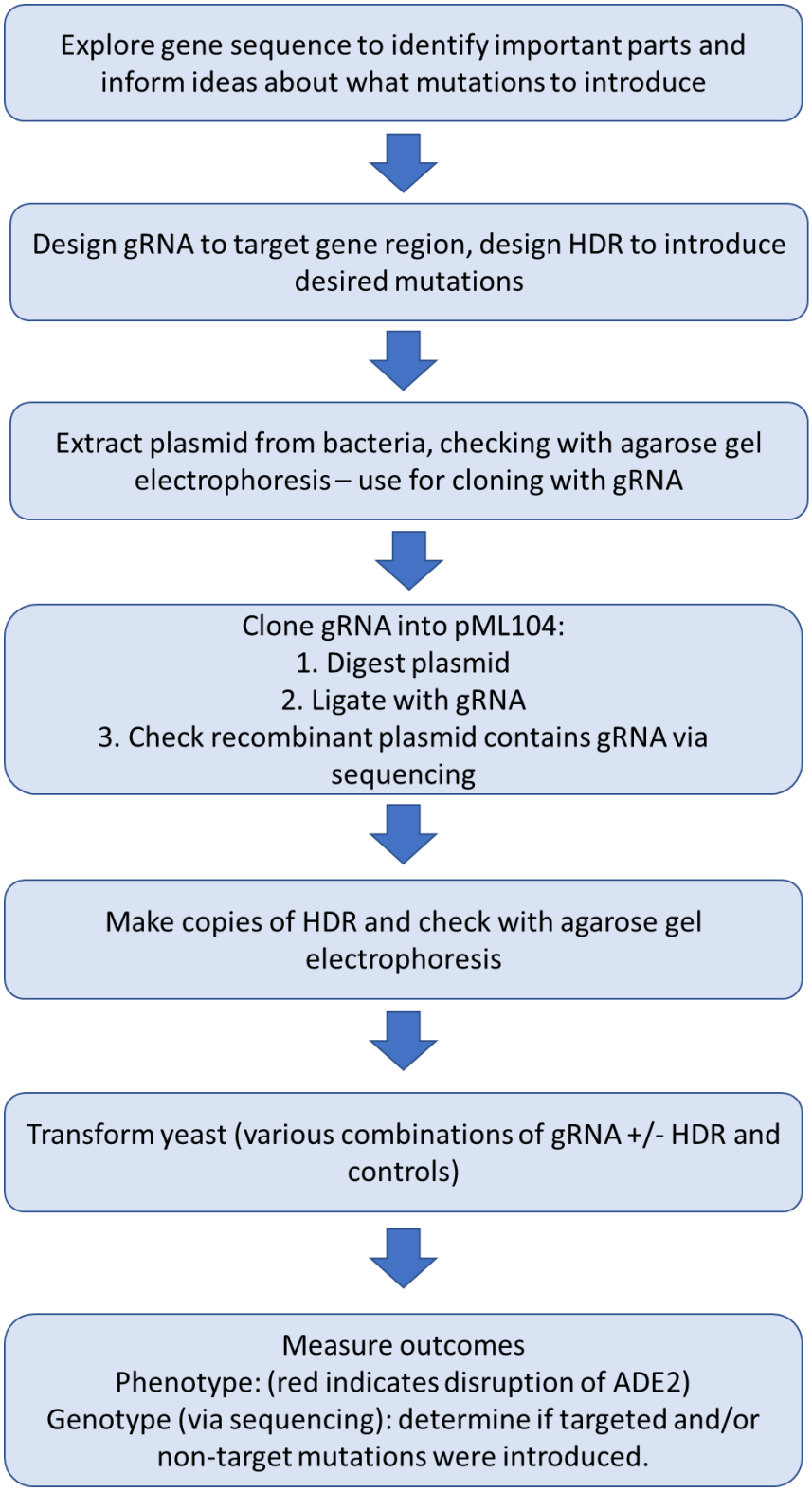


Figure from thermofisher.com/bioprobes

Experimental Overview



Cultures used in the lab

E. coli DH5α

- Wildtype cells are **not** antibiotic resistant
- Typically grown in/on LB media. This is a nutrient rich media typically used to grow bacteria. It contains tryptone, yeast extract, and sodium chloride: https://en.wikipedia.org/wiki/Lysogeny_broth
- Our bacterial cells were originally sourced from Zymo: https://www.zymoresearch.com/collections/mix-go-competent-cells/products/mix-and-go-competent-cells-dh5-alpha
- More information on the DH5a genotype: https://ecoliwiki.org/colipedia/index.php/DH5_alpha
- Wildtype colonies appear smooth and white
- If growing in/on media with Kanamycin, use final concentration of 50 ug/ml
- If growing in/on media with Ampicillin, use a final concentration of 50 ug/ml

Dam^{-}*/Dcm*$^{-}$ *E. coli* (C2925)

- Wildtype cells are **not** antibiotic resistant
- Typically grown in/on LB media.
- Sourced from New England Biolabs (https://www.neb.com/products/c2925-dam-dcm-competent-e-coli)
- Wildtype colonies appear smooth and white
- If growing in/on media with Kanamycin, use final concentration of 50 ug/ml
- If growing in/on media with Ampicillin, use a final concentration of 50 ug/ml
- *dam-/dcm-* genotype means that the genes for methylases that would normally methylate DNA affecting the BclI site are no longer active. This is important because the BclI restriction enzyme will not cut the BclI site if the plasmid DNA is methylated. We culture the pML104 plasmid in this *dam-/dcm-* strain of *E. coli* because after harvesting (extracting) pML104 from cultured *E. coli* will want to digest it with the BclI enzyme.

Saccharomyces cerevisiae

- Wildtype cells are **not** resistant to eukaryotic selection agents
- Background info on *S. cerevisiae*: https://www.jove.com/science-education/5081/an-introduction-to-saccharomyces-cerevisiae
- We are using strain BY4743. It is a haploid strain which is auxotrophic for uracil and sensitive to G418.
- Typically grown in/on YPD, which contains 1% yeast extract, 2% peptone and 2% dextrose
- Wildtype colonies appear white
- If growing in/on media with G418, use a final concentration of 200 µg/ml
- If using uracil as a selective agent, grow on uracil-deficient media and yeast must be transformed with a URA3-carrying plasmid

Yeast gene and protein nomenclature: yeast *genes* are presented in all caps for the wild type, and in lower case for the mutant. Thus, *ADE2* is wild type, and *ade2* is mutant. Gene names are italicized. Proteins are in all caps, and not italicized.

Adenine2 (ADE2). We will be editing the *ADE2* gene.

The *ADE2* gene encodes the enzyme Phosphoribosylaminoimidazole Carboxylase. This enzyme catalyzes a step in the *'de novo'* purine nucleotide biosynthetic pathway. Such a pathway is needed if purines are not abundant in the environment. In *ade2* mutants, the cells are deprived of adenine and red pigment accumulates, resulting in a pink phenotype.

Summary of ADE2 from https://www.yeastgenome.org/locus/S000005654#reference (refer to website for references listed below)

"The ADE2 gene encodes phosphoribosylaminoimidazole carboxylase, which catalyzes the sixth step in the de novo biosynthesis of purine nucleotides (pathway here; **2**). ADE2 transcription is regulated specifically by adenine and also by general amino-acid control (**3**). Gene expression is repressed in the presence of adenine, activated in the absence of adenine, and also slightly increased in conditions of amino-acid starvation (**5**, **3**, **6**). Maximal ADE2 expression is dependent on the transcriptional activators Gcn4p, Bas1p, and Pho2p, which bind to 5'-TGACTC-3' sites in the ADE2 promoter (**5**, **6**). Gcn4p upregulates ADE2 expression during amino acid starvation, a condition which inhibits Gcn4p degradation (**7**). Bas1p binds the ADE2 promoter in both repressing and derepressing growth conditions, but binding increases slightly when adenine is limiting. Under derepressing conditions, Pho2p also binds the ADE2 promoter in a Bas1p-dependent manner, and transcription is induced (**6**). Derepression by Bas1p and Pho2p also requires the purine biosynthesis pathway intermediates 5-phosphoribosyl-4-succinocarboxamide-5-aminoimidazole and 5-phosphoribosyl-4-carboxamide- 5-aminoimidazole, which are thought to promote interaction between these two activators (**8**, **6**).**Mutations in ADE2 lead to the accumulation of purine precursors in the vacuole, which causes the colony to be red in color. This pigmentation phenotype is widely utilized as a marker for genetic selection and screening (2, 4 and references contained therein)."**

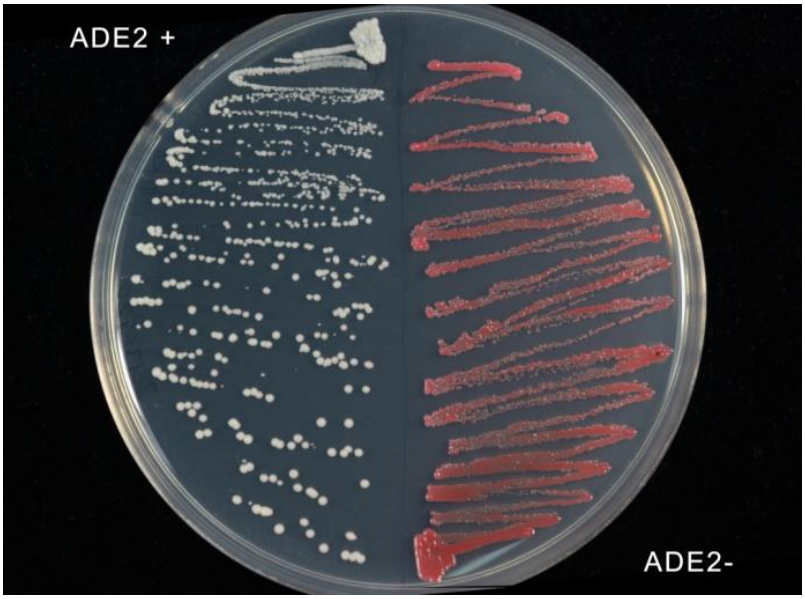


Image from parts.igem.org

Plasmid used in the lab – pML104

This is a cloning/expression plasmid (vector) that contains two genes relevant to CRISPR function. First, it encodes the CAS9 protein. It also contains a location where a guide RNA of our design can be inserted (in between the *Bcl*I and *Swa*I restriction enzyme sites). When transformed into the host cells (in our case, the transformed yeast cells), the expressed CAS9 protein combines with the 20 nt guide RNA to form a targeted nuclease that will make a double stranded (ds) cut within the targeted gene. *Note that since this plasmid must be grown in both prokaryotic E. coli, and eukaryotic yeast, it contains promoters and origins of replication for both types of cells.* It also contains resistance genes for a selectable genetic marker in both *E. coli* and yeast cells.

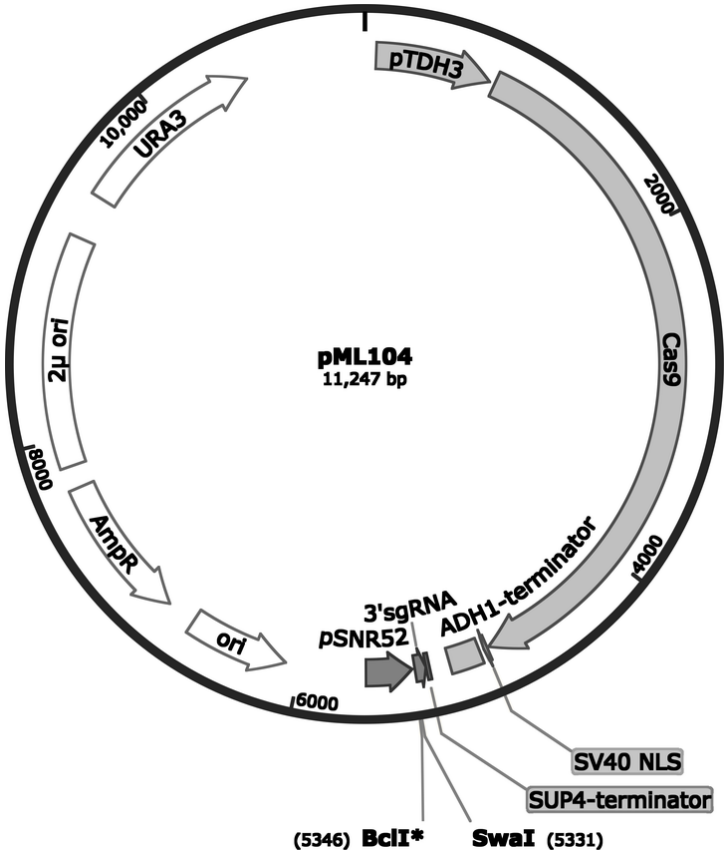


Figure from Addgene.

- Sourced from Addgene: https://www.addgene.org/67638/
- Expresses Cas9 and scaffold RNA (sgRNA) for genome editing in yeast
    - This plasmid is missing 13 nucleotides from the 5' end of the scaffold sequence that must be added along with the targeting gRNA sequence
- Uses *Bcl*I and *Swa*I restriction enzymes to clone gRNA
- BclI restriction enzyme will not cut the BclI site if the plasmid DNA is methylated. We culture pML104 in a strain of E. coli that is mutated such that they methylases that would normally methylate DNA affecting the BclI site are no longer active (the strain is called a *dam-/dcm-* strain, specifically strain C2925 from New England Biolabs)
- Confers ampicillin resistance in *E. coli*
- Contains URA3 marker for yeast selection
- 2u ori is an origin of replication recognized in yeast
- Ori is an origin of replication recognized in *E. coli*

Alkaline Lysis Plasmid Purification

Often we need to extract plasmid from bacteria in order to use the plasmid for other experiments.

Alkaline lysis purification is a common technique in which bacteria are lysed and DNA is denatured simultaneously in an alkaline SDS solution (at pH 12). The solution is then neutralized in an acidic, high salt buffer at which point the plasmid DNA re-natures. The solution is centrifuged, and while the plasmid remains in the supernatant, the chromosomal DNA is trapped in a precipitate formed by SDS in the presence of potassium, and is pelleted.

For more information see this site: https://bitesizebio.com/180/the-basics-how-alkaline-lysis-works/

Here is an explanation of the method from the Qiagen manual (Qiagen is the company from which we purchase the alkaline lysis kit materials):
"Bacteria are lysed under alkaline conditions. After harvesting and resuspension, the bacterial cells are lysed in NaOH/SDS (Buffer P2) in the presence of RNase A. SDS solubilizes the phospholipid and protein components of the cell membrane, leading to lysis and release of the cell contents while the alkaline conditions denature the chromosomal and plasmid DNAs, as well as proteins. The optimized lysis time allows maximum release of plasmid DNA without release of chromosomal DNA, while minimizing the exposure of the plasmid to denaturing conditions. Long exposure to alkaline conditions may cause the plasmid to become irreversibly denatured. This denatured form of the plasmid runs faster on agarose gels and is resistant to restriction enzyme digestion.

The lysate is neutralized and adjusted to high-salt binding conditions in one step by the addition of Buffer N3. The high salt concentration causes denatured proteins, chromosomal DNA, cellular debris, and SDS to precipitate, while the smaller plasmid DNA renatures correctly and stays in solution. To prevent contamination of plasmid DNA with chromosomal DNA, vigorous stirring and vortexing must be avoided during lysis. Separation of plasmid from chromosomal DNA is based on co-precipitation of the cell wall-bound chromosomal DNA with insoluble complexes containing salt, detergent, and protein. Plasmid DNA remains in the clear supernatant. Vigorous treatment during the lysis procedure will shear the bacterial chromosome, leaving free chromosomal DNA fragments in the supernatant."

At this point, the supernatant containing the plasmid is applied to a silica column. The DNA binds to the silica matrix, while RNA, proteins, and other cellular components do not bind and flow through the column when it is centrifuged.

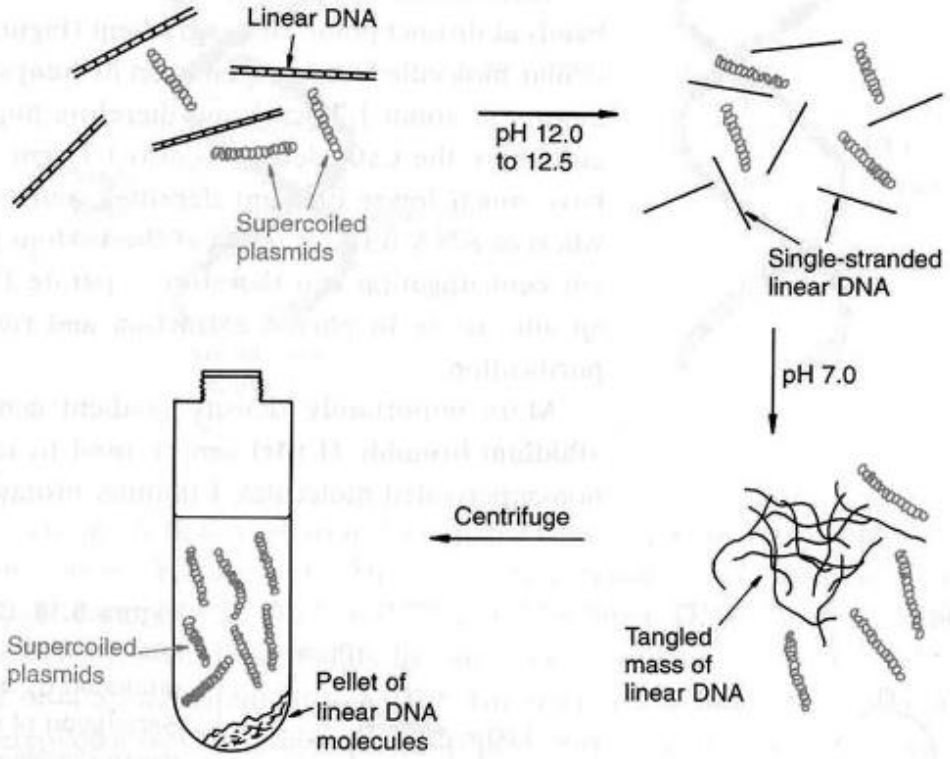


Figure from pediaa.com. Note that bacterial genomic DNA will be circular, but it is still very large and once it become separated it is hard to reconstitute, thus pellets out during centrifugation.

Spectrophotometric Analysis of DNA & RNA

Before you use DNA and RNA samples for other reactions or experiments it is often necessary to determine the concentration and purity of your DNA or RNA sample. You will use a UV-spectrophotometer to **quantify** and assess the **purity** of the DNA or RNA samples that you isolated.

DNA and RNA absorb light maximally at a wavelength of 260nm, and the absorbance of DNA at this wavelength is used to determine the concentration of the sample. The absorption of 1 OD (optical density = absorbance unit) at 260 nm is equivalent to approximately 50 µg/ml double-stranded DNA, approximately 33 µg/ml single-stranded DNA, 40 µg/ml RNA, or approximately 30 µg/ml for oligonucleotides.

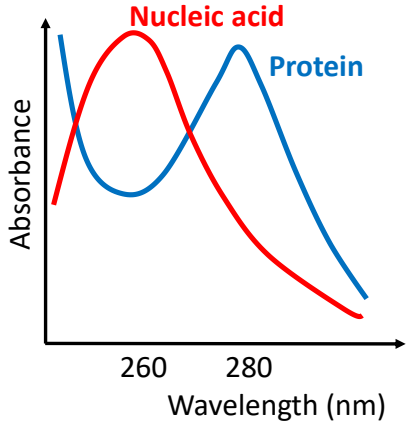


Figure 9. Absorbance spectra of nucleic acids and protein.

DNA absorbs less light at 280 nm, and the A260/A280 ratio of pure DNA should be1.8. However, proteins absorb light better at 280 nm than at 260 nm, and thus if there is a lot of protein contamination, the A260/A280 ratio drops below 1.8. Thus, this ratio is often used to estimate the purity of nucleic acid.

More information:
http://bitesizebio.com/25329/strengths-and-limitations-of-your-nanodrop/

Restriction enzyme cloning to add gRNAs to pML104 plasmid

Depending on the plasmid, there are different ways to add your desired gRNA sequence. We will be using the restriction enzyme cloning method to add unique guide RNAs, that you all have designed, into the pML104 plasmid.

Restriction Enzymes

Additional information on restriction enzymes:
Video: https://www.neb.com/products/restriction-endonucleases

Type II restriction enzymes (REs) are bacterial endonucleases that bind DNA and cleave it at a specific recognition sequence or restriction site. These enzymes are present in the same compartment of the bacteria as the chromosomal DNA, so why do these enzymes only digest foreign DNA and not the bacterial DNA? Because methyltransferase enzymes are present as well in the bacterial cells. The methyltransferase methylate adenines or cytosines in the recognition sequence in the host DNA and thus make the sequence resistant to the action of the RE. Names of restriction enzymes are derived from the name of the bacteria from which they were isolated—so the restriction enzyme EcoRI was isolated from *E. coli*, Sal I from *S. albus*, etc. When more than one enzyme modification system is found in the same strain, it is designated by a Roman numeral. So *Hind*III = strain *H. influenzae* Rd, modification system III. A restriction modification system is both the restriction enzyme and methyltransferase.

Most type II restriction enzymes have the following characteristics:

- Recognize symmetric DNA sequence (palindromic, reads the same in the 5’ to 3’direction on both strands).
- Recognition sequence is usually 4–6 nucleotides long.
- Usually cleave within recognition sequence—leave 5’P and 3’OH.

Ligation

To create recombinant plasmids using restriction enzyme digestion we need to ligate two or more DNA fragments together. If you have digested fragments such that they can only ligate in a specific orientation we call that **directional cloning**. This is usually accomplished using two different restriction enzymes (at least one of which creates sticky overhangs) to open a plasmid and digest an insert.

Ligation of two DNA fragments is achieved using T4 DNA Ligase. The role of T4 ligase *in vivo* is to repair single strand breaks in the phosphodiester bonds along a dsDNA molecule. You will be using T4 ligase to join or ligate the ends of two different DNA molecules. In order for T4 ligase to work, there must be 5' phosphates and 3' hydroxyls (OH) present on the ends of the two molecules; the enzyme also requires magnesium and ATP.

When setting up a ligation reaction it is important to consider the molar ratio of insert to vector (plasmid). Varying the ratio of insert: vector can affect the ligation efficiency, and thus the likelihood of getting your desired recombinant plasmid. Insert : vector ratios can vary from 1:1 to 10:1. Moles quantify the amount of a substance, so in the case of DNA fragments we need to consider the size (number of base pairs) of a fragment when determining moles. For example, the moles of DNA in 10 ng of a 3 kb plasmid is not the same as the moles of DNA in 10 ng of a 1 kb plasmid.

By increasing the moles of insert relative the moles of vector, you can increase the chances that the ends of insert will collide with the ends of the vector, and be ligated by ligase. This is why some people will prefer to set-up a 2:1 or 3:1 insert: vector ligation compared to a 1:1. However, each combination of insert and vector ligation can vary depending on compatibility of ends (e.g. blunt end vs. sticky ends).

We often have a known mass of insert and vector, but need to convert to moles to figure out how much of each to add to a ligation.

Converting mass of double-stranded DNA (dsDNA) to Moles:

moles dsDNA (mol) = mass of dsDNA (g)/((length of dsDNA (bp) x 617.96 g/mol) + 36.04 g/mol)

- moles dsDNA = mass of dsDNA (g)/molecular weight of dsDNA (g/mol)
- molecular weight of dsDNA = (number of base pairs of dsDNA x average molecular weight of a base pair) + 36.04 g/mol
- average molecular weight of a base pair = 617.96 g/mol, excluding the water molecule removed during polymerization.
- The 36.04 g/mol accounts for the 2 -OH and 2 -H added back to the ends.
- Bases are assumed to be unmodified.

This online tool can be used to help determine moles and calculate ratios (be sure to select the appropriate tool from the list):
https://nebiocalculator.neb.com/#!/ligation

Other useful information:

- Average weight of a DNA basepair (sodium salt) = 620 daltons
- A_{260} unit ds DNA = 50 µg/ml
- MW of a double-stranded DNA molecule = (# of base pairs) X (620 daltons/base pair)
- 1 µg of 1000 bp DNA = 1.52 pmol = 9.1 X 10^{11} molecules
- 1 µg of λ DNA (48,502 bp) = 0.03 pmol = 1.8 X 10^{10} molecules
- 1 pmol of 1000 bp DNA = 0.66 µg

Annealed Oligo + Restriction Enzyme Cloning

Oligo is short for oligonucleotide. An oligonucleotide is a short stretch of nucleotides (a polynucleotide).

Annealed oligo cloning can be an easy way to add small DNA fragments to a plasmid. With this method you design oligos that contain your insert sequence (e.g. gRNA sequence) and you add the restriction enzyme site overhangs to the ends of the oligos. Hybridizing the oligos (annealing) creates a double stranded DNA fragment with overhangs, which can be ligated to compatible overhangs after digesting your plasmid with the appropriate restriction enzymes. Here is an example. In the example below the oligos were designed so they have BamHI and SalI overhangs after annealing.

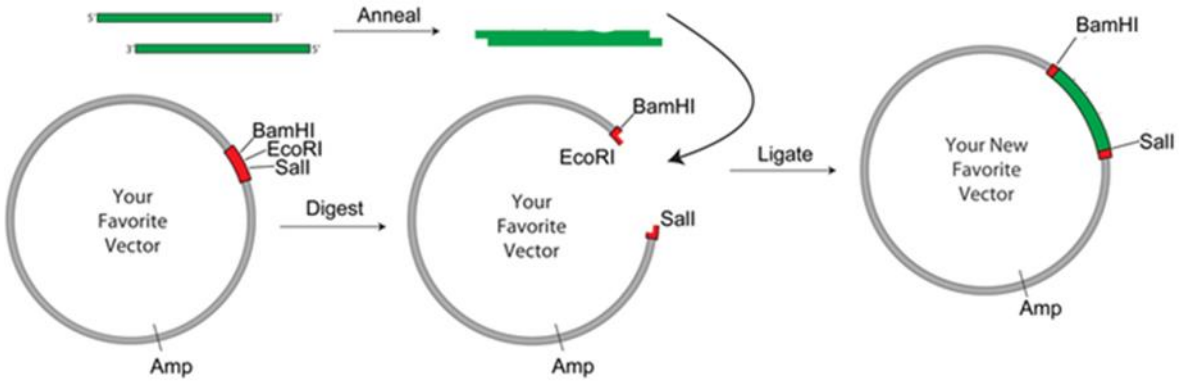


Figure modified from Addgene: https://www.addgene.org/protocols/annealed-oligo-cloning/

For our experiment we will be designing oligos that are a guide RNA sequence (we will call this gRNA2 and gRNA3, as we are already providing gRNA1).

The oligos will be designed such that they are complementary except for the necessary overhangs to ligate (with T4 DNA ligase) the annealed oligos into pML104. Recall that there are two restriction enzyme sites immediate upstream of the scaffold RNA sequence in the pML104 plasmid: BclI and SwaI. To insert gRNA2 into the digested plasmid, the top oligo must have a BclI overhang (SwaI digestion creates a blunt-end). In the example below, the researchers are adding a gRNA against the *TRP1* gene.

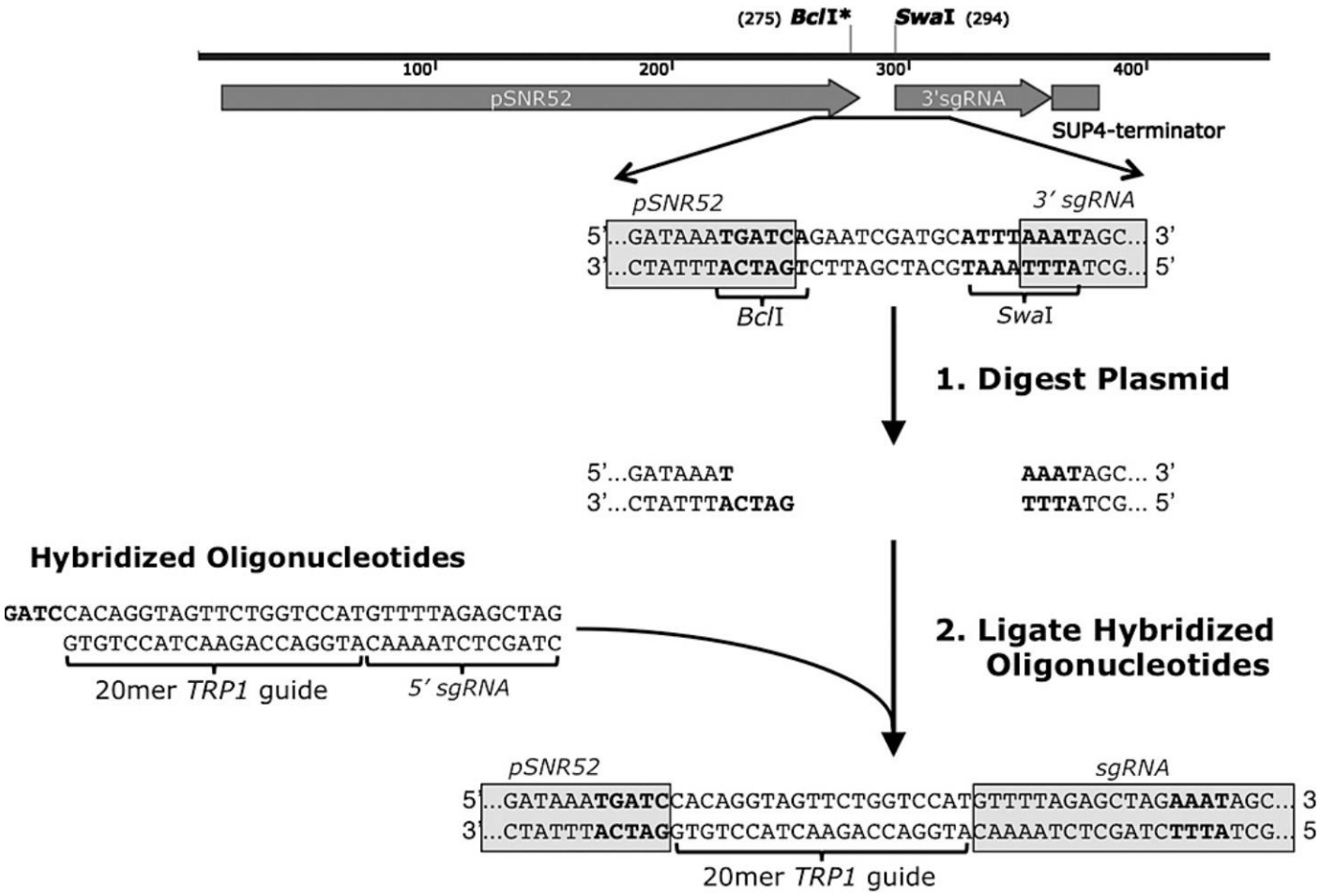


Figure from Laughery et al. 2015. *Yeast, 32(12):711-20. doi: 10.1002/yea.3098.*

Notice how a very short stretch of sequence will be released from the plasmid when it is digested with BclI and SwaI. We sometimes refer to this released fragment as a “stuffer”. After digesting the plasmid, it is common to use a clean-up method to separate the digested plasmid from the stuffer. If the stuffer is not removed, it can re-ligate into the plasmid during the ligation phase, competing with the desired insert and thus reducing the likelihood of recovering your desired recombinant plasmid.

Another important note is that the pML104 plasmid contains only the 3’ end of the scaffold guide RNA (sgRNA). The oligos that are designed to add our gRNA must also contain the 5’ end of the sgRNA sequence. The double-stranded 5’ sgRNA sequence is as follows:

```
5′ GTTTTAGAGCTAG 3′
3′ CAAAATCTCGATC 5′
```

Agarose Gel Electrophoresis for analysis of DNA samples

Additional resources:

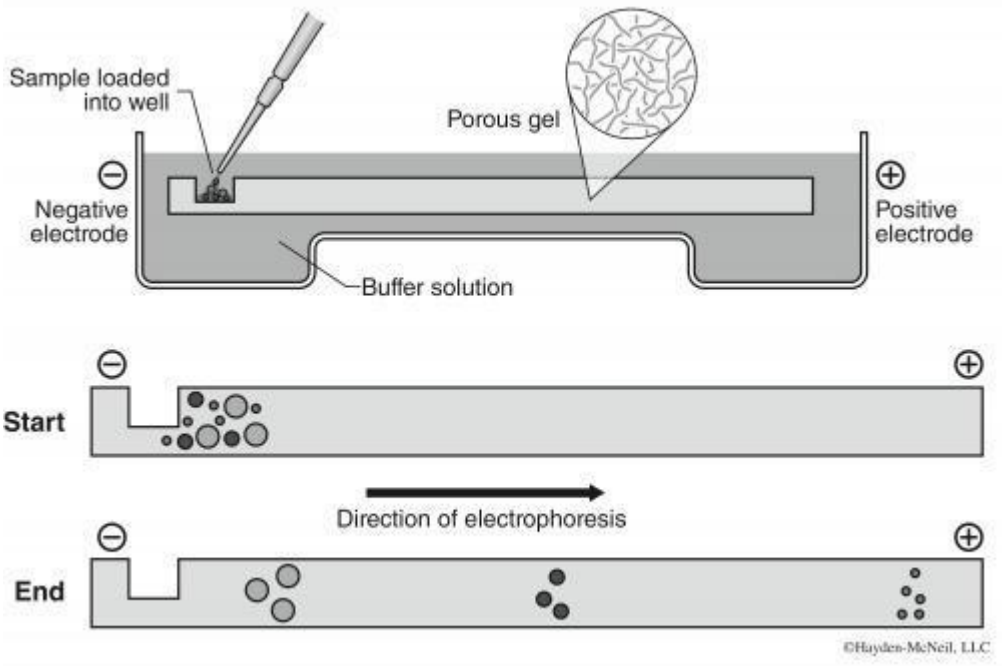
https://www.youtube.com/watch?v=TIZRGt3YAug

https://www.jove.com/video/3923/agarose-gel-electrophoresis-for-the-separation-of-dna-fragments

Electrophoresis is a commonly used technique that relies on the ability of charged particles in solution to migrate in an electric field. Typically, the molecules migrate through a porous, gel-like matrix so that they can be localized following electrophoresis. You will be using agarose gels as the matrix to separate DNA molecules of different sizes and conformations. Agarose gels are used both as an analytical method to determine the size or amount of DNA fragments and as a preparative method to purify DNA fragments.

Agarose is a component of agar, a complex polysaccharide derived from seaweed. When boiled in buffer, agarose dissolves, and then it solidifies upon cooling and has pores within it that act as a sieve to separate DNA fragments by size. The agarose gel is prepared in a pH-buffered solution that conducts an electric current through the gel, and a power supply generates the electric field through which the DNA molecules move.

DNA has a relatively uniform negative charge density due to its phosphate groups. When placed in an electric field in an agarose gel, negatively charged DNA molecules will migrate toward the positive electrode. The sieving properties of the agarose matrix cause larger (longer) molecules to migrate more slowly, and migrate a shorter distance in a given time compared to DNA molecules that are shorter in length molecules. The rate at which the DNA migrates is inversely proportional to its length. Whenever we run an agarose gel, standards of known fragment lengths (base pairs) and amounts (e.g. nanograms) are run as a guide to determine length and amount of DNA fragments (ng) in the experimental samples.



Side-view of an agarose gel being loaded, and DNA molecules migrating through the gel.

The ability to visualize the DNA on your gel after it runs is based on the intercalation of a fluorescent dye called SYBR Safe in between the stacked bases of the DNA. When this occurs, the SYBR safe will fluoresce at a visible wavelength when illuminated with UV light. Thus, in order to see your DNA bands, you must place your gel on a UV or blue light box. Note: UV light is harmful to your eyes and you must wear safety glasses when viewing your gels.

We will use a λ *Hind*III standard. The standard we used is a common one, and consists of DNA from a virus called lambda (λ) bacteriophage, digested with the restriction enzyme, λ *Hind*III. Each lambda phage head contains a 48,502 base pair (bp) DNA molecule. When this DNA is digested with the restriction enzyme *Hind*III, eight discreet fragments of DNA of different sizes are generated, as seen at right. Note that the a 564 bp and 125bp bands are not always seen when we run gels because they are faint in intensity, and the small size means they can (sometimes) migrate off the bottom of the gel.

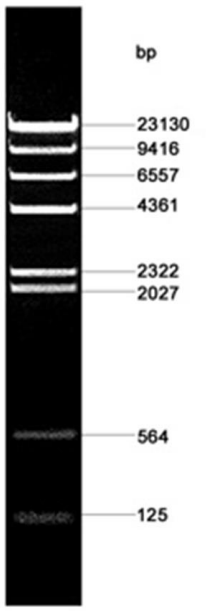


Figure 3. λ*Hind*III ladder run on 0.7%

Molecular weight (often substituted for size when dealing with agarose gels) is not actually the only factor that determines how far a DNA fragment will migrate in an agarose gel. Conformation also has an effect on migration. For example, supercoiled plasmid DNA will run faster than a cut or uncoiled plasmid of the same molecular weight (see next section for more information). Thus, it is important that, if you are using an agarose gel to determine the size of linear DNA fragments in an experimental sample, the standards you use should also be linear.

Each sized band comprises a known percentage of the total ladder. So, we know exactly how many ng of DNA make up each sized band in the ladder. See the descriptions on the following pages for more information on using the ladder to estimate amount of DNA in each band.

Advantages of running gel:

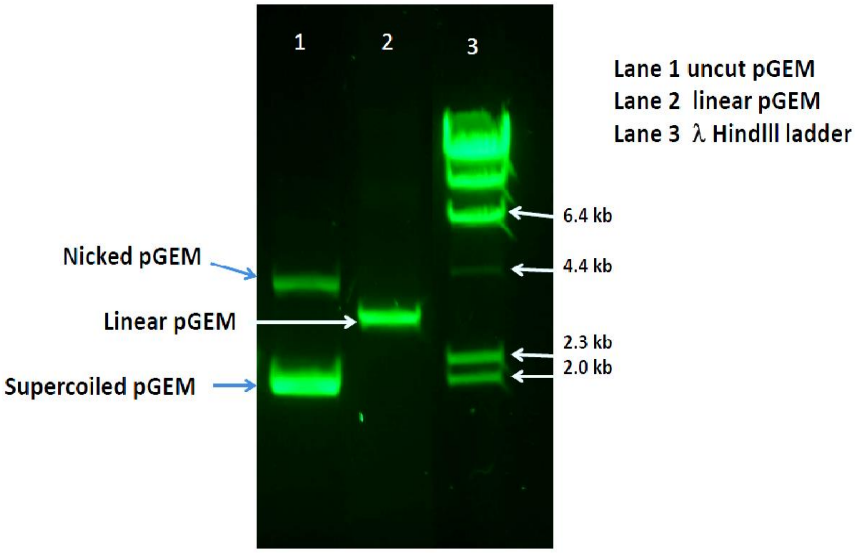
- You can determine the size of your DNA pieces.
- You can check the quality (integrity) of your DNA sample.
- You can estimate the amount of DNA in your DNA sample by comparing the intensity of the band to that of a band in the DNA standard/ladder.

Plasmids versus Linear DNA in Agarose Gels

In cells, plasmids are supercoiled, circular structures that are compact. When both strands of a plasmid are cut with a restriction enzyme at only one point on the plasmid, the plasmid becomes linear. When isolated from cells, the phosphodiester bond on one strand can be broken or nicked – the supercoiling then is released, and the plasmid takes on an open circle form.

Although the three different conformations of the same plasmid are the same number of base pairs, they run differently on a gel. The supercoiled form, which is the most compact, migrates most quickly. The open circle (sometimes also called nicked) form runs the slowest, and the linear plasmid is in between in terms of the rate of migration.
Comparison of cut and uncut plasmid DNA on an agarose gel. pGEM is a circular plasmid that is 3200 base pairs (bp, or 3.2 kilobases, kb) in length. Although all three forms of the pGEM plasmid are the same size (3200 bp), only the linear form runs “true to size” relative to the

standards which are also linear fragments of DNA.



Determining Size of DNA Fragments Separated by AGE

- The size of the linear pieces of DNA in a sample can be determined by comparing how far they migrated into the gel relative to how far the DNA bands of known size in a standard migrate into the gel.

- For example:
    - If your DNA sample band ran the same distance as the 2.3 kilobase (kb) band in the λ*Hind*III ladder, we would conclude your sample contains fragments that are about 2.3 kb in length.

    - If your DNA sample migrated to a position in the gel that is in-between the λ*Hind*III 2.3 kb band and 2.0 kb band we might conclude your sample contains fragments that are around 2.1 or 2.2 kb.

Estimating DNA Quantities for Discrete Bands on an Agarose Gel

Because the fluorescence of Sybr Safe that is intercalated into DNA is proportional to the total mass of DNA, you can estimate the amount of DNA in your sample by comparing it with gel bands from a "standard" amount of the λ HindIII digest. Before you do this, however, you must determine how much DNA is in each band of the λ*Hind*III, as follows:

- Recall the λ*Hind*III ladder contains various-sized fragments of DNA. Each fragment, generated by digesting the 48, 502 bp lambda DNA, represents a percentage of the total DNA molecule (see Table below).

- In BIMM 101 we will provide you with a λ*Hind*III solution that has been prepared such that **10 µl contains a total of 500 ng** of digested the λ*Hind*III DNA. If you load 10 µl of λ*Hind*III onto your gel, you can calculate how many ng of DNA are in each individual band by multiplying this total of 500 ng by the % of the total each fragment comprises. This has been done in the table below.

Percentage and amount of each fragment of DNA comprising the λ HindIII ladder.

	Number of Base Pairs	% of Total (48,502 base pairs)	ng DNA in 10 µl of λ HindIII Stock (500 ng total)
1	23,130	48%	240
2	9,416	19%	95
3	6,557	14%	70
4	4,361	na*	na*
5	2,322	5%	25
6	2,027	4%	20
7**	564	1%	5
8**	125	0.2%	1

**Do not calculate mass based on this band, but you can use it to determine size. The cohesive ends of the 23 kb and 4.4 kb band are known to join, so the intensity of the 4.4 kb band is an underestimate of how much is truly present in the sample (and likewise, the 23 kb band intensity is a slight overestimate). You can use the 4.4 kb band to estimate size.*

***Band 7 and 8 are sometimes not visible on the gels we are running.*

- To determine **how many ng of DNA you have in a band**, find the λ*Hind*III band that most closely matches your band in intensity and thickness, not how far down the gel it ran. If the intensity of your band is intermediate between two of the standards, estimate the amount or take the average of the two standards.

- You can now determine the concentration of DNA in your sample by dividing the amount of DNA in the band by the volume of DNA that was loaded on the gel.

Polymerase Chain Reaction (PCR)

Additional resources on PCR:
https://www.jove.com/science-education/5056/pcr-the-polymerase-chain-reaction
(Note that to view JOVE videos you have to be using UCSD internet or connected to UCSD through a VPN).

https://www.youtube.com/watch?v=2KoLnIwoZKU

https://bitesizebio.com/19132/pcr-basics-what-is-pcr/

DNA Polymerases used in PCR

Various types of polymerase can be selected for PCR. For example, Taq polymerase is a DNA polymerase I from the thermostable prokaryote *Thermus aquaticus*. It is the most commonly used polymerase for the ease and flexibility. However, sometimes a high-fidelity polymerase which has a lower error rate than Taq is desired. For example, Pfu DNA polymerase, which comes from the hyperthermophilic archaeon *Pyrococcus furiosus*, has a 3' to 5' exonuclease activity. This means that is the wrong nucleotide is added – one that is not complementary to the template – the polymerase can remove the mis-introduced nucleotide. This proofreading activity increases the likelihood that the copies made during PCR do not contain unintended sequence changes.

Controls for PCR

In PCR it is common to set-up a "no DNA" negative control, which contains everything except DNA template. This controls for any DNA contamination in your samples or set-up. We recommend that when resources are sufficiently available you set-up this negative control.

There are several factors that affect the success of a PCR reaction, both in terms of how much product is produced (yield) and the purity (or quality) of the product: amount and type of template DNA, primer sequences and concentrations, the annealing temperature, the concentration of magnesium, the number of cycles.

For more on optimization of PCR: https://www.neb.com/tools-and-resources/usage-guidelines/guidelines-for-pcr-optimization-with-thermophilic-dna-polymerases

For troubleshooting:
https://www.neb.com/tools-and-resources/troubleshooting-guides/pcr-troubleshooting-guide

Taq polymerase

When using Taq polymerase we use a reagent mixture called GoTaq Green Master Mix 2X from Promega: http://tinyurl.com/je6prfj
The mixture contains the following:

Buffer (pH 8.5)
400µM dATP, 400µM dGTP, 400µM dCTP, 400µM dTTP
3mM MgCl2.
Taq DNA polymerase

The 2X refers to the fact that everything in the mix is 2X concentrated compared to the concentration in a typical PCR.

In a 50 ul PCR reaction with 25 ul of 2X mastermix, what is the final concentration of dNTPs and Magnesium?

The master mix also contains yellow and blue dyes (making the mastermix look green), which function as loading dyes when reaction products are analyzed by agarose gel electrophoresis. The blue dye migrates at the same rate as 3–5kb DNA fragments, and the yellow dye migrates at a rate faster than primers.

There are several factors that affect the success of a PCR reaction, both in terms of how much product is produced (yield) and the purity/quality of the product. Other factors include the amount of template DNA, the concentration of magnesium, the annealing temperature, the number of cycles, the amount of primer, and the heat-stable polymerase. See this page for more information: https://tinyurl.com/y2qd28x2

Magnesium
Magnesium is required by *Taq* as a co-factor and it reduces the repulsive forces between the DNA template strand and the primer. Note that the Mastermix contains 3 mM magnesium, so the final concentration in a 50 µl reaction is 1.5 mM. Magnesium can also help stabilize the primer-DNA template complex by stabilizing the negative charges of the DNA backbone.

Template DNA concentration
The amount of template that is optimal for PCR really depends on the composition and complexity of the DNA. For example, a small plasmid is less complex than genomic DNA as a template, and thus it may be necessary to add more genomic DNA to the PCR if using it as a template. Template + primer combinations also influence success of the PCR, so testing different amounts of template with different primers, and annealing temperatures is sometimes necessary.

Primers

The sequence of primers determine what DNA will be copied during PCR. It is important to consider how the primer sequences are both complementary to the target DNA, and at the same time avoiding primers being complementary to one another. It is also best to choose primers that have a similar melting, and thus annealing temperature.

Annealing temperature

The temperature at which your primer anneals to the template will vary depending on primer sequence and length. It is common to use an annealing temperature that is about 5ºC lower than the melting temperature of the primers (melting temperature is the temperature at which half of the duplex between primer and template will dissociate). A crude way to calculate annealing temperature (Ta) is: Ta = 0.3 x (Tm of primer) + 0.7 x (Tm of product) – 14.9; where Tm of primer is the melting temperature of the less stable primer-template pair, and Tm of product is the melting temperature of the PCR product.

An annealing temperature that is too low can result in primers annealing to non-target sequences, where they are not 100% complementary, leading to PCR products that are not the desired sequence. An annealing temperature that is too high can have the opposite effect: it can prevent primers from annealing to the template, reducing the desired PCR product.

Reaction set-up:

Component	**Volume for a 50 ul reaction**	**Volume for a 25 ul reaction**	**Final concentration**
GoTaq Green Master Mix, 2X	25 µl	12.5 µl	1X
Forward (upstream) primer, 10 uM	0.5-5.0 µl	0.25-2.5 µl	0.1-1.0 µM
Revers (downstream) primer, 10 uM	0.5-5.0 µl	0.25-2.5 µl	0.1-1.0 µM
DNA template	0.5-2.5 µl (1 colony = 1 µl)	0.5-2.5 µl (1 colony = 1 µl)	<250 ng
Nuclease-free water	to 50 µl	to 25 µl	N.A.

PCR cycling conditions with Taq polymerase:

Step	Temperature	Time	Number of Cycles
Initial Denaturation	95°C	2 minutes	1 cycle
Denaturation	95°C	0.5–1 minute	
Annealing	42–65°C*	0.5–1 minute	25–35 cycles
Extension	72°C	1min/kb	
Final Extension	72°C	5 minutes	1 cycle
Soak	4°C	Indefinite	1 cycle

*Annealing temperature should be optimized for each primer set based on the primer T_m.

Making copies of HDR templates

In a yeast CRISPR-Cas9 experiment where specific edits are being made using HDR, many, many copies of the double-stranded or single-stranded HDR template would be added to the yeast cells along with a plasmid expressing the gRNA and Cas9. In the case of our experiments we will be using a double-stranded sequence. *Where do we get double stranded copies of the specific HDR sequence we have designed?*

New HDR templates can be synthesized commercially: that is, the entire top-strand sequence and the entire bottom strand sequence are synthesized and then base-pairing between the two strands creates a double-stranded HDR. One company that offers synthesis of DNA molecules is IDT (https://www.idtdna.com/pages). Note that synthetically made DNA and RNA molecules are often referred to as "oligos" (oligonucleotides), where "oligo" means few/small, and nucleotide refers to the nucleic acids of DNA and RNA molecules.

If we were to have the HDR synthesized, we could then make many, many, many more copies of the template for use in subsequent yeast transformations. We would make these copies using PCR. The top and bottom strand of the HDR template can then be used as templates for PCR if we need to make many more copies of the HDR template.

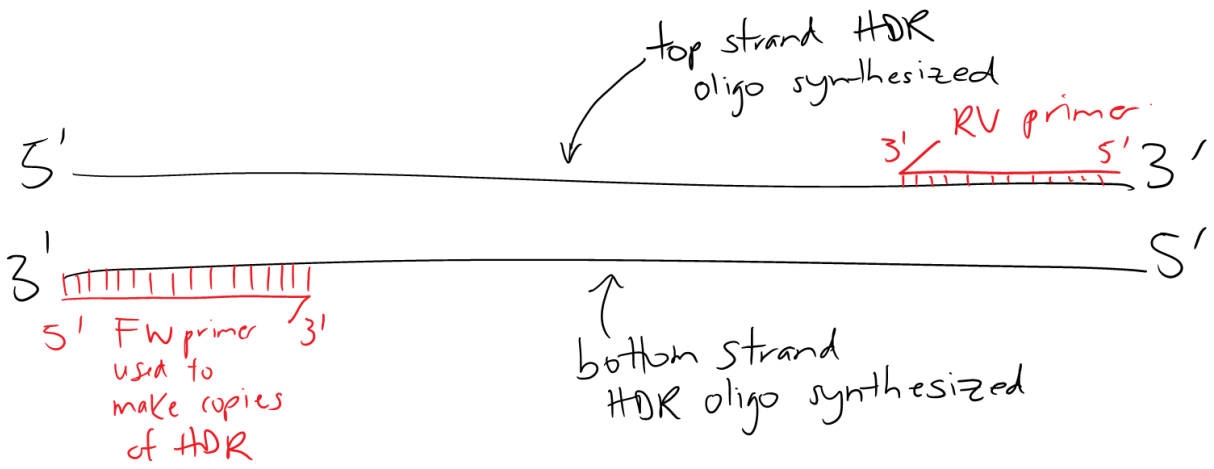


Figure: Making copies of an HDR sequence using PCR. In this example, we have a source of HDR that can be used as template for PCR.

However, if the HDR template is about 130 bp in length, this would require the synthesis of two 130 nt DNA fragments, which is quite costly.

A more economical approach is to synthesize two fragments that have 20 bp overlap, and can then be used to synthesize the full double stranded HDR template sequence and make many more copies using subsequent amplification cycles (similar to PCR, but no primers needed).

Essentially, the top fragment acts as a forward (FW) primer, and uses the bottom fragment as a template. Likewise, the bottom fragment acts as a reverse (RV) primer, using the top fragment as a template. In the case of a 130 bp HDR, each fragment is about 80 nucleotides long (see below).

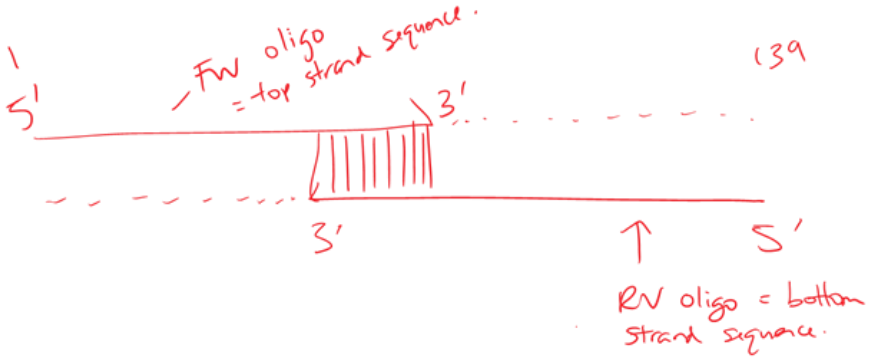


Figure: Overlapping fragments (oligos) that act as primers during PCR to synthesize and make copies of the complete double-stranded HDR sequence.

In this lab we use the overlapping oligos method.

Sanger DNA sequencing

There are several different methods currently used to sequence DNA. We are using a relatively older method to do our sequencing; it is the method that gives the longest reads and is still the most cost-effective for our purposes. This method is most commonly referred to as Sanger sequencing (for Fred Sanger, who developed it), but it is also referred to as the chain termination method. In this method, the DNA of interest is replicated, but fluorescently labeled dideoxynucleoside triphosphates are used in addition to non-labeled deoxynucleoside triphosphates as building blocks for the new DNA strands. Dideoxynucleoside triphosphates do not have the 3'OH necessary for formation of a phosphodiester bond. Thus, whenever a ddNTP is added to a growing DNA strand, no more nucleotides can be added to that strand.

In performing Sanger sequencing, the stretch of DNA being sequence is replicated by a DNA polymerase. Just like regular PCR, a primer needs to be part of the reaction so the polymerase can start synthesis. The primer is designed to be complementary to a region just upstream or downstream of the region being sequenced. In our case, we know the sequence of the plasmid and thus we can design a primer to the plasmid sequence just upstream of the promoter and gRNA sequence, and use that primer to start the replication process.

Reagents needed to perform Sanger sequencing:

- Template DNA
- Primer
- DNA polymerase
- dNTPs
- Fluorescently labeled dideoxy nucleotides (ddNTPs)—each ddNTP is labeled with a different dye, each of which emits light at a different wavelength.

All the above reagents are added to the same tube and the replication process is begun.

Remember that the binding of the primer and extension by DNA polymerase will be occurring simultaneously on millions of copies of the same plasmid with the same DNA insert. The new DNA strands will randomly incorporate the dNTPS and the ddNTPs, however, whenever a ddNTP is added, that new strand chain is terminated because there is no 3'OH. Thus, there is always a labeled ddNTP at the end of every new chain that is formed in the reaction, and there will be chains of all different lengths.

The mixture of all the different length products is then run through a column that separates the products by size, with the smallest pieces eluting from column first. Note that column separates products differing in size by just one nucleotide. As the products come off the column, light is used to excite the fluorescent dye linked to the ddNTPs, and the emission wavelength is detected. Computer algorithms ("base callers") interpret peak color to produce a DNA sequence.

For example, let's say the sequencing reaction generated the four shortest products shown below, which end in the dideoxynucleotide shown by the bold capital letter at the 3' end of the DNA fragment. Remember that each ddNTP is labeled with a different dye, so for example, a ddATP might be labeled with a green fluorescent dye, a ddTTP with a red dye, etc. As the products come off the column, the detector feeds the information about which color fluorescence is being emitted at that point, and the computer software interprets this into the DNA sequence.

5' T

5' T A

5' T A C

5' T A C G

These products would come off with the shortest first, which ends in a fluorescently labeled ddTTP. The computer software would conclude that the first nucleotide in the unknown sequence is a dTTP. The next product that would come off the column would end in a fluorescently labeled ddATP, and the software would conclude that the next nucleotide in the sequence is a dATP.

The following link provides an animation on Sanger sequencing:

http://www.dnalc.org/resources/animations/cycseq.html

PROTOCOLS

This section of the lab manual has a variety of protocols necessary to do the genetic editing in yeast using CRISPR-Cas9.

The protocol steps found in the manual may vary from how these same techniques are described in outside sources (e.g. if you are watching a video on gel electrophoresis, it may differ from the step-by-step instructions on how we do gel electrophoresis in BIMM 101). This this is not uncommon. The overall process is often the same between labs or informational sources, but the individual steps and small details (like volumes used) may vary.

If you are *learning* about the protocols it is important to understand the overall process, the purpose of different steps (how will modifications to different parts of the protocol affect outcomes?) what the expected outcomes are. By thinking about these things, you can apply your understanding to interpret outcomes, and try to make sense of unexpected results (troubleshooting).

The questions in grey boxes embedded within the protocols are meant for you to consider and answer as you are reviewing the protocol (e.g. before coming to lab, or during lab).

PROTOCOL 1: Explore ADE2 gene and choose a location to mutate

Goals:

- Use online bioinformatics tools to identify gene sequences and important features of a gene/protein sequence including promoters and conserved protein sequences.
- Use what is learned about a sequence to choose possible sites to mutate.
- Make predictions about how mutations in various sequences can change gene function.
- Produce an annotated ADE2 gene sequence to use in subsequent labs.

What you will need:

- ApE program, available for free online (http://jorgensen.biology.utah.edu/wayned/ape/) or some other sequence analysis program that you like to use

It is recommended that you complete the tasks below individually, but discuss them with your group members.
Results or answers to bolded tasks should be recorded in your lab notebook.

1. Got to the Yeast Genome Browser website: https://www.yeastgenome.org/

2. In the search bar, type ADE2 (this is the name of the gene we are studying)

a. On which chromosome is *ADE2* located?
b. How many chromosomes total are in a haploid *Saccharomyces cerevisiae*?

3. Under sequence, download the Genomic DNA +/- 1 kb sequence. Open this sequence in ApE.
   The +/- 1 kb means the sequence contains 1000 bp (1 kb) of sequence upstream and downstream to the start codon and stop codon, respectively. That is, bp 1001 should be the "A" of ATG.

For the next steps you will be using features in ApE. Consult Appendix B for some tips on using ApE.

4. The coding sequence of the gene is 1716 nucleotides in length. Use the ORF finder to identify the coding sequence and add it as a 'new feature' to your sequence (ORF = open reading frame, which is the same as "coding sequence").
   Go to menu, ORFs → Find next (click find next until you come upon the appropriately sized ORF)

5. Translate the ORF into an amino acid sequence: ORFs → translate. Select to see the codon spacing (Code), and the DNA sequence (DNA) as well as bp numbers (Line Width).

6. Use the following to identify possible promoter elements. Annotate possible promoter sequences in your ADE2 gene sequence file.
    - Go to http://rulai.cshl.edu/SCPD/
    - Select Genes, and ADE2 from the list that appears
    - **Make a screenshot of the result, and put it into your lab notebook. Below the screenshot, summarize what is being shown. Look up any necessary information to understand the results.**
        - References are usually provided; you can search for the referenced article to get more information.
        - Use online resources (http://stormo.wustl.edu/ScerTF) that can help you determine the consensus binding sites for any transcription factor (TF) in yeast.
        - The ScerTF site will lead you to sequence logo diagrams (http://weblogo.berkeley.edu/) that bring clarity to the bases that are conserved to bind TF proteins.
        - **Take a screenshot of the consensus sequence you find at the stormo site. Write a caption to summarize, in your own words, what this consensus sequence means/tells us about this region of the promoter of ADE2.**

    - **Annotate the ADE2 gene sequence by adding the identified promoter elements as features. Take a screenshot of this annotated version of ADE2 and paste it into your lab notebook.**

7. Use a BLASTP search to identify conserved sequences within the ADE2 gene:
    - Return to the Yeast Genome Browser page for ADE2
    - Go to 'analyze sequence' and under "S288C vs. other species" select the BLASTP vs. Fungi

        BLAST compares an input sequence to sequences in the database. BLASTP specifically compares a protein sequence (input) with other protein sequences. The results can show you where similarities and differences exist between the input sequence – ADE2 in this case – and other known sequences.
    - Choose several different genomes to compare our ADE2 sequence to; at least two that are not closely related.
    - Run NCBI-BLAST and look at the sequence comparisons.

        a. Is it possible to predict the location of functional domains based on the conserved sequences between ADE2 in different species of fungi?

    - **Take a screenshot one ADE2 gene sequence comparison that helps you identify conserved sequences. Put this screen shot into your notebook and document the boundaries of each conserved domain using the amino acid numbers of our ADE2 protein. Below, explain how you determined what parts were conserved sequences.**

8. Use the Pfam database to identify specific domains within the ADE2 protein sequence:
    - Go to https://pfam.xfam.org/
    - **Click on 'Sequence Search'** quick link, enter the ADE2 amino acid sequence and click Go.

      a. Could you have used the DNA sequence of the ORF to use for this search? Try. Why might it take longer for the search to finish compared to when using the amino acid sequence?

    - **Take a screenshot of Pfam result and put it into your lab notebook. Create a caption under the screenshot to describe the domains identified (names, what they do).**
    - **Which nucleotides of the ADE2 gene sequence, and which amino acid numbers, correspond to these domains?**

    - Annotate your ADE2 gene sequence file to indicate where the domains are located within the ORF.

      b. Do these domains correspond to the conserved sequences in ADE2 that you identified above (Q8)?

    - **Take a screenshot of the ADE2 gene sequence so the annotated promoter elements, ORF and protein domains are visible. Put this screenshot into your lab notebook.**

If you wanted to test whether or not these promoter and domain sequences were indeed important for expression and function of ADE2, respectively, what kinds of mutations would you introduce? What do you predict would be the outcome of these mutations (what would you measure)?

PROTOCOL 2: Designing gRNA and HDR templates

Goals:

- Be able to design and select a gRNA and associated HDR template to use in a CRISPR-Cas9 experiment to edit the ADE2 gene in yeast
- Design the gRNA such that it can be added to the pML104 plasmid using restriction enzyme digestion and ligation
- Explain how your designed gRNA and HDR will result in the desired edits

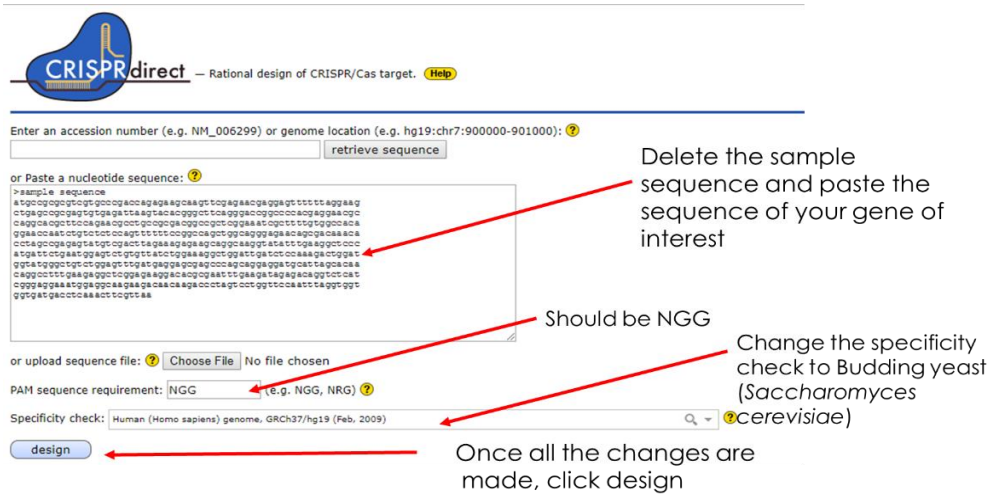
What you will need:

- ApE program, available for free online (http://jorgensen.biology.utah.edu/wayned/ape/) or some other sequence analysis program that you like to use
- The ADE2 gene sequence

It is recommended that you complete the tasks below individually, but discuss them with your group members.
Results or answers to bolded tasks should be recorded in your lab notebook.

Protocol 2a– designing gRNA

1. Go to this website: http://crispr.dbcls.jp/
2. Delete the sample sequence.
3. Select the portion of the gene sequence where you want the gRNA target and paste it into the search window. For example, if you want to target Cas9 to the middle of the gene, select 200 or so nucleotides from the middle of the gene sequence and paste it into the browser.
4. The PAM sequence should be NGG.
5. Change the specificity check to Budding yeast (*Saccharomyces cerevisiae)*
6. Once all the changes are made, click “design”.



7. Examine the results. **See the next page for details on understanding the results.**

> A. Put a screenshot of the results table in your lab notebook.
> B. Also type out which gRNA you are choosing (5' to 3' labels), and indicate the PAM beside the gRNA sequence.
> C. Explain why you chose this particular gRNA.

8. After choosing your gRNA, annotate the ADE2 gene sequence in ApE to indicate he location of the gRNA and PAM. You should have something like this:

Feature	Direc	Type	Location ↓
Start codon	>>>	misc_feature	551..553
gRNA1	>>>	misc_feature	589..608
PAM	>>>	misc_feature	609..611

```
      *   10    *   20    *   30    *   40    *   50    *   60    *   70    *   80    *   90    *
381 GGGTAATACTAAGTGATTGACTCTTGCTGACCTTTTATTAAGAACTAAATGGACAATATTATGGAGCATTTCATGTATAAATTGGTGCGTAAAAT
476 CGTTGGATCTCTCTTCTAAGTACATCCTACTATAACAATCAAGAAAAACAAGAAAATCGGACAAAACAATCAAGTATGGATTCTAGAACAGTTGG
571 TATATTAGGAGGGGGACAATTGGGACGTATGATTGTTGAGGCAGCAAACAGGCTCAACATTAAGACGGTAATACTAGATGCTGAAAATTCTCCTG
666 CCAAACAAATAAGCAACTCCAATGACCACGTTAATGGCTCCTTTTCCAATCCTCTTGATATCGAAAAACTAGCTGAAAAATGTGATGTGCTAACG
761 ATTGAGATTGAGCATGTTGATGTTCCTACACTAAAGAATCTTCAAGTAAAACATCCCAAATTAAAAATTTACCCTTCTCCAGAAACAATCAGATT
```

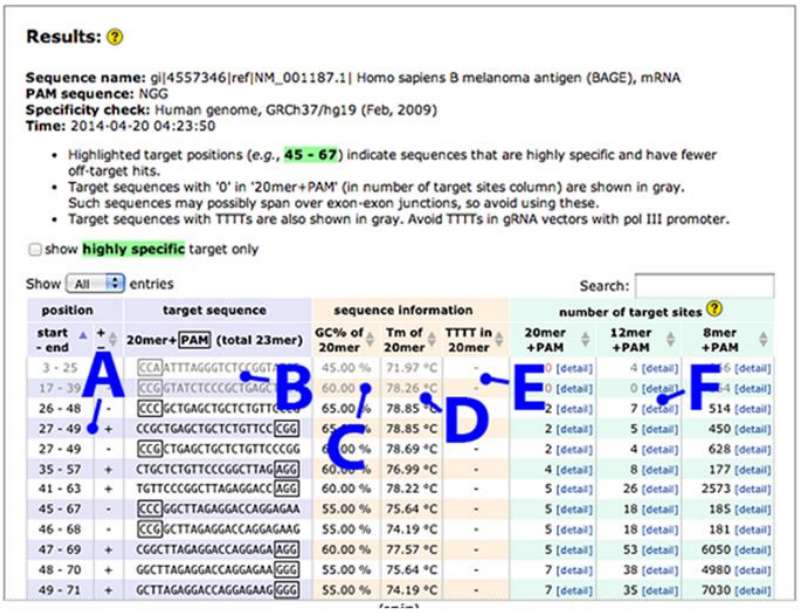
> A. Include a screen shot of the annotated region of the gene in your lab notebook
> B. Write/type out a stretch of the double stranded ADE2 gene sequence with 5' and 3' labels. Write the sequence so the strands are separated.
> C. Indicate the PAM sequence
> D. Write out the gRNA sequence (as an RNA molecule) and show it base-pairing to the genomic template. Make sure your gRNA also has 5' and 3' labels.
> E. Indicate where the double-stranded cut by Cas9 will occur in the genomic template
>
> *(your drawing should be similar to the sequence base-pairing shown on the bottom of page 21.)*
> *(if you do a drawing on paper, take a picture and upload it into your lab notebook)*

9. Apply the principals of annealed oligos and restriction enzyme digestion (previously in the lab manual) to design oligos to add your gRNA to pML104 plasmid.

Write out the double-stranded gRNA sequence in your lab notebook, highlighting the BclI overhang and the 5' scaffold sequence (see page 36 of lab manual)

10. Proceed to designing an appropriate HDR template.

Understanding the Results!



Results:

Sequence name: gi|4557346|ref|NM_001187.1| Homo sapiens B melanoma antigen (BAGE), mRNA
PAM sequence: NGG
Specificity check: Human genome, GRCh37/hg19 (Feb, 2009)
Time: 2014-04-20 04:23:50

- Highlighted target positions (*e.g.*, **45 - 67**) indicate sequences that are highly specific and have fewer off-target hits.
- Target sequences with '0' in '20mer+PAM' (in number of target sites column) are shown in gray. Such sequences may possibly span over exon-exon junctions, so avoid using these.
- Target sequences with TTTTs are also shown in gray. Avoid TTTTs in gRNA vectors with pol III promoter.

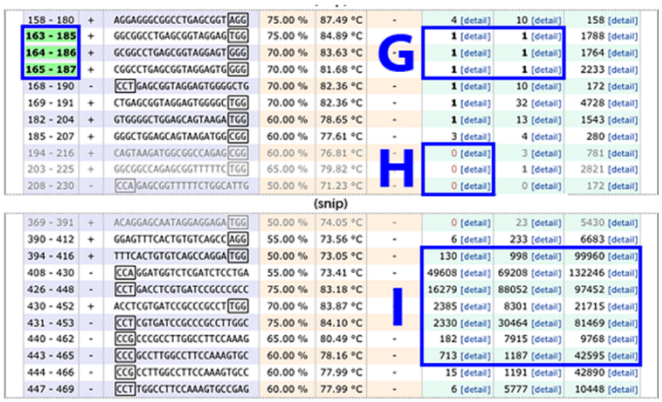
show **highly specific** target only

Show All entries Search:

position		target sequence	sequence information			number of target sites		
start - end	+ -	20mer+PAM (total 23mer)	GC% of 20mer	Tm of 20mer	TTTT in 20mer	20mer +PAM	12mer +PAM	8mer +PAM
3 - 25		CCAATTTAGGGTCTCCGGTA	45.00 %	71.97 °C	-	0 [detail]	4 [detail]	46 [detail]
17 - 39		CCGGTATCTCCCGCTGAGC	60.00 %	78.26 °C	-	0 [detail]	0 [detail]	4 [detail]
26 - 48	-	CCCGCTGAGCTGCTCTGTTC	65.00 %	78.85 °C	-	2 [detail]	7 [detail]	514 [detail]
27 - 49	+	CCGCTGAGCTGCTCTGTTCCCGG	65.00 %	78.85 °C	-	2 [detail]	5 [detail]	450 [detail]
27 - 49	-	CCGCTGAGCTGCTCTGTTCCCGG	60.00 %	78.69 °C	-	2 [detail]	4 [detail]	628 [detail]
35 - 57	+	CTGCTCTGTTCCCGGCTTAGAGG	60.00 %	76.99 °C	-	4 [detail]	8 [detail]	177 [detail]
41 - 63	+	TGTTCCCGGCTTAGAGGACCAGG	60.00 %	78.22 °C	-	5 [detail]	26 [detail]	2573 [detail]
45 - 67	-	CCCGGCTTAGAGGACCAGGAGAA	55.00 %	75.64 °C	-	5 [detail]	18 [detail]	185 [detail]
46 - 68	-	CCGGCTTAGAGGACCAGGAGAAG	55.00 %	74.19 °C	-	5 [detail]	18 [detail]	181 [detail]
47 - 69	+	CGGCTTAGAGGACCAGGAGAAGG	60.00 %	77.57 °C	-	5 [detail]	53 [detail]	6050 [detail]
48 - 70	+	GGCTTAGAGGACCAGGAGAAGGG	55.00 %	75.64 °C	-	7 [detail]	38 [detail]	4980 [detail]
49 - 71	+	GCTTAGAGGACCAGGAGAAGGGG	55.00 %	74.19 °C	-	7 [detail]	35 [detail]	7030 [detail]

A B C D E F

- **A**: Target Position
- **B**: Target Sequences, 20mer+3mer PAM (total 23 mer)
- **C**: GC content of the target 20mer.
- **D**: Calculated Tm of the target 20mer.
- **E**: Presence or absence of TTTT (four consecutive T's that cause pol III termination) in the target 20mer. Avoid TTTT in gRNA vectors with pol III promoter.
- **F**: Off-target* search results against genomic sequence. The number of target sites with perfect match is shown. The number displayed here includes both on-target and off-target sites. Smaller number (but not zero) is better for these columns to avoid off-target editing.



- **Off-target**: A different gene that contains a similar 23 base pair sequence to the gene we want to edit. Then because of the similar sequence, it might be also recognized by the guide RNA, therefore producing "off-target editing. It is important to use a sequence that occurs only in the gene we want to edit, and not elsewhere in the genome, to avoid off-target editing.

- **G**: Only one match in 20mer+PAM and 12mer+PAM search.

These targets are highly specific.
→ Recommended for CRISPR/Cas target. Target positions are highlighted with green (e.g., 163 - 185).

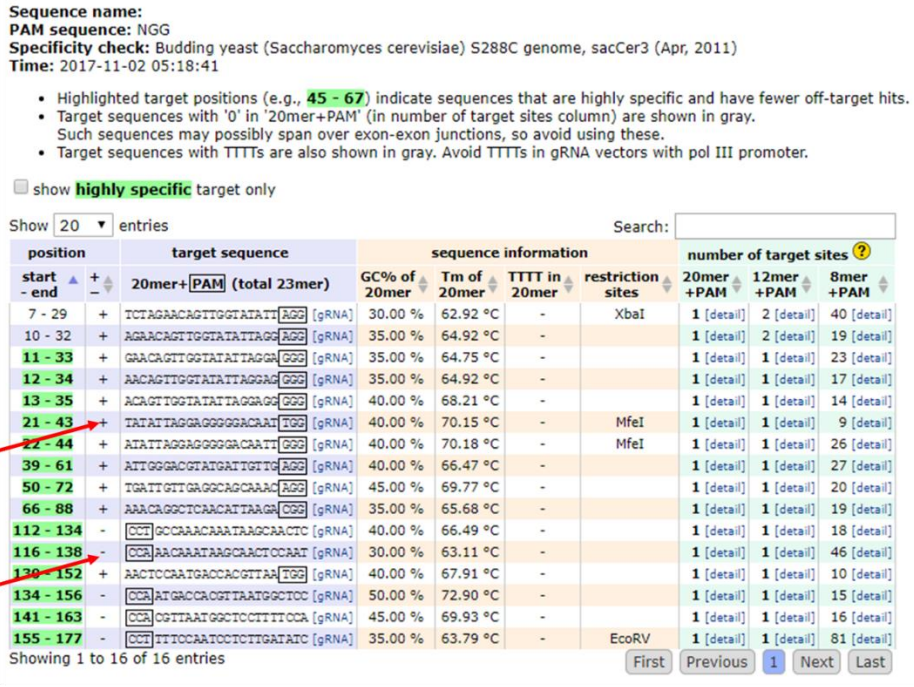
- **H**: No match in 20mer+PAM search. Possibly the sequence spans over exon-exon junction, so avoid using these.

→ Not recommended for CRISPR/Cas target.

- **I**: Very high number of off-target hits. Avoid using these sequences.

→ Not recommended for CRISPR/Cas target.

Any target sequence that is not followed by this three base sequence (NGG, where N = any base, and G = guanine bases) will not be recognized by the CRISPR/guide RNA complex!!!



Sequence name:
PAM sequence: NGG
Specificity check: Budding yeast (Saccharomyces cerevisiae) S288C genome, sacCer3 (Apr, 2011)
Time: 2017-11-02 05:18:41

- Highlighted target positions (e.g., 45 - 67) indicate sequences that are highly specific and have fewer off-target hits.
- Target sequences with '0' in '20mer+PAM' (in number of target sites column) are shown in gray. Such sequences may possibly span over exon-exon junctions, so avoid using these.
- Target sequences with TTTTs are also shown in gray. Avoid TTTTs in gRNA vectors with pol III promoter.

show highly specific target only

Show 20 entries Search:

position		target sequence	sequence information				number of target sites		
start - end	+ −	20mer+PAM (total 23mer)	GC% of 20mer	Tm of 20mer	TTTT in 20mer	restriction sites	20mer +PAM	12mer +PAM	8mer +PAM
7 - 29	+	TCTAGAACAGTTGGTATATTAGG [gRNA]	30.00 %	62.92 °C	-	XbaI	1 [detail]	2 [detail]	40 [detail]
10 - 32	+	AGAACAGTTGGTATATTAGGAGG [gRNA]	35.00 %	64.92 °C	-		1 [detail]	2 [detail]	19 [detail]
11 - 33	+	GAACAGTTGGTATATTAGGAGG [gRNA]	35.00 %	64.75 °C	-		1 [detail]	1 [detail]	23 [detail]
12 - 34	+	AACAGTTGGTATATTAGGAGG [gRNA]	35.00 %	64.92 °C	-		1 [detail]	1 [detail]	17 [detail]
13 - 35	+	ACAGTTGGTATATTAGGAGG [gRNA]	40.00 %	68.21 °C	-		1 [detail]	1 [detail]	14 [detail]
21 - 43	+	TATATTAGGAGGGGGACAAT [gRNA]	40.00 %	70.15 °C	-	MfeI	1 [detail]	1 [detail]	9 [detail]
22 - 44	+	ATATTAGGAGGGGGACAATT [gRNA]	40.00 %	70.18 °C	-	MfeI	1 [detail]	1 [detail]	26 [detail]
39 - 61	+	ATTGGGACGTATGATTGTTG [gRNA]	40.00 %	66.47 °C	-		1 [detail]	1 [detail]	27 [detail]
50 - 72	+	TGATTGTTGAGGCAGCAAAC [gRNA]	45.00 %	69.77 °C	-		1 [detail]	1 [detail]	20 [detail]
66 - 88	+	AAACAGGCTCAACATTAAGA [gRNA]	35.00 %	65.68 °C	-		1 [detail]	1 [detail]	19 [detail]
112 - 134	-	CCT GCCAAACAAATAAGCAACTC [gRNA]	40.00 %	66.49 °C	-		1 [detail]	1 [detail]	18 [detail]
116 - 138	-	CCA AACAAATAAGCAACTCCAAT [gRNA]	30.00 %	63.11 °C	-		1 [detail]	1 [detail]	46 [detail]
130 - 152	+	AACTCCAATGACCACGTTAA [gRNA]	40.00 %	67.91 °C	-		1 [detail]	1 [detail]	10 [detail]
134 - 156	-	CCA ATGACCACGTTAATGGCTCC [gRNA]	50.00 %	72.90 °C	-		1 [detail]	1 [detail]	15 [detail]
141 - 163	-	CCA CGTTAATGGCTCCTTTTCCA [gRNA]	45.00 %	69.93 °C	-		1 [detail]	1 [detail]	16 [detail]
155 - 177	-	CCT TTTCCAATCCTCTTGATATC [gRNA]	35.00 %	63.79 °C	-	EcoRV	1 [detail]	1 [detail]	81 [detail]

Showing 1 to 16 of 16 entries First Previous 1 Next Last

Forward Strand, sequence ends in TGG (PAM sequence)

Reverse Strand, sequence begins with CCA (complimentary to GGT PAM sequence)

Protocol 2b- Designing the HDR template

Things to consider:

- The HDR template should extend about 50-55 bp on either side of the cut site. This provides long enough homology arms to allow for HDR to occur.
- You will need to decide what types of mutations you want to introduce into the HDR in order to change gene function.
- Typically, mutations are introduced near the cut site (but it is OK to have mutations at locations extending 5' or 3' beyond the cute site)
- Will you include any mutations to alter the PAM sequence? Will these also affect the protein sequence (if coding) or expression (if promoter region)?
- If you plan to introduce mutations into the coding sequence it is necessary to ensure your HDR sequence is in-frame (see below on how to do this).

Designing mutations in the coding sequence: Follow steps 1-6 below

Designing mutations in the promoter or non-coding sequence: Follow steps 3-6

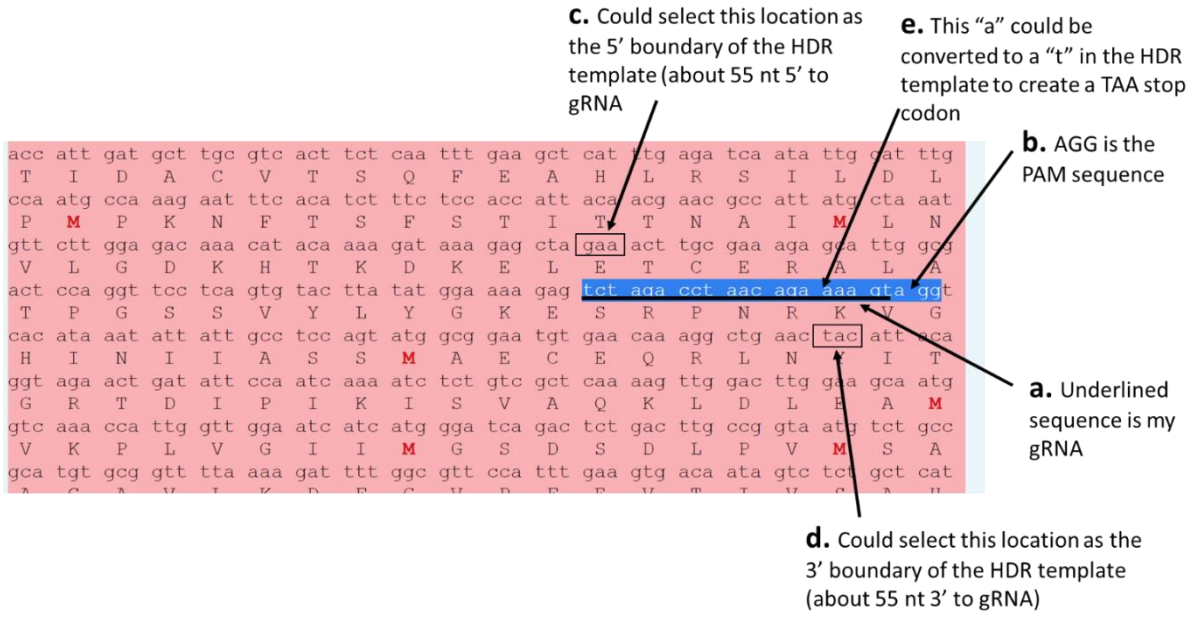
Designing an HDR template sequence

1. Use a translate tool to get the amino acid sequence aligned with the corresponding nucleotide coding sequence.
   - **Highlight the entire ORF within the ADE2 gene sequence in your ApE file**.
   - Go to ORFs → Translate
     - Select to see the DNA sequence (above or below the amino acid sequence). You can also play around with other viewing parameters such as if you want codon spacing (a space between each codon) and if you want to sequence numbers (line numbers). **Select "Copy highlight"** so you can see where your gRNA will anneal. You should get something like this:

```
551 ATG GAT TCT AGA ACA GTT GGT ATA TTA GGA GGG GGA CAA TTG GGA CGT ATG ATT GTT GAG
  1  M   D   S   R   T   V   G   I   L   G   G   G   Q   L   G   R   M   I   V   E

611 GCA GCA AAC AGG CTC AAC ATT AAG ACG GTA ATA CTA GAT GCT GAA AAT TCT CCT GCC AAA
 21  A   A   N   R   L   N   I   K   T   V   I   L   D   A   E   N   S   P   A   K

671 CAA ATA AGC AAC TCC AAT GAC CAC GTT AAT GGC TCC TTT TCC AAT CCT CTT GAT ATC GAA
 41  Q   I   S   N   S   N   D   H   V   N   G   S   F   S   N   P   L   D   I   E

731 AAA CTA GCT GAA AAA TGT GAT GTG CTA ACG ATT GAG ATT GAG CAT GTT GAT GTT CCT ACA
 61  K   L   A   E   K   C   D   V   L   T   I   E   I   E   H   V   D   V   P   T

791 CTA AAG AAT CTT CAA GTA AAA CAT CCC AAA TTA AAA ATT TAC CCT TCT CCA GAA ACA ATC
 81  L   K   N   L   Q   V   K   H   P   K   L   K   I   Y   P   S   P   E   T   I
```

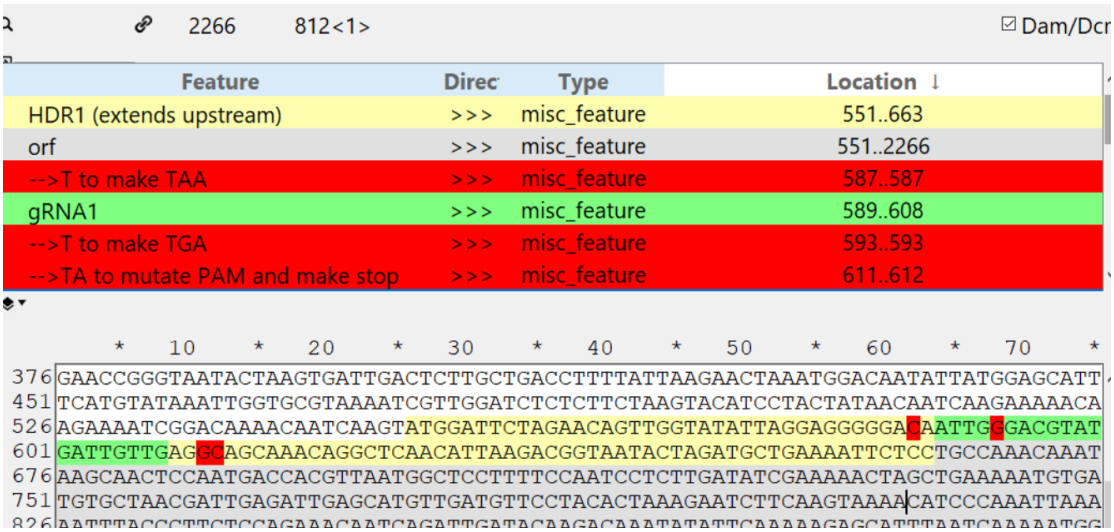
2. Find features so you can determine where to introduce sequence variations in the HDR. Find: **(a)** your gRNA sequence, **(b)** PAM sequence, **(c)** Extend about 50-55 bp upstream and **(d)** downstream from the gRNA sequence such that the HDR sequence is in-frame. Isolate this in-frame HDR sequence from your annotated ADE2 gene sequence. Copy the HDR sequence and paste it into a new window, and save as HDR template sequence for **your** gRNA.

   (e) Consider what mutations you want to introduce into the HDR template.



3. Go back to the *ADE2* gene sequence and annotate where the HDR template sequence spans. If you are designing a mutation in a non-coding sequence, select approximately 130 bp, with the predicted Cas9 cut site near the middle, to be the HDR template region.

4. Annotate the HDR template sequence to indicate **the nucleotide changes you will introduce in the sequence**. You should have something like the figure below:



> A. Take a screen-shot of the annotated HDR sequence and insert it into your lab notebook.
> B. If your mutation is in the coding sequence, take a snap shot of the translated sequence, showing DNA with amino acid, to show frame and mutations together and put in your lab notebook.
> C. Below the screenshot(s), describe the genetic changes that will occur if the HDR is used during repair, after a Cas9 cut.
> D. Explain the effect these changes are predicted to have on gene function, and why.

5. Make a copy of just the HDR template sequence, and put it into another ApE file. Make the nucleotide changes – this is now the sequence of the HDR template that would be synthesized to do the experiment.
6. Make a double-stranded sequence that would be the full HDR synthesized for use in a yeast experiment:
    a. Go to https://www.bioline.com/media/calculator/01_12.html
    b. Paste in your top strand HDR sequence
    c. From the menu at the bottom, select “display double stranded sequence” (make sure all other options are *not* selected)

> Make a screenshot of the double stranded sequence and put in your lab notebook

7. Confirm if your instructor/IA needs you to input your HDR sequences in another database for ordering.

PROTOCOL 3: Growing Liquid Bacterial & Yeast Cultures

Sometimes we need to grow our cultures (e.g. bacteria or yeast) in liquid media, for example to extract plasmids from the bacteria. Be sure to use proper sterile technique to set-up a clean, uncontaminated culture. Depending on the culture, you may need to add a selective agent (e.g. antibiotic).

In the case of our experiment, we would need to grow up a culture that carries the pML104 plasmid. We would extract the plasmid from the bacteria and this would provide a large amount of pML104 to use in subsequent work (e.g. restriction enzyme digestion and ligation)

Below we describe the protocol to set-up liquid cultures.

Aseptic technique:
https://www.youtube.com/watch?v=wttvhJU9PZM&feature=youtu.be or
https://www.jove.com/science-education/10040/aseptic-technique-in-environmental-science
(note you have to be connected through UCSD to access JOVE)

Inoculating a liquid culture:
https://www.youtube.com/watch?v=5XbLrXtWNOk&feature=youtu.be

1. Take a glass tubes containing 10 ml of appropriate media (LB for *E. coli,* YPD for yeast). Label the side of the tubes clearly with what you are inoculating. For example, if you are inoculating the liquid culture with *E. coli* containing a plasmid, the label could read: "*E.coli* w/pML-gRNA1 plasmid". Ensure your names/group number are also on the tube.

2. **If using ampicillin to select for E. coli carrying a specific plasmid**: you need to have a final concentration of 50 μg/ml ampicillin. The stock solution is typically 50 mg/ml.

   How much will you add to achieve your final desired concentration?

   If you are setting up a liquid culture of yeast:
   If the yeast are being selected by the presence of the URA3 gene, the yeast should be grown in uracil-minus media.

3. Find the stock plates with the needed culture. These plates would have been set-up so that the culture could grow for at least 24 hours prior to using for setting up liquid culture. They may look something like this:

4. Sterilize your inoculating loop in the flame until it is bright orange. Allow the loop to cool for at least 10 seconds.

5. When you are ready to add the sample, open the tube and pass the top of the tube through the flame.

6. Touch your inoculating loop to a colony from the appropriate plate and transfer it to the corresponding tube containing liquid media. Rotate the loop to dislodge the colony into the liquid culture.

7. Flame the loop and continue the process for the next tube, if setting up multiple cultures.

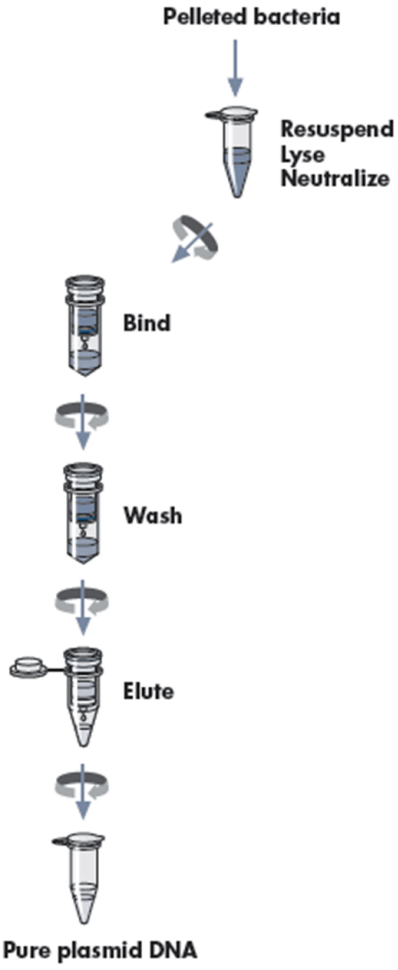
8. Place in a shaking incubator at 37^{o}C (*E. coli*) or 30^{o}C (yeast).

PROTOCOL 4: QIAprep Alkaline Lysis Plasmid Purification

To extract plasmid DNA from bacterial cells we are using a commercial kit from Qiagen, the "QIAprep Miniprep". The Qiaprep Miniprep Handbook can be downloaded here:
https://tinyurl.com/yc5cb986
Or you can find it by searching online for "Qiaprep Miniprep handbook"

This kit utilizes a common technique in which bacteria are lysed and DNA is denatured simultaneously in an alkaline SDS solution at pH 12 (buffer P2). The solution is then neutralized in an acidic, high salt buffer (buffer N3) at which point the plasmid DNA re-natures. The solution is centrifuged, and while the plasmid remains in the supernatant, the chromosomal DNA is trapped in a precipitate formed by SDS in the presence of potassium, and is pelleted.

See Alkaline Lysis Plasmid Purification in the Background section for more details.



Right: QIAprep Spin Procedure. From the QiaPrep Miniprep Handbook

What solutions are used at each step?

PROTOCOL:

1. Re-suspend overnight cultures by briefly vortexing – the cells will have settled to the bottom of the tube and you need to re-suspend them before moving onto the next step**. Also check to make sure that your culture is turbid (cloudy)! You don't want to use a culture that did not grow!**

   If your culture was pelleted and frozen, skip step 1-3.

2. Transfer 10 ml of each overnight culture into a labeled 15 ml blue-capped plastic conical tube. Spin at 6000 rpm in the clinical centrifuges for 5 min.

   Coordinate with at least one other group so we don't have long waits for using the centrifuge.
   If you aren't sure how to use these centrifuges, as your IA for assistance.

3. Remove and discard supernatants into a beaker with bleach in it (you can carefully pour off the supernatant and the pellet will remain in the tube). Remove any excess liquid with a pipette, without disturbing the pellets.

4. Add **500 µl** of buffer **P1** (on ice; contains RNase) to each cell pellet. Resuspend pellets by gentle vortexing and/or by pipetting. **Use a different pipette tip for each pellet.** Ensure they are completely resuspended.

   Why is it recommend that you use a different pipette tip to resuspend each pellet?

5. Transfer each resuspended pellet to a labeled 2 ml microfuge tube.

6. Lyse the bacteria in each tube by adding **500 µl** of buffer **P2** to each tube, and invert 10 times. ***Do not vortex.*** Solution will clear when lysis is complete (should take no more than a few minutes).

   P2 buffer contains NaOH/SDS. What is the purpose of these ingredients?

7. Add **700 µl** of Buffer **N3 to each tube** and invert 8-10 times, immediately but gently.

   N3 buffer contains acetic acid, guanidinium, and high K+ conc. What is the purpose of these ingredients?

   You should see a big precipitate here and the solution should turn transparent.

8. Centrifuge to remove cellular debris—10 minutes at maximum speed in the minicentrifuges. After spinning **do not** discard the supernatants.

9. Transfer each of the supernatants into a clean, labelled microfuge tube.

There will be a pellet of cellular debris left in the original microfuge tube. The cellular debris contains guanidine hydrochloride, which creates chlorine gas when mixed with bleach, so do not pour the cellular debris into bleach. Consult your IA/Instructor about where to dispose of this waste.

10. Label the tops of the Qiagen columns, one for each culture. **Label the columns themselves not the collecting tubes.** Each column should be placed into a 2 ml collection tube.

11. Pipette **700 µl** of the **supernatant** to the inside of the labeled Qiagen columns (not the collection tubes). You will have supernatant remaining in the 1.5 ml microfuge tube – do not discard this, you will use it!

12. Centrifuge the column in the collection tube for 60 seconds in a mini centrifuge. Discard the liquid that comes through the column into the collection tube (the "flow-through").

**Do not discard the liquid into the bleach, for the same reasons noted above.*

13. Pipette the remaining supernatant into the appropriate column. Centrifuge the column in the collection tube for 60 seconds in a mini centrifuge. Discard the flow-through **into the "Mini Prep Waste" container in the fume hood**.

**Do not discard the liquid into the bleach, for the same reasons noted above.*

Where is the plasmid in this step of the protocol? What is in the flow-through?

14. Wash each column by adding **750 µl** of buffer **PE** into the column and centrifuging for 60 seconds at maximum speed.

Buffer PE contains ethanol. What is the purpose of the ethanol?

15. Discard flow through into the "Ethanol Waste" container in the **fume hood**, and re-centrifuge empty columns for 2 minutes at high speed.

    This dries the column and removes any residual wash buffer.

16. Place your columns into clean, labeled 1.5 ml microcentrifuge tubes. Be sure to label the 1.5 ml tubes.

17. Add **50 µl** of **warm EB** buffer to the middle of the membrane in the column, let sit for **2** minutes, and centrifuge for 1 minute.

EB is low salt buffer; elutes DNA from column in this buffer. Warming the buffer helps increase yield.

18. Determine the concentration and purity of your plasmid extractions using the Nanodrop (see below for more information on spectrophotometric analysis of DNA and RNA). Record the concentration, and A260/A280 value. Ask your Instructional Assistant to demonstrate how to use the Nanodrop (there is no protocol for it, you can take note of how it is used when you measure your sample).

    These are high copy number plasmids, so your concentrations will probably be between 200 and 300 ng/ µl (low copy number might be closer to 50 ng/ul). If your concentration is much lower than 200 ng/µl talk to your instructor about what to do next.

    What does “high copy number” mean?
    What genetic feature on the plasmid would make it a low versus high copy number?

19. Check your plasmid extractions using agarose gel electrophoresis. This can be done during this lab or the next lab (check with your IA/Instructor).

    When checking plamsids on a gel, you ideally do not want to load more than 100-200 ng of plasmid, so sometimes it is a good idea to check the concentration first using the nanodrop (or load a dilution alongside undiluted).

20. Make sure your plasmid DNA samples are clearly labelled (including group number). Your IA should have a freezer storage box set up to store the samples.

PROTOCOL 5: Measuring DNA mount and purity using a nanodrop

Before you use DNA and RNA samples for other reactions or experiments it is often necessary to determine the concentration and purity of your DNA or RNA sample. You will use a UV-spectrophotometer (nanodrop) to **quantify** and assess the **purity** of the DNA or RNA samples that you isolated.

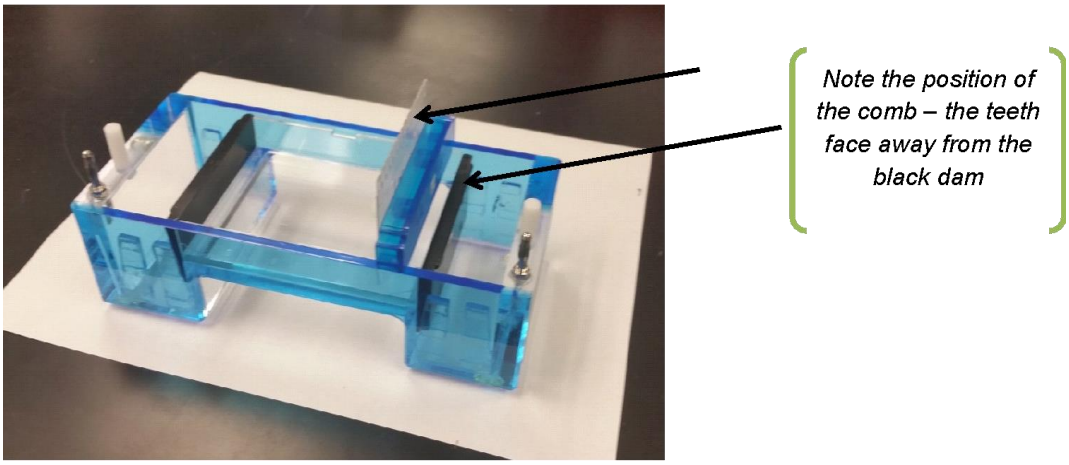
For more information see Spectrophotometric Analysis of DNA & RNA in the Background section

1. Ensure the nanodrop is set to record for the correct type of sample (e.g. dsDNA)
2. Gently wipe the upper and lower inside portion of the pedestal with water and a kimwipe.
3. Pipette 2 μl of blank sample onto the lower measurement pedestal (e.g. water or buffer, depending on your sample), and lower the pedestal. Blank the machine. A “blank check” may be required, in this case wipe the pedestal, pipette more blank, and re-measure.
4. Wipe the pedestals. Pipette 2 ul of your sample, ensuring there are no air bubbles in the 2 ul drop. Lower the pedestal, measure and record values (concentration, A260/A280).
5. Wipe the pedestals clean after use using a Kimwipe.

PROTOCOL 6: Making and Running an Agarose Gel

Refer to Agarose Gel Electrophoresis for analysis of DNA samples Background Information.

1. Set up the gel apparatus with one of the combs. Make sure:
    - The teeth of the comb are about 1 cm in from the top of the gel.
    - The bottoms of the teeth are not touching the chamber.
    - Select a comb with the appropriate number of wells for your samples.



Agarose gel electrophoresis set-up in lab.

We have two well sizes that we use in this lab: small and large. The well sizes are dictated by the width of the teeth of the comb used to make the wells.

For the wider wells, we typically load about 10-12 ul of sample. For the smaller wells we typically load between 5 ul-8 ul of sample.

Find the two different combs in your lab locker. How many wells will each comb make if used when casting a gel?

The agarose gel is made by mixing powdered agarose in TAE buffer, and melting the agarose in the buffer. Typically, we use a 0.8-1.0% (grams/100 ml) gel.

How would you make 50 mL of 0.8% agarose? In other words, how many g (or mg) of agarose would you mix and melt into 50 ml of TAE?

2. There may be pre-melted agarose in a warm water bath to keep it melted. Confirm that Sybr Safe has been added to the gel by the IA. If not, add it yourself (5 μl to 50 mL of melted agarose.

3. Pour the molten agarose into the chamber before it begins to cool and solidify. If it is your first time making a gel in BIMM 101, have your IA or instructor check your apparatus set-up before pouring the gel.

4. Wait for the gel to completely solidify (about 20 minutes). It should appear opaque throughout. Have your IA check the first time. **Do not pull out the comb until the gel is completely solidified!**

5. Once solidified, remove the comb and dams, and then flood gel with TAE buffer (about 250 ml) so that gel is entirely covered.

6. Load ladder (e.g. λ*Hind*III) and DNA samples mixed with 6x sample buffer* into the wells.

   When loading the gel, it is not necessary to poke the pipette tip down into the well. Instead, place the tip just above the well, and slowly eject the sample so it sinks into the well.

   <u>*6x Sample buffer:</u>
   Sample buffer is a mixture of dye and glycerol. It is typically added to DNA samples being run on a gel to make the DNA sample heavier, so it will sink into the well. The buffer contains a low molecular weight dye, bromophenol blue, that will help you judge when the gel is done. This dye does not bind to DNA and is not used to visualize the DNA —the Sybr Safe does that.
   We usually recommend adding about 2-3 μl of 6x sample buffer to about 10 μl of sample.

   <u>DNA samples</u>
   If you do not know the approximate concentration of your DNA sample(s), you may want to run some dilutions of the sample(s). If the sample is very concentrated and you load a large amount of DNA fragments (e.g. 200+ ng) into a well, the band will be very saturated with Sybr Safe, and hard to estimate both an accurate size and amount.

7. Place the lid on the gel box, then plug the leads into the power supply. Turn on the power supply and set for 150 volts. Press “run”.

   If another group at your bench is not ready when you are, you can start your gel. Then when they are ready, they can just turn off the power, plug in their leads and re-start the power supply.

8. Once the gel has completed running you can visualize the bands using the blue light illuminators. Check with your IA if you aren’t sure how to use them. Take a picture of the gel for your records and analysis.

PROTOCOL 7: Restriction Enzyme Digestion of pML104

Refer to Restriction enzyme cloning to add gRNAs to pML104 plasmid in the Background section.

Assume you have a plasmid sample that is 120 ng/ul (in reality we would have used our analysis of the plasmid from the gel, and or measuring concentration by nanodrop to know our starting plasmid sample concentration).

The following digestions should be set-up: 1) SwaI only, 2) Double digest (*Swa*I and *Bcl*I), 3) No digestion

What is the purpose of the single digests and the no digestion controls?
What if you also set-up, in a separate tube, a BclI only digest? Would that be useful? Why or why not?

SwaI only and double digest:

1. Set up reaction as follows:

   For a 50 ul reaction

DNA (pML104 plasmid)	1 ug
10x NEBuffer 3.1	5 ul (1x)
*Swa*I	1 ul (10 units)
Nuclease-free water	up to 50 ul

2. After adding everything to the tube, spin the tube briefly in the bench-top centrifuge to collect all the contents in the tube. Then *gently* flick the tube to mix the reagents. The point is to make sure the enzyme and all ingredients are mixed together, but not scattered on the walls of the tube.

3. Incubate at 25°C for 30-45 minutes.

4. Remove 5 μl. This 5 μl aliquot is the *Swa*I only control sample.

5. Add 1.0 μl (or 10 units) of *Bcl*I to the remaining 45 μl *Swa*I reaction. This will become your double digest.

6. Incubate at 50°C for 15 minutes.

Checking digests:

7. Run 5 μl of the double digest alongside the SwaI only and undigested plasmid (be sure not to load more than about 100 ng of undigested plasmid on the gel!)

8. If the double digest was successful, proceed to column clean-up of the double-digested pML104 or store in freezer for clean-up at a later date. The column-cleaned plasmid will be used for ligation.

Do you expect the double and single digested and undigested pML104 plasmids to be the same size? Migrate the same distance on the gel?

PROTOCOL 8: Column clean-up of PCRs and digested plasmids

After PCR or restriction enzyme digestion there may be a need to column-clean your sample. For example, PCRs contain salts, primers, and enzymes that need to be removed before it can be used for other protocols such as transformations. After a restriction enzyme digestion we may wish to remove a "stuffer" fragment to improve efficiency of insert ligation.

The column clean-up kit utilizes the binding of DNA greater than about 100 bp (in the presence of highs salts) to a silica column, while the unwanted components and shorter DNA fragments (such as primers and "stuffers") wash through. The DNA (e.g. PCR product or plasmid) will be eluted in a pre-warmed low salt buffer. We are using this protocol to clean the PCRs of HDR1 and HDR2 PCR, as well as to remove the stuffer from the pML104 digestion.

Protocol

The elution buffer should be warming at 37^{o}C because warm elution buffer tends to result in greater yield of the clean product. Confirm with your IA that the buffer is warming.

1. Select your samples to be cleaned.

2. Add 3 volumes of binding buffer to the samples. For example, if you have about 100 μl of PCR sample, you would add 300 μl of binding buffer.

3. Apply each sample to a separate column in a collection tube, and spin for 1 minute at max speed (remember to label the column!)

   If all of the sample volume does not fit in the column, pipet what fits, spin, discard the flow-through, and then pipet the remaining sample and spin.

4. Discard the flow-through; apply 700 µl of wash buffer to the column and centrifuge for 1 minute.

5. Discard flow-through. Centrifuge empty column for 1 minute to dry.

6. Place the column into a new, labeled 1.5 ml tube. To elute DNA add 30 ul **warm** elution buffer to the center of the column. Let sit for one minute, then spin for 1 minute.

7. Nanodrop your samples to determine the concentration, and therefore amount, of DNA recovered from the clean-up.

PROTOCOL 9: Ligation of gRNA oligos and digested pML104

Refer to **Ligation** in the Background section.

9a. Hybridizing gRNA oligos

Protocol adapted from: https://www.addgene.org/protocols/annealed-oligo-cloning/

1. Each oligo should be resuspended in annealing buffer (10mM Tris, pH7.5-8.0, 50mM NaCL, 1mM EDTA) to a concentration of 2 ug/ul. If oligos come
2. Mix 2µg each in a total volume of 50µL of annealing buffer.
3. Place tube in a thermocycler programmed to start at 95°C for 2 minutes and then gradually cool to 25ºC over 45 minutes (or heat to 95°C for 2 minutes in a heat-block, remove from heat and let cool at room temperature for 45 minutes).
4. Dilute 5µL of annealed oligos with 45µL nuclease-free water and quantify the concentration (should be about 8ng/µl – check with nanodrop).
5. Hybridized oligos will be used to ligate into digested pML104. One 50 µl tube of 8 ng/µl will be enough for multiple groups.

9b: Ligation of hybridized gRNA oligos and digested pML104

You will need:

- Diluted, hybridized gRNA oligos were previously prepared (see 7a) and should be about 8 ng/ul.
- Column-cleaned, digested pML104. *Assume the concentration of this sample is 120 ng/ul.*

1. Mix the annealed oligos with cut plasmid to make a 2:1 insert: vector molar ratio (see below).
    - Use 30 ng of digested plasmid.
    - Recall information earlier in the lab manual, and useful tools such as: https://nebiocalculator.neb.com/#!/ligation
    - Set-up a 20 ul ligation reaction (see below) between pML104 and the hybridized gRNA# oligos
    - Plan to set-up these two control reactions:

      Control 1: is there any undigested plasmid in the double-digested sample (in other words, was all of the plasmid digested)?

      Control 2: was the stuffer completely removed by column clean-up?

How will you set-up these control reactions? That is, what should these control reactions contain?

COMPONENT	**20 µl REACTION**
T4 DNA Ligase Buffer (5X)	4 µl
Vector DNA	X µl
Insert DNA	X µl
T4 DNA Ligase	1 µl
Nuclease-free water	to 20 µl

2. Incubate for 15-30 minutes at room temperature, or overnight at 4°C.
3. Proceed to transforming bacteria (or save ligation in the freezer for transformation on another day).

PROTOCOL 10: Transforming *E. coli*

This procedure introduces plasmids into competent *E. coli* DH5α cells (DH5α refers to a particular strain of E. coli. You can find more information about this strain here: https://microbewiki.kenyon.edu/index.php/DH5-Alpha_E.coli). A very small volume of competent cells can be transformed and plated on media containing antibiotics, and subsequently you can observe colony growth and identify colonies containing your plasmid of interest.

Why are antibiotics added to the media?

Which antibiotic should be in the plates for your transformations?

Video overview of making and transforming competent cells:
https://www.addgene.org/protocols/bacterial-transformation/

The cells we are using have been previously treated to make them competent to take-up foreign DNA. This process involves treating *E. coli* cells harvested in the log stage of growth with divalent cations, in particular calcium. Once the cells are competent, plasmid can be added and taken up by the cells. We have purchased highly competent cells (Zippy™ cells) - you just need to dilute them and add your plasmid (e.g. ligation products).

Often when transforming cells, it is a good idea to include a positive and negative control. A positive transformation control is usually an uncut plasmid, such as uncut pML104. This will be our positive control, because we know this plasmid should transform the DH5α *E. coli* cells very well. We will use water as a negative control.

What do you expect from these controls in terms of colony growth after the transformation?

- Competent cells are **incredibly fragile** - never vortex them or pipette them vigorously.
- It is very important to **keep the cells on ice** during the procedure *until* it is time to plate. Warming of the cells prior to plating will reduce transformation efficiency.
- You will transform an aliquot of cells with each ligation reaction you set up, plus the positive and negative controls described above.

Transformation Protocol

Set-up five transformations:

- Ligation of pML-gRNA
- Uncut pML104 (10 ng/ul plasmid)
- Ligation control #1
- Ligation control #2
- Water (negative control)

1. **Two groups** will share one vial of cells. One group, take one vial of cells and, **on ice,** add 1.2 ml of ice-cold dilution buffer. Mix by gently flicking with your finger.

2. Each group should then aliquot 90 µl of the diluted cells into 5 **pre-chilled, labelled,** microfuge tubes. Ensure your tubes are clearly labeled for each different reaction.

3. Add 5 µl of the appropriate sample to the respective tube. Gently flick the tube to mix the cells and water. Do not vortex. Keep on ice.

4. Incubate transformation reactions on ice for 5 to 10 min.

5. Spread the bacteria on LB plates with ampicillin using aseptic technique.

    - Use an ethanol and flame-sterilized glass or metal plating rod to spread the bacteria around the plate. *Be sure to let the flame-sterilized rod cool sufficiently before spreading.*

More info: http://www.jove.com/video/3064/aseptic-laboratory-techniques-plating-methods
(click on “Spread-plating: Turntable with Glass or Metal Rod” from the right-hand menu)

- It is important to make sure the liquid is really spread around the plate so the bacteria do not over-grow on one part of the plate (the goal is to get distinct colonies).

6. Ensure the bottom (media side) of the plates are clearly labelled. Stack your plates and tape them together. Invert and incubate for 20-30 hours at 37°C. The IAs will remove plates the next day and store them at 4°C until the next lab.

What do you expect to see from each transformation? Why?

Analyzing *E. coli* transformations

- How many colonies did you get from each transformation?
- Calculate transformation efficiency for the positive control sample:

Transformation efficiency provides an indication of effectiveness of uptake of DNA into the competent cells. It represents the total number of bacterial cells that took up plasmid divided by the amount of plasmid used in the experiment. Transformation efficiency is expressed as colonies/one microgram of DNA.

Therefore, transformation efficiency = total number of colonies/relative amount of plasmid in the volume of cells spread on the plate, normalized to 1 µg plasmid.

Here is an example:

5 µl of 0.2 ng/µl pGEM was added to 75 µl of cells, and all 75 µl were plated. 2000 colonies grew on the plate.

1. (5 µl) x (0.2 ng/µl) = 1 ng of plasmid
2. 1 ng = 0.001 µg plasmid
3. Transformation efficiency = 2000 colonies/0.001 µg plasmid
   = 2×10^6 colonies/µg plasmid

Is it reasonable to calculate a transformation efficiency for the ligation reaction? Why or why not?

Record the transformation efficiency from three other groups. How do they compare to your value? What are possible reasons for the differences?

- Once a positive colony is identified by both colony PCR and sequencing, we will grow up the bacteria and extract plasmid. The recombinant plasmid (pML-gRNA) will eventually be used to transform yeast.

PROTOCOL 11: Colony PCR

Refer to Polymerase Chain Reaction (PCR) in the Background section if needed.

Often times after a transformation with a ligation we may have many colonies. Which colony carries our desired recombinant plasmid? Growing up colonies in liquid culture, extracting plasmid, and sending for sequencing is one way to determine if the plasmid carried by an individual colony contains the desired sequences. However, if there are many colonies, we may need to check multiple colonies before finding one that has our desired plasmid. To avoid having to do dozens (or more!) of plasmid extractions, one way to preliminarily screen bacteria is to do a colony PCR.

Using primers that will anneal and amplify a sequence from the desired recombinant plasmid, we use a colony as the "template" for PCR. During the high-temperature denaturation step in PCR, the bacterial cells will break open, releasing the plasmid DNA. The plasmid DNA can then be used as template for PCR. If we see a positive result from the PCR we can choose to grow up a specific colony for plasmid extraction and sequencing.

PCR method:

- The forward primer will anneal to the gRNA sequence.
- The reverse primer will anneal to the pML104 plasmid, about 400 bp downstream of the scaffold RNA sequence:

5′- CCCAAGAAAAAGCGCAAGGT -3′

Sketch the pML-gRNA plasmid. Where will the M13 RV primer anneal? Where will the FW primer anneal? How big is the expected PCR product?

- Set-up 25 μl reactions, using a final primer concentration of 0.4 μM
  You should make a table in your lab book for the volumes you will be adding to each reaction. ***Note you will need to figure out what volume of stock primer to add to achieve the desired 0.4 μM final concentration (see table on next page that will help)***

- You should also plan set-up the following controls:
  - Colony that carries pML104 only (from your transformation control plate)
  - No DNA template control (water instead of template)

- You should choose 5 colonies to use for colony PCR (see next page for more info on how to pick colonies and add them to PCR). You will also streak the colony that you put into the PCR tube onto a new plate. This way, if a colony is found to carry the desired plasmid, you can grow the colony from the re-streak.

- Cycling conditions:

95°C 2 min, 30 cycles of: 95°C 30 seconds, 50-55°C anneal* 30 seconds, 72°C extension 15 seconds

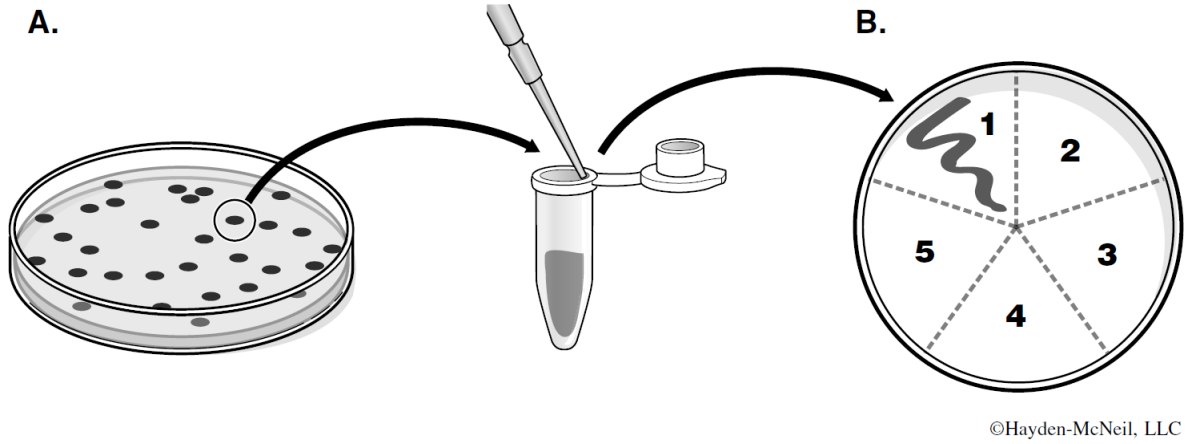
*annealing temperature might vary depending on the melting temperature of the gRNA primer (forward primer). The Tm of the reverse primer is 59°C.

PCR Set-up:

Component	**Volume for a 50 ul reaction**	**Volume for a 25 ul reaction**	**Final concentration**
GoTaq Green Master Mix, 2X	25 µl	12.5 µl	1X
Forward (upstream) primer, 10 uM	0.5-5.0 µl	0.25-2.5 µl	0.1-1.0 µM
Revers (downstream) primer, 10 uM	0.5-5.0 µl	0.25-2.5 µl	0.1-1.0 µM
DNA template	0.5-2.5 µl (1 colony = 1 µl)	0.5-2.5 µl (1 colony = 1 µl)	<250 ng
Nuclease-free water	to 50 µl	to 25 µl	N.A.

1. Get an LB+Amp plate. Divide into 5 sections.

2. Pick-up a sterile pipette tip and pick-up the colony with the tip, and smear the colony on the wall of the PCR tube (a). Then take that same tip and streak it on section 1 (for colony 1) of the LB plate (b).



You can also choose to FIRST add the PCR components to the tubes, and then inoculate the PCR tubes with the colony.

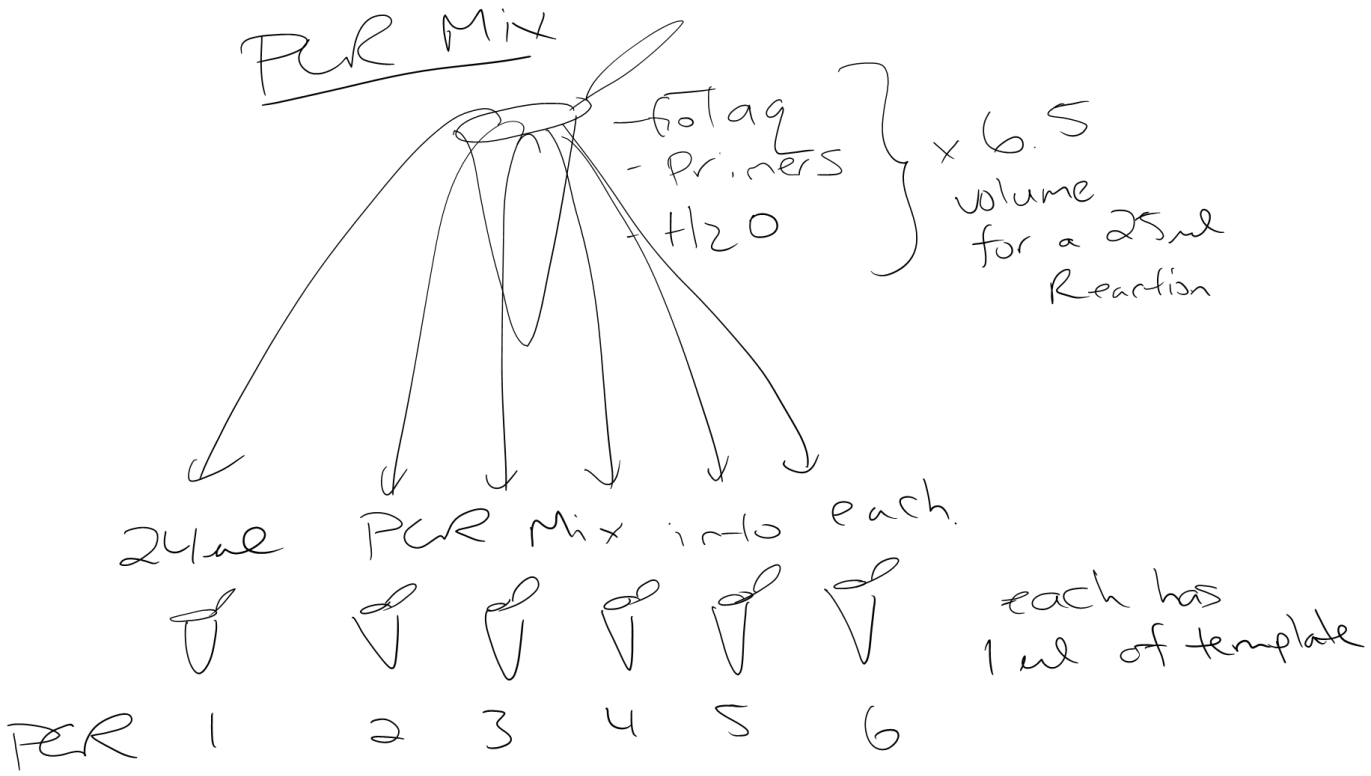
If you put the colonies in tubes before the PCR mastermix, be sure to use a different pipette tip to aliquot the mastermix into each PCR tube to avoid cross-contamination.

3. Incubate the re-streak plate at 37°C for at least 24 hours.

4. You can now add the PCR components to the PCR tubes.

If all the reactions need the same primers, you can set up a PCR Mix that contains enough volume of all the PCR reagents in one tube, and then aliquot the specific volume needed to all your PCR tubes.

For example, if I have 6 reactions that need GoTAq mastermix I would calculate to make a PCR Mix for 6.5 reactions (the 0.5 additional is to accommodate for some small loss during pipetting the aliquots to each individual PCR tube).



5. Put the PCR tubes in the PCR machine (or give to your IA/Instructor who will put them in the machine). Verify the cycling conditions for your records.

6. Check the colony PCRs using AGE (this may be done during a subsequent lab). You can load 10 ul of the PCR onto the AGE.

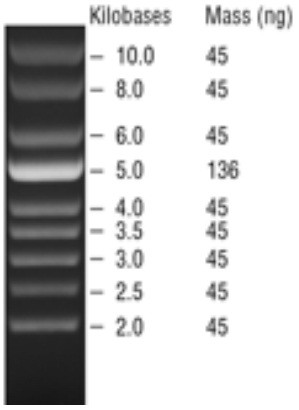
7. Choose one positive colony (the PCR was result suggests colony carries desired recombinant plasmid), to grow in liquid culture for plasmid extraction and sequencing (Refer to PROTOCOL 3: Growing Liquid Bacterial & Yeast Cultures).

PROTOCOL 12: Extract possible pML104-gRNA plasmids and send for Sequencing

1. Use the QiaPrep Alkaline Lysis Mini Prep protocol to extract plasmid from liquid cultures you previously set up.

2. Use agarose gel electrophoresis to check your extractions and estimate the concentration of the plasmid samples.

    - Run 5 μl of each plasmid extraction, mixed with 2-3 μl of 6x sample buffer
    - Load 10 μl of ladder into one or two wells. If we have access to a supercoiled DNA ladder, use it as a size standard (see below for more information on the supercoiled DNA ladder)



NEB supercoiled plasmid ladder. The band containing the 5 kb plasmid standard is brighter than the other bands. The mass reported is true when loading 10 ul.

What is the benefit of using a supercoiled ladder, compared to the λ HindIII ladder we typically use?

3. Use your gel image to estimate the concentration of your extracted plasmids.

4. Prepare a 10 μl aliquot of 20-75 ng/μl of each plasmid for sequencing. Give your IA your microfuge tubes of plasmid, VERY CLEARLY LABELLED using the labelling system your IA or Instructor recommends.

5. The samples will be sent for sequencing with a primer that anneals upstream of the gRNA insert sequence:

    pML104 sequencing primer: 5’ caggaaacagctatgac 3’ (called M13 RV)

PROTOCOL 13: Analyzing Sequencing Results

There are questions scattered throughout the protocol. Determine if your IA/Instructor want you to answer these questions in your lab notebook or in a separate assignment that you will submit.

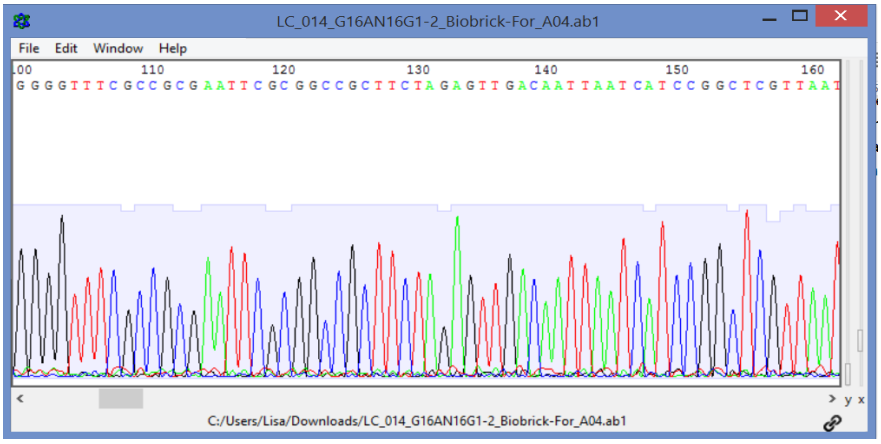
Use ApE: http://biologylabs.utah.edu/jorgensen/wayned/ape/
or a DNA analysis program that you like to use (instructions below are for ApE).
**If you are using computers in the York Computer Labs you do not need to download a program.

- For each sample you should have two sequence files:
    - Chromatogram file (.abi)
    - Sequence file (.seq)
- Have access to the gRNA sequence and the annotated pML104 plasmid sequence files you have previously created.

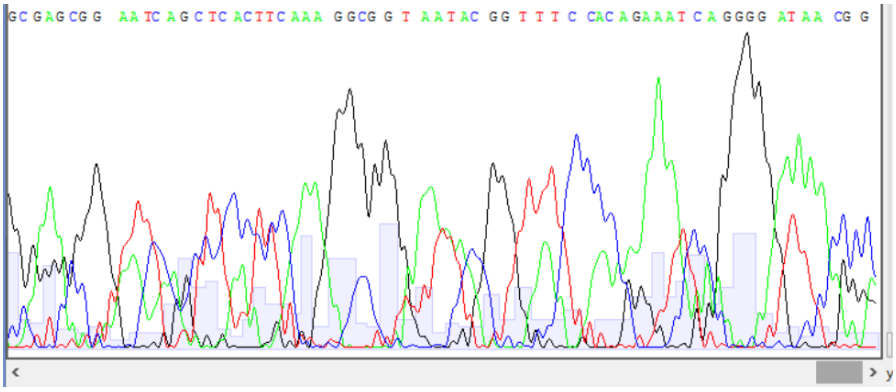
A. Checking the chromatogram

1. Find your chromatogram sequence file (.abi file). Open the chromatogram file in ApE

- Your chromatogram should look something like the one below. Note how the first 70 peaks or so do not look very sharpened and that there may be some noise along the baseline. You should find that from about nucleotide 50 or 70 onward, the peaks are well resolved and there is little noise (like the chromatogram below). This is a good sequencing reaction.



If you find that multiple peaks overlap past the 70-nucleotide mark, this could be an indication that the sequencing reaction was of poor quality. See second chromatogram below: note how the peaks all overlap. If you don't think your sequence looks good, ask your IA or the Instructor.



What you hope to find is that after about 70 nucleotides you have a nice, clean chromatogram with well-defined peaks. If so, you can use the nucleotide sequence determined from this chromatogram to find the features we care about. Follow steps in the next section.

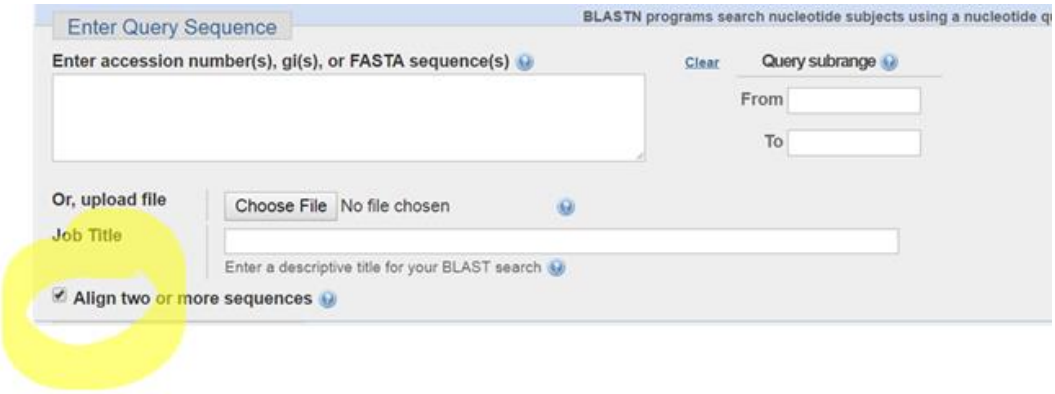
Q1: Overall, how "clean" were each of your sequencing reactions? Explain.

B. Use ApE to analyze and annotate your sequence

1. Open the .seq file for the corresponding chromatogram, in ApE.

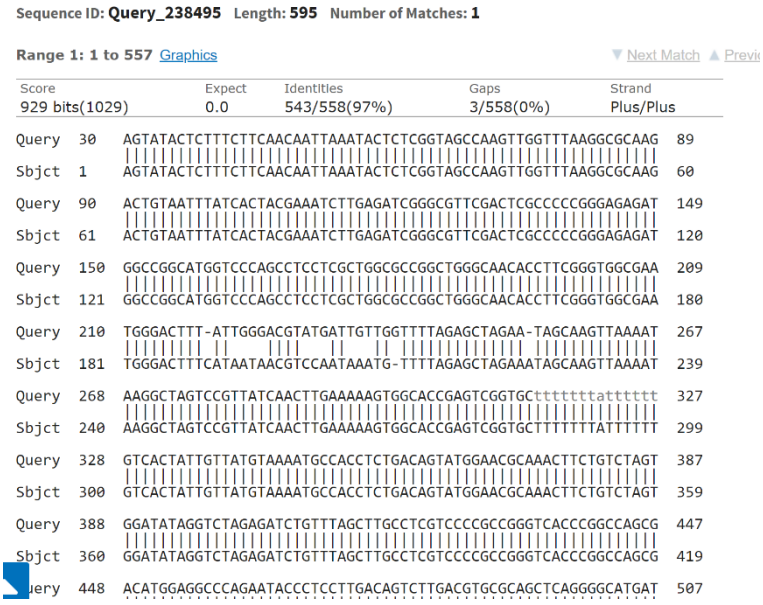
2. Search for the various sequences within your sequencing reaction and add them as features**.**
   - **gRNA sequence**
   - **sequences upstream of the gRNA**
   - **Scaffold gRNA sequence**
   - **Terminator sequence**

3. You can use the align two sequences tool through NCBI BLAST to determine how you're your pML104-gRNA sequence aligns to the known pML104 sequence:
   - Go to https://blast.ncbi.nlm.nih.gov/Blast.cgi
   - Select Nucleotide Blast
   - Make sure the box beside Align two or more sequences is checked. You should see two search boxes, one for Query and one for Subject:



- In the “Query” box, paste in your sequencing reaction result.
- In the “Subject” box, paste in the region of the plasmid you expect to find in your sequencing reaction.
- Select “somewhat similar sequences” (blastn) under Program Selection
- Click on “BLAST”

- Once you get the results screen, select the “alignment” tab. If your sequence result is indeed the plasmid containing the gRNA, you should get something like this (the numbers of query to subject alignment will differ):



- The Identities value at the top of the alignment tells you what percentage of your sequence (query) matches the subject sequence.

Notice the region where the query and subject do not align well. What is a likely reason for this non-matching region?

Where would we *not* expect to see mis-matches?

PROTOCOL 14: Making double-stranded HDR templates

Goal: Make double-stranded HDR templates, which will then be cleaned and eventually used to transform yeast along with pML104-gRNA plasmid.

HDR1 – make copies with PCR
The full HDR1 template has already been copied, but may not be at a high enough abundance to do a yeast transformation (we want at least 200-500 ng for a yeast transformation). You can set-up a PCR with existing HDR1 as the template for PCR, primers that anneal to the absolute 5' and 3' ends of the template, and will extend inward, making copies in an exponential fashion (typical PCR).

PRIMER SEQUENCES:
Forward: 5' CGGACAAAACAATCAAGTATG 3'

Reverse: 5' GGAGAATTTTCAGCATCTAG 3'

New HDR templates - synthesis by the overlapping oligo method
This is not like a typical PCR, which is exponential, because the top and bottom HDR oligos will anneal and act as primer and template, extending to create a double-stranded HDR.

For more information about the two methods of making copies see Making copies of HDR templates in the background section.

Here is an example of two overlapping oligos that would be used to make double-stranded HDR1. Bolded nucleotides are different than the wildtype sequence because these are the mutations designed into the HDR sequence. The top strand is called the "top oligo" and the bottom strand is called the "bottom oligo"

5'ATGGATTCTAGAACAGTTGGTATATT**G**GGAGGGGGA**T**AATTG**T**GACGTATGATTGTTGAG**TA**AGC**T**AACAGGCTC 3'
3'AACTC**AT**TCG**A**TTGTCCGAGTTGTAATTCTGCCATTATGATCTACGACTTTTAAGAGG 5'

- We use Taq polymerase for these reactions.
- Reactions should have primer (or top and bottom oligo) concentrations of 1 μM forward (top) and 1 μM reverse (bottom).
- Set-up a 100 μl reaction to ensure there will be an abundance of double-stranded HDR to use for yeast transformations.

Component	Volume for a 50 ul reaction	Volume for a 25 ul reaction	Final concentration
GoTaq Green Master Mix, 2X	25 μl	12.5 μl	1X
Forward (upstream) primer, 10 uM	0.5-5.0 μl	0.25-2.5 μl	0.1-1.0 μM

Revers (downstream) primer, 10 uM	0.5-5.0 μl	0.25-2.5 μl	0.1-1.0 μM
DNA template	0.5-2.5 μl (1 colony = 1 μl)	0.5-2.5 μl (1 colony = 1 μl)	<250 ng
Nuclease-free water	to 50 μl	to 25 μl	N.A.

- Cycling conditions for HDR1: 95ºC 2 min, 40 cycles of: 95ºC 30 seconds, 48ºC anneal 30 seconds, 72ºC extension 15 seconds
- Conditions for synthesis of new HDRs: 95ºC 2 min, 2 cycles of: 48ºC anneal 30 seconds, 72ºC extension 15 seconds

- Once the cycle is complete, check 5 μl using AGE (this may occur during the next lab). Successful reactions will be column-cleaned for use in yeast transformation (refer to PROTOCOL 8: Column clean-up of PCRs and digested plasmids).

PROTOCOL 15: Transforming yeast

Wildtype yeast cells have been pre-treated to become competent to uptake foreign DNA using the Zymo Yeast Transformation II Kit:
https://files.zymoresearch.com/protocols/_t2001_frozen-ez_yeast_transformation_ii.pdf

The competent cells were frozen at -70°C, and are being provided for your transformation.

What you will need:

- pML-gRNA, at least 500 ng (but ideally 1 ug)
- HDR1 clean PCR product, at least 200 ng

What transformations should be set-up? That is, what should yeast be transformed with to help address our goals and research questions?
What are the controls?
What do you expect to observe on the plates of yeast colonies that grow after the transformation?

Yeast transformation protocol

1. Obtain 50 μl of competent yeast cells for each transformation you are doing. Keep cells on ice.

2. Add plasmid and HDR (if appropriate) to the 50 μl of competent cells. Mix gently by flicking.

3. Add 500 μl of warm EZ 3 solution to each tube of cells and mix thoroughly by flicking.

4. Incubate at 30°C for at least 1.5 hours with gently shaking (about 125-150 RPM).

5. Add 500 μl of warm liquid YPD. Incubate at 30°C for at **least 1.5 hours** with gently shaking (about 125-150 RPM).
   This is a recovery phase.

6. Make sure the plates you will need for spreading are warming at 30°C.

7. You will spread the yeast from a single transformation onto multiple plates (e.g. about 330 μl onto each plate). Recall that yeast transformed with pML104 are plated on uracil minus (URA-) plates. Use aseptic technique.

8. Incubate the plates at 30°C for 3-4 days to allow for growth of transformants.

What do you expect to see on the plates after 3-4 days? Do you expect differences between the various transformations?

Analyzing transformants & setting-up liquid cultures

- Take note of how many colonies have grown from all your transformations, as well as colony phenotype. You can use the dissecting microscopes to get a closer view of colony morphology.

- Inoculate 2 ml of liquid culture for **four** different colonies.
- *Take-note that pML-transformed yeast need to be grown in uracil minus media.

- Grow, shaking at 30^{o}C, for 24-36 hours. You will be extracting DNA from these cultures to check if the target gene was edited.

PROTOCOL 16: Genomic DNA extraction from yeast

1. Label a microfuge tube for each culture.
2. Mix culture (vortex – but make sure cover is on tightly before vortexing!)
3. Transfer 200 μl of culture into labeled tube.
4. Centrifuge for 45 seconds at 10,000 RPM to pellet the yeast.
5. Pipette supernatant out of the tube (discard supernatant into bleach).
6. Add 50 μl of 10 mM NaOH to the pellet.
7. Close the tube tightly and vortex until you can't see the pellet anymore (look closely).
8. Place the microfuge tube in the 99°C heat block so that the level of the liquid is below the sand. Incubate for 5 minutes 30 seconds.
9. After the incubation, immediately place the tube on ice. Use as template for PCR or store in freezer for later use.

PROTOCOL 17: PCR of *ADE2* for genetic screening

To determine if the *ADE2 gene* in yeast transformants was edited we will amplify a region of the ADE2 gene and send it for sequencing.

Depending where within the gene you expect to find mutations, we will use different primers to amplify different regions of the gene. Here are the primers, and respective melting temperatures, we have access to (sequences are written 5' to 3'):

D-FW	TTGTTGCATGGCTACGAACC	59
E-RV	CTTTACAACGAAGTTACCTCTTCCA	59
F-FW	TTGAAGTCGAGGACTTTGGCA	59
G-RV	CCGCCATACTGGAGGCAATAA	60
H-FW	CGAAAGAGCATTGGCGACT	58
I-RV	CATTTGATGTAATCATAACAAAGCC	55
J-FW	CCTTTTGATGCGGAATTGAC	54
K-RV	CCAATTGTCCCCCTCCTAAT	56

Locate these primers within the ADE2 gene sequence.

Where do they anneal? Which primer set would be best for identifying mutations caused by the gRNA used in our experiments?

What size PCR product do you expect?

PCR cycling conditions: 95°C for 2 minutes, 35 cycles of: 95°C 30 seconds, 45-60°C anneal 30 seconds, 72°C 30 seconds; 2 min final extension; hold at 15°C.

1. For each genomic DNA sample, set up a 50 μl PCR, using the appropriate primers at a final concentration of 0.5 μM. Use 2.5 μl of genomic DNA as template.

2. Check PCRs using agarose gel electrophoresis (today, or during a subsequent lab).

3. Column-clean successful PCRs.

4. Prepare a 10 μl sample that is 10-20 ng/μl to send for sequencing. Provide this to your IA, labelled as instructed. What primer will be used for sequencing?

PROTOCOL 18: Analysis of genomic DNA sequencing

- For each sample you may have two sequence files (available on course website):
    - Chromatogram file (.abi)
    - Sequence file (.seq)

- You will need access to the wildtype ADE2 sequence, annotated to show expected mutations as a result of HDR:
  ADE2 with gRNA1-HDR1 annotations:
  https://drive.google.com/file/d/1VVxymy9yCkWeIEYJBSnFH9ZZfCs-Ce86/view?usp=sharing

  ADE2 with gRNA11-HDR11 annotations:
  https://drive.google.com/file/d/10VWPr-Onpo64atJ_niDy1Zh41LdVomuu/view?usp=sharing

What do you expect to see in the returned sequences if HDR occurred? What if HDR did not occur?

Things to consider when checking your sequences:

- Are there any differences detected between the wildtype and sequencing results that were not expected? If so, are they likely true changes or an artifact the sequencing? Explain.

- Once you have analyzed all of the sequences, consider the following:
    - Are there any differences in frequency of edits depending on gRNA?
    - Frequency of all HDR edits incorporated? Any trends in which HDR edits were more often incorporated? What does this tell us about the HDR process?
    - What types of mutations occurred by NHEJ?
    - Were there any unexpected results? What is a possible explanation for them?
    - What would you do in the future to elaborate on these results?

PART 1: ANALYZING TWO SEQUENCES

The purpose of this exercise is to analyze two sequence results and compare them to wildtype. After completing this, you will be prepared to do a multiple sequence alignment to look at the results from several to dozens of sequencing results.

- Access the wildtype ADE2 sequence annotated with gRNA1-HDR1 (see links above).
- Access these two sequencing results, from a transformation with pML104-gRNA1 + HDR1. These are both chromatogram files.

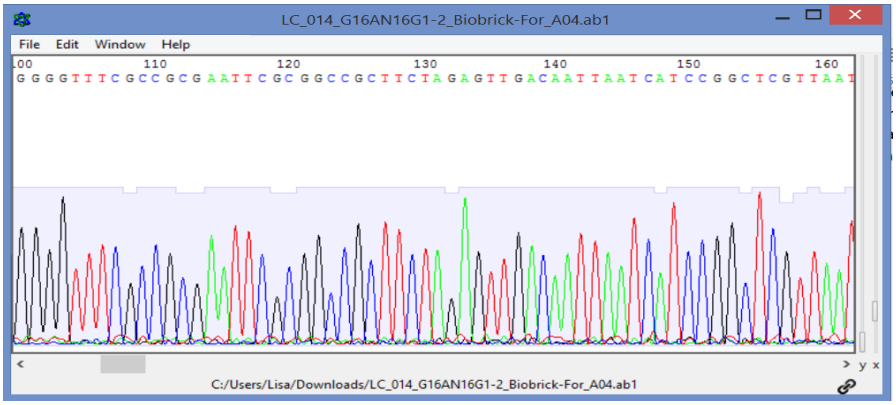
  62A sequence:
  https://drive.google.com/file/d/1WWGqL4QILgkEnZEn3ejNULiZCqJ2Shzn/view?usp=sharing

  12A: https://drive.google.com/file/d/1QRVCFeHnv5UuCf_rIwKGf7PDVorm8n-j/view?usp=sharing

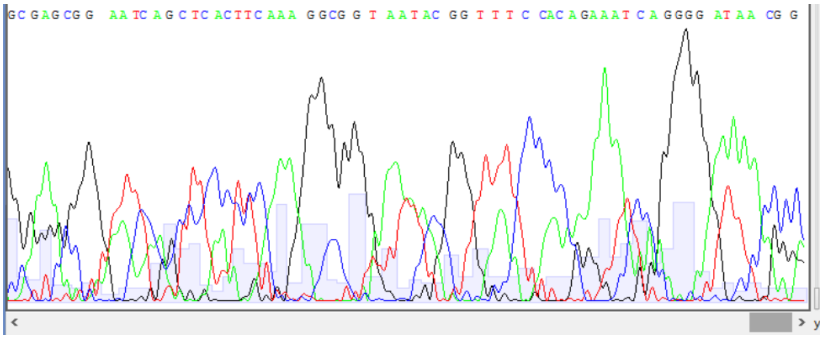
A. Checking the chromatogram

1. Open the chromatogram file (.ab1) in ApE

- Your chromatogram should look something like the one below *after* the first 50-70 nucleotides. It is common for the first 50-70 peaks to sometimes appear not very sharp, but after that region a 'good' sequencing reaction will have sharp peaks, as seen below.



If you find that multiple peaks overlap past the 70-nucleotide mark, this could be an indication that the sequencing reaction was of poor quality. See second chromatogram below: note how the peaks all overlap. If you don't think your sequence looks good, ask your IA or the Instructor.



What you hope to find is that after about 70 nucleotides you have a nice, clean chromatogram with well-defined peaks.

Overall, how “clean” were each of your sequencing reactions? Explain.

B. Use ApE to compare the sequencing results to wildtype

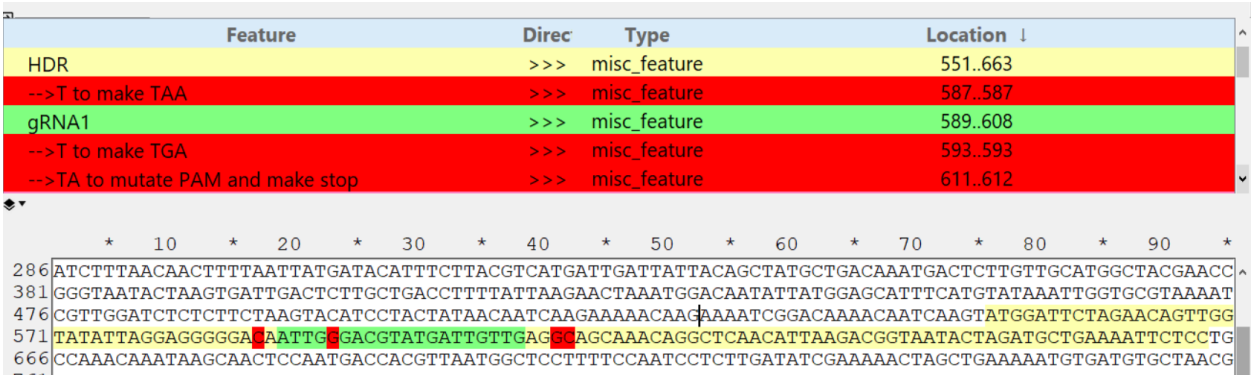
1. With the chromatogram file open, convert it to a nucleotide sequence: File → New DNA from basecalls and it will create a sequence file in ApE.
    - Make sure you convert both 62A and 12A.

OR, if you have access to the .seq file you can drag it/open it in ApE.

2. Open the relevant annotated wildtype sequence (see links above). In the annotated file you should see features indicating the

- **gRNA sequence**
- **PAM sequence**
- **sequences upstream and downstream of the gRNA that correspond to the homology arms of the HDR (if HDR used in the transformation)**
- **Mutation locations in the HDR, and what mutations they are (annotated features)**

You should have something like this:



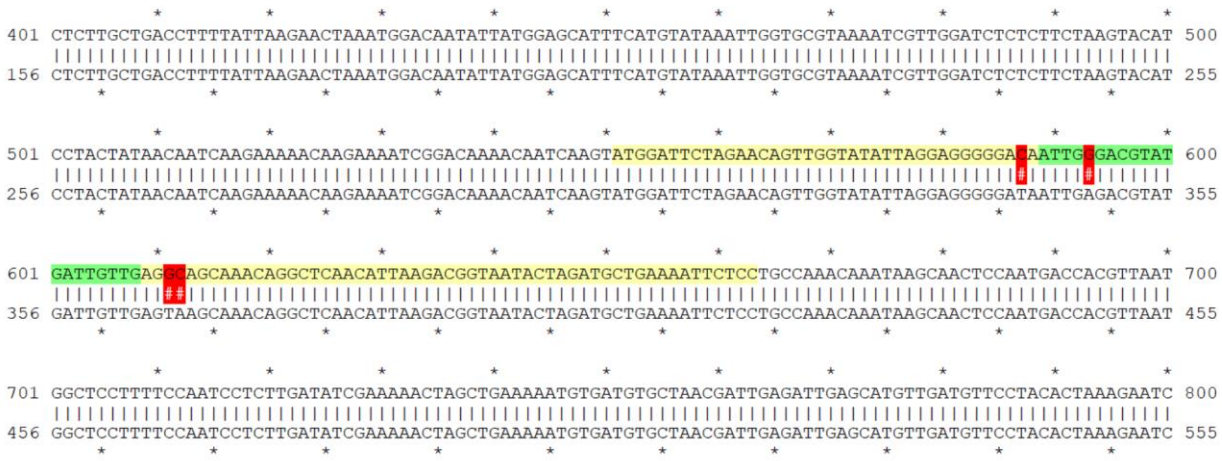
3. **If you want to compare one sequence at a time:**

   Use the “align two sequences” tool in ApE to determine how the sequencing result aligns to the wildtype ADE2 gene sequence:

   - Tools → Align two sequences
   - Select the two sequences, one being the wildtype ADE2 with annotations (as above) and the other being the sequencing result
   - Make sure “copying with highlighting” is selected
   - Click “OK” to do alignment

You should get something like the alignment below. Here the top sequence is wildtype ADE2, and the bottom is the sequencing result. Notice the red highlighting with # in-between the nucleotides that are different. This shows where point mutations exist in the ADE2 sequencing reaction in comparison to the wildtype sequence.

The next step is to confirm these sequence differences were expected: were they part of the HDR template? Are they the result of NHEJ?



Notice the region where the query and subject do not align well. What is a likely reason for this non-matching region? Where would we *not* expect to see mis-matches?

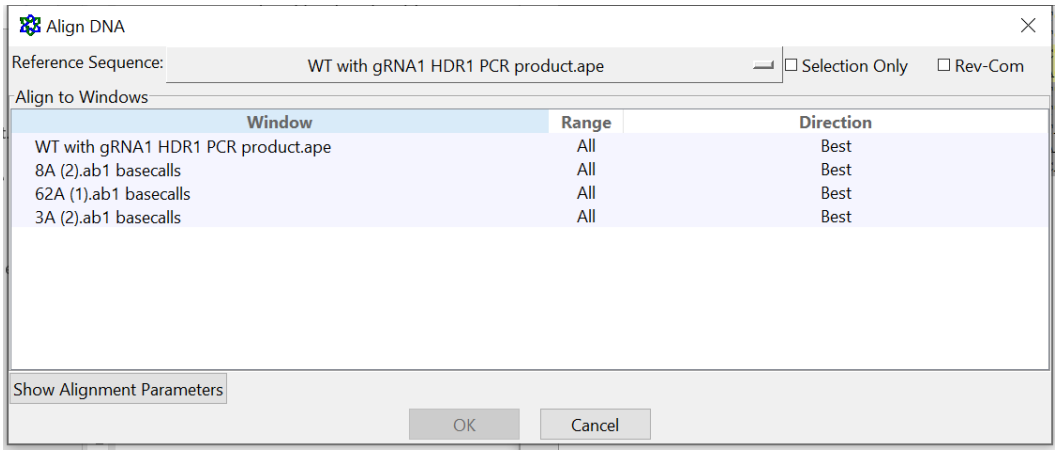
4. To perform a multiple sequence alignment (where you can compare multiple sequencing results to the wildtype at the same time):

Creating a Multiple Alignment in ApE

1. Have all of the sequence files you want to align open as nucleotide sequences.

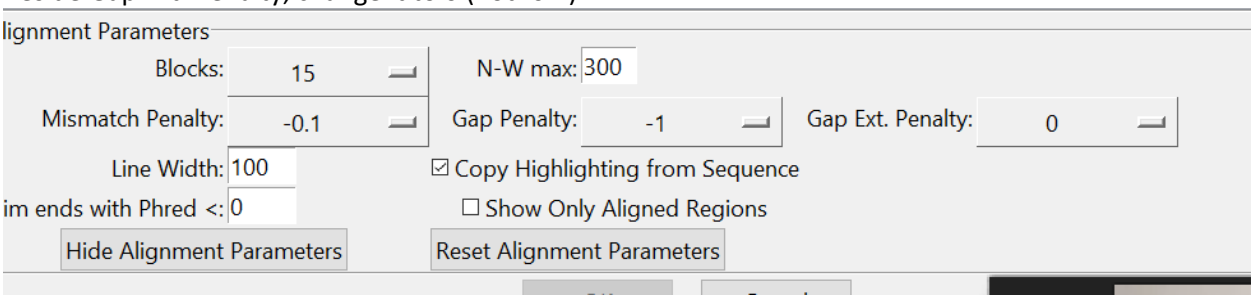
2. Go to **Tools Align Sequences**

   A window like this should open:



Be sure your **wildtype** sequence is the Reference Sequence. Then, click on all the sequences you want in your alignment (do not select the WT b/c you already have it as the reference). You can hold down Ctrl to select multiple sequences.

- CLICK ON **SHOW ALIGNMENT PARAMETERS.**
  Beside Gap Ext. Penalty, change it to 0 (not -0.2)



- Click OK. An alignment window will open. The first nearly 1000 nucleotides of the wildtype will not align with the sequencing result (if it is gRNA1-related sequence) because the wildtype sequence contains 1000 nt upstream of the start codon. It will look something like this at the top of the alignment:

```
Fri Dec 18, 2020 10:42 PST
both gRNA-HDR1 and gRNA-HDR11 on ADE2 WT.ape from 1 to 3716
Alignment to
8A (2).ab1 basecalls--  Matches:635; Mismatches:3; Gaps:3079; Unattempted:0
62A (2).ab1 basecalls--  Matches:653; Mismatches:6; Gaps:3058; Unattempted:0

                *         *         *         *         *         *         *         *         *
  1>CTACTAGAAAGAAG-GGCGTCAAAAATCTGCTATGCTATTACGAAATTACTCTATAGGTTAGAATGTCATCCCATAGGTATGGCCCTTTTAAAGAAGTAT>99
  1>~~~~~~~~~~~~~G-GGC-------------------------------------------------------------------------------->4
  1>~~~~~~~~~~G---NGGC--------------ATGC-------------------------------------------------------------->9

    *         *         *         *         *         *         *         *         *         *
100>CTCCACAATCAATTGCGAGAAGCCTATCTAGAATCAAAACGACACTTTATTTCCAAAAAGGGAGATTCGACCAACACTTCCTCTACCATTGCATCATCAT>199
  4>---------------------------------------------------------------------------------------------------->4
  9>---------------------------------------------------------------------------------------------------->9

    *         *         *         *         *         *         *         *         *         *
200>CTTTCGCTGGCGCATCTGTTCCTCTATCTTCAAACGAATCAGGAATGCTAAACGGCTTGAAGCAAATTAACGAACAACAAGAATCTACATTAGAAACCAC>299
  4>---------------------------------------------------------------------------------------------------->4
  9>---------------------------------------------------------------------------------------------------->9

    *         *         *         *         *         *         *         *         *         *
300>TCAAAAGGAAGACTAGTAACGCCGTATCGTGATTAACGTATTACATAAGTTACAGGATTCATGCTTATGGGTTAGCTATTTCGCCCAATGTGTCCATCTG>399
  5>-------------------------------------------------------------------------------ATGTG--------->9
  9>---------------------------------------------------------------------------------------------------->9

    *         *         *         *         *         *         *         *         *         *
400>ACATTACTATTTTGCATTTTAATTTAATTAGAACTTGACTAGCGCACTACCAGTATATCATCTCATTTCCGTAAATACCAAATGTATTATATATTGAAAG>499
  9>---------------------------------------------------------------------------------------------------->9
  9>---------------------------------------------------------------------------------------------------->9

    *         *         *         *         *         *         *         *         *         *
500>CTTTTGACCAGGTTATTATAAAAGAAACTTCATGCTCGAAAAAGATCATTTCGAAAAGTTGCCTAGTTTCATGAAATTTTAAAGCAGTTTATATAAATTT>599
  9>---------------------------------------------------------------------------------------------------->9
```

- Scroll down to the region that is targeted by the gRNA. You should see a clean alignment, with mostly matching nucleotides. If HDR or other mutations have occurred, these will be indicated by red highlighted nucleotides in the sequencing result rows.

Below is an example. Here, the part I am interested in is the region immediately around the expected cut site, so around bp 999 through 1098 of the wildtype sequence (top sequence).

```
    *         *         *         *         *         *         *         *         *         *
999>GTATGGATTCTAGAACAGTTGGTATATTAGGAGGGGGACAATTGGGACGTATGATTGTTGAGGCAGCAAACAGGCTCAACATTAAGACGGTAATACTAGA>1098
162>GTATGGATTCTAGAACAGTTGGTATATTCGGAGGGGGACAATTGGGACGTATGATTGTTGAGTAAGCAAACAGGCTCAACATTAAGACGGTAATACTAGA>261
165>GTATGGATTCTAGAACAGTTGGTATATTCGGAGGGGGATAATTGTGACGTATGATTGTTGAGTAAGCTAACAGGCTCAACATTAAGACGGTAATACTAGA>264
```

Take a screenshot of the alignment for your lab notebook. Make a proper figure caption, indicating what is being annotated and what mutations were found in the sequencing results.

Making sense of the nucleotide numbers in an alignment:

- Each ROW is a different sequence. Above,
  - Row 1 is WT
  - Row 2 and 3 are different sequencing results.
- The nucleotide (nt) numbers of the sequence in each row are displayed.
- For the alignment above, the region of interest is around nucleotide (nt) 1000 of the wildtype. Because the sequencing results are from a PCR reaction that was sent for sequencing, the nucleotide numbers do not align. For example, imagine you had a 1000 bp PCR product, and the primer used for sequencing anneals at nt 100, and the region of interest is around nt 200, it would make sense that the sequencing result nt 160 onward will be aligning with the wildtype in our region of interest (as above).

PART C – Analyzing class sequencing results

- Access the folder with the sequencing results from all the colonies returned from the at-home experiments.

https://drive.google.com/drive/folders/1nhwqPvC4QhvoQXMNMWsdkBEnaSmMP8yA?usp=sharing

- Use the multiple sequence alignment tool to compare the results from colonies transformed with pML104-gRNA1 + HDR1 to wildtype to characterize the HDR mutations introduced.

Take a screenshot of the alignment for your lab notebook. Make a proper figure caption, indicating what is being annotated and what mutations were found in the sequencing results.

You may want to also make a summary regarding the number of sequences will full or partial HDR. Discuss with your group, IA, Instructor some ways to summarize these results.

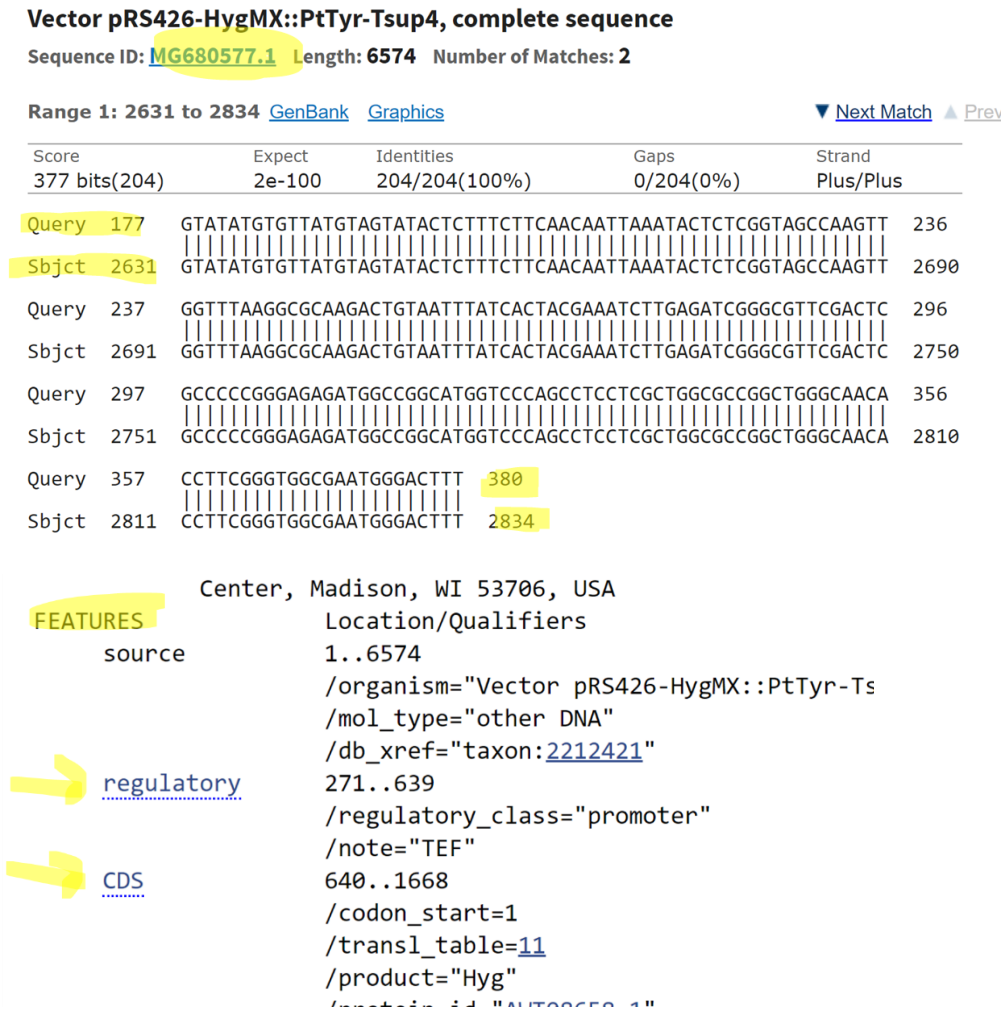
- Repeat the alignment process to characterize the mutations introduced by NHEJ (pML104-gRNA transformations).
- Repeat the alignment process to characterize mutations introduced by transformations with pMl104-gRNA**11** and HDR**11**

  gRNA11 results:
  https://drive.google.com/drive/folders/1X6_HFGs4sAOCdxaueVrNWzRXAUETCADv?usp=sharing

- **Take similar screenshots for these additional alignments, and make summaries, in your lab notebook.**

APPENDIX A – Identifying Sequences with BLAST

- Go to https://blast.ncbi.nlm.nih.gov/Blast.cgi and select “Nucleotide BLAST”
- Copy and paste the sequence into the Query box
- Click “BLAST”
- If you are looking for features:
    - Click on the alignment that interests you
    - Take note of the nucleotide numbers that your query sequence matches that of the of the subject sequence
    - Click on the accession number/sequence ID of the subject sequence
    - Scroll through to see the list of features and the corresponding nucleotide positions. Find the position(s) from your query match. If you click on the feature hyperlink, the page will take you to the sequence.



Vector pRS426-HygMX::PtTyr-Tsup4, complete sequence

Sequence ID: MG680577.1 **Length:** 6574 **Number of Matches:** 2

Range 1: 2631 to 2834 GenBank Graphics ▼ Next Match ▲ Prev

Score	Expect	Identities	Gaps	Strand
377 bits(204)	2e-100	204/204(100%)	0/204(0%)	Plus/Plus

```
Query  177   GTATATGTGTTATGTAGTATACTCTTTCTTCAACAATTAAATACTCTCGGTAGCCAAGTT  236
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2631  GTATATGTGTTATGTAGTATACTCTTTCTTCAACAATTAAATACTCTCGGTAGCCAAGTT  2690

Query  237   GGTTTAAGGCGCAAGACTGTAATTTATCACTACGAAATCTTGAGATCGGGCGTTCGACTC  296
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2691  GGTTTAAGGCGCAAGACTGTAATTTATCACTACGAAATCTTGAGATCGGGCGTTCGACTC  2750

Query  297   GCCCCCGGGAGAGATGGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACA  356
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2751  GCCCCCGGGAGAGATGGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACA  2810

Query  357   CCTTCGGGTGGCGAATGGGACTTT  380
             |||||||||||||||||||||||||
Sbjct  2811  CCTTCGGGTGGCGAATGGGACTTT  2834
```

```
            Center, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1..6574
                     /organism="Vector pRS426-HygMX::PtTyr-Ts
                     /mol_type="other DNA"
                     /db_xref="taxon:2212421"
     regulatory      271..639
                     /regulatory_class="promoter"
                     /note="TEF"
     CDS             640..1668
                     /codon_start=1
                     /transl_table=11
                     /product="Hyg"
```

APPENDIX B: Features of ApE

Toolbar

Many of the toolbar buttons are associated with two related actions. The first action is activated by simply clicking on the toolbar button. The second action is activated by holding down the shift key while clicking on the toolbar button or by clicking on the toolbar button with the right side button of a three-button mouse. In this list, the primary action is listed first, followed by the secondary action. The default toolbar has the following buttons:


New: Open a new, blank DNA window
Duplicate: Open a new DNA window containing a copy of the currently selected DNA


Open: Open a new DNA file. Currently supported file types are: DNA strider, FASTA and plain text


Save: Save current DNA file (no file name prompt, unless the file has not been previously saved)
Save As: Save the current DNA file with prompt for new file name.


Cut: Cut selected DNA to clipboard.
Cut rev-com: Cut the reverse complement of selected DNA to clipboard.


Copy: Copy selected DNA to clipboard.
Copy rev-com: Copy the reverse complement of selected DNA to clipboard.


Paste: Paste DNA from the clipboard.
Paste rev-com: Paste the reverse complement of DNA from the clipboard.


Select From-To:
Jump To:


Find:
Clear Find:


Uppercase:
Lowercase:


Swap case:


Reverse Complement:


Enzyme Selector:


Graphic Map:
Graphic Map + Unique Enzymes:

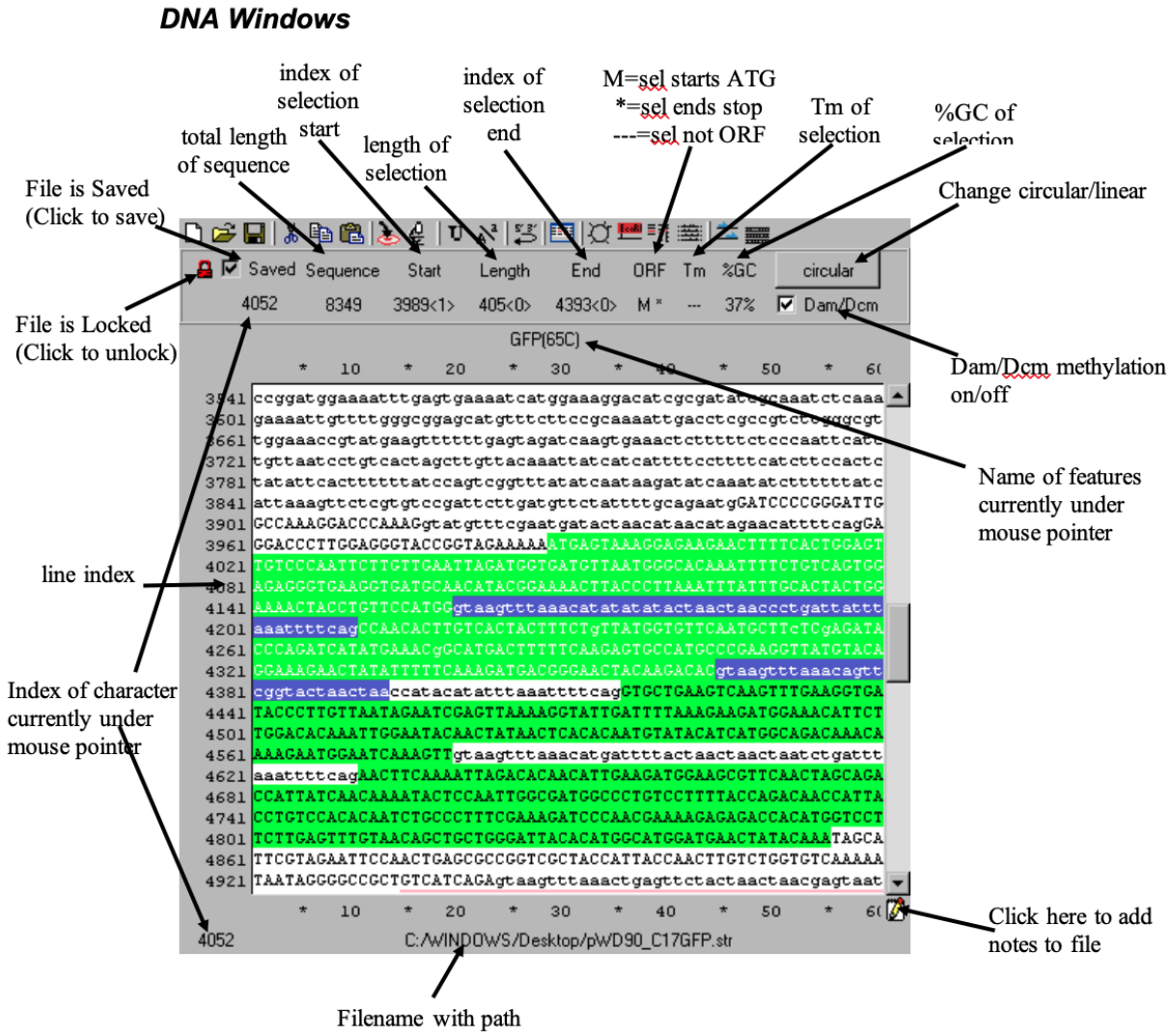

Highlight Enzyme Sites:
Clear Enzyme Site Highlight:


Digestion Window (Agarose Gels):


Text Map:


Find PCR Primers:


Align Sequences:

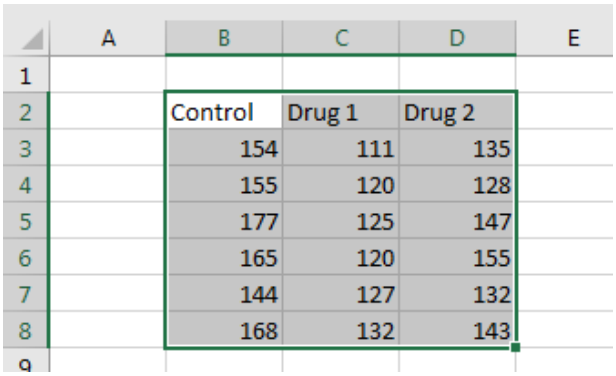


APPENDIX C: Creating Graphs in Excel

Bar Graphs

Bar graphs are used when the independent variable is a category like a country or condition. In the example below, we are going to graph the average blood glucose concentration in mg/dL from 6 control mice versus two groups of 6 mice each treated with different drugs. In this example, we will also learn how to add errors bars.

First, organize the data in Excel as shown below:

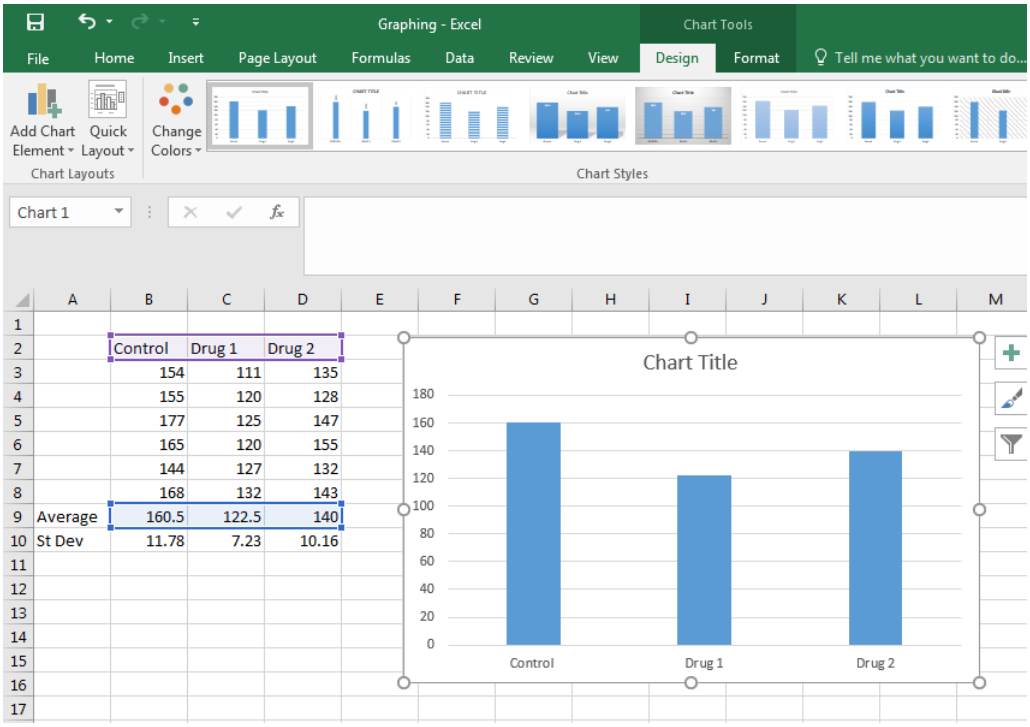


	A	B	C	D	E
1					
2		Control	Drug 1	Drug 2	
3		154	111	135	
4		155	120	128	
5		177	125	147	
6		165	120	155	
7		144	127	132	
8		168	132	143	
9					

Next, find the average and the standard deviation for each group. (If you do not know how to do this, ask for help.)

	A	B	C	D	E
1					
2		Control	Drug 1	Drug 2	
3		154	111	135	
4		155	120	128	
5		177	125	147	
6		165	120	155	
7		144	127	132	
8		168	132	143	
9	Average	160.5	122.5	140	
10	St Dev	11.78	7.23	10.16	

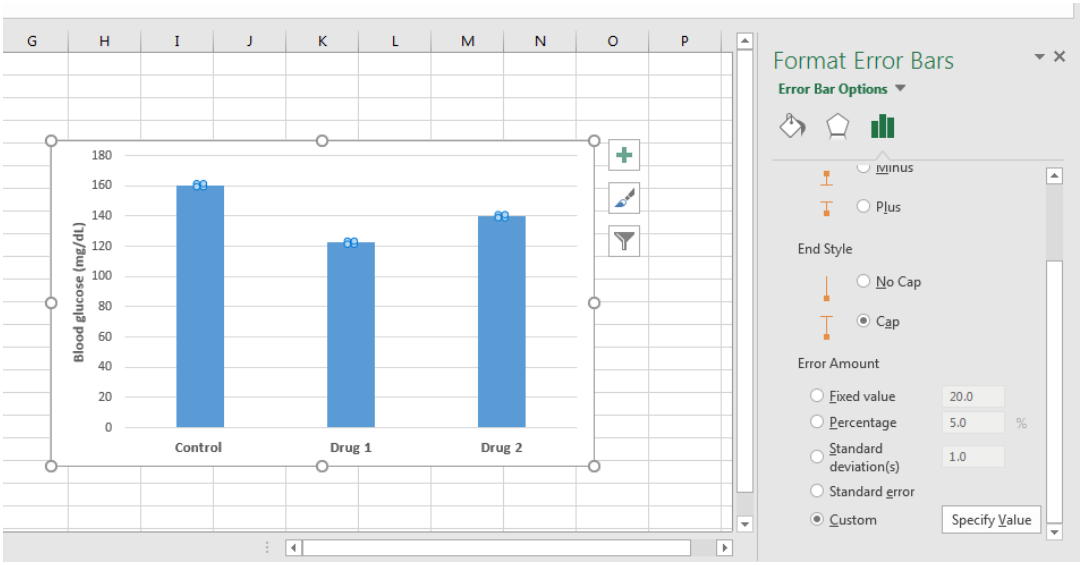
Now highlight the titles of the columns, then while holding down control, highlight the average values. Then Select a bar graph from the chart wizard. Note that by selecting both the column titles and the averages the individual bars will be labeled.



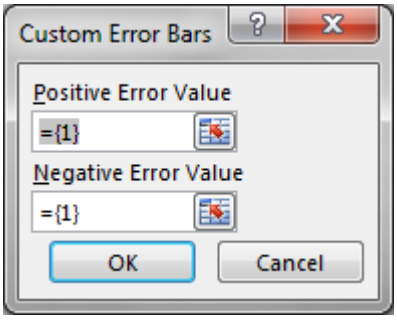
Label the axis etc. using the chart design tool as you did for the line graph.

Adding Error Bars

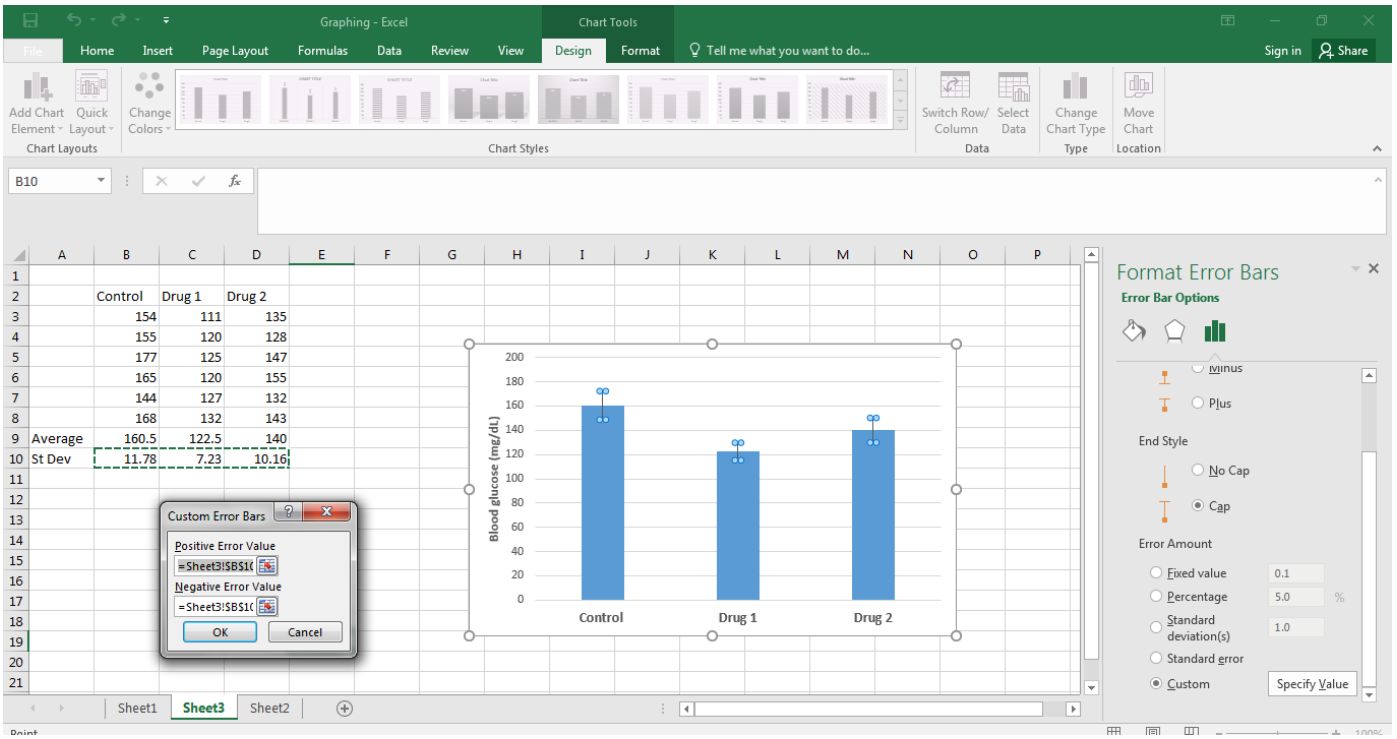
Using the chart design menu, select “Error bars”. Excel may automatically add error bars once you do this, but you need to make sure the correct values are used. Make sure you select “Custom” on the error bar menu as shown below:



Then click on “Specify value”. This should bring up a small window like the one shown below:



For positive error value, select the standard deviation values for each mean. Then do the same thing for the negative error value. Click OK, and the error bars should appear on the bar graphs.



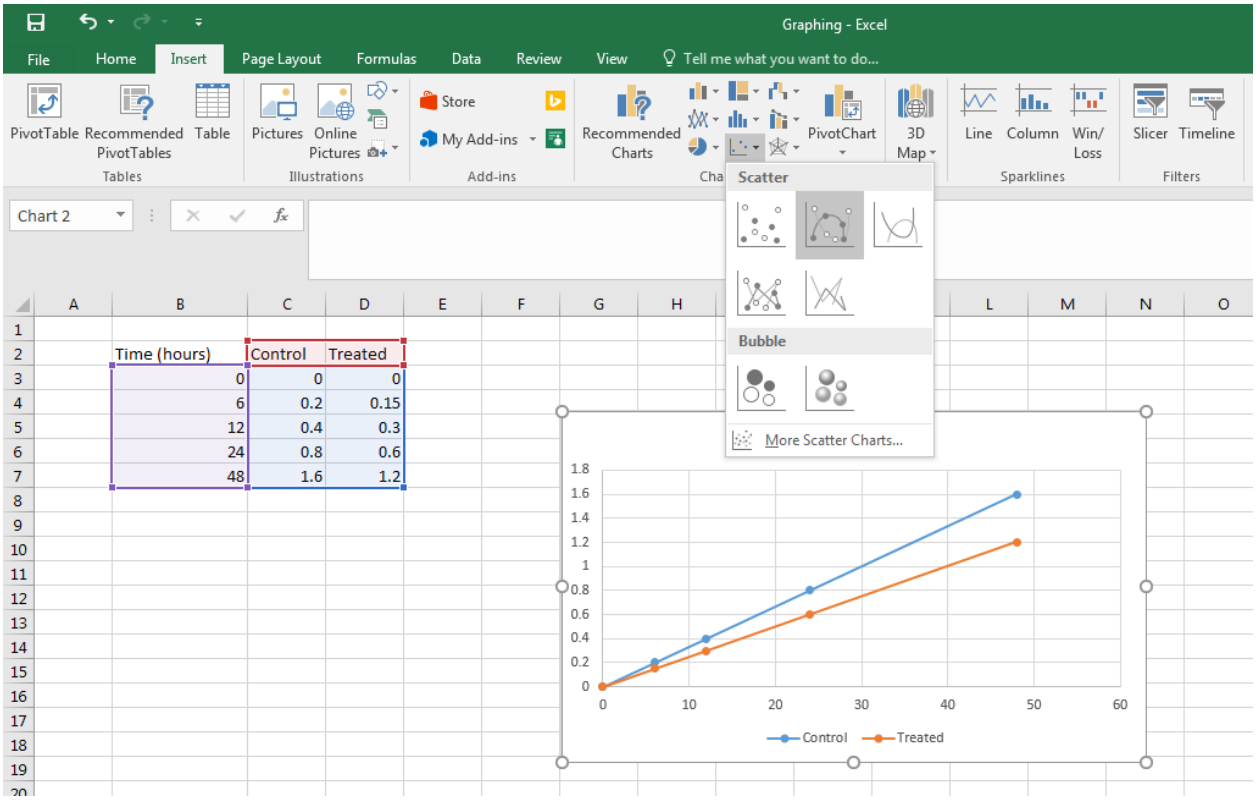
Creating a Line Graph

Line graphs are used to plot data as a function of a continuous, ordered variable like time or concentration. In the example below, we will plot the absorbance of a control and treated sample as a function of time.

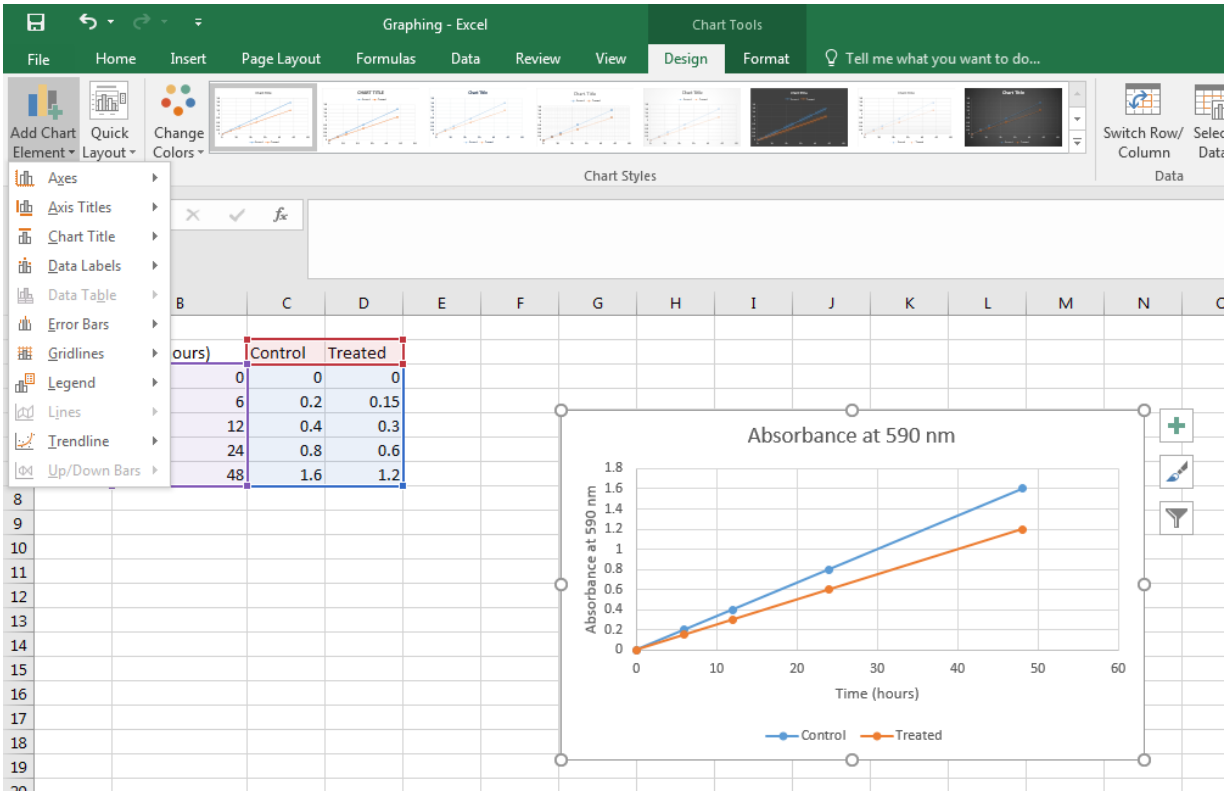
First, organize the date in columns, with the independent variable first, and the dependent variable in columns to the right

	Time (hours)	Control	Treated
3	0	0	0
4	6	0.2	0.15
5	12	0.4	0.3
6	24	0.8	0.6
7	48	1.6	1.2

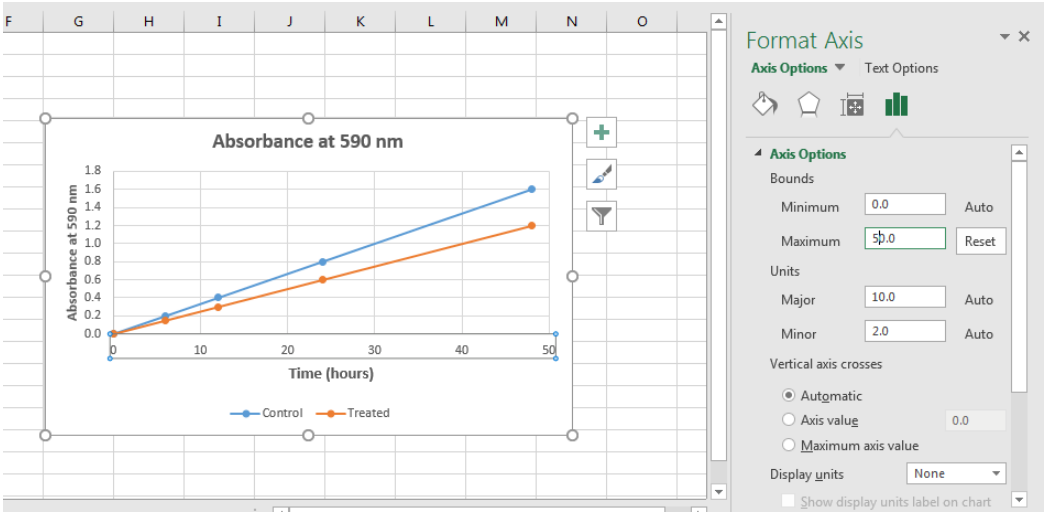
Select the data and the headings, and then go to Insert → Charts → Smoothed marked scatter. You should see a graph like the one below appears. Note that these directions will vary somewhat depending on which version of Excel you use and you should try doing this exercise in the version you have on your laptop.



Now click anywhere in the chart. A chart design menu should appear. Use the menu to add new chart elements, like a title and to label the X and Y axes.



You can adjust many aspects of the graph by double clicking or right clicking in specific parts of the graph. For example, if you click on either axis, you can change the way the axis is formatted. Let's say in the graph above you don't like the way the x axis extends to 60 hours, and you want to shorten it to 50 hours. Just click the x axis values, and a menu like the one shown below will appear. In this example, the x axis maximum value was changed to 50 hours instead of 60, and the graph changed accordingly.



You can also change the font or text size by clicking on the title or axis label. In the example above, the title and axes labels are now bold-faced.

APPENDIX D: A Quick Guide to Making Figures & Formatting

Figure Caption:

The caption is the few sentences that **go below the figure**. Figures should include an informative figure caption that describes the data that is being presented in the figure in the first sentence. For example, an acceptable figure caption of a DNA gel might say "Figure 1: Agarose gel electrophoresis of DNA samples." The sentences after the title help the reader interpret the results by including information about the samples and any other relevant information (such as what the samples are). For example, if a particular data point is highlighted, the identity of this data point should be discussed A reader should not wonder about what they are looking at. Please do not draw conclusions in the figure, save that for the discussion.

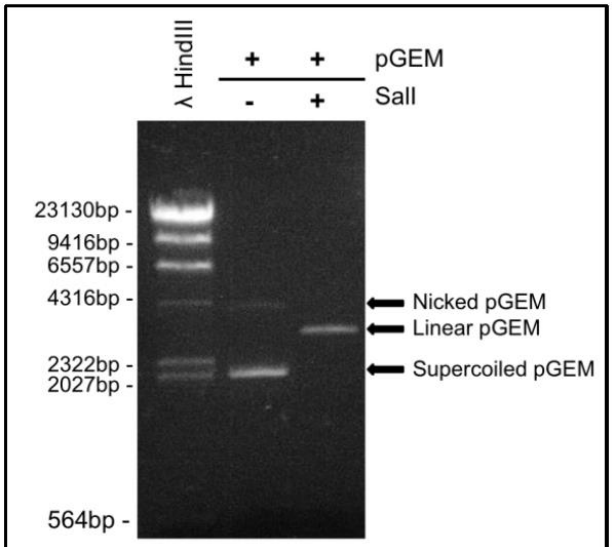


Figure 1: Agarose gel electrophoresis of pGEM restriction digest. Left lane contains λ HindIII ladder. Center lane contains undigested pGEM. Right lane contains SalI digested pGEM.

If your figure includes statistical analysis, you should also include an explanation of that in the figure caption. For example, if you present a graph with error bars, you should say "error bars are _____". If your figure includes statistical results, e.g. a * to signify a significant difference, as determined by a t-test or an ANOVA (for example), you should have something in the figure caption that indicates the meaning of * (e.g. * t-test $p<0.01$)

Your figure should have sufficient labeling so the reader knows what they are looking at. If possible, do not give things ambiguous labels (like 1, 2, 3). Instead, give them descriptive labels such as luxAB PCR, luxAB digest, etc. If you need to use somewhat vague labeling, be sure to include a detailed legend and/or description in the figure caption.

Wrapping text around a figure (in Word): Insert your figure (as an image, e.g. copy it paste it as an image). Click on the figure. Go to Format. Select Wrap text. Select "Square".

Cropping a figure to save space: If you paste a figure and it has a lot of extra white space around the edges, crop the figure so it takes up less space. Click on the figure. Click on Format. Click on "crop" (under the size options). Drag the edges in until you have removed the unnecessary edges.

Making a figure of a gel or blot: If presenting an agarose gel or other blot (western blot, northern blot, etc.), it is important to include enough relevant information so that the reader can make the same conclusion as you. For example, if showing data trying to convince your

audience that a fragment of DNA is of a particular size, it would be important to include the ladder and the sizes of the different fragments in the ladder. This allows the reader to look at the figure and independently determine the size of the DNA fragment. See Figure 1 for an example of a figure featuring an agarose gel.

In addition, the lanes of the well should be clearly labeled. The most common way to label lanes is to include the information on the top of the gel. However, it would also be acceptable to label the lanes with numbers and describe the contents of each lane in the figure legend.

Bar charts are a simple way of showing the mean value of data, and when error bars are added they can show the distribution of data around the mean. Box plots (also called box and whisker plots) show the median, first quartile and third quartile, while also showing the distribution of the data. Some box plots can also be used to show outliers, or the whiskers can be extended to the largest and smallest data points. Histograms show the distribution of the data by grouping the data into bins and reporting the number of times (the frequency) a data point falls into that bin. Box plots and histograms can be complicated to make in programs like excel, but there are many online tools that can allow you to make usable graphs if these particular types of plots are

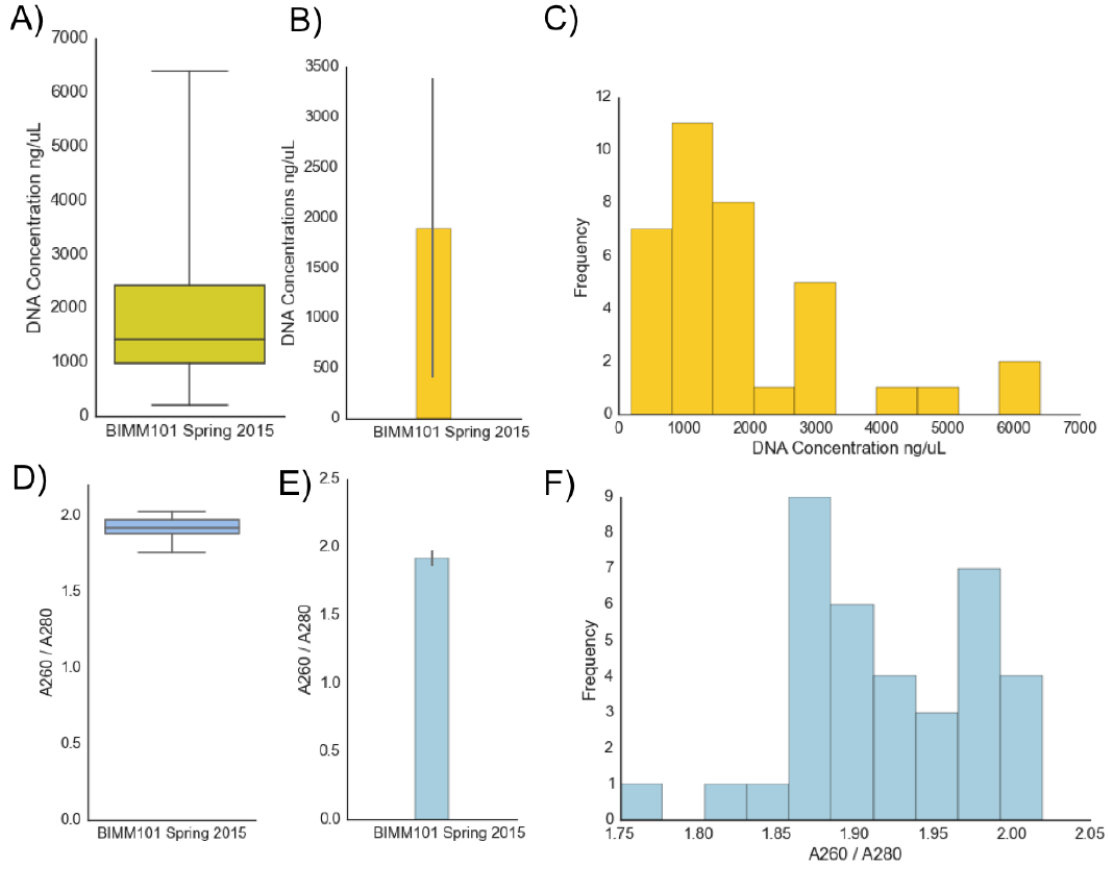


Figure 2: Various types of graphs showing the distribution of class data from *Vibrio fisheri* DNA extraction. (A) Box plot of DNA concentration. (B) Bar plot of DNA concentration. Error bars represent 1 standard deviation. (C) Histogram of DNA concentrations. (D) Box plot of A260/A280 ratio. (E) Box plot of A260/A280 ratio. Error bars represent 1 standard deviation. (F) Histogram of A260/A280 ratio.

> Error bars: Error bars can only be applied to data points that represent at least 3 different measurements. Error bars are typically used to show either the standard deviation (S.D.) or the standard error of the mean (S.E.M.). If using error bars in a figure,

be sure to say in the figure legend what they represent. For example, “Error bars represent 1 standard deviation.” See Figure 2B and Figure 2E for an example of error bars

APPENDIX E: Comparing Means with a t-test

Consider the following data sets:

Data set A: 7, 10, 3, 4, 11, 15, 12, 13, 14, 5, 5, 8
Data set B: 5, 5, 10, 11, 4, 4, 2, 1, 12, 3, 2, 7, 6

Do you think these data sets are the same or different?

You may think that the data sets look pretty similar. Or maybe the individual numbers in set #1 are a little larger, but you are not sure. Suppose this data shows how many years a cancer patient lived after treatment with treatment A (data set 1) and treatment B (data set 2). Imagine also that a loved one has the same type of cancer and has been offered either treatment A or treatment B. Look at the data sets more closely. How do you feel about the data now? Which treatment would you recommend to your loved one?

You probably favor treatment A, because there seems to be an abundance of double-digit numbers from data set one, while there are only 3 from the other data set. But how can you be more confident that treatment A is actually better than treatment B?

In order to answer that question, we have to think a little more about the numbers. While we can never be 100% certain, we can assign a probability that the data sets are truly different based on the experimental data and not simply by random chance.

While what we want to know is the probability that one treatment is better than the other given what we learn from the data, we first need to come up with a large list of explanations and then see how well the data fits each of them. Of course, this is very complicated and would take a long time to figure out. Instead, we will try to answer a different but simpler question: are the datasets are the same and are any apparent differences due to chance fluctuations?

The two possible explanations (#1 that any differences in the data sets are due to random chance data and #2 that the data sets are truly different) can also be stated as two opposing hypotheses called the null hypothesis and the alternative hypothesis.

Rejecting or proving the null hypothesis is the first step toward understanding data. Statistical tests like the Student's t-test you will be using below generate a probability or "p value" that gives a measure of how likely the differences in data sets are, assuming the null hypothesis is true. In other words, a P value is the probability of obtaining an effect at least as extreme as the one in your sample data, assuming the truth of the null hypothesis.

In terms of finding a probability value that shows that two data sets are significantly different, the cut-off number that has been accepted by the scientific community is 5%. Thus, a p value of <0.05 means that there is less than a 5% chance that your data sets are the same, and you can say the data sets are "significantly different". Another way of stating this is that if the p value is less than 0.05, the null hypothesis can be rejected—in other words, something other than chance must have influenced the results. The lower the p value is, the greater is the probability that the data sets are significantly different.

For example, let's say you do a drug study on diabetics, and you measure glucose levels in treated versus untreated patients. The glucose levels in the patients treated with the drug look lower overall when compared by eye to levels the untreated patients. When you do a statistical test comparing the glucose levels in the two patient groups, you get a p value of <0.04. This means that if the drug had no effect, you would obtain the observed difference in less than 4% of studies due to random sampling error.

Note that just because p-values are low and the data sets are significantly different, it does not mean the alternative hypothesis is true! All we can say is that the chance of us getting our data by random chance is low.

There are many, many different statistical tests that can be used to determine if sets of data are really different or not—in fact, one of the biggest problems that non- statisticians have is selecting the right one. When comparing two data sets, one of the most common and easy to perform tests is referred to as the "student's *t*-test", or "*t*-test" for short.

In doing a *t*-test, you must first decide if your data sets are paired or unpaired. The unpaired, or independent samples *t*-test is used when two separate sets of independent but identically distributed samples from each of two separate populations are being compared. For example, suppose you are evaluating the effect of a medical treatment, and you enroll 100 subjects into the study, then randomize 50 subjects to the treatment group and 50 subjects to the control group. In this case, we have two independent samples and would use the unpaired form of the *t*-test.

Dependent samples or paired t-tests typically consist of a sample of matched pairs of similar subjects, or one group of subjects that has been tested twice (a "repeated measures" *t*-test). A typical example of the repeated measures *t*-test would be where subjects are tested prior to a treatment, and the same subjects are tested again after the treatment. In order to do a paired t test, the number of samples in the two sets of data must be the same.

What does the *t*-test measure?

The *t*-test assesses whether the means of two groups are statistically different from each other, relative to the spread or variability of the data within the groups. The formula for the t-test is a ratio. The top part of the ratio is just the difference between means of the two groups. The bottom part is a measure of the variability or dispersion of the scores.

The formula for the t test is as follows:

$$t = \frac{\bar{X}_T - \bar{X}_C}{\sqrt{\frac{var_T}{n_T} + \frac{var_C}{n_C}}}$$

http://www.socialresearchmethods.net/kb/stat_t.php

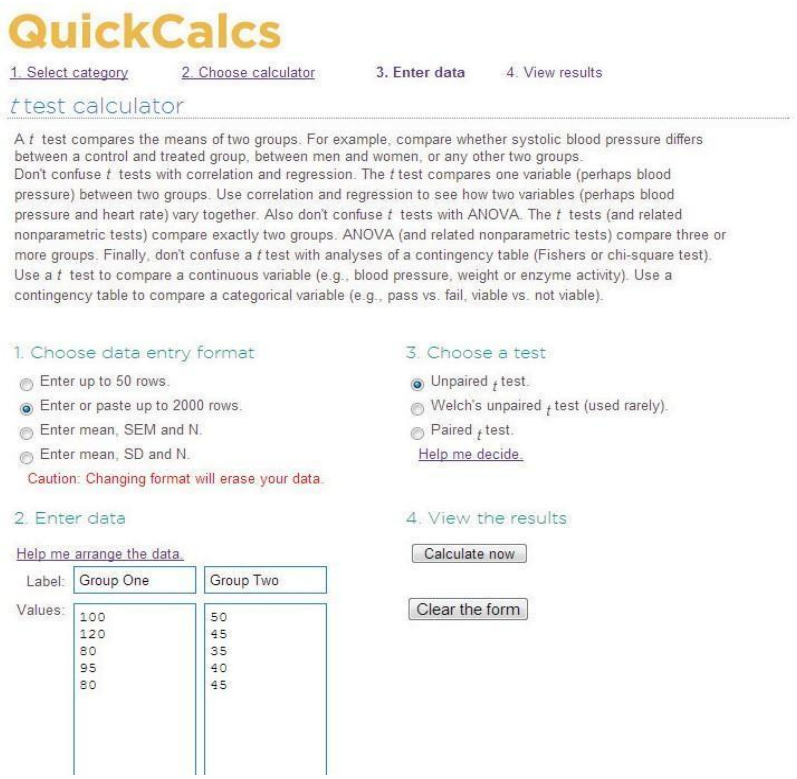
Where X is the mean of the group, var is the variance within the group, and n is the number of samples in the group. Once you calculate the t-value, you then have to use a table of significance to determine the p value. However, statistical computer programs routinely provide the significance test results and save you the trouble of looking them up in a table.

How to do a t-test

1. Go to Graphpad: www.graphpad.com and click on “QuickCalcs”

2. Choose “continuous data” (but take note all the other statistical tests you can do, for other types of data) and then “t-test to compare two means” (again, take note of all the types of statistical tests you can do)

3. Under choose data entry format, select “enter or paste up to 2000 rows”. Copy and paste your data into the two columns (e.g. control in Group One and treatment in Group Two)



4. Select “unpaired T-test,” and hit “calculate now”—the results of the t-test should be displayed.

5. Record the significance value (this will be shown as the “p value”) as well as the mean and standard deviation of each group. What does standard deviation measure? (If you don’t know, look it up.)

1. Look at the significance value (the “p value”) – can you reject the null hypothesis?

2. You can also perform a t-test in excel using this function:
   =ttest(array 1, array 2, 2, 3)

   Array 1 is all of the data from one group (e.g. control), array 2 is all of the data from your comparison group, 2 is for a two-tailed test, 3 is for assuming unequal variance.
   More information can be found here: http://www.statisticshowto.com/how-to-do-a-t-test-in-excel/

APPENDIX F: Analyzing pML104 plasmid sequence (Computer Lab)

The goal of this exercise is to explore the features of the pML104 plasmid. **You will use this plasmid sequence again in future labs, so save a copy for easy access.**

- Program needed: A Plasmid Editor (ApE)

Available for free download at: http://biologylabs.utah.edu/jorgensen/wayned/ape/
or another DNA analysis program that you like to use

**If you are using laptops or computers in the York Hall Labs or Computer Labs you do not need to download a program, ApE is already installed.

The plasmid has a variety of genes and sequence elements, some of which are denoted schematically in the maps shown earlier in the lab manual. For the purposes of constructing a new plasmid with a specific guide RNA sequence, you'll need to identify specific sequence elements at the nucleotide level in these plasmid sequences.

For each plasmid you are going to identify:

- Promoter driving expression of guide+scaffold RNA
- Terminator
- Scaffold sequence
- Other sequences among the promoter-gRNA-scaffold region.

See below for details.

How to add features in ApE: Highlight the desired sequence, click on Feature → New Feature → give it an appropriate name, can select desired color.

See Appendix B for more info on the functions in ApE.

After completing the analysis below, you should have an annotated map of pML104 plasmid (no guide RNA), and an expected pML104-gRNA annotated map.

Create a document with your annotated plasmid maps, and answering any other questions that appear below.

Upload your document to Course website by the end of lab.

Annotating the pML104 67638 plasmid sequence:

- € Download the pML104 plasmid sequence from the Course website site (this sequence is partially annotated)

The plasmid sequence is already partially annotated, e.g. the Cas9 protein coding sequence, the URA3 coding sequence, origins of replication. You can look at the sequence as a circular map by clicking on the circle icon in the line-up at the top of the ApE window. The steps below ask you to find other sequences of importance in the plasmid.

1. To add a guide RNA sequence to this plasmid, restriction enzyme digestion and ligation is used. Specifically, the gRNA sequence is added into the *BclI-SwaI* cut plasmid. The recognition sites for these enzymes are below. You can find restriction enzyme sites in ApE using the Enzymes Selector menu option. Find them in the pML104 sequence and add them as features.

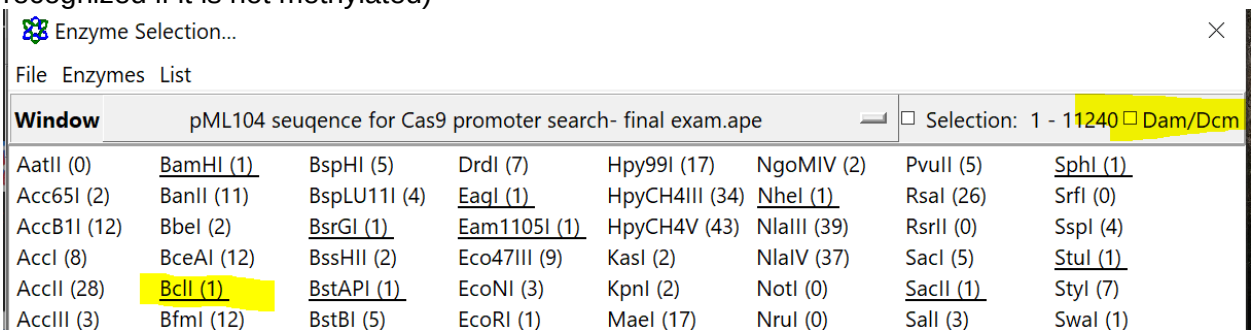
BclI:

5′...T▼GATCA...3′
3′...ACTAG▲T...5′

SwaI:

5′...ATTT▼AAAT...3′
3′...TAAA▲TTTA...5′

When searching for BclI site, using the "Enzyme Selector" function of ApE, make sure the Dam/Dcm box is **unchecked** (because BclI restriction enzyme site sequence will only be recognized if it is not methylated)



2. The gRNA will be added immediately upstream of the scaffold RNA sequence. However, this plasmid is missing 13 nucleotides from the 5' end of the scaffold. The 3' portion of the scaffold RNA (3' sgRNA) sequence is already in this pML104 plasmid (the 5' portion will be added along with the target-specific gRNA sequence). Here is the full scaffold RNA sequence. The bolded nucleotides are the 5' sgRNA sequence and the non-bolded is the 3' sgRNA is already in the plasmid.

5′ GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTG
AAAAAGTGGCACCGAGTCGGTGGTGC 3′

Find the 3' sgRNA and add it as a feature in your pML104 sequence.

3. Next, we want to identify the promoter that drives the expression of the guide RNA/scaffold. The promoter is annotated in the plasmid sequence file, compare the promoter sequence

against the yeast gnome using the yeast genome browser BLAST function, to identify which gene the promoter came from.

- Go to a yeast (Saccharomyces cerevisiae) genome database site, https://www.yeastgenome.org/blast-sgd, and paste the copied promoter sequence into the query box. Do a BLASTN search of the yeast reference genomic sequence. You will see a list of sequence matches between your query and the yeast genome. Choose a good match and click on the genome browser link for this sequence match.
- Update the promoter feature in your plasmid map to indicate which gene the promoter came from.

Which gene did the promoter come from? What is the function of this gene? Do you think the promoter will result in high or low expression of the gRNA in yeast? Why? What RNA polymerase recognizes this promoter?

Note that the transcription termination sequence is already annotated.

4. Using a similar approach, search for (and annotate in your ApE file) the yeast promoter that drives transcription of the Cas9 gene.

Note that the Cas9 gene is transcribed using the bottom strand as the coding sequence, and the ApE file is the top strand only. Knowing this, identify the sequence that will correspond to the start codon on the transcribed bottom strand.

Which gene did the promoter come from? What is the function of this gene? Do you think the promoter will result in high or low expression of the gRNA in yeast? Why?

5. Identify the SV40 NLS sequence at the end of the Cas9 coding sequence.

What is the purpose of this sequence?

6. Save your sequence, click on the circle icon again, and view your well annotated plasmid sequence! You will use this annotated sequence in future lab exercises. **Include this plasmid map in your document for this assignment.**